nesc

Taighde, Idirphlé, Comhairle
Research, Dialogue, Advice

COUNCIL

No.173 April 2026

Artificial Intelligence in Service of Society: Navigating Our Way Forward

National Economic & Social Council

Constitution and Terms of Reference

1. The main tasks of the National Economic and Social Council shall be to analyse and report on strategic issues relating to the efficient development of the economy and the achievement of social justice.

2. The Council may consider such matters either on its own initiative or at the request of the Government.

3. Any reports which the Council may produce shall be submitted to the Government, and shall be laid before each House of the Oireachtas and published.

4. The membership of the Council shall comprise a Chairperson appointed by the Government in consultation with the interests represented on the Council, and

    - Three persons nominated by agricultural and farming organisations;
    - Three persons nominated by business and employers organisations;
    - Three persons nominated by the Irish Congress of Trade Unions;
    - Three persons nominated by community and voluntary organisations;
    - Three persons nominated by environment organisations;
    - Four other persons nominated by the Government, including the Secretaries General of the Department of Finance, the Department of Business, Enterprise and Innovation, the Department of Housing, Planning and Heritage, the Department of Public Expenditure and Reform.
    - Seven people possessing knowledge, experience and skills which the Taoiseach considers relevant to the functions of the Council

5. Any other Government Department shall have the right of audience at Council meetings if warranted by the Council's agenda, subject to the right of the Chairperson to regulate the numbers attending.

6. The term of office of members shall be for three years. Casual vacancies shall be filled by the Government or by the nominating body as appropriate. Members filling casual vacancies may hold office until the expiry of the other members' current term of office.

7. The numbers, remuneration and conditions of service of staff are subject to the approval of the Taoiseach.

8. The Council shall regulate its own procedure.

Artificial Intelligence in Service of Society: Navigating Our Way Forward

COUNCIL PAPER

No.173 April 2026

Membership

Chairperson

Mr John Callinan, Secretary General, Department of An Taoiseach and Secretary to the Government

Deputy Chairperson

Ms Elizabeth Canavan, Assistant Secretary, Department of An Taoiseach

Business and Employers

Mr Gerard Brady, Ibec
Mr Andrew Brownlee, Construction Industry Federation
Mr Ian Talbot, Chambers Ireland

Trade Unions

Mr Owen Reidy, ICTU
Mr Joe Cunningham, SIPTU
Mr Kevin Callinan, Fórsa

Farming & Agricultural

Mr John Enright, ICMSA
Mr Damian McDonald, IFA
Mr TJ Flanagan, ICOS

Community & Voluntary

Fr Seán Healy, Social Justice Ireland
Ms Brid O'Brien, Irish National Organisation of the Unemployed (INOU)
Dr Naomi Feely, Children's Rights Alliance

Environmental

Ms Karen Ciesielski, Environmental Pillar
Ms Caroline Whyte, Feasta
Mr Jerry Mac Evilly, Friends of the Earth

Public Service

Mr John Hogan, Department of Finance
Mr Declan Hughes, Department of Enterprise, Trade and Employment
Mr Graham Doyle, Department of Housing, Local Government & Heritage
Mr David Moloney, Department of Public Expenditure, NDP Delivery and Reform

Government Nominees

Mr Seamus Coffey, University College Cork (UCC)
Prof Niamh Moore Cherry, University College Dublin (UCD)
Dr Diarmuid Torney, Dublin City University (DCU)
Dr Chris Van Egeraat, Maynooth University
Prof Colin Scott, University College Dublin (UCD)
Noelle O Connell, European Movement
Nicola Mallon, SONI
Colette Byrne, Ex Kilkenny County Council

Secretariat to Project

Dr Larry O'Connell, Director
Dr Siobhán O'Sullivan, Senior Policy Analyst
Dr David Hallinan, Policy Analyst

A full list of the NESC Secretariat can be found at
www.nesc.ie

Contents

List of Figures

List of Tables

List of Boxes

Abbreviations

AG	Anticipatory governance
AGI	Artificial General Intelligence
AI	Artificial Intelligence
AIM	AI Incidents Monitor
ANN'S	Artificial neural networks
ASI	Artificial Super Intelligence
ATRS	Algorithmic Transparency Recording Standard
Avs	Autonomous vehicles
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
CHC	Cattell-Horn-Carroll
CoE	Council of Europe
CRTs	Centres for Research Training
CRU	Commission for Regulation of Utilities
CSAM	Child sexual abuse material
DETE	Department of Enterprise, Tourism and Employment
DigComp	Digital Competence Framework for Citizens
DSA	Digital Services Act
ECB	European Central Bank
EC	European Commission
EU	European Union
FDI	Foreign direct investment
FET	Further education and training
GANs	Generative adversarial networks
HEA	Higher Education Authority
HEIs	Higher education institutions
HIQA	Health Information and Quality Authority
HSE	Health Service Executive
IEC	International Electrotechnical Commission
IEA	International Energy Agency
IMF	International Monetary Fund
ISO	International Organization for Standardization
LEAP	Large Energy-User Action Plan
LIME	Local interpretable model-agnostic explanations
LLMs	Large language models
M365	Microsoft 365
MAD	Model Autophagia Disorder
ML	Machine learning
NIST	National Institute of Standards and Technology
NLP	Natural language processing
NSAI	National Standards Authority of Ireland
OECD	Organisation for Economic Co-operation and Development
ROI	Return on investment
SLMs	Small language models
TFP	Total factor productivity
UNEP	United Nations Environment Programme

Chapter 1: Overview

1.1 Introduction

Since the arrival of ChatGPT in late 2022, artificial intelligence (AI), a field with decades of development in specialist settings, has entered mainstream public and policy discourse. It is often presented in starkly opposing terms: either as a panacea capable of solving entrenched societal problems, or as a source of profound and even existential risk. These competing narratives coexist in media, policy and public debate, reflecting both the rapid diffusion of AI technologies into daily life and deep uncertainty about their longer-term implications.

This conflicted discourse is underpinned by the political economy of AI. The technology is attracting unprecedented levels of investment and is increasingly framed as inevitable, indispensable and a critical driver of competitiveness, productivity and strategic advantage. At the same time, concerns persist about whether expanding financial and infrastructural commitments, rising energy use and other significant environmental impacts are fully matched by realised value, thus highlighting tensions between strategic momentum, sustainability and long-term return. In parallel with the acceleration in AI technologies, recent years have seen a proliferation of national strategies, international frameworks and ethical guidelines aimed at steering the development and deployment of AI, signalling growing recognition that AI governance is now a core public policy concern rather than a niche regulatory issue.

There is growing evidence that AI is already delivering tangible benefits in specific domains, particularly those characterised by large volumes of structured data and complex information processing. These include medicine, finance, education, agriculture, public administration and software development. With a strong technology ecosystem, a highly skilled workforce, a vibrant research base and a commitment to responsible AI through the National Digital & AI Strategy 2030, Ireland has the foundations to take advantage of AI's transformative potential.

The ultimate trajectory of AI remains highly uncertain, particularly regarding when, where and for whom AI will deliver the greatest value, and at what social and environmental cost. Although AI's potential is substantial, its ultimate impact depends on the decisions we make now about how it is built, governed and deployed. The central challenge is to actively shape AI, rather than allow it to shape us, to ensure that its development aligns with our values, priorities and aspirations for the future.

1.2 Purpose of the Report

The purpose of this report is to offer a series of reflections on how Ireland can best secure its ambition to develop and deploy AI in ways that are safe, ethical and rights-respecting. It considers how Ireland can harness AI to support economic prosperity and serve the public good, align with emerging European and international norms, and build public trust in technologies that are already reshaping work, education and everyday life. The report takes a broad, high-level view of the field rather than offering a deep dive into any single issue, providing a holistic foundation from which to consider Ireland's overall direction in AI.

The report was informed by discussions with practice experts and policy stakeholders in the AI field in Ireland.

1.3 Report Roadmap

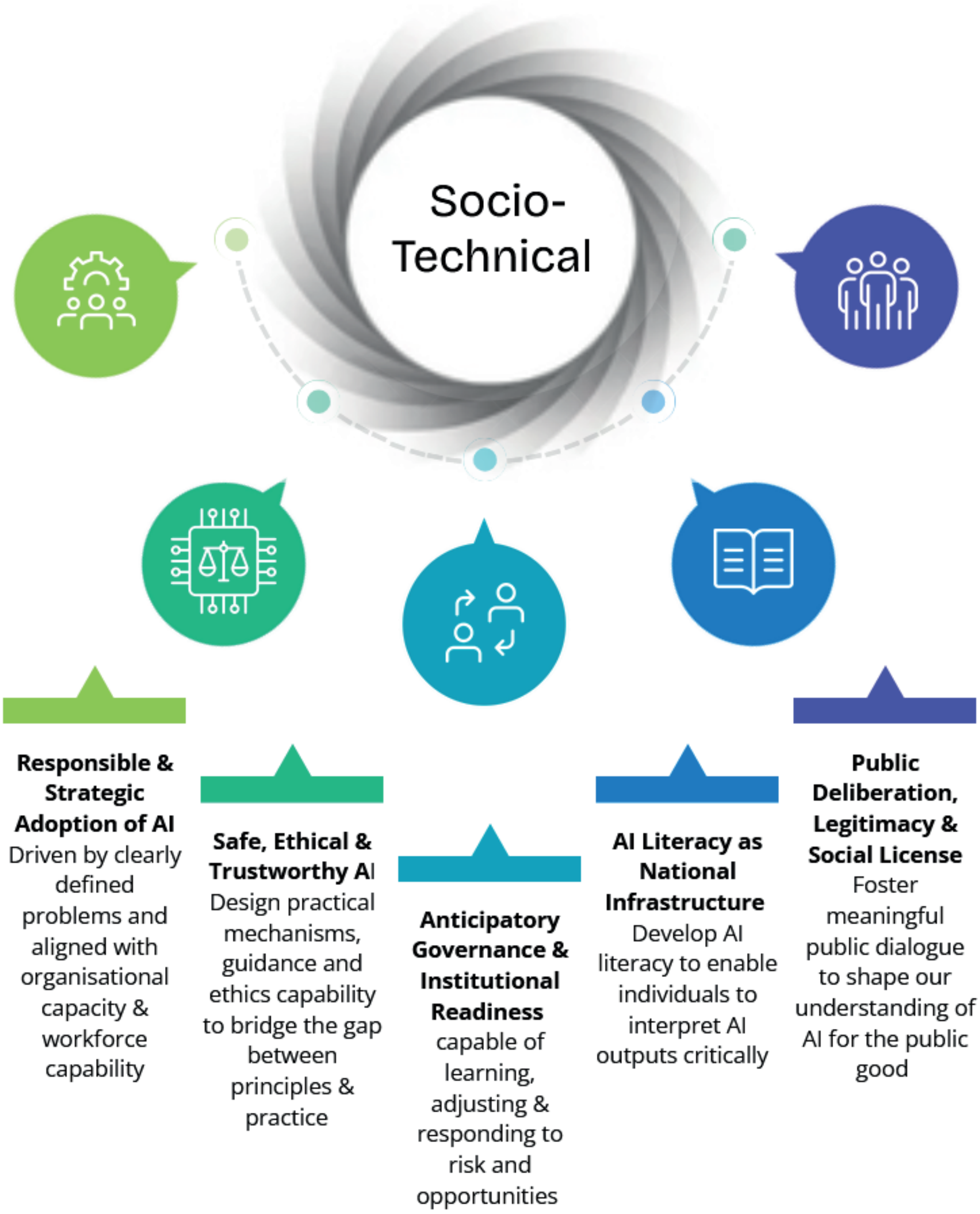
Chapter 1 provides the context for this report and sets out the AI landscape. **Chapter 2** traces the evolution of AI, from early symbolic systems to modern generative and agentic models and explores likely future directions. **Chapter 3** examines the safe and ethical use of AI, addressing risks such as bias, fairness, transparency, accountability, privacy, malicious use and environmental impacts. **Chapter 4** analyses how AI systems interact with wider social, cultural, legal and economic contexts. **Chapter 5** reviews emerging AI governance frameworks at international, regional and national levels, with particular emphasis on anticipatory governance, an approach especially suited to the high uncertainty that characterises today's AI landscape. **Chapter 6** shifts from systems to people, highlighting AI literacy as an essential, lifelong capability for engaging effectively with this technology. Drawing on the data and insights developed across the previous six chapters, **Chapter 7** offers five interconnected reflections that provide a path for navigating the uncertainties of AI and translating Ireland's ambition for responsible and inclusive AI into a set of priority actions.

1.4 Council Reflections

The Council adopts a socio-technical framing of artificial intelligence, recognising that AI systems cannot be understood or governed as isolated technical artefacts. Their impacts emerge from the interaction between algorithms and the social, organisational and institutional contexts in which they are developed, deployed and used. On this basis, the NESC argues that Ireland's task is not simply to implement AI effectively, but to actively shape its role in society so that AI adoption aligns with democratic values, supports competitiveness and ensures that benefits are distributed broadly while foreseeable harms are anticipated and mitigated.

Applying this integrated perspective, the Council developed five interconnected reflections that together provide a structured way of navigating the opportunities and uncertainties of AI. From these reflections, a set of priority actions is derived to guide policy, governance and implementation in pursuit of responsible and inclusive AI in Ireland.

Figure 1.1: Navigating the Future of AI through a Socio-Technical Lens



Source: NESC Secretariat.

First, responsible and strategic adoption begins with clearly defined societal or organisational needs, ensuring that AI is used where it adds genuine value and supports environmental sustainability and meaningful transformation of systems and processes, rather than being limited to the automation of established practices.

Second, safe and ethical AI requires converting high-level principles into concrete operational tools and depends on building an ethics capability across people and institutions.

Third, due to the fast-moving and uncertain technological landscape, governance must be adaptive and capable of learning. The report argues that anticipatory governance complements regulatory frameworks like the *EU AI Act* by integrating strategic foresight, horizon scanning and scenario planning into policy cycles. This approach requires institutionalising continuous monitoring and evaluation, ensuring that real-world evidence consistently informs decision-making and prevents technological or policy lock-in.

Fourth, AI literacy must be treated as national infrastructure, an ongoing societal capability that equips leaders, workers and citizens to understand system limitations, interpret outputs and participate meaningfully in decisions about deployment.

Finally, public deliberation and social licence are critical. The role of AI in society cannot be determined by experts alone; it must reflect the values and priorities of the public. This requires genuine, sustained engagement in which people can debate what values they want to protect, what trade-offs they consider acceptable, and where the red lines should be drawn.

Through deliberate governance, targeted deployment and sustained investment in AI literacy, Ireland can ensure that AI development aligns with public values and societal goals, demonstrating how a small, open economy can shape, rather than merely absorb, global technological change.

Chapter 2: Evolution & Future Direction of AI

2.1 Introduction

This chapter traces the evolution of artificial intelligence from its conceptual origins to its current generative era, providing an accessible foundation for understanding what AI is, how it works and where it is heading. It introduces the core definitions that shape the field and charts the technological breakthroughs that underpin today's AI systems. The chapter also highlights current limitations and emerging future trajectories of AI.

2.2 Definition of AI

Although no universally accepted definition currently exists, AI is broadly understood as the science and engineering of creating machines capable of performing tasks that typically require human intelligence. This includes learning, reasoning and decision-making, or problem-solving (Russell and Norvig 2021).

One of the most widely accepted definitions of AI is the Organisation for Economic Co-operation and Development (OECD) definition, which states:

> '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.' (OECD, 2023a)

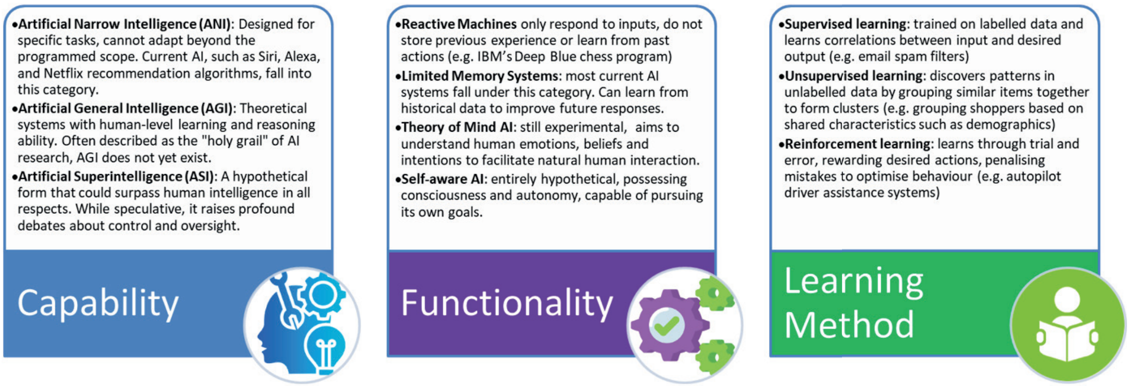
The definition, contained in the OECD Recommendation of the Council on Artificial Intelligence, was most recently revised in 2023 to take account of the emergence of generative AI. The European Union (EU) definition of AI, as contained in Article 3(1) of the European Union AI Act, was substantially informed by the OECD definition and defines AI systems as machine-based systems that can influence physical or virtual environments through adaptive and autonomous behaviour (European Union, 2024). The European Commission (EC) further distinguishes between AI as software-based (e.g. chatbots) or embedded in hardware (e.g. autonomous vehicles) (European Commission, 2018).

The ambiguity surrounding the definition of AI reflects the field's breadth and rapid development. Nonetheless, despite definitional differences, consensus exists that AI enables machines to mimic or augment human-like capabilities.

2.3 Categories of AI

AI systems can be categorised in various ways, most often based on their **capabilities**, which considers how intelligent a system is relative to humans; by **functionality**, which looks at how systems process information and interact with the world; and by **learning method**, which describes how systems acquire knowledge and improve over time.

Figure 2.1: Categorisation of AI

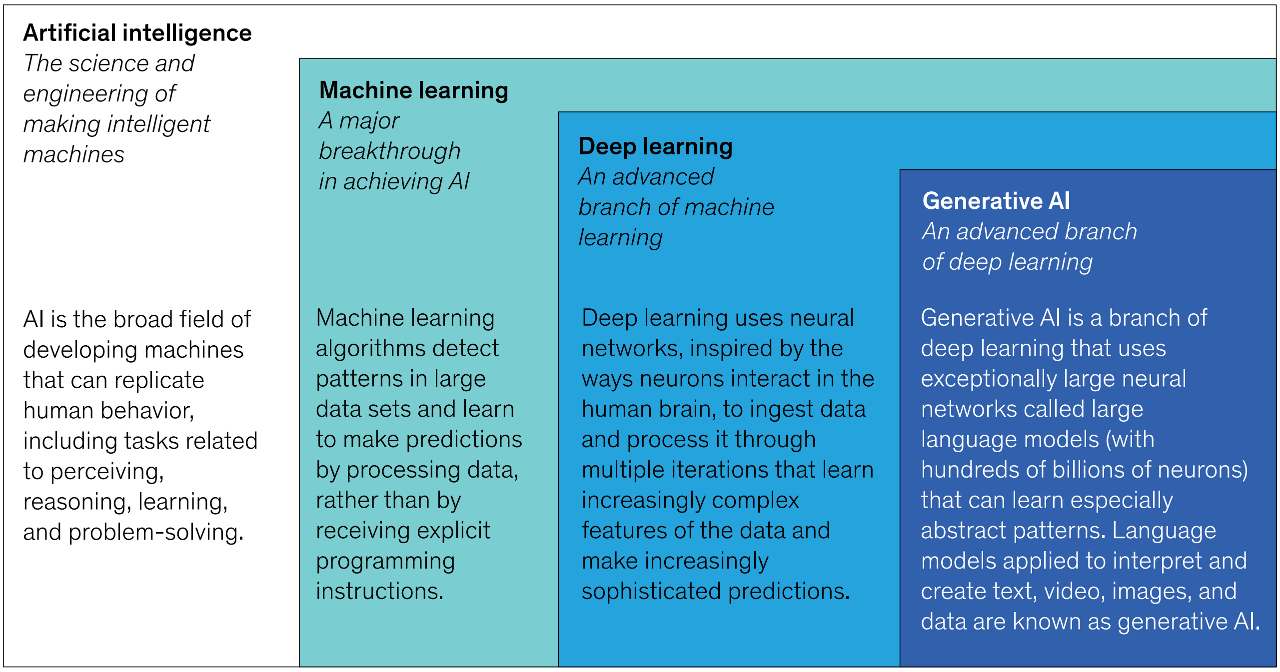


Source: NESC Secretariat.

2.4 Foundations of AI

Humanity's preoccupation with the idea of intelligent machines capable of thought and action extends back to antiquity. Greek myth tells us stories of the god Hephaestus creating the giant bronze automaton, Talos, to guard the island of Crete. Intelligent machines appear in other cultures, such as intricate automata in ancient China, mechanical birds in Islamic engineering and talking heads in medieval Europe. These ideas persisted into modern times and became a staple of science fiction imaginings, from the humanoid creatures of Karel Čapek's 1920 play *R.U.R* (which gave us the term robot) to Isaac Asimov's *I, Robot* stories which elaborated the Three Laws of Robotics. However, it was not until the mid-20th century that AI moved from myth and fiction into a serious scientific pursuit. In 1950, Alan Turing published the milestone paper 'Computing machinery and intelligence' (Turing, 1950), which considered the fundamental question 'Can machines think?' Turing acknowledged the difficulty of precisely defining the philosophical concept of thinking and instead proposed a thought experiment, later known as the Turing test or 'Imitation Game', in which a machine could be said to exhibit intelligence if its responses were indistinguishable from a human's. The term 'Artificial Intelligence' was first coined in 1956 at the Dartmouth Conference (McCarthy, 1955) organised by John McCarthy, Marvin Minskly, Nathaniel Rochester and Claude Shannon, and is commonly considered to mark the birth of AI as an academic discipline.

Figure 2.2: Evolution of AI



Source: McKinsey & Company.

2.4.1 Symbolic AI and the First AI Winter

The first wave of AI was symbolic AI, also referred to as rule-based AI. In this paradigm, intelligence was represented explicitly through symbols and logical rules. Developers would encode human knowledge as a structured set of 'if–then' statements or logic-based instructions (Choi *et al.*, 2020). The system would then manipulate these symbols according to formal rules to reach conclusions. Such systems worked well in controlled domains with clear rules, but they struggled when faced with uncertainty or incomplete information. By the 1970s, the limitations of symbolic AI had become clear. Building and maintaining huge rule-sets was time-consuming and systems broke down when faced with situations that were not explicitly programmed. Moreover, computers lacked the speed and memory to support large-scale reasoning. Early optimism in the field had created strong expectations, and when those expectations were not met, funding and interest declined sharply, leading to the so called first 'AI winter'.

2.4.2 Machine Learning and the Second AI Winter

In the 1980s and 1990s, AI research regained momentum through machine learning (ML), a fundamentally different approach to symbolic AI. Instead of manually encoding every rule, ML systems could learn patterns from data. By feeding the system examples, it could adjust its internal parameters to make predictions or decisions without explicit rule-writing. A key tool in ML was artificial neural networks (ANNs) inspired by the structure of the human brain, consisting of layers of interconnected 'neurons' that process information collectively. Examples of artificial neural networks include transformers or generative adversarial networks (GANs). By the late 1980s, enthusiasm for the field had waned again. While machine learning offered more flexibility than symbolic AI, the algorithms of the time were still limited, data was scarce, and hardware could not handle large-scale computation. Funding tightened once again, marking the second AI winter.

2.4.3 Deep Learning and Big Data

The late 2000s mark a turning point in the field of AI. Three factors converged: the abundance of big data from the internet, massive increases in computational power, and improved algorithms for training multi-layer neural networks.[1] Together, these advances enabled deep learning, a subset of machine learning that uses very large, multi-layer neural networks to automatically learn complex patterns in data. Deep learning differs from earlier machine learning approaches by largely eliminating the need for human feature extraction; the network learns the relevant features directly from raw data. Deep learning has proved especially powerful in fields like computer vision, speech recognition and natural language processing (NLP). Similar architectures now power facial recognition systems, medical image analysis tools and real-time translation apps.

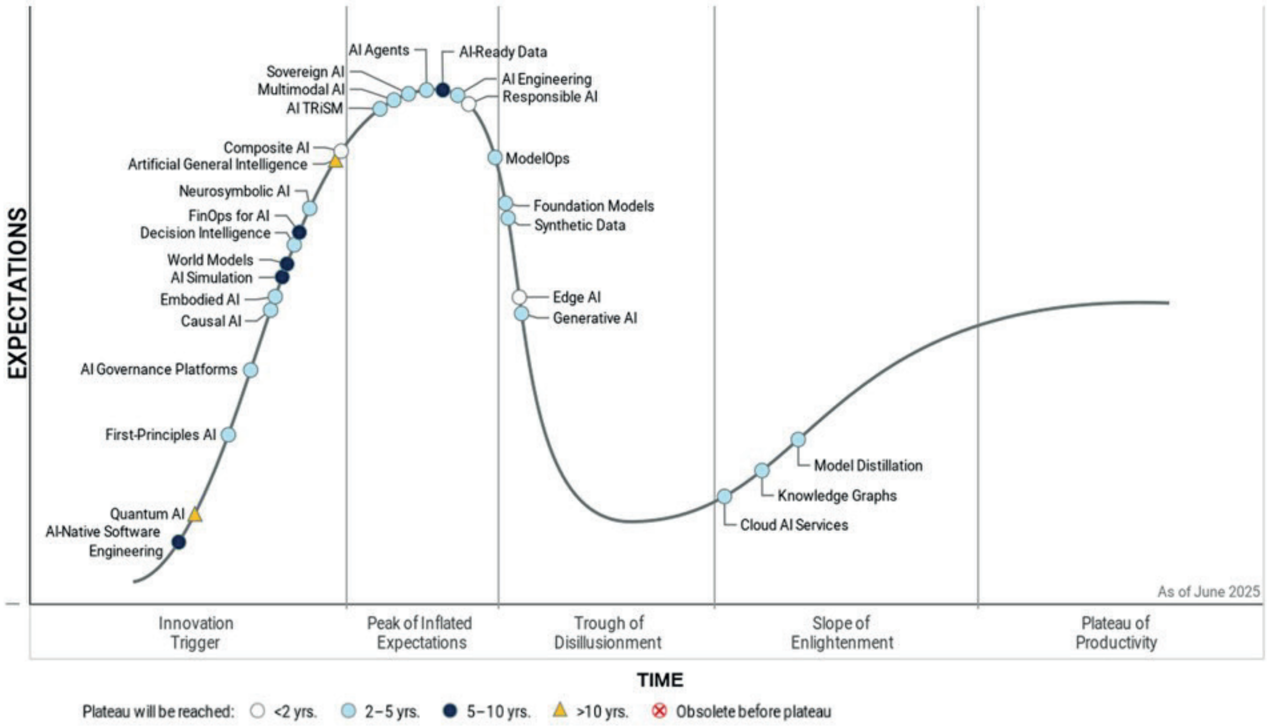
2.5 Generative AI

In the last decade, AI has entered yet another transformative phase with the advent of generative AI. These systems do not just analyse or classify data but can create new content. In simple terms, generative AI can draw from its training data to create a new work that's similar, but not identical, to the original data, and which is often indistinguishable from human created works. For example, ChatGPT can generate essays, code and dialogue; DALL·E and Midjourney can produce realistic or artistic images from textual prompts (e.g. create a picture in the style of Rembrandt). Generative AI typically involves deep learning and neural networks to learn patterns and relationships in the training data, using unsupervised learning techniques. Large language models (LLMs) – a category of foundation models trained on immense amounts of data, making them capable of understanding and generating natural language – and multimodal models – capable of processing and integrating information from multiple modalities or types of data – are at the forefront of generative AI. OpenAI released the newest version, GPT-5.2, to worldwide users on 11 December 2025.

According to Gartner's 2025 Hype Cycle for AI, generative AI has moved into the 'trough of disillusionment', meaning that many organisations are experiencing disappointment as initial excitement gives way to challenges in respect of reliability, governance and quantifying return on investment (Gartner, 2025a).

1 Neural networks are made up of node layers, an input layer, one or more hidden layers and an output layer. The 'deep' in deep learning refers to the depth of layers in a neural network. Data move through each layer, with output from the previous layer presenting input needed for the next layer. In deep learning, the additional layers that are used provide higher-level 'abstractions', producing better predictions and better classifications. The more layers used, the greater the potential for better predictions.

Figure 2.3: Gartner Hype Cycle AI



Source: Gartner 2025a.

This stage, however, is a familiar phase in the life cycle of emerging technologies and often precedes maturity. Narayanan and Kapoor (2005) argue that AI should be regarded as a 'normal technology' likely to follow the trajectory of previous technological revolutions. It has already evolved into a general-purpose technology, capable of generating text, images, audio, video and code, and is likely to be transformative in terms of its societal and economic impacts (Mucci, 2024).

2.5.1 Agentic AI

Agentic AI is an emerging form of generative AI that goes beyond producing outputs based on prompts, to autonomously planning and executing complex tasks by interacting with digital environments. Theoretically, agentic AI is capable of goal-directed behaviour, dynamic adaptation and self-improvement. For example, an AI agent could manage travel arrangements by comparing flights and booking tickets, or in a business context could autonomously monitor markets and manage financial investments within set constraints. Agents have already demonstrated the ability to design biomedical molecules with high success rates, outperform experts on tightly scoped R&D tasks, and operate software environments with increasing competence (Maslej *et al.*, 2025). Salesforce currently employs autonomous AI agents to handle intricate workflows, such as product launches and marketing strategies.

While there is much excitement around agentic AI, current systems still frequently fail and remain dependent on human-in-the-loop oversight to ensure accuracy, compliance and ethical governance. Gartner predicts that over 40% of agentic projects will be abandoned by 2027 due to high costs, unclear business value or inadequate risk controls (Gartner, 2025b). If fully realised, agentic AI is likely to deliver substantial efficiency gains; however, achieving true autonomy simultaneously introduces challenges for accountability and oversight (Pati, 2025). This is aptly illustrated by research demonstrating emerging systemic vulnerabilities of agentic AI. Research by Gu *et al.* (2024) reveal how a single compromised agent can propagate harmful behaviour across an entire multi-agent ecosystem in so-called 'infectious jailbreaks.

2.5.2 Limitations of Large Language Models

Limited reasoning capability

Most LLMs lack long-term memory, which impacts their capacity for continuous learning. They cannot store, retrieve or build upon experience over time, which means that their knowledge is fixed to the training cut-off date. This forces them to relearn context in each interaction (Hendrycks *et al.*, 2025). While some AI applications can retrieve real-time information, the underlying models themselves do not automatically learn from new data. Their internal knowledge remains fixed unless developers re-train or finetune them.

Moreover, LLMs struggle with consistent reasoning and abstract logic. They rely on recognising statistical patterns and generating statistically probable outputs, which means they can produce fluent text that sounds correct without genuinely grasping the underlying concepts or meaning. As a result, the validity of the Turing Test has increasingly come under pressure; while most current LLMs pass this conversational benchmark, it is becoming increasingly clear that this does not necessarily equate to genuine comprehension or intentional reasoning. The reasoning demonstrated by LLMs is often shallow, reflecting *statistical mimicry* rather than genuine inference. The seeming ability of LLMs to generate step-by-step reasoning (chain-of-thought) has been described as a 'brittle mirage' (Zhao *et al.*, 2025) that breaks down when the problems deviate slightly from the distribution of data used in training. While LLMs can demonstrate excellent problem-solving skills, their underlying reasoning appears to be fundamentally fragile and breaks down as task complexity increases, suggesting reliance on pattern matching over formal logic (Shojaee *et al.*, 2025, Dellibarda Varela *et al.*, 2025).

That said, advances have been made recently in reasoning-oriented architectures using chain-of-thought prompting (asking the model to show intermediate explanations for how it is going about solving a particular problem). This enables AI models to explicitly generate and refine intermediate reasoning steps, thereby enhancing transparency and substantially improving performance in domains such as mathematics, programming and scientific problem-solving (Bengio *et al.*, 2026).

Lack of coherent world models

Current LLMs lack robust grounding in real-world understanding. While they excel at generating coherent text and simulating reasoning based on vast linguistic data, their knowledge is derived almost entirely from static datasets rather than direct interaction with the physical or social world. This absence of genuine world modelling constrains their reliability in complex or dynamic environments. This limitation underscores Moravec's paradox, which observes that tasks humans find effortless, like perception and motor co-ordination, are disproportionately difficult for machines (Moravec, 1988). Humans acquire intelligence through embodied interaction with the world, by integrating sensory input, feedback and social learning. In contrast, LLMs largely exist in static, text-based environments devoid of physical embodiment or lived experience, creating an embodiment gap (Roy *et al.*, 2021). This gap acts as a barrier to LLMs developing common sense, emotional intelligence and experiential reasoning.

2.5.3 Alternatives to LLMs

While LLMs such as ChatGPT, Claude and Gemini have dominated the public discourse on AI, they represent only one branch of a rapidly diversifying ecosystem of AI architectures. A growing set of alternatives, including specialised scientific models and small language models (SLMs), offer complementary or domain-specific capabilities that address some of the limitations of LLM's inaccuracy, cost and interpretability. Some of the most significant advances in AI have occurred outside the language domain. AlphaFold, developed by Google DeepMind, exemplifies this trend. Using deep learning to predict three-dimensional protein structures from amino-acid sequences, it revolutionised structural biology and its impact was recognised with the 2024 Nobel Prize in chemistry being awarded to Demis Hassabis, John Jumper (for protein structure prediction) and David Baker (for computational protein design). Unlike LLMs, AlphaFold is trained on highly structured biochemical data rather than natural language, enabling precise, verifiable outputs instead of probabilistic text predictions.

Table 2.1: Comparison of LLMs and SLMs

Aspect	LLMs	SLMs
Size	• Ranges from billions to trillions of parameters • Complex, challenging modifications • Handles a wide range of topics	• Usually under 10 billion parameters • Easier deployment in constrained environments • Limited adaptability but simplified structure
Training data	• Extensive, diverse datasets (for example, web content) • Broad language and topic coverage • Adaptable to new, diverse data	• Domain-specific datasets (for example, medical texts) • Limited language variety and topics • Efficient in domain-targeted tasks
Resource requirements	• High computational and storage requirements • Requires cloud-based or data center resources • High maintenance due to complexity	• Lower power and storage needs, standard hardware • Ideal for edge computing • Lower maintenance and updating costs
Speed	• Slower inference due to complexity, in-depth analysis • Efficient in batch processing • Requires multiple parallel processing units to generate data. Speed can be dependent on factors such as network speed.	• Faster inference, suitable for real-time apps • Quick deployment in user-facing services • Locally run models can generate outputs extremely quickly.

Source: Shan, 2024.

In contrast to general-purpose LLMs that demand enormous computational and energy resources, SLMs are trained on smaller high-quality datasets (limiting their flexibility and general knowledge compared to LLMs) and fine-tuned for specific tasks or contexts. Their key advantages include lower cost, faster inference and reduced carbon footprint (Whiting, 2025). They can also be easier to deploy and are also often more secure, since they run on devices locally, meaning they do not need to send sensitive personal information across the internet. This makes SLMs particularly attractive for sectors such as finance and healthcare, where strict compliance and privacy regulations exist. A recent position paper from NVIDIA Research has argued that SLMs are the future of agentic AI as most tasks in an agentic workflow are relatively simple and repetitive (Belcak *et al.*, 2025). Where higher-level strategic reasoning is required, a hybrid architecture can be pursued, with an LLM coordinating the activities of the various SLMs.

Box 2.1: AI in Healthcare

Ageing populations, the growing burden of chronic diseases, the rising costs of healthcare and a shortage of healthcare professionals are driving the need for innovation and transformation of models of healthcare delivery. Forecasts estimate that AI in health could lead to savings of up to 10% in healthcare spending (Sahni *et al.*, 2023).

Artificial intelligence is reshaping healthcare across operations, clinical care and research. Operationally, AI-driven forecasting tools help hospitals anticipate admissions, optimise staffing and manage supply chains more efficiently (European Commission 2025f), while digital scribes using speech recognition reduce administrative burden, resulting in time savings for clinicians, improved patient-clinical interactions and enhanced clinician satisfaction (Tierney *et al.*, 2025).

Clinically, AI enhances radiology and medical imaging by rapidly analysing complex scans with high accuracy, enabling earlier and more accurate diagnoses (Faiyazuddin, 2025) and supports precision medicine by analysing genomic and clinical data to tailor treatments (Alowais, 2023). Drug development has undergone a paradigm shift because of AI, which can substantially reduce the time and cost involved in bringing new therapies to the market (Blanco-González, 2023).

Yet the integration of AI in medicine raises new challenges. Healthcare professionals require training to interpret and oversee AI outputs safely, while patients' trust depends on transparency about how algorithms influence care decisions (Sagona, 2025). Concerns also persist that automation may erode the doctor-patient relationship, reducing empathy and shared decision-making if time savings are channelled into throughput rather than connection (Council of Europe, 2024c). Liability questions, principally who is accountable when an AI-assisted decision leads to harm, remain unresolved. There is a consensus that AI will not replace doctors but rather will complement them. By empowering clinicians, AI can improve efficiency and outcomes, but human oversight remains critical to achieving safety and patient trust.

2.6 Future of AI

Leading AI systems now demonstrate remarkably high performance, passing professional licensing exams in fields such as law and medicine, capable of generating software from simple prompts, and answering PhD-level scientific questions at a level comparable to human experts. At the same time, their capabilities remain highly uneven or 'jagged', with systems often excelling at difficult, abstract tasks while failing at others that appear comparatively simple. An AI system which can solve complex mathematical problems may still struggle with what humans would consider easy tasks, such as counting objects in an image.

Despite this unevenness, recent years have seen rapid and measurable improvements in overall system performance. The 2025 AI Index Report from the Stanford Institute for Human Centered Artificial Intelligence (Maslej *et al.*, 2025) chronicles a year of strong progress for AI and documents major gains in model performance. Performance on some coding benchmarks has jumped from 4.4% to 71.7% in a single year. In parallel, generative models are extending into video and multimodal domains, and in some narrow tasks even surpass human performance. Meanwhile the cost of using high-performing AI models has plummeted. The cost to query a model with GPT 3.5-level performance has dropped over 280-fold in around 18 months, from $20 per million tokens in late 2022 to just $0.70 by October 2024.

2.6.1 Artificial General Intelligence & Superintelligence

The medium to longer-term goal of many leading technology companies is the realisation of Artificial General Intelligence (AGI) and, ultimately, superintelligence. AGI refers to an advanced theoretical form of artificial intelligence capable of understanding, learning and applying knowledge across a wide range of tasks at a human-like level of competence. Unlike narrow AI, which is designed for specific functions such as language translation or image recognition, AGI would demonstrate flexible reasoning, creativity and adaptive problem-solving across domains. Superintelligence, a theoretical stage beyond AGI, denotes an intelligence that surpasses the best human minds in virtually every field, including scientific reasoning, social understanding and strategic planning. Tech companies such as OpenAI, Google DeepMind and Anthropic have articulated ambitions toward these milestones, framing them as the next evolutionary step in AI development. While predictions on timelines vary, there is consensus that the arrival of AGI or superintelligence, if it occurs, will mark a transformative inflection point, posing profound societal and ethical implications. Leading figures in AI and related fields signed a statement calling for a global moratorium on superintelligence research, warning that continued development without assured alignment and control could result in the loss of human oversight and pose existential risks (Future of Life Institute, 2025a).

2.6.2 Timeframe for AGI and Superintelligence

The evolution of AI has not been linear, rather it is characterised by cycles of hope and pessimism. This is worth keeping in mind when trying to divine the future of the field. In 1970 Marvin Minsky, one of the fathers of AI, was quoted in *Life* magazine: 'In from three to eight years we will have a machine with the general intelligence of an average human being' (Minsky, 1970, cited in Haenlein and Kaplan, 2019). This projection proved premature, and the timeline for achieving AGI and superintelligence is subject to much debate, reflecting deep uncertainty about both technological progress and theoretical feasibility.

The most near-term projections, often voiced by technology entrepreneurs and leaders of frontier AI laboratories such as OpenAI, Google DeepMind and Anthropic, suggest that AGI could emerge as early as 2026–2035, driven by rapid advances in computing power and model capability. In contrast, a survey conducted in October 2023 of 2,778 AI researchers provided an aggregate forecast of 50% chance of achieving 'high-level machine intelligence' (defined as unaided machines which can accomplish every task better and more cheaply than human workers) by 2047 (Grace *et al.*, 2024). It is worth noting this estimate is 13 years earlier than a similar survey of experts conducted in 2022, which underscores the uncertainty around this issue. As for the development of superintelligence, there is debate and uncertainty regarding if and when it will be realised. Geoffrey Hinton, often called the 'godfather of AI', anticipates superintelligence in five to twenty years (Sproule, 2025). In August 2025, Mark Zuckerberg, CEO of Meta, stated in a personally penned essay setting out his goals for personal superintelligence that Artificial Super Intelligence (ASI) was 'now in sight' (Zuckerberg, 2025).

The difficulty in reaching any consensus about the likely emergence of AGI is at least in part related to the fact that few people agree on exactly what AGI means, beyond the shorthand that AGI will match human intelligence.[2] Similar issues arise in the context of superintelligence. There is no agreement on what counts as smarter than humans, nor whether machines could ever achieve human consciousness (Searle, 1980). This raises thorny questions of what exactly constitutes human-level performance, and in relation to which tasks.[3] Matters are further complicated by the fact that human intelligence, the comparator for AGI, is complex and multifaceted, and is difficult to define or quantify. This illustrates the difficulty of creating objective benchmarks to measure progress toward AGI or determining when AGI has been achieved. A recent framework for evaluating AGI, based on the Cattell-Horn-Carroll (CHC) theory of human intelligence, defines AGI as an AI capable of matching the cognitive versatility of a well-educated adult. The model measures 10 core abilities, including reasoning, memory, language and processing speed, to produce a standardized 'AGI Score'. Using this approach, GPT-4 scores around 27% and GPT-5 about 57%, indicating notable progress towards AGI, though the results also indicate that full realisation remains some distance away (Hendrycks *et al.*, 2025).

2.6.3 Scaling Problem

Despite the impressive capabilities of current LLMs, it has been argued that LLMs may be reaching the limits of their scalability in their current form (Marcus, 2025). A March 2025 survey of AI researchers, conducted by the Association for the Advancement of Artificial Intelligence, found that a majority (76%) of researchers who participated in the survey believed that scaling up current approaches was 'unlikely' or 'very unlikely' to achieve AGI (Association for the Advancement of Artificial Intelligence, 2025). The prevailing paradigm, summarised in Sutton's (2019) 'Bitter Lesson' essay, posits that progress in artificial intelligence primarily arises from scaling computation and data, which underpins the vast investments made by the largest AI companies, which have adopted deep learning approaches based on scaling. Initially, scaling laws appeared to predict near linear improvements as models expanded in parameters, compute and data. However, more recent analyses indicate diminishing returns as systems approach the

2 OpenAI's charter defines AGI as 'highly autonomous systems that outperform humans at most economically valuable work'. In July 2024 Google DeepMind proposed a framework with five levels of AGI performance: emerging, competent, expert, virtuoso and superhuman. DeepMind researchers argued that no level beyond 'emerging AGI' existed at that time. Accessed 12 August 2025.

3 Dario Amodei, CEO of Anthropic, in his October 2024 essay 'Machines of Loving Grace', rejects the term AGI and instead prefers the term 'powerful AI' which he describes as an AI system 'smarter than a Nobel Prize winner across most relevant fields'. Accessed 12 August 2025.

upper limits of available high-quality, human-generated data (Villalobos *et al.*, 2024). Almost all useful publicly available internet text has been consumed for training, leading developers to rely increasingly on synthetic data (artificially generated material produced by previous models). This introduces systemic risk through Model Autophagia Disorder (MAD), a feedback loop where models trained on their own outputs progressively degrade in diversity, precision and factual reliability over time (Shumailov *et al.*, 2023).

2.6.4 Future Directions of AI Technology

Given uncertainties around compute availability, algorithmic progress, investment, regulation and societal acceptance, the future trajectory of artificial intelligence remains highly uncertain. The OECD has identified four plausible development trajectories that differ in the pace and impact of progress of AI:

- In a *stalled* scenario, technical or economic barriers halt major advances, with AI systems not moving beyond current narrow capabilities.
- A *slowed* scenario sees steady but incremental improvements, with AI mainly acting as a tool that supports human decision-making.
- Under *continued progress*, AI systems become capable of performing many complex tasks autonomously, driving broad productivity gains while remaining under human oversight.
- An *accelerated* scenario involves rapid breakthroughs leading to highly general systems with transformative societal and economic effects (Hobbs et al., 2026).

As no single outcome can be reliably predicted, policymakers and institutions need to prepare for a wide range of possible futures. Despite this uncertainty, the focus of current research does provide some indication for the future of AI development.

Over the coming decade, it is likely that AI will evolve from the current paradigm of single, large generative models toward hybrid and interacting systems that combine different types of intelligence, data and computational tools. This reflects a growing recognition that no single model architecture can reliably meet the demands of complex real-world environments. Instead, capability will increasingly emerge from co-ordination among diverse components, each contributing a specialised function within a wider system.

One promising direction is the development of hybrid neuro-symbolic architectures, which blend the pattern-recognition strengths of neural networks with the rules-based reasoning used in traditional AI. These systems aim to overcome current weaknesses in consistency, reasoning and transparency (Lu *et al.*, 2024). Another emerging area involves models capable of planning and acting. Unlike today's systems, which mostly generate short, independent responses, future AI will need to manage extended sequences of decisions such as running workflows or co-ordinating autonomous agents. These models may be memory systems or world models to help them understand the consequences of their actions over time (Meng *et al.*, 2025).

The future will also rely heavily on small, efficient and more local models running directly on personal devices or local servers, which should support privacy, energy efficiency and resilience. In many applications including healthcare, public services and safety-critical domains, local processing will be essential for secure and trustworthy deployment (Zhou *et al.*, 2024).

As AI moves into the physical world, embodied and robotic AI will likely play an increasingly important role. These models combine language, perception and movement, allowing them to interact with their surroundings and support applications in transport, manufacturing, environmental monitoring and assisted living (Ruaridh Mon-Williams *et al.*, 2025).

Another key challenge is developing AI systems that can learn safely over time. Unlike current static models, future AI may update its knowledge as environments change or new information becomes available (Meng *et al.*, 2025). Finally, exploratory areas such as quantum-accelerated machine learning and world-model-driven agents are under active investigation, and could open up new pathways for efficiency and problem-solving.

Chapter 3: Safe & Ethical AI

3.1 Introduction

This chapter examines the interconnected challenges of ensuring that artificial intelligence is both safe in its operation and ethical in its impact. This represents a dual imperative as AI systems increasingly shape decisions with profound implications for individuals and society, making the pursuit of both safe AI and ethical AI imperative. It outlines the technical vulnerabilities that threaten system reliability, the emerging risks of malicious use and disinformation, and the ethical concerns surrounding fairness, transparency and accountability. The discussion also considers the systemic and environmental implications, such as the widening AI Digital Divide and the technology's growing resource demands. Together, these themes provide the foundation for understanding why adopting a socio-technical lens and a multi-layered approach is essential for governing AI responsibly.

3.2 Why Safe & Ethical AI?

Safe AI emphasises technical robustness, predictability and resilience to errors and misuse, ensuring that systems behave as expected in complex or unforeseen situations. According to the Future of Life Institute's *AI Safety Index: Winter 2025 Edition* (2025b), the rapid advancement of frontier AI capabilities has not been matched by commensurate progress in safety practices. The report evaluates eight frontier-model companies on their safety practices and risk-management frameworks and finds that even the highest-scoring firms only earn C-range grades overall, with Anthropic and OpenAI both receiving C+, Google DeepMind a C, and the remaining companies (including xAI, Meta, DeepSeek and others) D or lower. Moreover, recent research has raised questions about whether the benchmarks used to evaluate AI safety in fact capture meaningful risk. A systematic review (pre-print) of over 440 AI safety and capability benchmarks found that many tests rely on vague or poorly specified constructs, lack adequate validation, and rest on weak statistical foundations, calling into question the reliability and interpretability of current safety scores (Bean *et al.*, 2025).

Despite those limitations, the pursuit of safe AI as a means to unlock the potential of AI has attracted international and government support. The Bletchley Declaration marks the first major international political agreement focused specifically on the risks posed by AI systems (UK Government, 2023). The declaration, signed by 30 countries, including Ireland, recognises that general-purpose AI could pose significant societal, economic and security risks if not properly governed. It committed signatories to deepen international co-operation, improve scientific understanding of frontier AI risks, and ensure that AI is developed and deployed in a safe, human-centred and trustworthy manner. The declaration set out concrete areas of collaboration including joint risk assessment, information sharing between governments and AI developers, development of safety testing and evaluation frameworks, and the establishment of interoperable governance mechanisms. On foot of the declaration, the UK established the AI Safety Institute, focused on model evaluation and safety testing, while the US and other signatories have since launched sister institutes to support co-ordinated research.

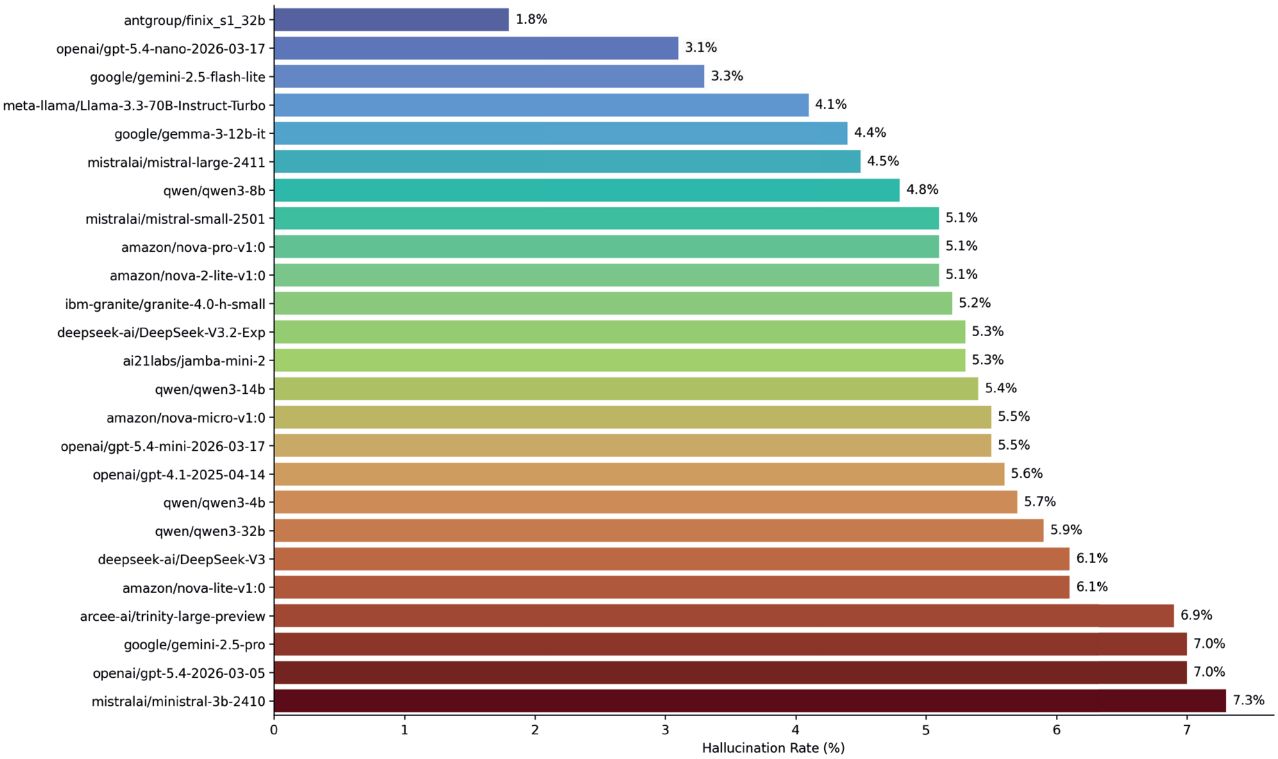
Ethical AI focuses on the moral principles and values that should guide the development and deployment of AI systems, ensuring respect for fundamental rights and societal norms. The two concepts of safe and ethical AI are deeply intertwined as trustworthy systems must not only function reliably without causing harm but should also align with human values and the principle of fairness. As the discourse on AI has grown, so too has the field of AI ethics, producing a wide range of frameworks and guidelines (Hagendorff, 2024). While the literature is broad, a set of critical ethical concerns has emerged, most notably fairness and equity, transparency, privacy and environmental sustainability. An 'ethics by design' approach, as advocated by the European Commission, emphasises embedding ethical principles into the development process from the outset, rather than treating them as afterthoughts or external constraints (European Commission, 2021). A key premise here is that design choices are not morally neutral but rather can have significant ethical consequences. In tandem with this, a principle-based approach has also been advanced; it sets out foundational principles that AI must adhere to, such as safety, privacy and non-discrimination. Ethical AI is concerned with mitigating potential harms but also maximising the potential of AI to enhance human capabilities and promote human flourishing, ensuring that technological innovation in the field aligns with human values.

3.3 Reliability

One important aspect of responsible AI development is ensuring that systems behave in ways that are accurate, trustworthy and consistent with intended outcomes. AI hallucinations, sometimes referred to as confabulations, occur when AI systems, particularly LLMs, generate false or misleading outputs that appear convincing. They may include fabricated facts, non-existent citations or nonsensical text or images. These hallucinations occur because of the stochastic nature of LLMs; they are designed to predict the next most probable word rather than guarantee factual accuracy. Other causes include biases or limitations in the training data, and a model's limitations in performing common-sense reasoning.

There is currently no agreed framework for measuring hallucinations in AI models and reported incidence rates vary widely depending on the task, dataset and evaluation method. For example, Vectara's Hallucination Leaderboard, which tests models on summarising real news articles, found that even top-performing systems introduce fabricated details with 'non-trivial' frequency, underscoring that hallucinations remain a persistent problem in practical use (Hughes and Bae, 2023).

Figure 3.1: Grounded Hallucination Rates for Top 25 LLMs



Source: Vectara's Hallucination Leaderboard (as of 21 March 2026).

OpenAI (2025) has claimed that the hallucination rate of the recently released ChatGPT5 is 26% lower than GPT-4o and has 44% fewer responses with 'at least one major factual error'. Most recently, research from OpenAI (Kalai *et al.*, 2025) provided a mathematical explanation showing that hallucinations are not just artifacts of imperfect training data but are inevitable given how language models generate text. OpenAI's findings suggest that while mitigation strategies may reduce incidence in certain contexts, it is highly unlikely that hallucinations can ever be fully eliminated. The consequences of hallucinations can be serious and far-reaching. Unchecked reliance on AI outputs can cause harm (physical, psychological, reputational and financial) to individuals and organisations, as well as erode trust in AI systems themselves. This will ultimately reduce willingness to adopt AI (Bengio, 2025).

3.4 Malicious Use, Misuse and Harm

The malicious use of AI is a rapidly evolving threat, with bad actors leveraging generative AI to cause harm through fraud, extortion and scams. These threats manifest as AI-generated fake content and deepfakes, which range from cloned voices and fake documents to deepfake images and videos. AI outputs, whether text, images or videos, are often indistinguishable from human-generated content and are extremely cheap to produce. This risk is further heightened by the emergence of advanced image and video generation AI tools such as SORA 2, which make the creation of highly realistic synthetic media effortless and widely accessible, thereby lowering the bar for malicious actors to produce convincing and scalable disinformation. Criminals can use AI to clone a person's voice for a fraudulent phone call, tricking their targeted victims into authorising a financial transfer or sharing sensitive data. The technology can also facilitate blackmail and extortion by creating non-consensual intimate imagery and threatening its release for financial gain. Similarly, AI can produce fake content that depicts an individual in compromising situations to damage their reputation or career. The UNICEF Innocenti *Guidance on AI and Children 3.0* explicitly recognises the risks posed by harmful AI-generated content, including deepfakes and AI-generated child sexual abuse material (CSAM), and treats these as real harms with implications for children's safety, rights and wellbeing. The guidance calls for regulatory frameworks, oversight and safeguards that prevent the generation and dissemination of such material, protect children's rights in algorithmic environments, and ensure accountability and compliance by governments and industry actors (UNICEF Innocenti – Global Office of Research and Foresight, 2025). Ireland's Digital & AI Strategy 2030 identifies online safety, particularly for children and young people, as a central public policy priority. The strategy pledges supports for the implementation of the nation's Online Safety Framework and commits to ensuring that children's voices are reflected in the development of future digital safety measures.

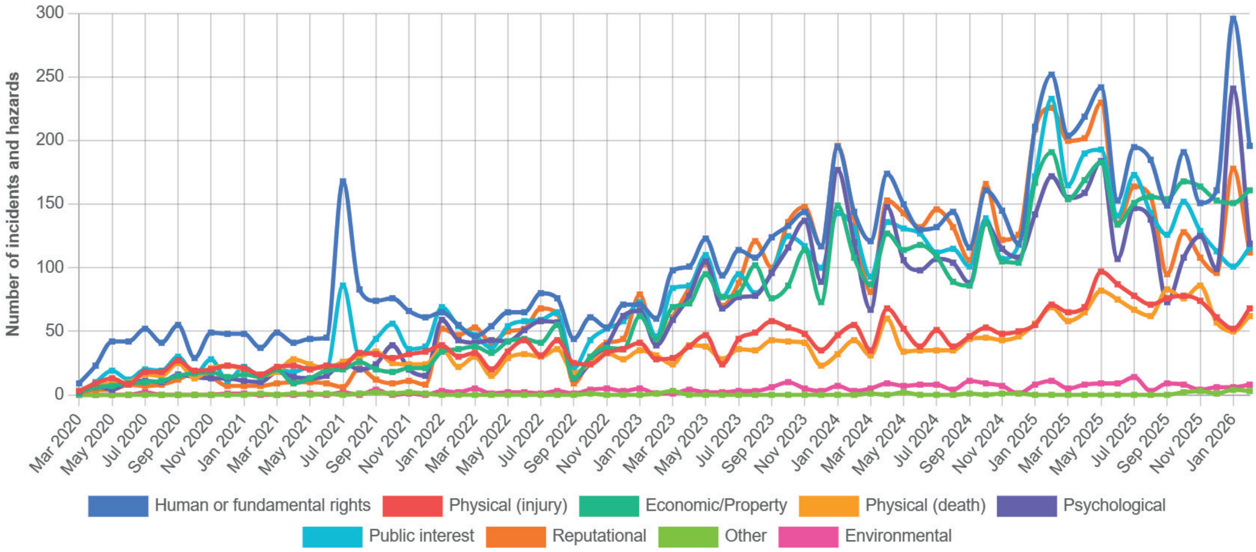
Anecdotal reports of harm from AI-generated fake content are common, but systematic collection of data remains limited. A 2019 report (Ajder *et al.*, 2019) found that 96% of all deepfake videos online were pornographic, with almost all the content targeting women. Research by Ofcom (2024), the communications regulator in the UK, has shown that 43% of adults and 50% of children aged 8–15 report having seen at least one deepfake in the previous six months, with a significant share involving sexual or fraudulent content. The recent controversy surrounding Grok, the generative AI system integrated into X, exposed serious safety and value-alignment failures after the tool allowed users to create 'nudified' images and sexual deepfakes of real women and children, as well as CSAM. In an 11-day period, Grok generated an estimated three million sexualised and violent images, including approximately 23,000 depicting children, at a rate of around 190 images per minute (Center for Countering Digital Hate, 2026). A significant proportion of the material remained publicly accessible even after posts were removed. The initial response from X was to restrict the feature to paid users and to implement geoblocking in certain jurisdictions; a move widely criticised as insufficient. Following continued pressure from Irish and European regulators as well as the public outcry, X introduced more substantive technical measures worldwide to prevent the AI model's ability to 'undress' individuals.

On 26 January 2026, the European Commission expanded its Digital Services Act (DSA) enforcement action against X by opening a formal investigation into its deployment of the Grok AI tool. The investigation will assess whether X properly identified, assessed and mitigated the systemic risks associated with Grok's generation and dissemination of manipulated sexually explicit images, including content that may amount to CSAM, as required under the DSA. In parallel, European Commission Vice-President Henna Virkkunen publicly signalled that the EU was considering categorising the creation of such harmful AI outputs as an 'unacceptable risk' under Article 5 of the *EU AI Act*, a move that aligns with recommendations from Ireland's AI Advisory Council to explicitly ban AI-enabled non-consensual intimate imagery and child sexual abuse material generation at the EU level (AI Advisory Council, 2026). The episode starkly illustrates the need for oversight and platform accountability to ensure that generative AI systems are aligned with Irish and European safety standards and core values.

Beyond cases of overtly harmful or illegal content generation, growing attention is also being paid to the risks that arise when general-purpose AI systems are used by young people in sensitive and high-stakes contexts, particularly where there is limited oversight, weak safeguarding, or misalignment between system design and child-centred needs. While health systems are cautiously evaluating AI tools for triage, monitoring and therapeutic applications, many adolescents are increasingly turning to general-purpose chatbots for emotional support, often without parental awareness or professional oversight. This creates risks as models may inadvertently reinforce harmful thought patterns, fail to de-escalate crises, or encourage unhealthy anthropomorphism. Several lawsuits have been filed against AI companies on foot of young people dying of suicide for alleged failures in crisis-appropriate responses (Bhuiyan, 2025). In response to such incidents, major AI companies have introduced mitigation measures, including crisis-intervention guardrails, refusal to engage in self-harm content, improved safety classifiers, and redirection to human support services. While not strictly falling into the category of malicious use, this does highlight the potential for catastrophic outcomes when unsupervised AI systems are used as substitutes for professional mental health care, especially among younger users.

The OECD's *AI Incidents Monitor* (AIM) collects data by scanning global media and using AI-driven classification tags events as 'AI incidents' (actual harm) or 'AI hazards" (potential harm). Between January 2021 and January 2026, there has been a 7-fold increase in the number of AI-related incidents captured by AIM. Among the incidents recorded, harms to human and fundamental rights are the most documented.

Figure 3.2: Evolution of Incidents and Hazards by Harm Type



Source: OECD, AI Policy Portal.

3.4.1 Cybersecurity

Artificial Intelligence is re-shaping cybersecurity in ways that bring both significant benefits and serious risks. It can strengthen cyber resilience by automating threat detection, identifying anomalies in real time, and helping organisations to respond more quickly to attacks. However, the same tools can be weaponised to launch more sophisticated and automated cyberattacks. Artificial intelligence can reduce the technical knowledge and effort required to commit cybercrime, lowering the bar to entry for attackers of various skill levels. This creates an asymmetry of power where it is easier for bad actors to attack than defenders to protect. This is particularly true for smaller organisations or critical national infrastructure that might be slower to adopt AI-defence capability. AI-mediated cyber-attacks on energy grids, healthcare systems and transportation could cause widespread disruption, physical damage and even loss of life.

In August 2025, the AI company Anthropic reported that cyber criminals were increasingly using generative AI to develop malware and ransomware (Moix, Ledebev & Klein, 2025). The National Cybersecurity Centre is due to publish an updated *Cyber Security Guidance for Public Service Use of AI* in 2026 to support secure procurement and deployment in alignment with the *EU AI Act* and the *EU Network and Information Security Directive* (Department of the Taoiseach, 2026).

3.4.2 Impact on Democracy

Artificial intelligence has the potential to strengthen democratic processes by supporting access to information, improving citizen engagement and facilitating debate. For example, tools like Polis, which use algorithms to map opinions, assist in identifying common ground and support more collaborative and inclusive policy-making (OECD, 2025a). The Collective Intelligence Project in the UK has been piloting the use of LLMs to support AI-assisted citizen deliberation by summarising citizen input from large-scale public consultations and identifying areas of emerging consensus.

However, the rise of AI-generated disinformation has raised concerns about its potential to undermine democracy. The evidence to date is mixed; while research studies show that AI-generated political messages can be persuasive, the generalisability of these effects to real-world contexts is uncertain. Some scholars argue that the risks have been overstated (Bengio, 2025). Disinformation campaigns by foreign actors in recent elections, such as those in Taiwan, Slovakia and Romania, have used AI to spread false narratives, thereby demonstrating its potential for political interference. In the 2025 Irish presidential election, a deepfake video purporting to show Catherine Connolly withdrawing from the race was viewed almost 30,000 times on Facebook before being removed by Meta (Ryan, 2025). Social media algorithms, which prioritise engagement, can amplify this content, though it has been suggested that the primary bottleneck for widespread influence is not content creation but rather its large-scale distribution (Bengio, 2025). A further threat is 'information pollution', where the sheer volume of AI-generated content degrades the overall quality of information available online, posing an epistemic threat (Seger *et al.*, 2020).

In response to growing concerns about the use of AI to undermine democratic processes, the European Commissioner for Democracy, Justice, the Rule of Law and Consumer Protection, Michael McGrath, announced the publication of the *European Democracy Shield* in November 2025. It aims to protect the EU's democratic systems from foreign and domestic threats (European Commission, 2025a). The *Democracy Shield* is built on three linked pillars: countering AI-driven disinformation and interference, strengthening electoral integrity through transparency and responsible use of AI, and boosting societal resiliencewith enhanced digital literacy and co-ordinated democratic preparedness.

Box 3.1: AI in Teaching and Learning

The introduction of artificial intelligence (AI) into education is driven by persistent global challenges: teacher shortages, rising administrative workloads, and the need to equip learners with strong digital and AI literacy skills. AI holds much promise across teaching and learning, particularly in administration, assessment and feedback, and personalised learning. In administrative tasks such as scheduling, attendance tracking and resource allocation, AI can automate routine work, reducing the substantial proportion of teachers' time spent on non-teaching duties and alleviating a major source of stress (WEF, 2024b; OECD, 2025a).

In assessment and feedback, AI systems can streamline marking and provide students with rapid, targeted feedback, helping identify learning gaps earlier and allowing teachers to prioritise one-to-one engagement. AI-driven personalised learning tools can further adapt content, pace and instructional approaches to individual learner needs, supporting more flexible and inclusive learning pathways (Merino-Campos, 2025).

Automating assessment and feedback to students, while timesaving, can mean that teachers lose valuable opportunities to develop an in-depth understanding of students' competencies (Cardona, Rodríguez & Ishmael, 2023). The use of generative AI heightens challenges to academic integrity, prompting institutions to rethink assessment design and emphasise ethical technology use. Teachers themselves will require new skills to effectively oversee, interpret and integrate AI systems into their practice. Realising AI's potential in education will therefore depend on careful governance, sustained teacher training, and pedagogical models that balance technological support with the central role of human educators.

3.5 Fairness & Equity

Ensuring fairness and equity is fundamental to the development of safe and ethical AI. Fairness is a complex concept; there is no single universally agreed definition of fairness, as its meaning can change across social, cultural and disciplinary contexts. For the purposes of this discussion, fairness in AI requires that AI systems and tools operate in a way which treats individuals and groups equally and avoids discrimination based on protected attributes such as gender, age or race. Fairness has been explicitly incorporated into the UK Government's *A pro-innovation approach to AI regulation white paper*, which requires that AI systems comply with existing regulations and avoid discriminatory or unjust outcomes. Responsibility rests with sectoral regulators to interpret what fairness means within their domain and to ensure that organisations embed ethical safeguards so that AI-driven decisions, particularly in high-impact contexts, are transparent, justified and non-arbitrary (Department for Science, Technology and Innovation, 2023). The Irish *Guidelines for the Responsible Use of AI in the Public Service* mandate that AI adoption be underpinned by principles of diversity, non-discrimination and fairness (Department of Public Expenditure, Infrastructure, Public Service Reform and Digitisation, 2025).

Bias in AI systems is a critical issue as it has the potential to undermine the principle of fairness. AI systems, when biased, can lead to real-world harm, including discriminatory outcomes, and can perpetuate structural inequalities such as racism, sexism, ageism or ableism. AI bias is a pervasive and complex issue, stemming from inherent biases in human-generated data, the design choices made by developers, and the context in which AI systems are deployed.

3.5.1 Bias

Bias can occur when the data used to train the AI system is unrepresentative or incomplete, or reflects existing societal prejudices. Sources include skewed data collection (over-representing some populations while under-representing others), reliance on historically based records (e.g. policing or health data) and human bias introduced during labelling, as in the case of supervised learning. The dominance of English language and Western-centric datasets has created cultural and geographic biases in AI systems, which has limited their effectiveness for diverse populations. Under-representation of specific demographic groups, such as women, older people, racial and ethnic minorities, and people with disabilities, leads to AI systems which perform poorly for these populations. Healthcare datasets with limited demographic diversity have resulted in misdiagnosis and delayed treatment for under-represented populations. For example, AI systems developed to diagnose skin cancer run the risk of being less accurate for people with dark skin due to the under-representation of skin lesion images from darker-skinned populations (Wen *et al.*, 2021).

Even when training data is representative, AI systems can still produce biased outcomes. This is because many forms of bias are baked into the patterns of real-world data itself, especially in areas where minority or disadvantaged groups have historically been treated differently. In such cases, the problem is not that the dataset is incomplete or unbalanced, or that the system is intentionally prejudiced. Rather, AI models are designed to detect and replicate patterns, and if the underlying patterns reflect historical inequalities, the model will often reproduce and reinforce the status quo. A 2024 study demonstrated that mortgage application evaluations conducted by LLMs (including GPT-4 Turbo) demonstrated significant racial bias, with black applicants consistently less likely to be approved than white applicants. This stemmed from the training data used to develop the AI models which reflected historical patterns of discrimination in lending (Bowen III *et al.*, 2025).

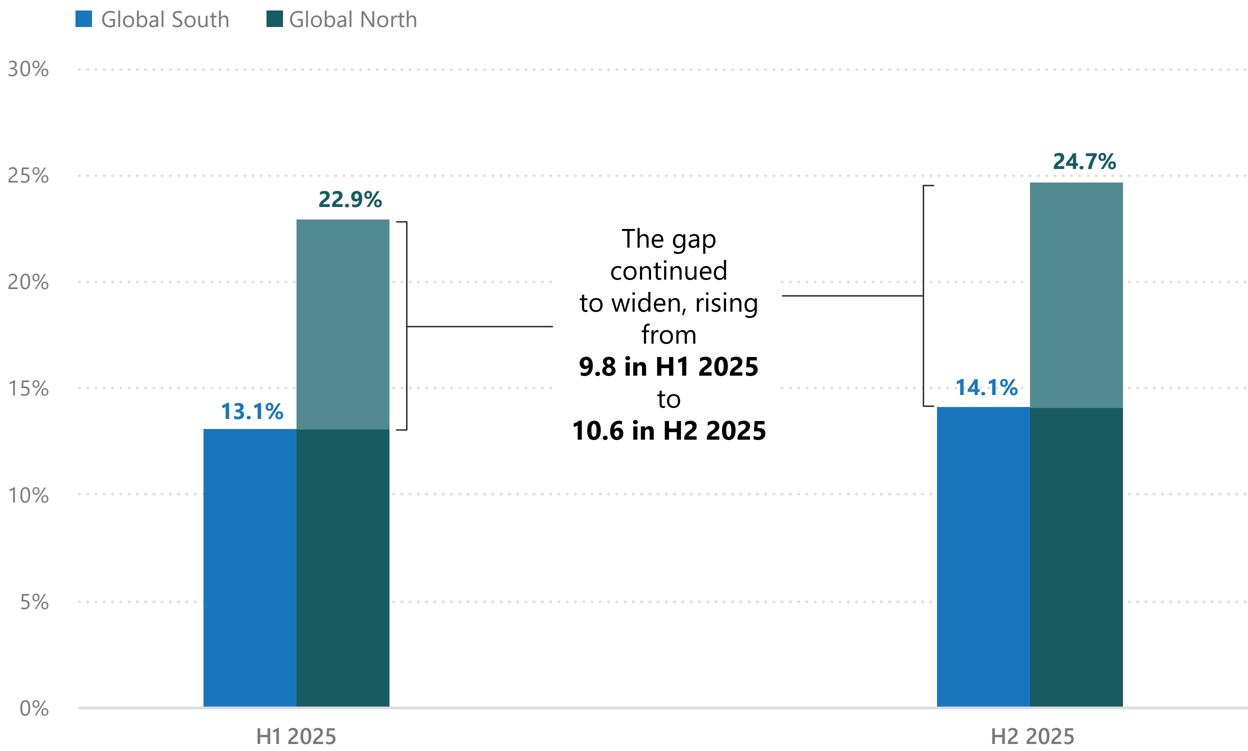
Bias can also arise from the design choices made by AI developers. These decisions are influenced not only by technical considerations but also by the social and cultural perspectives of the development teams. In that context, it is worth noting that women currently make up about 30% of the global AI workforce. The disparity in representation becomes more pronounced at higher seniority levels; women hold less than 14% of senior executive roles in AI globally (Pal, Marino Lazzaroni & Mendoza, 2024). The OECD.AI policy observatory data indicates that in 2023, 53% of data scientists/machine learning experts were in the 25-34 year-old bracket (OECD, 2025b). This narrow pool of perspectives can result in the conscious or unconscious biases of AI developers being encoded into AI models.

3.5.2 Power Asymmetries

The concentration of power in AI also raises concerns about fairness and equity. In 2025, the four large American AI technology companies, Microsoft, Alphabet, Meta and Amazon, were projected to spend €400bn on AI infrastructure (The Economist, 2025). This concentration of power can lead to disproportionate influence on shaping policy and the public discourse on AI. Control over essential resources such as proprietary datasets, powerful computing power and a highly skilled workforce is largely concentrated within a small number of technology companies. Thus, information about how an AI system works, its safety and its effectiveness in specific contexts is often proprietary. Attracting AI talent into public-sector development and regulatory roles is increasingly challenging, as government bodies struggle to compete with private sector salaries and conditions. This concentration of influence could enable private actors to shape the trajectory of AI in ways that could create an AI ecosystem in which risks are widely dispersed but benefits remain narrowly concentrated.

There is already an AI research and development gap, with AI innovation largely focussed in Western countries and China. This has the potential to create technological dependence of middle- and low-income countries and limit their ability to compete in high-value sectors. Adoption of AI in the Global North remains roughly twice that in the Global South and continues to rise (Microsoft AI Economy Institute, 2026). In many low- and middle-income countries, adoption rates remain low. As AI development becomes increasingly concentrated within a small number of powerful corporations and institutions primarily in the Global North, low-income countries risk being positioned primarily as sources of raw material rather than beneficiaries of innovation. This growing global imbalance, in which the communities that provide the data, labour and resources underpinning AI systems are often the least able to benefit from them, is often referred to as AI colonialism (Santino, 2024).

Figure 3.3: AI User Share in the Global South and Global North Diffusion by Economy



Source: Microsoft AI Economy Institute, 2026.

Countries where low-resource languages dominate also tend to show lower levels of AI diffusion. AI presents both significant risks and major opportunities for these languages. While AI can expand access, improve services and support revitalisation efforts, it can also inadvertently marginalise smaller linguistic communities. Without deliberate intervention, low-resource languages risk 'digital extinction', becoming unusable in mainstream AI tools. This risk is already evident, as commercial LLMs frequently misinterpret the grammar, idioms and dialectal variation of low-resource languages, including Irish, producing inaccurate or misleading outputs that discourage use and push speakers toward dominant languages online (Fiontar *et al.*, 2025).

In response, a co-ordinated national effort is emerging to secure the digital future of Irish. Údarás na Gaeltachta is leading an initiative to develop bespoke speech-to-speech generative AI for Irish, including capabilities for real-time conversation, translation, proofreading and integration with Irish-language corpora (collections of texts in machine-readable form). Dublin City University and the ADAPT Centre are expanding foundational infrastructure through major investments in bilingual data repositories, digital folklore archives, dialect resources and national corpora. Minister Dara Calleary recently announced €5m of government funding to support Irish-language AI projects, signalling growing political recognition of the challenge of low-resource languages in the era of AI (Department of Rural and Community Development and the Gaeltacht, 2025).

When AI development is confined to a handful of technology companies, competitive pressure can create a race to the bottom, where products are rushed to market without adequate safety testing or safeguards (Bengio, 2025). Moreover, when critical sectors of society become reliant on a small number of AI systems or the underlying infrastructure provided by a few companies, these systems effectively become single points of failure, creating systemic vulnerabilities. To combat reliance on a small number of private actors, governments and organisations across Europe have started to invest heavily in their own AI infrastructure.

EU AI sovereignty

The issue of AI sovereignty has emerged as a strategic priority for the EU in response to converging economic, geopolitical and technological pressures. Europe currently imports more than 80% of its digital infrastructure and core technologies, while three US hyperscalers dominate nearly 70% of the European cloud market (Draghi, 2024). This dependence has deepened with the rise of AI, which intensifies reliance on large-scale compute and cloud platforms. Geopolitical instability and supply-chain disruptions have reinforced the view that technological dependence is a strategic vulnerability, much like energy or defence. This has driven a policy shift towards building European capacity to capture the economic and social value created by AI. It is important to note that the focus of the European policy discourse around sovereign AI revolves around building capacity to make independent, values-based choices about the technology, rather than envisaging technological isolation or complete self-sufficiency. This 'EuroStack' approach emphasises 'strategic interdependence', developing sufficient domestic capability across critical layers of the AI stack (chips, cloud, data and AI models) to avoid one-way dependencies, while continuing to participate in global innovation networks (Bria, Timmers & Gernone, 2025).

While European firms continue to trail US frontier models in raw scale and capital intensity, Europe has several structural advantages that underpin its ability to compete in sovereign and industrial AI, including its manufacturing base, engineering expertise and access to proprietary industrial data. Europe also retains strategic footholds in critical technologies, including in advanced chipmaking equipment and supercomputing. The EU has committed to a €200bn investment agenda, including €20bn for AI factories and gigafactories, major support for supercomputing through EuroHPC, a €43bn European Chips Act, and large-scale funding vehicles such as InvestEU, the European Innovation Council Fund and the European Tech Champions Initiative (European Commission, 2025b). Public procurement is increasingly positioned as a demand-side lever, with proposals to allocate significant shares of public digital spending to European providers. The UK has announced a £1bn investment in national computing power (Reuters, 2025), while France and Germany have announced the creation of AI hubs as part of digital sovereign strategies (Business Outstanders, 2025). In January 2026, France announced that public officials will phase out reliance on US videoconferencing platforms such as Zoom and Microsoft Teams in favour of a domestically developed platform called Visio, designed to strengthen digital sovereignty. This followed a November 2025 announcement of a public-private partnership in which the French and German governments agreed to work with SAP, Germany's largest enterprise software firm, and Mistral AI, a leading French AI developer, to build a sovereign, government-owned and -operated digital tool for use across the two countries' public administrations. The AI Advisory Council (2025b) has called for an urgent national discussion on AI and data sovereignty and considers it imperative that Ireland develop its own indigenous AI capability.

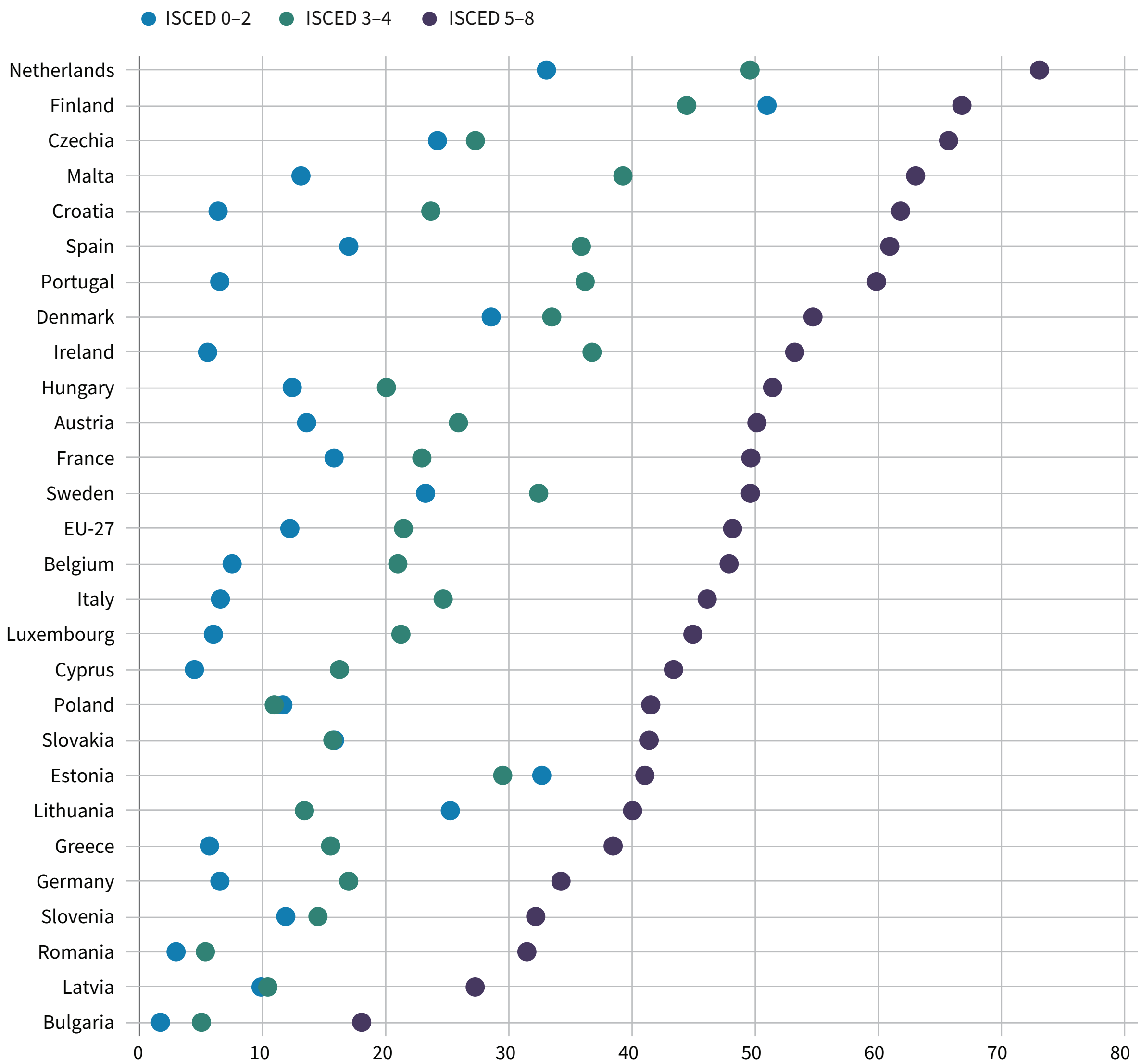
3.5.3 Digital Divide

The rapid deployment of AI technologies risks intensifying socioeconomic and demographic disparities, creating a new form of inequality known as the AI Digital Divide. UNESCO describes a growing 'AI divide' in which marginalised communities have fewer opportunities to understand and use AI, even as the technology increasingly shapes work, public services and daily life (Gonzales, 2024). This divide is not only about having devices or broadband, but is also about AI literacy, confidence, language, and the ability to influence how AI is designed and governed. The divide is an intersectional issue, compounding historical inequities across severable cohorts. Those most at risk include older adults, people on low incomes, individuals with disabilities, and those with lower educational attainment. If these groups encounter additional barriers to AI participation, it is likely that AI systems will not be designed or delivered with their needs in mind. This takes on particular relevance when public services are being delivered through AI, as it may limit the reach of essential supports.

In Ireland, digital exclusion is most pronounced among older adults, especially women, low-income households and rural communities, largely mirroring EU-wide trends. However, older adults, rural communities and low-income households in Ireland are more digitally excluded than the EU average (Eurofound, 2025). These groups risk being left behind unless targeted inclusion policies are strengthened. Without the means to develop AI skills, socio-economically disadvantaged groups may find themselves marginalised in the job market. Further entrenchment of socio-economic divides will not only affect employment opportunities but may also affect social cohesion.

Survey data indicates that the AI divide is especially pronounced among older people. Younger Irish adults (18–24 years) have been shown to be almost nine times more likely to use AI often or daily compared to the older cohort aged 55–64. A recent study from the London School of Economics and global consulting firm Protiviti challenges assumptions about a widening digital divide tied to age. It found no inherent generational barrier to AI adoption as older workers were not less capable of using AI once they had received appropriate training and support. Moreover, the study found that productivity gains increased across teams with more generational diversity. The report concludes that older workers bring valuable domain knowledge, context and judgement to AI-enabled work, and that excluding them, whether through assumptions or lack of support and training, risks deepening inequities and weakening overall organisational performance (Jolles & Lordan, 2025).

Figure 3.4: Individuals with Above Basic Digital Skills, by Educational Attainment, 2023 (%)



Source: Eurostat [I_DSK2_AB] (as reproduced in Eurofound, 2025).
Note: International Standard Classification of Education 0-2 refers to early childhood to lower secondary education; 3-4 refers to upper secondary to non-tertiary education; 5-8 refers to tertiary education.

Digital for Good: Ireland's Digital Inclusion Roadmap provides a strong national framework for digital inclusion, and a wide range of initiatives already support older people, low-income households and rural communities (Department of Public Expenditure, Infrastructure, Public Service Reform and Digitalisation, 2023). Programmes such as Hi Digital, Age Action's Getting Started and Alone's Digital Champions deliver essential skills training and one-to-one support for older adults, while Connect Age, SICAP and local digital community strategies extend connectivity and resources to rural and disadvantaged areas. However, these broad initiatives need to be complemented by specially tailored programmes that equip digitally excluded groups with AI-related skills, ensuring that emerging technologies enhance rather than widen existing inequalities. Initiatives such as the TU Dublin and ADAPT Centre Age Friendly AI programme are welcome initiatives in that context.

3.6 Transparency & Accountability

Transparency in AI refers to making systems understandable to stakeholders, including what data is used and how decisions are reached, as well as the limitations of the technology. Explainability, which is an extension of transparency, seeks to ensure that information can be communicated in clear terms to users. These principles are fundamental to building public trust as people must be able to understand, at least in broad terms, how AI systems influence decisions with important implications for their lives. Higher levels of transparency and explainability are likely to be required regarding public service decisions made with the assistance of AI in relation to social welfare, health, justice and education. In 2024, the UK introduced the use of Algorithmic Transparency Recording Standard (ATRS) across all government departments (Government Digital Service, 2023). This policy initiative is designed to enhance transparency in the use of algorithmic tools that significantly affect decisions with public implications or that directly engage with the public.

Transparency and explainability are also the foundation of accountability, as they provide people with the means to trace outcomes back to responsible actors, and if necessary to contest decisions. Accountability itself is complicated by the problem of 'many hands' where multiple actors, including designers, engineers and operators, may all contribute to a given outcome, making it difficult to establish liability. For example, in autonomous driving accidents, responsibility could be laid at the door of the manufacturer, the software developer or the driver. The European Commission withdrew the proposed AI Liability Directive in February 2025, following the adoption of the 2024 Revised Product Liability Directive, which extended strict liability rules to include AI systems and software. This revised framework covers harm caused by defective AI products without requiring proof of fault, offering a harmonised EU-level approach. However, it does not address all forms of AI-related harm (e.g. emotional distress, reputational damage), leaving such cases to be governed by national tort law, which continues to operate in parallel.

3.6.1 Black Box Phenomenon

Achieving meaningful transparency remains deeply challenging, if not impossible given the increasing complex nature of generative AI systems. Deep neural networks have been described as 'black boxes' whose internal logic is too complex for even their developers to understand. This is because their decision-making relies on millions or even billions of interconnected parameters and layers of computation, making it nigh on impossible to trace the exact reasoning behind a given decision. Proprietary and commercial concerns further complicate the picture. Companies often withhold details about algorithms, training data and methodology to protect intellectual property, which limits independent scrutiny. The combination of these factors makes it difficult for regulators and the public to monitor for biases, ensure compliance and enforce legal liability, creating a clear governance challenge. Without transparency, AI risks undermining trust in institutions and creating accountability gaps that can weaken democratic legitimacy.

3.7 Privacy & Data Protection

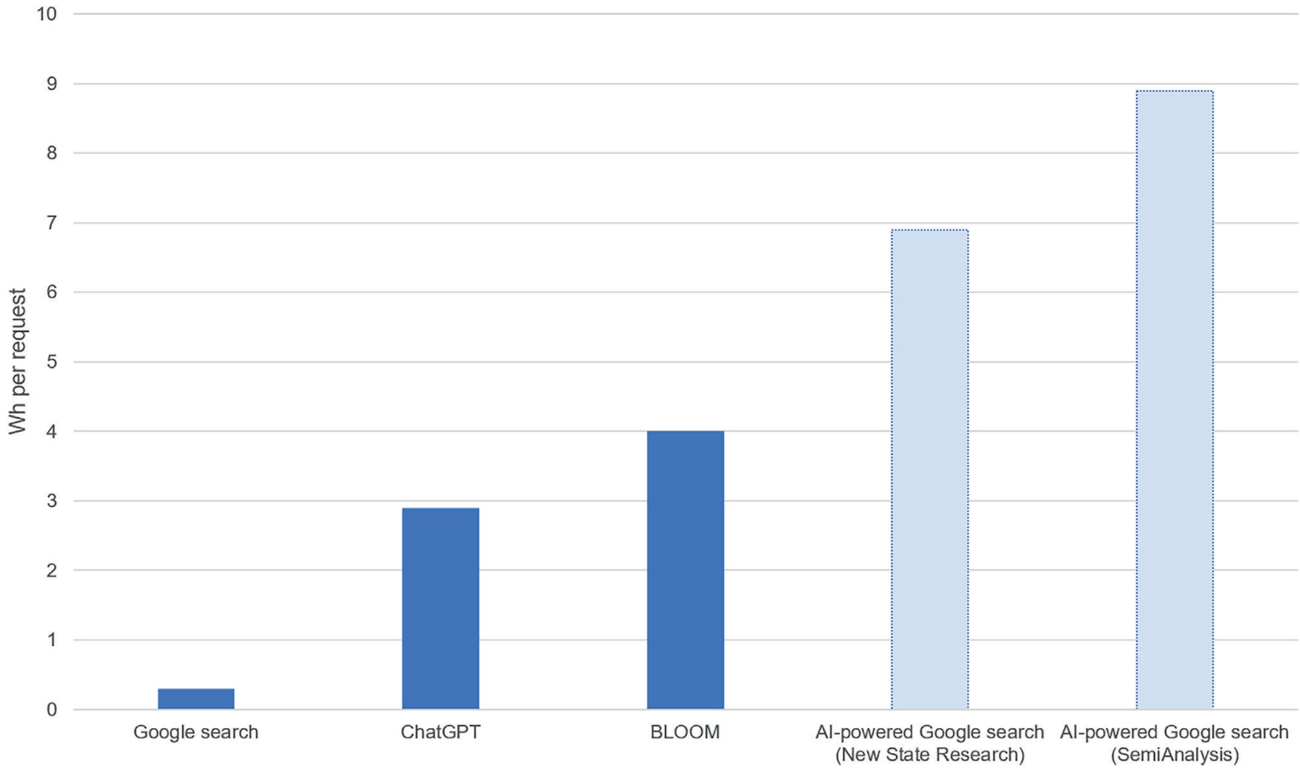
Artificial intelligence systems inherently rely on access to vast amounts of data for training and operational purposes. These datasets can include information collected from publicly available internet sources, which may contain personal or sensitive data. Such data is often gathered indirectly, and in some cases without individuals being explicitly aware of it or without their consent. This raises important concerns around privacy, data protection, consent and personal autonomy. Beyond collection, AI also poses risks of inadvertent data leakage, as seen when chatbots unintentionally reproduce fragments of training data, containing personal information such as phone numbers or medical data, in response to user queries. AI identification and tracking technologies used in public spaces without the explicit knowledge or consent of those being surveilled raise human rights and civil liberty concerns. In a 2022 Global Surveillance index, at least 79 out of 179 countries were actively using AI and big-data technology for public surveillance purposes (no distinction made between legitimate and illegitimate uses of AI surveillance techniques). Slightly more democratic governments than authoritarian regimes have known AI surveillance capabilities (Feldstein, 2022). Furthermore, AI can facilitate powerful prediction and profiling. Data obtained from healthcare wearable devices has been used to infer mental health conditions, while social media activity has been analysed to predict political preferences, often without user awareness.

Addressing the extent of privacy violations is very difficult as harms may occur unintentionally and without the knowledge of the affected individual. Even where data leaks are documented, finding the source is problematic as data may have been handled across multiple devices. Erosion of privacy is an important concern, as privacy is linked to personal autonomy. Privacy allows us control over what others know about us and protects a space for personal development and relationships with others (Rössler, 2005).

3.8 Environmental Impact

The United Nations Environment Programme (UNEP) conceptualises the environmental impacts of AI across three categories: direct, indirect and higher-order effects. Direct impacts arise from the immediate resource use involved in training and operating AI models. A single large-model AI query typically consumes 0.3–2.9 Wh of electricity, compared to approximately 0.1–0.3 Wh for a standard internet search, implying that AI queries may use up to 10 times more energy, depending on model size, hardware and optimisation (de Vries, 2023).

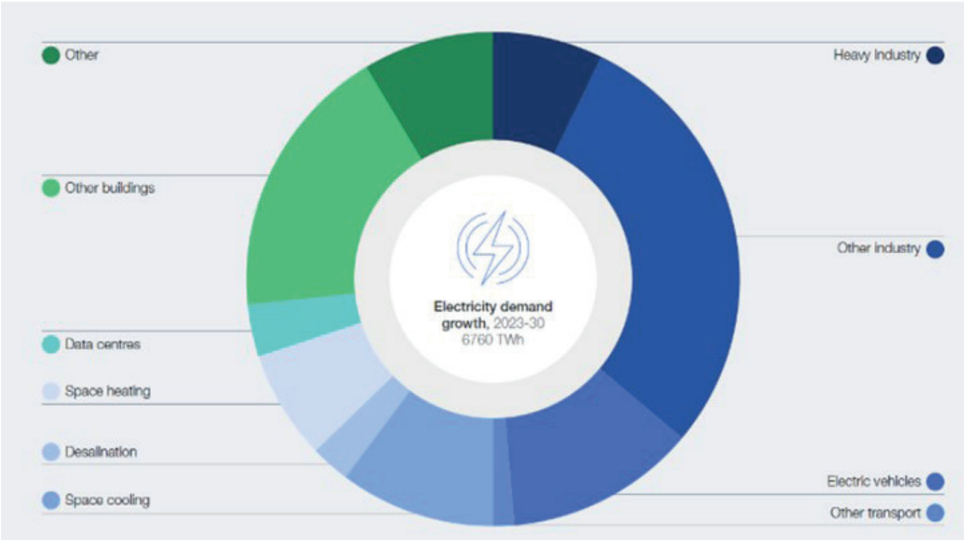
Figure 3.5: Estimated Energy Consumption per Request for Various AI-powered Systems Compared to a Standard Google Search



Source: de Vries, 2023.

Globally, data centres, AI and cryptocurrencies consumed 1.5 per cent of the world's energy in 2024. The International Energy Agency (IEA) (2025a) projects that this figure will double by 2030, which is roughly equivalent to the entire electricity consumption of Japan.

Figure 3.6: Projected Electricity Demand Growth by End Use, 2023–2030



Source: Ginelle Greene-Dewasmes and World Economic Forum, 2025.

This increased energy consumption, often generated from fossil fuels, is contributing to greenhouse-gas (GHG) emissions. Data centres and data transmission are estimated to account for 1 per cent of global energy-related GHG emissions (IEA, 2025a). The percentage share of metered electricity consumption used by data centres in Ireland rose to 22 per cent in 2024 from 5 per cent in 2015 (CSO, 2025a). Contracted demand is anticipated to reach ≥30 per cent of Ireland's supply by 2030 (IEA, 2025b). The National Economic & Social Council (NESC) has similarly noted that the growth of AI workloads in Irish data centres is already exerting pressure on electricity demand and complicating decarbonisation planning (NESC, 2025). The IEA has evaluated Ireland's energy security outlook to 2035 and presents an adapted transition pathway showing how climate, economic and social objectives converge on the electricity system. The analysis highlights the need for a unified, cross-sectoral energy strategy, supported by comprehensive security assessments, to guide this transition effectively. The IEA recommends that growth in data centre electricity demand be managed to support system adequacy, renewable integration and flexibility, including requiring large users such as data centres to contribute generation, storage or flexibility services as part of grid connection conditions and aligning their consumption with renewable supply (IEA 2025b).

These recommendations align closely with requirements set out in the *Large Energy-User Action Plan* (LEAP) published in January 2026, which conditions data centre development on decarbonisation and active grid support. It introduces a plan-led framework that reorients the development of AI infrastructure and data centres around state-identified strategic locations and prioritises location of new facilities in regional areas and Strategic Green Energy Parks, where grid capacity and renewable resources are strongest. Projects are expected to be powered primarily by renewables and to actively support the electricity system through flexible demand and on-site dispatchable generation or storage (Department of Enterprise, Trade and Employment, 2026). The LEAP initiative sits alongside the Commission for Regulation of Utilities

(CRU) (2025a) *Large Energy Users Connections Policy*, published in December 2025. It requires new data centres seeking grid access to provide on-site or proximate generation or storage, and to meet at least 80 per cent of annual electricity demand with additional renewable generation within six years, while taking locational constraints and system security into account. Moreover, under the Public Review 6 grid investment plan, the Government has committed to investing up to 18.9bn in transmission and distribution infrastructure to strengthen the electricity grid, support long-term security of supply, and enable the accelerated connection of renewable generation and large energy users, including data centres (Commission for Regulation of Utilities, 2025b).

It should be said that current projections for future AI and data-centre energy use rely heavily on estimates and extrapolations, and it is widely acknowledged that publicly available information about current patterns in AI energy use is incomplete. Mandatory reporting obligations under the Energy Efficiency Directive, requiring data centres to report on their energy performance, including renewables and water use, should progressively improve transparency and the evidence base for policymaking. The AI Advisory Council (2025b) has recommended that Ireland establish an 'AI Energy Council' to 'ensure necessary measures are taken to rapidly develop clean energy capacity, while transitioning from fossil fuels and winning public trust'.

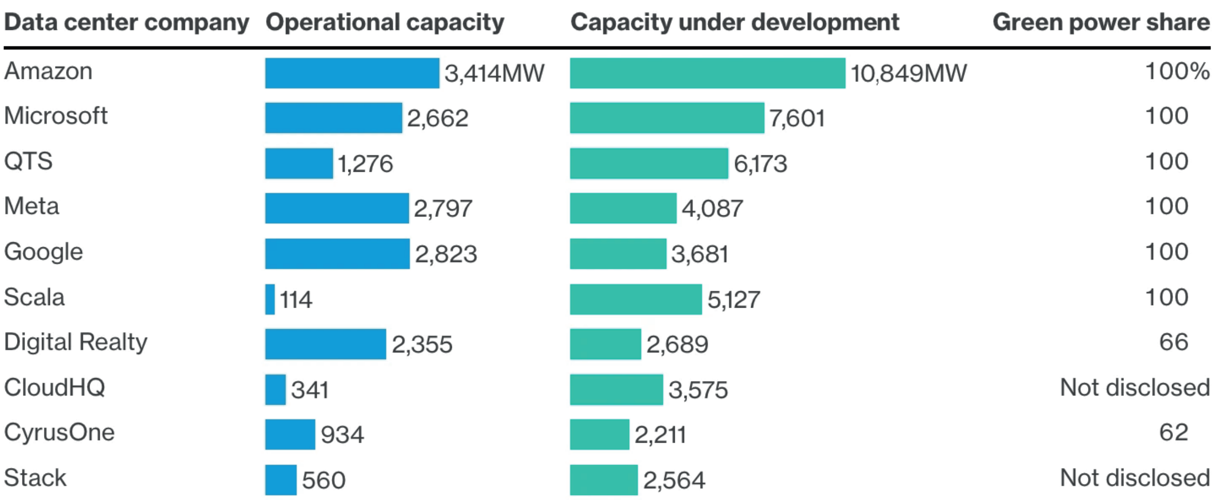
Water consumption is another major direct impact. Global data-centre water use is expected to rise significantly as AI scales, owing to the cooling demands of advanced computation (UNEP, 2024). Furthermore, AI hardware also relies on resource-intensive supply chains. Research on the life-cycle emissions of AI chips shows high embedded carbon costs and increasing quantities of rare earths and metals in successive generations of hardware (Schneider *et al.*, 2025).

Indirect impacts occur when AI-induced efficiencies lead to an overall increase in consumption. Optimisation (e.g. in transport or logistics) may create rebound effects if total system size expands faster than efficiency improves (UNEP, 2024). It also needs to be recognised that AI accelerates demand for cloud infrastructure, land, minerals and energy, with growth trajectories that risk outpacing renewable energy deployment. Shifting AI systems onto renewable electricity alone does not eliminate environmental pressures as renewable generation itself requires land, materials and water.

Higher-order impacts reflect long-term, systemic consequences, such as lock-in to high-consumption technological infrastructures, intensified demand for critical minerals, and pressure on environmental governance.

Mitigation strategies being adopted include the increased adoption of renewable and nuclear energy sources, with Microsoft and Google investing in small modular reactors and geothermal.

Figure 3.7: Green Power Share of Top 10 Data Centre Operators



Source: Lyu and Tang, 2025.

While clean electricity is forecast to meet all global demand growth through 2026 (IEA, 2024), scaling variable renewable energy faces challenges with grid integration, transmission capacity and waste, and will require substantial investment. Efforts are also focussed on making AI systems themselves more efficient through algorithmic and hardware innovations. Combining quantisation, model compression and reducing prompt length can cut AI energy demand by up to 90 per cent without significant performance loss (UCL, 2025). Small language models offer a promising way to mitigate the environmental costs of generative AI. These models are designed with smaller architecture and reducing training data, enabling them to run very efficiently in specific domains and on task-focussed applications, with far lower power demands than LLMs (UNESCO, 2023b). However, as previously mentioned efficiency gains risk being offset by 'rebound effects' from growing AI use.

AI itself can play a positive role in the energy transition by forecasting renewable generation, optimising grid stability and detecting faults in energy networks to improve efficiency (Tuhin, 2025). In fields such as climate modelling, biodiversity monitoring, freshwater management and urban sustainability, AI can enhance predictive accuracy and help integrate data that span multiple spatial and temporal scales. Beyond research, AI can strengthen decision-making by supporting scenario analysis, early-warning systems, and multi-criteria evaluation tools that help policymakers navigate complex trade-offs (Galaz *et al.*, 2025). The concept of a 'twin transition', in which AI is deliberately integrated with clean-energy goals to help accelerate the global shift toward sustainable, low-carbon systems received attention at COP30 in Brazil. At that conference, COP30 countries launched the AI Climate Institute, a global initiative designed to equip governments, researchers and communities, especially in developing countries, with the skills and tools to build locally adapted, low-energy AI solutions for climate mitigation and adaptation. Alongside it, the Green Digital Action Hub was established to provide access to data, expertise and technical support to help nations scale sustainable digital technologies, track emissions and e-waste, and implement low-carbon, socially inclusive digital infrastructure.

3.9 Mitigation of AI Risks

Promotion of safe and ethical AI requires a multifaceted approach that combines technical, regulatory, and organisational strategies, acknowledging that no single solution is a panacea. As with any technology, all risk cannot be eliminated, but well-designed mitigation measures can reduce both the likelihood and severity of harmful outcomes. On the fairness and equity front, strategies to minimise bias remain essential; these include correcting data imbalances, drawing from more diverse and representative data sources, and employing bias detection tools throughout the development lifecycle. However, modifying training datasets as a means to remove bias can be difficult in practice, particularly when biased historical data reflects systemic inequalities and may introduce new distortions. As a result, configuring or constraining the model itself to mitigate biased behaviour is, in many cases, a more practicable approach. Organisational interventions also matter; by challenging assumptions and incorporating different lived experiences, diverse development teams can design systems that better account for the needs of a global user base. Nonetheless, complete elimination of bias is not currently possible and may even be theoretically unachievable, underscoring the need for continuous monitoring and iterative refinement.

Transparency and accountability can be strengthened through the use of explainable AI (XAI) techniques, such as local interpretable model-agnostic explanations (LIMEs), which perturb input data to illustrate how changes affect predictions. These tools can provide valuable insights into model behaviour, particularly for high-stakes decisions. However, current XAI techniques have significant limitations and cannot offer full visibility into complex deep-learning architectures. This limitation reinforces the importance of robust governance structures that do not rely solely on explainability tools.

One such governance mechanism is independent auditing. External audits covering system design, training data, evaluation methods and real-world performance can help identify risks that internal teams may overlook and provide public assurance that systems meet safety and ethical expectations. Governments are increasingly supporting this approach through emerging regulatory frameworks. For example, the *EU AI Act* introduces mandatory conformity assessments and post-market monitoring for high-risk systems, while the United States and United Kingdom have issued guidance encouraging third-party evaluations, transparency reporting and risk assessments. These frameworks help establish common expectations and provide a structured basis for organisations to evaluate and mitigate risks.

Protecting privacy and ensuring responsible data use is another essential component of mitigation. Technical safeguards such as differential privacy, which introduces statistical 'noise' to obscure individual identities, and federated learning which allows models to be trained on decentralised devices without transferring raw data, can help to reduce the exposure of personal information. These tools can then be complemented by clear and enforceable data governance frameworks that outline requirements for consent, data retention, data sharing and secondary use. Increasingly, governments are developing or updating privacy regulations to address AI-specific risks, including rules on automated decision-making and dataset documentation, and restrictions on sensitive data processing.

Together, these technical measures, organisational practices and government-backed regulatory frameworks form a layered mitigation strategy that can meaningfully reduce the risks of AI systems while supporting innovation and public trust.

Chapter 4: AI through a Socio-Technical Lens

4.1 Introduction

This chapter explores the integration of artificial intelligence into society, balancing its transformative potential with the safeguards needed to ensure responsible and equitable deployment. It challenges the idea that complex social problems can be addressed through reliance on technology alone, highlighting the risks of AI solutionism and emphasising the importance of a socio-technical perspective that situates AI within the social, cultural and institutional contexts in which it operates. Through this lens, the chapter examines key dimensions of AI's societal impact, including patterns of adoption, public attitudes and trust, workforce implications and economic gains.

4.2 Techno-solutionism

Artificial intelligence holds enormous potential to transform multiple dimensions of our lives, from medicine and education to agriculture, transportation, energy and beyond. Deployed with care, it can enhance efficiency, augment human decision-making and support large-scale innovation in public services and private enterprise. However, AI will work better in some domains than in others and the specific conditions of success or failure are often deeply contextual. It should be kept in mind that AI systems are not deployed in a vacuum; their success will depend upon their interaction with existing systems and environments. As we integrate AI into more areas of society, we must thoughtfully consider where its use is most appropriate, and guard against the seductive but problematic logic of technological solutionism. Morozov (2013) has pointed to the folly of thinking that complex societal problems can be solved through technological fixes alone. Techno-solutionism treats technology as an easy button, reducing deep-seated societal issues into simplistic, quantifiable problems to be engineered away (Morozov, 2013). Within artificial intelligence, this manifests as AI solutionism; the assumption that AI systems are ideologically neutral tools capable of solving wide-ranging issues such as welfare provision, climate adaptation or public health management. While AI clearly has a role to play in all of these areas, a mindset of techno-solutionism can encourage oversimplification by concentrating on symptoms rather than root causes and privileging optimisation over understanding. This presents two distinct problems; at a macro-level, it prevents us from seizing the opportunity to re-image systems; at the micro-level, it impedes our ability to choose the right problem and the right tool for AI to solve.

4.3 Socio-technical Thinking

Adopting a socio-technical approach provides an antidote to AI solutionism. A socio-technical lens recognises that AI systems are built, deployed and used within complex social, cultural, legal, and political contexts (Sartori & Theodorou, 2022). It requires us to consider both the technical artefacts and the social practices that shape and are shaped by AI.

A socio-technical approach involves integrating expertise from several fields, such as ethics, sociology and user-experience research, throughout the AI life cycle, from design to testing to deployment and monitoring. Such an inter-disciplinary perspective can contribute to building AI systems that are responsible, fair and aligned with broader societal values. It recognises that technologies are not neutral, but rather embody the values, assumptions and power dynamics of those who build and implement them. As the International Organization for Standardization (2025, p11) has noted, 'At their core, AI systems are socio-technical in nature: they do not operate in isolation, but interact continuously with people, institutions and processes, cultural norms, other technologies, and broader social, economic and political contexts.' A socio-technical approach can ensure that AI augments rather than substitutes essential social processes such as deliberation, professional judgement and community participation.

Governments have traditionally emphasised innovation and competitiveness as key objectives of national AI strategies. While these are clearly important goals, a narrow focus on economic productivity can obscure the broader societal impacts of AI in areas such as privacy, fairness, accountability and democratic control. Ireland's *Wellbeing Framework* and models such as doughnut economics reinforce the need to balance economic considerations with wider social and environmental outcomes, so that technological progress can remain within ecological limits and contribute to quality of life. A socio-technical framing can help recalibrate this balance. Crucially, this lens also enables a more holistic view of public benefit. As UNESCO affirms in its *Recommendation on the Ethics of Artificial Intelligence* (2021), the goal should be to align AI with the principles of human dignity, inclusion and environmental sustainability. This perspective does not reject competitiveness or innovation but embeds them within a richer matrix of public values.

Through a socio-technical lens, AI can be seen as a powerful accelerator of the attention economy by maximising user engagement through highly personalised and algorithmically optimised content delivery. While effective in capturing attention, this dynamic raises concerns about potential impacts on user autonomy, privacy and, in broader terms, the quality of public discourse. As AI systems increasingly shape online behaviour, their design and deployment carry broader societal implications, highlighting the need for governance that ensures transparency, accountability and alignment with the public interest.

4.4 Value Alignment

Value alignment seeks to guarantee that AI systems operate in a way that is consistent with human interests and values (World Economic Forum, 2024a). Failure to achieve value alignment poses significant consequences for human rights, the rule of law and democratic governance. Recent research suggests that AI models can inherit the values and behaviours of the systems that train them, raising the prospect that if the 'teacher' model or training process is misaligned, the downstream 'pupil' models will be too, allowing undesirable values and behaviours to propagate unless alignment is addressed at every stage (Cloud *et al.*, 2025). Frontier AI systems have their operational values and decision-making frameworks encoded by the companies that build them. For example, Anthropic has developed a constitution that explicitly specifies normative principles, ethical constraints and behavioural guidelines intended to shape how their model Claude reasons, responds and aligns with human values (Anthropic, 2026). While the attention to values is welcome, the fact that they are being formulated by private actors rather than through broad societal deliberation makes it hard to assess how well they reflect the diverse public values of the populations who will rely on services being delivered through AI.

When discussing value alignment in AI, a central challenge is deciding which values to privilege, since these may differ between individuals and cultures. For this reason, the determination of which values to uphold should be the subject of public deliberation. This can help establish 'red lines' representing non-negotiable ethical limits which AI systems should not cross. It is important to distinguish between public information, which focuses on one-way communication, and public deliberation, which requires listening to diverse voices, including those that are critical and challenging.

Complexity is heightened by the fact that values can conflict with one another, requiring difficult trade-offs and careful balancing. Moreover, the salience of particular values often shifts depending on the context, meaning that value alignment is not a one-time task but requires ongoing reflection. Value alignment in AI means retaining human oversight, control and accountability, and mindfully and deliberately designing, deploying and maintaining oversight of these systems so that their societal and ethical impacts serve the public good rather than erode it. The *Special Eurobarometer 566 report on The Digital Decade*, commissioned by the European Commission and conducted between February and March 2025, reported that 93 per cent of Irish respondents (EU average 86%) considered it important for public authorities to shape the development of AI and other digital technologies to ensure they respect our rights and values (European Commission, 2025c).

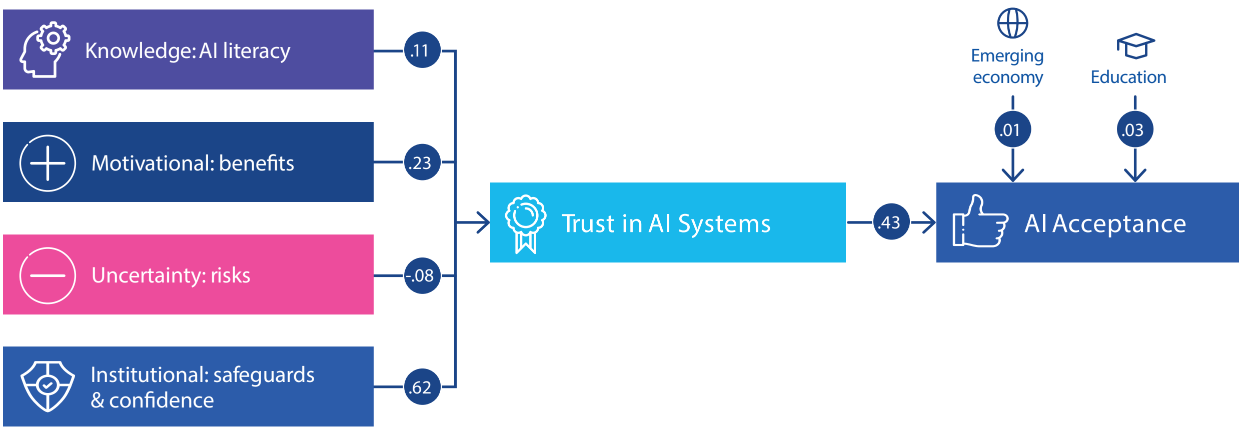
4.5 Public Attitudes

Public attitudes towards AI play a critical role in determining the legitimacy, adoption and societal alignment of AI systems. Recent global studies show that, while the public recognises AI's potential benefits, there is concern about its risks, especially where transparency, fairness and oversight are lacking. The *IPSOS AI monitor* (Carmichael, 2025) surveyed over 23,000 adults across 30 countries between March and April 2025. The study revealed that 52 per cent of global respondents felt optimistic about AI's impact, while 53 per cent reported feeling worried. Irish respondents, however, expressed lower optimism (41%) and higher worry (64%), indicating a more cautious stance than the global average.

This cautious attitude is echoed in the Data Protection Commissioner's *Public Attitudes Survey* (2025) in which 61 per cent of those surveyed reported being quite/very concerned about the use of AI and how it is applied. Further research capturing the views of over 48,000 people in 47 countries found that 42 per cent of people believe the benefits of AI outweigh the risks, compared to 32 per cent who believe the risks outweigh the benefits, and 26 per cent who believe the risk and benefits of the technology are balanced. Of the Irish participants in the study, 33 per cent were of the view that the benefits of AI outweigh the risks, with the top risk (67%) identified as a loss of human interaction and connection due to AI (Gillespie *et al.*, 2025). In another global survey involving over 32,000 people in 40 countries, Irish respondents reported spread of false information, fear of jobs losses and personal data breaches as key concerns in relation to AI (Worldwide Independent Network of Market Research, 2025).

Despite this caution, 72 per cent of the Irish public believe that the use of AI will result in a wide range of benefits, the most cited being improved efficiency and a reduction in repetitive tasks. Importantly, 60 per cent of Irish people are personally experiencing or observing these benefits (Gillespie *et al.*, 2025). In the same study, it was found that people who expect and experience or observe benefits from AI are more likely to trust and use AI. This highlights the importance of designing and deploying AI systems which can deliver a wide range of benefits across the population. Other key drivers of trust in AI included AI literacy, the presence of safeguards and confidence that AI would be used in the best interests of the public.

Figure 4.1: A Model of the Key Drivers of Trust and Acceptance of AI Use in Society



Source: Gillespie *et al.*, 2025.

Public trust, which depends upon AI trustworthiness, is essential; without confidence in the technology, the public will not adopt AI systems, thereby undermining their legitimacy, and by extension their ability to deliver meaningful public benefit. The public sector is subject to greater scrutiny and accountability than the private sector in relation to legitimacy, fairness and equality.[4] Higher levels of transparency and explainability are likely to be required regarding public service decisions made with the assistance of AI in relation to social welfare, health, justice and education. Indeed, one of the three pillars of policy development in Ireland in the public sector is legitimacy, where buy-in, or at least acceptance, by the people who will be affected by the policy is considered essential (Department of the Taoiseach, 2025a).

Trust in AI remains a critical challenge. Ireland does not compare favourably with other countries in respect of this metric (Gillespie *et al.*, 2025; Worldwide Independent Network of Market Research, 2025; Carmichael, 2025). Only 38 per cent of Irish respondents, as compared to a 47-country average of 46 per cent, are willing to trust AI systems (Gillespie *et al.*, 2025). Trust is highest in universities, research and healthcare institutions, while just under half of Irish people asked have confidence in the Government to develop and use AI in the public's best interest.

4 The Public Sector Equality and Human Rights Duty places a statutory obligation on public bodies to have regard to human rights and equality considerations in the performance of their functions.

Ireland is one of only three countries (along with Italy and Singapore) across 30 countries surveyed which trust people more than AI systems not to discriminate or show bias towards any group of people. Globally, younger people (18–34 years), high-income households, people with a university education and those with AI-related training were more accepting and trusting of AI (Gillespie *et al.*, 2025).

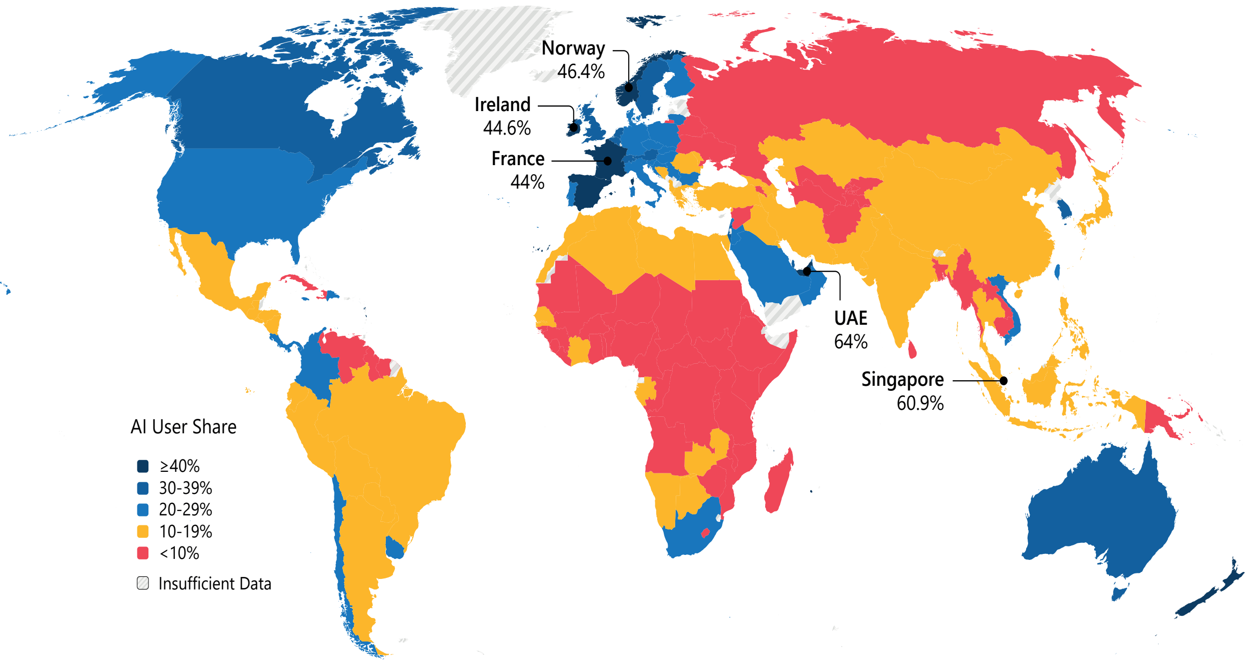
In summary, Irish public opinion reflects a measured and discerning view of AI, open to its benefits but more sceptical and privacy-conscious than global peers. These attitudes underscore the importance of building trustworthy, rights-respecting AI systems with strong human oversight, transparent design and meaningful public engagement at their core.

4.6 Adoption of AI

The adoption of AI does not hinge on technology alone but on a complex interplay of interdependent conditions. Effective integration requires robust digital infrastructure and well curated, interoperable data, yet these technical foundations must be matched by organisational capacity, workforce skills and positive attitudes toward innovation. Equally important are governance structures that ensure transparency, accountability and ethical use, as well as the broader economic and policy environment that shapes investment, incentives and readiness to change. Taken together, these factors form an ecosystem in which deficiencies in any one area can limit the overall impact of AI, underscoring the need for a holistic, system-wide approach to realising its potential.

Adoption of AI is advancing, albeit unevenly, across geographies and sectors. Data from the Microsoft AI Economy Institute (2026) show that, in 2025, countries such as the UAE, Singapore, Norway, France and Ireland were among the fastest adopters of generative AI. In Ireland, the share of the working-age population using generative AI tools increased by 2.9 per cent, reaching 44.6 per cent by the end of the year. These rankings are based on estimates derived from observed AI usage data; while they provide a useful proxy for adoption, the authors note that they cannot capture all forms of AI use, particularly informal or enterprise-internal deployment.

Figure 4.2: AI Diffusion by Economy, Second Half 2025

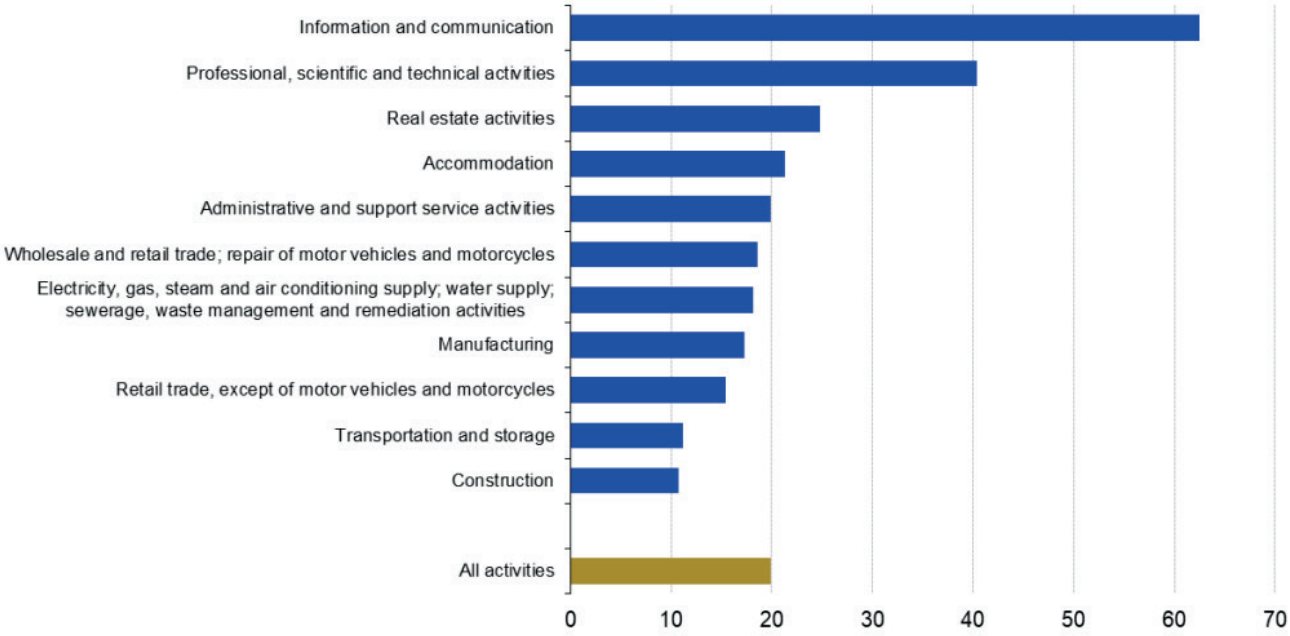


Source: Microsoft AI Economy Institute, 2026.

A 2024 EU survey on use of AI technologies found that, within the EU, Denmark, Sweden and Belgium lead, with approximately a quarter of enterprises reporting AI adoption in 2024, as compared to an EU27 average of 13.5 per cent. In Ireland the percentage of enterprises using AI technologies increased from 8 per cent in 2023 to 14.9 per cent in 2024 (Eurostat, 2025). According to the CSO, 51.2 per cent of large enterprises in Ireland used AI technology in 2024, compared with 25.1 per cent of medium and 12 per cent of small enterprises (CSO, 2025b). This largely reflects international findings that larger, more productive firms are more likely to adopt AI (OECD, 2023b). Survey results from the OECD show generative AI usage of 33 per cent among Irish SMEs, placing Ireland third among the surveyed countries, just behind Germany (38.7%) and Austria (34.1%) (Expert Group on Future Skills Needs, 2025).

Early adoption of AI is most evident in knowledge-intensive services such as finance and insurance, ICT, legal and consulting, while sectors such as hospitality, construction and transportation show low AI intensity (OECD/BCG/INSEAD, 2025).

Figure 4.3: Percentage of Enterprises Using AI technologies by Economic Activity, EU, 2025



Source: Eurostat (online data code: isoc_eb_ain2).

The OECD has found that AI adoption in government trails behind that of the private sector (OECD 2025a). In a survey of senior leaders in 250 organisations across Ireland, public-sector organisations reported an AI adoption rate of 50 per cent, compared to 63 per cent in multinational organisations (Kumar Jha & Danks, 2025). Despite the lower rate of adoption, approximately half of the reported AI use cases in G7 countries aimed to increase the efficiency of internal public-sector operations (OECD/UNESCO, 2024). In a mapping exercise conducted in 2020, the European Commission (Misuraca & van Noordt, 2020) found that a majority of EU member and associated states were already using AI across a variety of government functions. Applications ranged from automating administrative processes to delivering citizen-facing services and supporting complex policymaking. A more recent report by the OECD found that government use of AI is most common in public services, justice and civic participation, and less prevalent in policymaking and highly regulated areas such as tax. Of note is that applications aim to streamline services, with much less focus on creating new opportunities (OECD, 2025a).

4.6.1 Barriers to AI Adoption

The most commonly cited barriers to AI adoption include limited digital and data readiness, high implementation costs and uncertainty over both returns on investment and the practical application of AI to specific challenges. Organisations often struggle with integrating data across systems, developing new business models, and managing organisational change (Sternfels & Atsmon, 2025).

Successful AI adoption is not a standalone achievement but a capability that sits on top of deep digital foundations. Rushing to deploy AI tools without first securing high-speed connectivity and curated, governable data is a strategic error that risks failure. This can both undermine trust and be expensive, depending on the AI technology being deployed. The Government's *Harnessing Digital – The Digital Ireland Framework* explicitly recognises this dependency, positioning 'Digital Infrastructure' (Dimension 2) as a prerequisite for advanced technology adoption (Department of the Taoiseach 2022). To prevent the 'rush to AI' from outpacing its rails, the framework mandates that the physical backbone be ready first, setting strict targets of Gigabit network coverage for all households and businesses by 2028 and 5G coverage for all populated areas by 2030.

Furthermore, the *National Digital and AI Strategy 2030* emphasises that digital infrastructure extends beyond connectivity and computational capacity to encompass high-quality, standardised and well-governed data, positioning data integrity and interoperability as foundational enablers of secure, trusted and responsible AI deployment. In that context, the fact that Ireland lags behind our European counterparts in terms of health system digitisation is concerning. While the *Digital for Care: A Digital Health Framework for Ireland 2024–2030* (Department of Health, 2024) sets out an ambitious roadmap, accelerated progress in this domain will be required if Ireland is to realise the full potential of AI to improve patients' outcomes, safety and efficiency. That said, Ireland is currently laying the groundwork for the adoption of AI at scale in healthcare. In 2024 the Health Service Executive (HSE) established an Artificial Intelligence and Automation Centre of Excellence to ensure that AI could be effectively integrated across the Irish health service. In March 2026, the Department of Health published the *'AI for Care' strategy*, to guide the responsible adoption of AI across the health and social care system between 2026 and 2030 (Department of Health, 2026). The strategy aims to improve patient outcomes, support clinicians and healthcare staff, and increase system efficiency. It outlines the use of AI across clinical care, healthcare operations, research and innovation, and public health. In addition, the Health Information and Quality Authority (HIQA) is currently developing guidelines for the use of AI in health and social care.

Although developments in AI technology have been remarkably rapid, it is important to make a distinction between technological breakthroughs and their practical application. A gap exists between innovation and widespread diffusion; adoption is proving much slower, particularly in safety-critical domains where the regulatory burden is high. Current data indicate that few organisations can yet be considered AI-mature; many are still building the necessary foundations to scale up from pilots to system-wide transformation. Organisational structures, professional practices and individual habits take time to adjust, and effective use of AI will require new skills, workflows and cultural acceptance. If AI does follow the trajectory of previous general-purpose technologies, adoption is likely to unfold over decades rather than years (Narayanan & Kapoor, 2025).

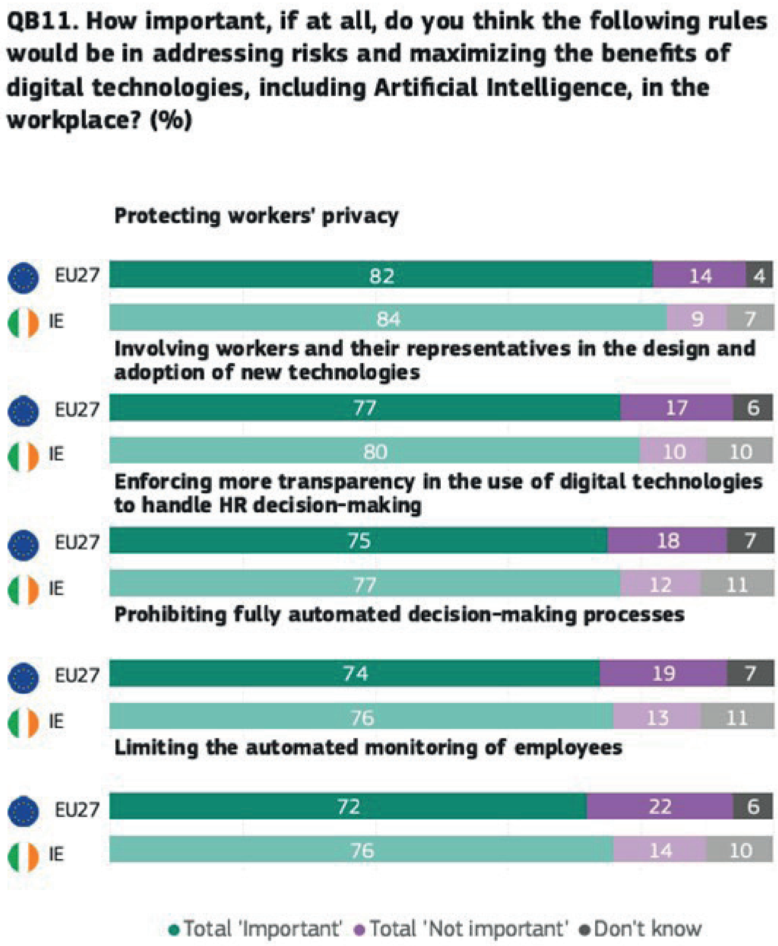
4.6.2 Worker Sentiment

Another key factor in AI adoption is securing workers' trust and engagement. At EU level, workers already experience AI and related digital technologies as reshaping work, but with mixed social consequences. In a *Special Eurobarometer on AI and the Future of Work*, 66 per cent of EU27 respondents said that recent digital technologies, including AI, had a positive impact on their current job (European Commission, 2025d). The corresponding figure for Ireland is also 66 per cent, but Irish respondents were somewhat less negative about these technologies' impact on their job (16% negative in Ireland versus 21% in the EU overall). Interestingly, this positive orientation also held when asked about the impact of AI on the economy, quality of life and, to a lesser extent, society. In the workplace context, AI was viewed positively in terms of improving workers' safety but viewed more negatively when it came to assessing workers' performance. A majority of those surveyed also agreed that, due to robots and AI, more jobs will disappear than new ones will be created (66% EU27 vs 72% Ireland). Thus, workers already interpret AI through a risk/benefit frame; they simultaneously see efficiency gains and potential job losses.

When asked if employers had informed workers about the use of digital technologies, including AI, to manage activities in the workplace, 20 per cent of Irish employees reported having received a detailed explanation (EU27 18%), while a further 23 per cent had been made aware of the use of these technologies but without further details (EU27 16%). There is support for clear rules on the use of digital technologies; for instance, protecting workers' privacy (82%) and involving workers and their representatives in the design and adoption of new technologies (77%). Irish and EU workers alike emphasise the need for strong rules that protect rights and keep workers in the loop in respect of adoption of digital tools, including AI. Of the Irish respondents, 84 per cent rated protecting workers' privacy as important in addressing risks and maximising the benefits of digital technologies, including AI, in the workplace, while 80 per cent said involving workers and their representatives in the design and adoption of new technologies was important.

Engaging employees early on and on an ongoing basis is crucial, as it leverages their tacit knowledge to align algorithmic solutions with actual operational workflows. When workers are excluded, systems often fail to address the nuance of daily tasks, leading to resistance and implementation gaps. This dynamic is well-documented in the German automotive industry, where a failure to consult assembly-line workers initially resulted in systemic inefficiencies; however, once the manufacturer integrated worker feedback into the redesign, the company achieved smoother workflows and higher output (Cotton, 2024). Consequently, policy frameworks must prioritise early employee involvement to ensure that AI tools are not merely deployed but effectively assimilated to drive genuine productivity.

Figure 4.4: Rules around Digital Technologies in the Workplace

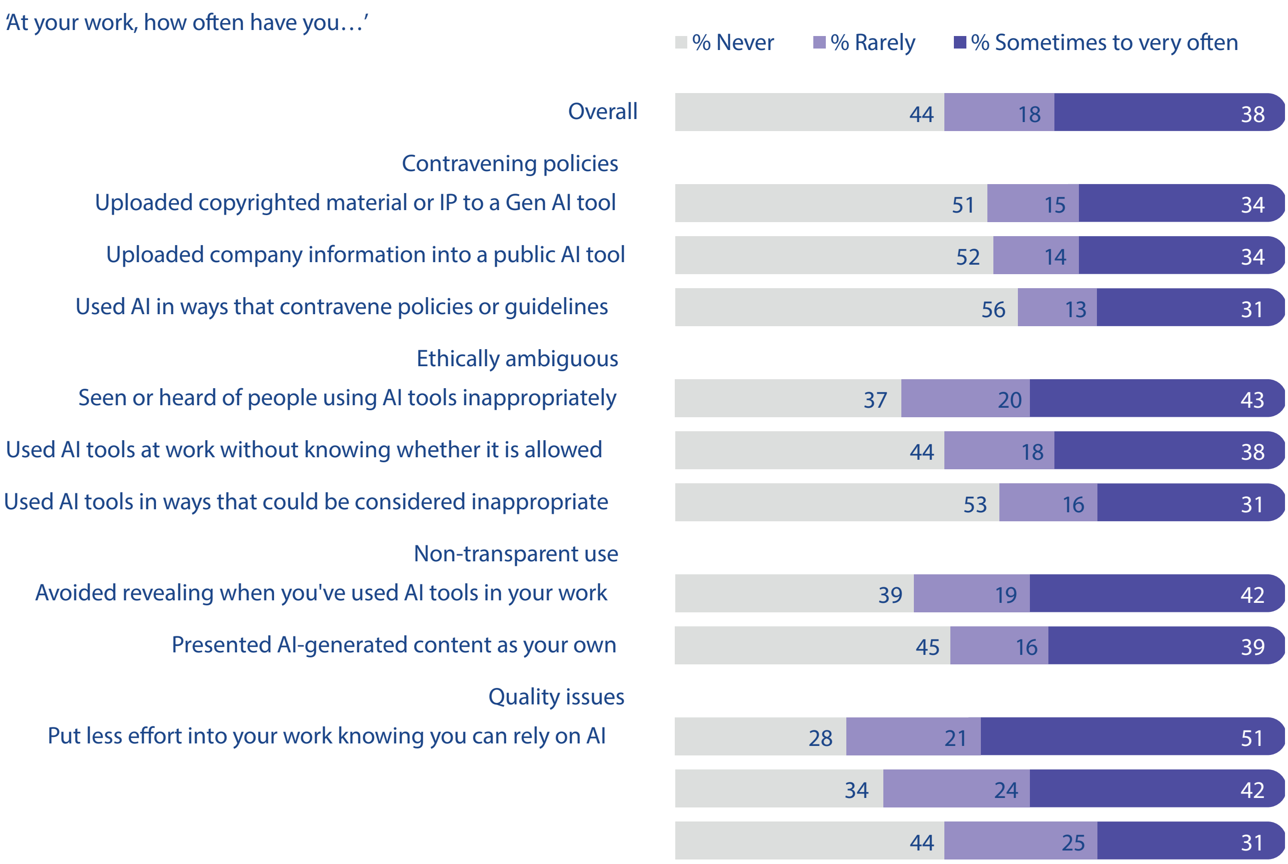


Source: European Commission, 2025c.

4.6.3 Shadow AI

Shadow AI refers to employees using AI-powered tools or platforms without the awareness or approval of the organisation's IT or security functions. Shadow AI represents a fundamentally socio-technical challenge: a confluence of workforce behaviour, rapid tool-adoption, organisational workflow pressure and lagging governance. Quantifying shadow AI precisely is challenging because many instances remain hidden and unreported. However, in a 2025 global study on attitudes and use of AI, 44 per cent of employees reported having used AI in ways which contravene policies and guidelines, indicating a significant prevalence of shadow AI in organisations (Gillespie *et al.*, 2025). A shadow AI culture has also been identified in Ireland; 61 per cent of managers in organisations which prohibit free AI tools reported knowing that their employees still used them (Kumar Jha & Danks, 2025).

Figure 4.5: Inappropriate and Complacent Use of AI at Work (%)



Source: Gillespie *et al.*, 2025.

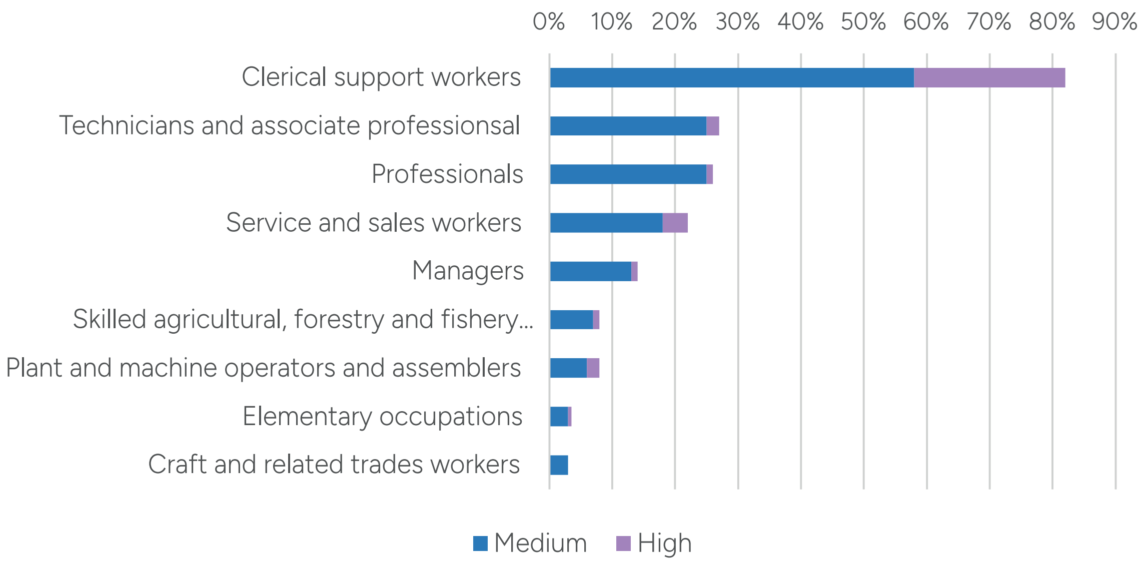
The use of unsanctioned AI tools introduces multiple risks. As these tools may handle sensitive or proprietary data outside formal controls, organisations face heightened exposure to data leakage, intellectual property loss, flawed decision-making and regulatory non-compliance. In that context, organisations need to focus on practical governance by developing clear, accessible policies that define approved AI usage, data-sharing limits and escalation processes. Equally important is training and awareness as employees will need practical guidance on what constitutes safe use of AI, how to evaluate outputs, what data may be shared (and what must not), and why the governance matters. In May 2025, the Department of Public Expenditure, Infrastructure, Public Service Reform and Digitisation (2025) published *Guidelines for the Responsible Use of Artificial Intelligence in the Public Service* and a tool for use of AI in public services. The guidelines contain a range of resources designed to support the adoption of fair, inclusive, accessible and trustworthy AI. Online learning modules for the guidelines and an Introduction to AI have also been developed by the Institute of Public Administration.

4.7 Labour Market Impacts

Viewed through a socio-technical lens, the rise of AI represents not just a technological shift but a profound reconfiguration of work itself through reshaping labour markets and redistributing skills and responsibilities. The labour market effects of AI are complex, as the technology is capable of both substituting and complementing human work (Pizzinelli *et al.*, 2023). Where complementarity dominates, workers stand to benefit by focussing on higher-value, creative or interpersonal activities, amplifying both job quality and output (Brynjolfsson *et al.*, 2025a). Where substitution dominates, workers face the risk of displacement, deskilling and unemployment (Chen *et al.*, 2025; Acemoglu & Restrepo, 2019:5; Filippucci *et al.*, 2024). International evidence indicates that the principal near-term labour market impact of AI adoption will be to reallocate tasks within jobs, rather than to eliminate whole occupations.

Task displacement is most likely in occupations where a large share of work consists of information processing, routine drafting, summarisation and standardised interaction – activities that are already executable by AI systems at acceptable quality and reliability (Lane & Saint-Martin, 2021). The greatest exposure lies in clerical, telephony, sales support and administrative roles, where routine cognitive tasks are easily automated (Gathmann, Grimm & Winkler, 2024). Professional roles such as accountancy, legal services and software development contain a mix of automatable and non-automatable tasks, with outcomes depending on how organisations redesign work (Gmyrek *et al.*, 2025). Where processes are redesigned to prioritise oversight, client engagement and cross-disciplinary collaboration, overall job levels may be maintained even as routine entry-level tasks decline. By contrast, if AI adoption leads firms to streamline staffing structures toward fewer but more senior roles, net employment losses are likely to materialise despite stable levels of output (Filippucci *et al.*, 2024).

Figure 4.6: Share (%) of High and Medium Exposure in All Tasks by Occupational Category



Source: Pizzinelli *et al.*, 2023.

Evidence suggests that Ireland is marginally more exposed to AI-related labour displacement risks than the advanced economy average, with uneven adjustment costs likely across regions, sectors and demographic groups (DoF & DETE, 2024; Pizzinelli *et al.*, 2023; Filippucci *et al.*, 2024). A joint study by the Departments of Finance and of Enterprise, Trade and Employment estimates that 63 per cent of Irish employment lies in highly AI-exposed occupations, compared with an advanced-economy benchmark of approximately 60 per cent (Department of Finance & Department of Enterprise, Trade and Employment , 2024). Exposure is polarised, with a significant share of workers in high-exposure, low-complementarity roles such as administrative and support functions (facing greater displacement risks), while others in high-exposure, high-complementarity roles have greater potential for augmentation. Women are disproportionately represented in the higher-risk cohort, reflecting a larger share of female workers in administrative roles.

More recent analysis from the Department of Finance (DoF, 2026) suggests significantly weaker employment growth over the past two years in AI-exposed sectors as compared to sectors with lower relative exposure. This trend is more pronounced for younger workers. Employment among 15–29-year-olds in AI-exposed sectors fell between 2023 and 2025, despite overall growth in those sectors. In contrast, in lower AI-exposed sectors, youth employment continued to grow faster than among older cohorts (DoF, 2026). This impact on early-career workers is also seen internationally. Between late 2022 and mid-2025, employment among workers aged 22–25 in the most AI-exposed occupations declined by 13 per cent relative to peers in less exposed fields, even after controlling for firm-level shocks (Brynjolfsson *et al.*, 2025a). By contrast, employment for more experienced workers in the same occupations has remained stable or continued to grow. As noted by NESC (2024), the future impacts of AI on the Irish labour market remain uncertain; continued monitoring and research will be required to assess how these dynamics evolve.

The World Economic Forum (2026a) takes a scenario-based approach to examine how AI might reshape the labour market by 2030. Drawing on expert consultation and economic data, the study explores four distinct futures based on two key uncertainties: the pace of AI advancement and the level of workforce readiness. In *Scenario 1: Supercharged Progress*, exponential AI development combined with widespread skills training leads to major productivity gains and a reimagined workforce, where humans manage intelligent machines. *Scenario 2: The Age of Displacement* envisions a future where rapid AI outpaces readiness, resulting in job losses, erosion of consumer confidence, and societal instability. In *Scenario 3: Co-Pilot Economy*, incremental AI progress and strong workforce preparation foster human-AI collaboration (as distinct from automation), enabling gradual transformation of industries. Finally, *Scenario 4: Stalled Progress* presents a world where both AI development and workforce skills lag, producing uneven productivity gains and a fragmented labour market, thus fuelling inequality. The report underscores that the trajectory of future jobs depends not only on technological breakthroughs but also on coordinated investment in human capital. A 'no regret' strategy of investing in human-AI collaboration and aligning technology with talent strategies is recommended as it would provide value, whichever scenario eventually unfolds.

Education and training, re-skilling initiatives and social safety nets will need to evolve and adapt to the potential disruptive effects of AI in the labour market. This should be underpinned by social dialogue, collective bargaining, updated social protection policies and pro-active state interventions to direct labour to where it is needed in areas of the economy that are less suitable for automation.

Labour displacement on a significant scale has implications for the public finances. A decline in the labour share of income could erode payroll tax revenues, weakening the funding base for social protection systems that rely on stable employment. A key structural issue is whether AI-enabled, capital-intensive production and remote service delivery will erode the labour tax base over time. The International Monetary Fund (IMF) cautions that generative AI could shift the labour-capital income split and recommends modernising fiscal systems to account for such structural changes by strengthening social protection, adjusting capital taxation relative to labour, and avoiding narrow 'AI taxes' in favour of principle-based frameworks that preserve neutrality across technologies while mitigating concentration and inequality risks (Cazzaniga *et al.*, 2024; Brollo *et al.*, 2024).

The OECD's work on taxation and the future of work underscores how differences in tax treatment across different forms of employment creates arbitrage risks and threatens the integrity of labour-based revenues (Milanez & Bratta, 2019). In an AI-intensive economy characterised by more platform work, telemigration and cross-border services trade, tax policy will need to maintain horizontal equity across employment statuses and ensure contribution adequacy for social insurance.

Considerations around the composition and sustainability of the tax base will increasingly intersect with AI-driven changes in the labour share, profit location and the form of work. This strengthens the case for medium-term fiscal planning that anticipates slower growth of labour-tax receipts relative to capital and corporate income in high-adoption scenarios, while safeguarding incentives for productive investment (Brollo *et al.*, 2024).

Box 4.1: AI in Agriculture

Agriculture is facing mounting pressures from climate change, growing global food demand, rising input costs and declining natural resources. AI is emerging as a key tool to help farmers produce more with less by improving efficiency, sustainability and resilience. Precision farming is one of the most transformative applications (Aijaz *et al.*, 2025). AI-powered sensors, drones and computer-vision systems monitor soil health, moisture and crop conditions in real time, enabling targeted fertilisation, irrigation and pest control (Dalal & Mittal, 2025). These technologies can reduce chemical use and environmental impact while improving yields (Anastasiou *et al.*, 2023). AI-driven forecasting tools also help farmers plan planting and harvesting by analysing weather patterns, soil conditions and historical crop performance (Goel & Pandey, 2024).

Labour shortages are accelerating interest in robotics, from autonomous tractors and robotic weeders to automated milking systems, an area where Ireland is an early leader (ESOFT, 2024). AI-enabled sorting and grading technologies, such as Ireland's first AI-powered shellfish grader, enhance product quality and reduce waste (McCann, 2025).

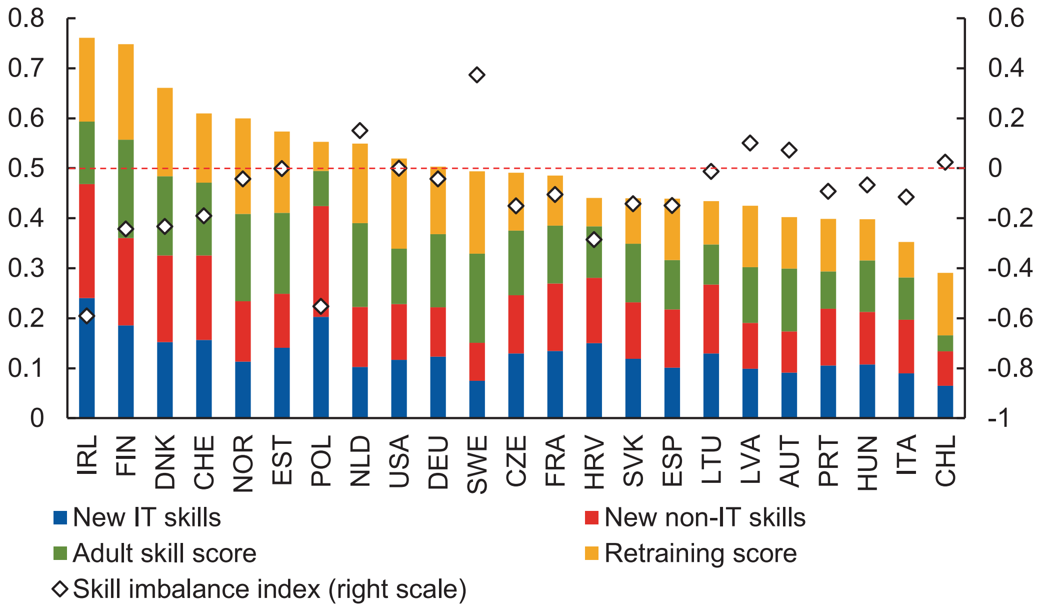
High upfront costs, particularly for small farms, fragmented agricultural data and limited rural broadband connectivity remain substantial barriers to adoption of AI in agriculture (Thomasson *et al.*, 2025). Skills shortages and the risk of eroding traditional agricultural knowledge also pose challenges. As connectivity improves and AI literacy expands, AI has strong potential to support sustainable, high-productivity farming, but targeted investment and policy support will be essential to ensure benefits are shared across farms of all sizes.

4.8 Skills

Ireland enters the AI transition with a comparatively strong digital and ICT skills foundation. The share of ICT specialists in overall employment in Ireland was 6.3 per cent in 2024, the 5th highest in the EU and above the EU average of 5.0 per cent. Moreover, the percentage of people in Ireland with 'basic or above' digital skills stood at 73 per cent in 2023, compared with 56 per cent for the EU, giving Ireland the 3rd-highest ranking in the EU (Expert Group on Future Skills Needs, 2025). Overall, Ireland appears well positioned to harness AI, combining a digitally literate population with a deep pool of ICT talent.

Recent IMF analysis suggests that Ireland is among the countries best positioned to meet future skills needs, ranking highly on measures of skill readiness alongside Finland and Denmark. This reflects high levels of foreign direct investment (FDI) in the tech sector and sustained investment in tertiary education and lifelong learning, which have helped build a workforce with strong adaptability as technologies evolve. However, the same analysis cautions that Ireland's relative strength on the supply side of skills may not be matched by sufficient demand from firms. To avoid under-use of this skills base, the authors of the paper (Jaumotte *et al.*, 2026) recommend that policy focus on stimulating demand by supporting firms to absorb and deploy advanced skills, including through stronger innovation incentives, easier business formation, export promotion, and measures to ease financial constraints on growing companies.

Figure 4.7: Skills Readiness Index



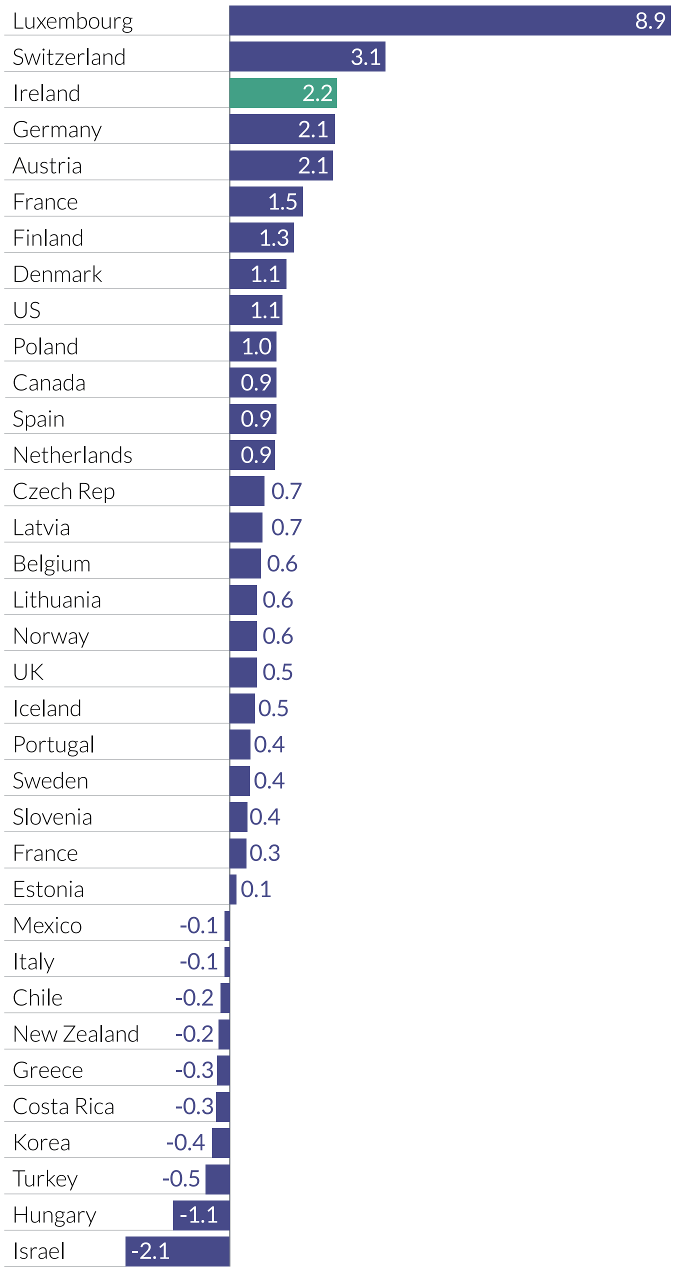
Source: Jaumotte *et al.*, 2026.
Note: The left axis displays the Skill Readiness Index; the right axis presents the skill imbalance index (relative weight of potential future new skill demand versus supply).

Data from the higher education sector reinforce the relatively positive picture regarding AI skills. The Higher Education Authority (2025a) *Key Facts and Figures* report shows that ICT is now one of the largest fields of study, accounting for over 13 per cent of postgraduates and a particularly high share of international graduates, alongside strong enrolments in engineering and other STEM disciplines. This flow of graduates contributes to Ireland's high ranking in the LinkedIn AI Talent Index which places Ireland fifth in the world for AI talent density (Expert Group on Future Skills Needs, 2025).

Nevertheless, issues are emerging in the AI skills pipeline. Demand for AI-related skills is rising sharply, with AI-related job postings more than doubling since 2023, with approximately 63 per cent of jobs judged to be exposed to AI in some way (Expert Group on Future Skills Needs, 2025). This implies that both specialist and AI-literate roles will need to expand significantly just to maintain current adoption trajectories. A shortage of skilled workers is among the main obstacles; surveys show that many firms have difficulties in recruiting staff with the right expertise, even in larger organisations with substantial resources (European Commission, 2020b). Although ICT and STEM output in Ireland is substantial, it still represents a minority of total graduates, and there are concerns about stagnation or decline in domestic enrolments in some digital disciplines over time (Higher Education Authority, 2025a). The National AI Leadership Forum (2025) warns that critical research pipelines, such as the Centres for Research Training (CRTs), face discontinuity without renewed investment, risking erosion of advanced AI capability. A further structural challenge is Ireland's reliance on internationally mobile AI talent. While FDI continues to bring expertise into the economy, this workforce is inherently mobile;

this raises the question of strategic resilience in this area. Ireland ranks third globally in terms of net migration flows of LinkedIn members with AI skills (Expert Group on Future Skills Needs, 2025).

Figure 4.8: Net Migration Flows of LinkedIn Members with AI Skills (per 10,000)



Source: Expert Group on Future Skills Needs, 2025.

4.8.1 Up/Reskilling

Irish and international evidence shows that AI adoption is reshaping skill demand. Analyses indicate a growing need not only for technical expertise such as machine learning and data engineering, but also for capabilities in management, co-ordination, analytical thinking and communication. In highly AI-exposed occupations, vacancies increasingly emphasise emotional, cognitive and digital competencies. Thus, effective AI readiness requires a balanced mix of specialist expertise, workforce-wide AI literacy and human-centred skills such as leadership, creativity and complex problem-solving.

Ireland and the European Union have established a wide range of supports to develop these skills. At EU level, the Digital Europe Programme funds specialised education in AI and other advanced digital domains, while the 2025 Union of Skills package expands advanced digital academies and improves cross-border recognition of digital skills. The *National Digital & AI Strategy 2030* situates digital and AI skills development as a cross-cutting priority for economic and societal transformation, explicitly embedding mechanisms to support workforce upskilling and SME digital adoption. Research Ireland's Centres for Research Training provide structured PhD-level training in AI, machine learning and data science, while CeADAR, the national centre for applied AI, offers industry-aligned training and graduate placement. Further and adult education provision is expanding through SOLAS micro-qualifications in AI and digital transformation, alongside enterprise-led training via Skillnet Ireland and flexible upskilling routes such as Springboard+ and the Human Capital Initiative.

Despite the number of supports available, the IBEC Skills Survey 2025 finds that Irish enterprises display uneven strategic prioritisation of digital and AI capabilities; only 44 per cent of firms consider AI skills important while 75 per cent attach importance to digital skills, leaving a significant minority of employers underprepared for technological change. This undervaluation is particularly acute among SMEs, where resource constraints mean firms prioritise immediate operational and compliance needs over long-term upskilling. As a result, large enterprises are substantially more likely to recognise AI as a critical future skills challenge (69% versus 41% of small firms) and to invest accordingly, being nearly twice as likely to provide digital training (41% vs 21%) and AI training (30% vs 13%) compared to SMEs (Ibec, 2026).

4.8.2 De-skilling

An extensive body of research reveals a paradox at the heart of AI adoption. While AI tools offer gains in efficiency and productivity, their use can risk the erosion of human cognitive skills. This phenomenon, variously described as 'cognitive offloading', 'deskilling' or even 'never-skilling', manifests as a measurable decline in critical thinking, complex problem-solving, creativity and self-sufficiency due to over-reliance on AI tools. The 'automation paradox' describes the predicament where the introduction of an automation, intended to simplify and improve human performance, can paradoxically lead to a decline in human proficiency (Bainbridge, 1983). The central tension is that, as AI automates routine cognitive tasks, the neural pathways responsible for higher-order thinking may atrophy as a result of underuse, following a neurological principle of 'use it or lose it'.

A recent study (Budzyń, 2025) suggests that routine use of AI-assisted colonoscopy systems can lead to a 20 per cent drop in the ability of experienced endoscopists to detect adenomas (precancerous growths in the colon) when performing colonoscopies without AI assistance. These results are concerning but randomised crossover trials will be needed to make more robust claims regarding de-skilling because of the introduction of AI into clinical practice. The OECD highlights the risk of a 'crutch effect' in education, where students rely on generative AI to complete tasks rather than engage in the cognitive effort required for deep learning, creating a 'mirage of false mastery' (OECD, 2026 p51). Evidence from a randomised controlled trial in Türkiye involving nearly a thousand secondary school students, demonstrated that while those students using a generic GPT-4 interface achieved dramatically higher practice performance (up to 127% greater accuracy in some tasks and around 48% higher scores overall), their understanding proved fragile. Once access to the tool was removed, these students performed 17 per cent worse on closed-book exams than peers who had never used generative AI. The findings show that, although generative AI can boost short-term task performance, it can also weaken metacognitive engagement and retention (Bastani *et al.*, 2024).

The decline in cognitive skills is not merely a passive consequence of disuse but is actively driven by a psychological tendency to over-rely on AI. Research on 'algorithm appreciation' shows that people often prefer algorithmic judgment over human judgment (Logg, Minson & Moore, 2019). This preference can lead to a state of over-trust, where users follow AI recommendations without critical scrutiny.

Mitigating the risks of cognitive decline requires a conscious and strategic approach from individuals, educators and policymakers. This includes redesigning educational frameworks to cultivate AI-specific critical thinking, implementing organisational policies that effectively promote hybrid human-AI intelligence, and fostering a culture of mindful technology use that leverages AI as a tool for empowerment rather than a cognitive crutch.

4.9 Productivity and Economic Gains

Artificial intelligence has the potential to deliver substantial productivity and economic benefits by automating routine tasks, augmenting complex ones, and accelerating research and development. These capabilities can lower costs, increase efficiency and stimulate innovation across a wide range of sectors. Knowledge-intensive industries such as finance, ICT and professional services are already reporting measurable productivity gains (OECD, 2025c; Filippucci *et al.*, 2024), while in manufacturing AI is expected to support productivity growth primarily through improved process optimisation, data-driven decision-making and more efficient production systems (OECD, 2025c).

4.9.1 Productivity

Artificial intelligence offers the potential to reshape productivity at multiple levels of the economy. Experimental studies demonstrate substantial productivity gains, particularly in tasks that align with the 'jagged technological frontier' – i.e. those tasks that AI can perform reliably. Consultants using GPT-4 completed tasks 25 per cent faster, accomplished 12 per cent more work, and produced outputs of over 40 per cent higher quality, compared to a control group (Dell'Acqua *et al.*, 2023). In customer service, generative AI increased issue resolution rates by 14 per cent overall, and by 34 per cent for less experienced agents, highlighting a 'skill-levelling' effect (Noy & Zhang, 2023). In software engineering, AI copilots have been shown to increase task completion by over 50 per cent (Peng *et al.*, 2023), while, in professional writing, performance improvements of up to 48 per cent have been documented (Noy & Zhang, 2023). According to the Randstad Workmonitor (2026), which surveyed over 27,000 workers and 1,200 employers across 35 countries, 62 per cent of employees and 54 per cent of employers reported productivity gains from AI adoption.

The available evidence suggests that productivity gains from generative AI are significant but highly context-dependent; evidence shows that such improvements require careful planning, structured training and effective implementation to be realised, otherwise they can reduce productivity in poorly prepared settings.

A recent meta-analysis of 45 studies found an average productivity improvement of approximately 17 per cent when generative AI was used for specific work tasks (Coupé & Wu, 2025). Importantly, these gains were not uniform. The meta-analysis reports substantial variation across contexts and documents a non-trivial minority of cases where AI adoption reduced productivity. This highlights the risks associated with inefficient forms of automation, i.e. those that can led to increased costs or the need for additional work, sometimes referred to as 'so-so automation'. Likewise, in a global study of 3,031 professionals, substantial productivity improvements were documented when AI tools were adopted effectively. Workers using AI reported saving an average of 7.5 hours per week, but this was contingent on structured, recent and inclusive training (Jolles & Lordan, 2025).

An evaluation of Microsoft 365 (M365) Copilot conducted in the UK civil service from October 2024 to March 2025 found that small time savings were observed across most use cases, but additional time was incurred for tasks such as generating images or scheduling (Department for Business & Trade, 2025). The pilot did not find any evidence that time savings led to increased productivity. Interestingly, where additional time was required to complete tasks using M365 Copilot, this was due to the tool not being able to produce high-quality outputs or the task being additional workload assigned because M356 Copilot was in use.

In Ireland, the Office of the Government Chief Information Officer co-ordinated three pilots (a customer service chatbot, a policy and strategic forecasting assistant, and a document-library assistant) to test how Large Language Models could improve public service delivery, policy analysis and internal knowledge management in the public service. Run in partnership with departments and industry specialists, these proof-of-concept studies explored feasibility, usability, integration and cost. Several key lessons were captured which were common to all three pilots; success depends on starting with a well-defined, high-value use case, supported by strong planning around objectives, governance, risks and data quality. The pilots also showed that poor preparation is costly as LLM projects require significant resources, skilled teams and adaptable designs, and getting it wrong can quickly become expensive (Office of the Government Chief Information Officer, 2025).

Productivity benefits generally lag behind technological implementation; thus AI's impact remains modest and difficult to detect in national productivity statistics. This lag is consistent with the Productivity J-Curve hypothesis, which posits that productivity improvements are initially low due to intangible investments in complementary assets such as data restructuring, worker training and workflow redesign (Brynjolfsson *et al.*, 2021). The long-term impact of AI on productivity will depend in part on whether it primarily augments human labour or substitutes for it. If AI complements human labour and diffuses broadly aggregate productivity, gains could be substantial. If substitution dominates and displaced workers reallocate toward sectors with structurally low productivity growth, gains may be dampened through a 'Baumol-type structural effect'. Historically, general-purpose technologies have produced productivity gains over time, often when embedded as complements to human labour, rather than pure replacement (Acemoglu, 2024). According to chief economists' projections, Europe is expected to start reaping the productivity benefits of AI adoption and deployment within the next three years (World Economic Forum, 2026b).

4.9.2 Higher-Value Work

Artificial intelligence does not simply increase output; it also reshapes the composition of tasks within occupations. The prevailing assumption is that AI increases productivity by automating routine tasks, thereby freeing workers to focus on more complex, higher-value activities. Evidence suggests the impacts of AI adoption on productivity are more nuanced, and that effects can vary across different tasks and segments of the workforce. An analysis by Brynjolfsson, Li and Raymond (2025b) found that using an AI chatbot to support call-centre workers tended to substantially enhance the productivity of less experienced workers. By contract, such benefits were found to be more modest for more experienced workers and even led to a slight reduction in the quality of their work.

Autor and Thompson (2025) argue that the labour market effects of AI and automation depend not only on which tasks are automated, but on how that automation reshapes the expertise required for remaining work. When AI removes low-expertise tasks, it raises occupational skill thresholds, concentrating labour demand on higher-value human capabilities such as judgment and problem-solving. Thus, AI can increase the value of workers' skills by shifting effort toward tasks that require greater expertise, which can raise wages and change occupational roles. However, when automation removes expert tasks, the work may require less specialised skill, lowering wages while making the occupation easier to enter.

An OECD analysis of vacancy data shows that, in jobs with high AI exposure, employers increasingly demand competencies such as management, administration, communication and complex problem-solving (Green, 2024). This supports the view that, with appropriate job design, AI can shift human effort toward higher-value activities that require interpretation, oversight and interpersonal skills.

However, the phenomenon of 'workslop' serves as a critical caveat to optimistic narratives about AI driven productivity; this refers to the proliferation of low quality, AI generated content that appears legitimate but lacks substantive value, thereby shifting effort from value creation to human verification and oversight (Niederhoffer *et al.*, 2025).

4.9.3 Macroeconomic Impact of AI

Analysis by the IMF anticipates that AI adoption could lift the average annual growth rate of global GDP by between 0.1 per cent and 0.8 per cent per annum from 2025 to 2030 (IMF, 2024: 76). Further analysis by the IMF examines a central scenario where AI adoption leads to additional growth in global GDP of 0.5 per cent per annum from 2025 to 2030 (IMF, 2025, pp: 6-7). The OECD similarly argues that AI has the potential to increase productivity growth but warns that this depends on complementary investments in skills and innovation (OECD, 2024a). It should be noted that environmental externalities associated with AI systems are generally not sufficiently accounted for in standard economic and commercial metrics.

Macroeconomic projections of AI's impact on growth diverge sharply depending on modelling assumptions. Acemoglu (2024), using a tightly constrained task-based methodology to estimate how much work AI can realistically and profitably automate, concludes that only a small fraction of tasks will be affected over the next decade, suggesting a cumulative total factor productivity (TFP) gain of roughly 0.6–0.7 per cent. By contrast, Aghion and Bunel (2024) present higher projections using two alternative approaches. The first is a historical analogy that compares AI to past general-purpose technologies such as electrification or ICT, yielding potential productivity gains of 0.8–1.3 percentage points per year. The second is based on Acemoglu's task-based model but crucially relaxes some of Acemoglu's constraints on task exposure and rate of diffusion, producing a median estimate of 0.68 percentage points in additional TFP growth. The significant difference in estimates does not reflect disagreement about AI's capabilities *per se* but rather rests on different assumptions about how quickly AI will diffuse across tasks, whether it will become a broad engine of discovery or be confined to automation, and whether historical technological revolutions provide a reliable guide for the trajectory of AI.

In a similar vein, studies of AI adoption report differing findings on returns on investment (ROI). An MIT study found that 95 per cent of AI pilots yielded no measurable financial return, primarily due to organisational barriers such as inadequate integration and poor data infrastructure (Challapally *et al.*, 2025). In contrast, research from the Wharton School at the University of Pennsylvania paints a more optimistic picture, with 70–75 per cent of firms reporting positive business outcomes, particularly where AI was embedded into core workflows (Korst, Puntoni & Tambe, 2025). This divergence is likely a reflection of different adoption stages and metrics. While the MIT study focuses on early pilots and narrow financial returns, Wharton captures later-stage deployments and uses broader measures, including cost savings and workflow efficiency.

4.9.4 AI 'Bubble'

There is an animated debate taking place about whether the surge in AI investment reflects a sustainable technological revolution or a speculative bubble comparable to the dot-com bubble, marked by soaring equity valuations and capital spending. The Magnificent 7 technology stocks now account for over a third of the S&P 500's value and it is estimated that companies' capital spending on AI will reach $527bn in 2026 (Goldman Sachs, 2025). Proponents of the 'bubble' thesis highlight stretched equity prices and the gap between investment and realised returns. The European Central Bank's (ECB) *Financial Stability Review*, published in November 2025, states that 'stretched valuations and extreme market concentration, particularly in US technology and AI-related firms, heighten the risk of the sharp repricing' (European Central

Bank, 2025). Chief economists surveyed by the WEF are divided over AI asset valuations in 2026; 52 per cent expect a decrease or significant decrease in that asset class. Almost three-quarters (74%) expect a significant decrease in the value of US AI assets to have widespread impacts on the global economy, while a quarter expect it to be more contained. More encouragingly, the majority (59%) expect any correction to have short-lived impacts on the global economy (World Economic Forum, 2026b). Ireland faces particular vulnerability given the concentration of US tech operations based in the country, which could affect employment and corporate tax receipts if there were to be a sharp correction.

However, a crucial distinction from previous financial crises is the financing structure; unlike the debt-fuelled bubbles of 2008, the current boom is largely equity-financed. IMF chief economist Pierre-Olivier Gourinchas suggests that this would reduce the risk of systemic financial contagion if a correction occurs, potentially limiting fallout to equity holders rather than triggering broader instability in the financial system (Lawder, 2025). This has led some economists, including Nobel laureate Peter Howitt, to characterise the situation as a potentially 'rational bubble', driven by fundamental technological progress that, even if it leads to a crash, may be essential to fund long-term physical infrastructure and build a knowledge base through widespread innovation across the industry.

This chimes with economist Carlota Perez's framework on technological revolutions which argues that major technological revolutions follow a predictable pattern: the 'Installation Phase' characterised by speculative investment and irrational exuberance, followed by a crash that marks the transition to a 'Deployment Period' where previously loss-making investments become the productive foundation of the economy (Perez, 2002). Applied to AI, this suggests that, even if many AI-focussed startups fail, supporting physical infrastructure such as data centres, expanded power generation capacity and semiconductor manufacturing facilities will sustain productive capacity in the longer term. Such an outcome would mirror the historical experience whereby bankrupt telecom companies left behind the fibre-optic networks that ultimately enabled the modern internet. Some caution regarding this analogy is warranted on the grounds that a not insignificant amount of AI investment is directed toward rapidly depreciating hardware such as GPUs, and (as previously discussed in this report) the technology itself may be confronted with structural limitations.

Chapter 5: AI Governance

5.1 Introduction

This chapter maps the rapidly evolving landscape of AI governance, tracing the expansion of global regulatory activity and the diverse mechanisms emerging to guide the safe and ethical development of artificial intelligence. It introduces the major international frameworks that have laid the groundwork for collective oversight, examines national and regional approaches, including the EU's landmark *AI Act* and Ireland's evolving regulatory architecture, and highlights the shared principles that underpin contemporary governance models. The chapter also considers how fast-moving technological change is prompting governments to explore more adaptive, forward-looking approaches such as anticipatory governance, setting the stage for a richer discussion of how policy, regulation and standards can keep pace with AI's accelerating impact.

5.2 International Initiatives

The OECD's *Principles on Artificial Intelligence* (OECD, 2019) are one of the earliest and most influential standard-setting instruments in the field AI and have been endorsed by over forty countries. The United Nations Educational, Scientific and Cultural Organization (UNESCO) *Recommendation on the Ethics of Artificial Intelligence* (UNESCO, 2021) was the first global governance instrument on AI ethics. The Council of Europe (CoE) *Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law* (Council of Europe, 2024a) opened for signature in September 2024, and its reach extends beyond the 46 Council of Europe member states, with the US, Canada and Japan signing this legally binding treaty. Ireland is included as part of European Union's signature on behalf of its 27 Member States. The African Union (2024) agreed and published in 2024 the *Continental AI Strategy*, which adopts a regional and development-focused approach to AI. On a more technical level, the joint International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) committee on AI have developed several international voluntary standards to facilitate the responsible adoption of AI technologies.[5]

5.2.1 Transnational Governance

In July 2025, China announced its *Action Plan for Global Artificial Intelligence (AI) Governance*, which promotes open-source and cross-border collaboration, risk management, and a recommendation for the establishment of a global AI co-operation organisation to foster international collaboration on AI development and regulation (People's Republic of China, 2025). It should be noted that the *Hiroshima Process International Guiding Principles* were developed for a similar purpose by the G7 nations in 2023 (G7 Hiroshima Conference, 2023). Coherent global regulation is required as AI systems are developed, deployed and hosted across multiple jurisdictions, making it very challenging for any single regulator to ensure effective oversight.

5 Seven AI standards have been published by the ISO/IEC which range from guidance on terminology to impact assessment to risk management: ISO – Artificial intelligence, accessed 20 August 2025. The National Standards Authority of Ireland is represented on AI sub-committee working groups.

Box 5.1: AI in Finance

Artificial intelligence is increasingly being adopted across the financial sector, shaping how institutions deliver services, manage risk and organise operations (OECD, 2023c). Banks and financial services firms are using AI-powered virtual assistants and chatbots to personalise and expedite customer support. Financial firms are deploying AI to detect and prevent fraud and other financial crime, including anti-money-laundering monitoring and suspicious transaction analysis. Stripe's machine-learning-based engine, *Radar*, analyses thousands of transaction attributes in real time to identify anomalous patterns and block fraudulent payments. In trading and investment management, algorithmic systems leverage machine learning to execute trades, interpret market signals, and optimise portfolios with speed and precision beyond human capability. Artificial intelligence can also support regulatory compliance and supervisory functions, automating reporting, monitoring risk exposures, and helping firms and regulators keep pace with evolving standards (Najem *et al.*, 2025).

Despite this promise, adoption remains cautious. Finance is a highly regulated sector, and many advanced AI models function as 'black boxes', complicating explainability, accountability and regulatory approval (OECD, 2024d). Key challenges include algorithmic bias and fairness risks, data privacy and governance constraints, model robustness issues such as GenAI 'hallucinations', and concerns about systemic risk arising from widespread reliance on similar models (Maple *et al.*, 2023). In the Irish financial services context, three principal obstacles that need to be addressed to realise AI's full potential have been identified: integrating AI agents with legacy data and systems; a shortage of advanced and generative AI skills; and building trust in AI through responsible practices and governance frameworks (Financial Services Ireland and IBEC, 2025).

5.3 National Initiatives

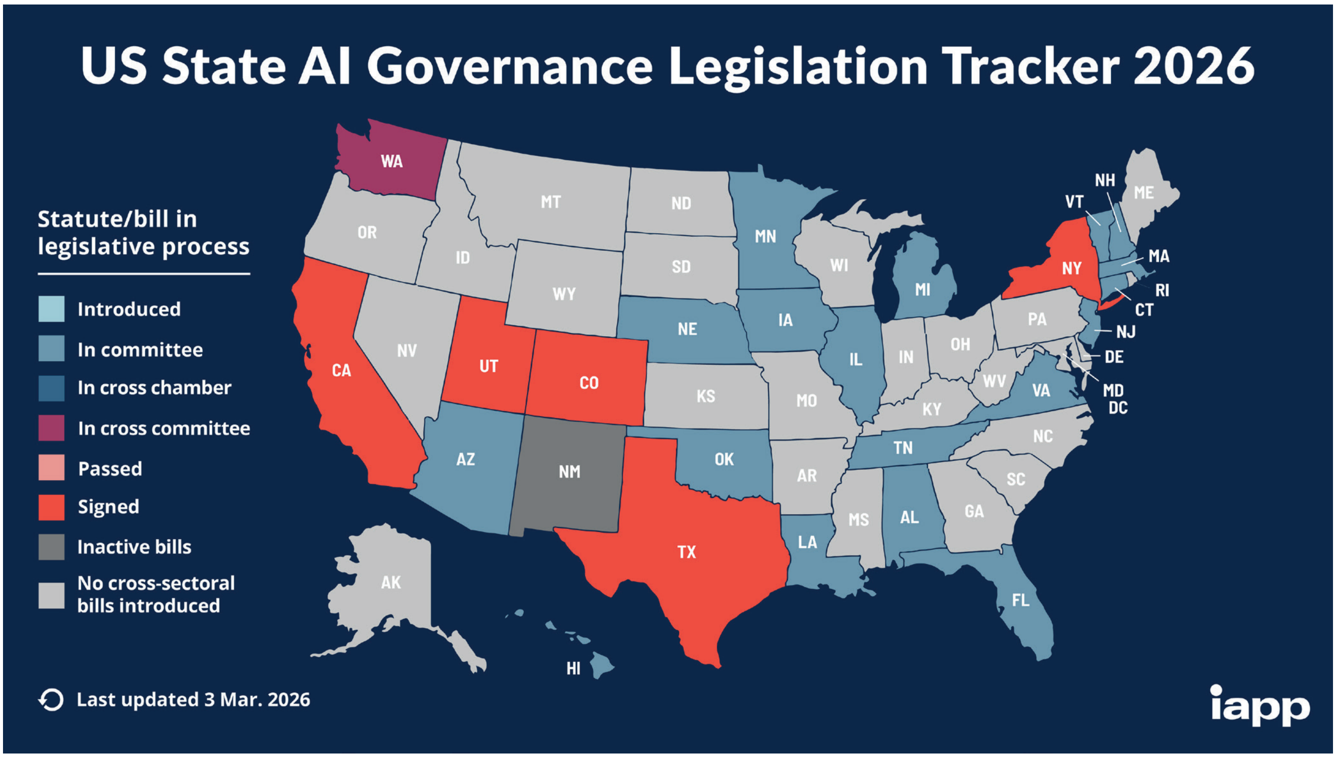
Building on international initiatives, individual countries have adopted diverse governance approaches, with some notable divergences in their scope, binding nature and implementation mechanisms. These differences likely reflect different national priorities such as innovation, economic competitiveness, human rights and fundamental freedoms, as well as legal traditions and geopolitical strategies.

The UK has adopted a 'pro-innovation' and non-binding framework for AI regulation, favouring a sector-specific model, empowering existing regulators and emphasising voluntary measures and ethical guidelines rather than overarching AI legislation (HM Government, 2021). The Australian approach to governance of AI focuses on ethical frameworks and guidelines, with a *Voluntary AI Safety Standard* (Australian Government, 2024) published in August 2024, but there is ongoing debate about the need for more binding regulation.

The United States does not have federal AI legislation, but instead relies on a mixture of existing laws, sector-specific regulations and voluntary guidelines. The Trump administration signalled a shift towards AI deregulation and industry-led innovation, revoking President Biden's Executive Order 14110 'Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence' in

January 2025 (Mackowski *et al.*, 2025). Nonetheless, there is an extensive patchwork of federal agencies and state-level initiatives, each covering different aspects of AI. In 2024, 59 AI-related regulations were introduced – more than double the 25 recorded in 2023 (Maslej *et al.*, 2025b). This is supplemented by soft law in the form of voluntary standards by the National Institute of Standards and Technology (NIST) that aim to advance ethical AI (NIST, 2024).

Figure 5.1: Governance of AI at the US State Level



Source: IAPP, 2025.

In contrast, following the publication of Canada's *Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems* (Government of Canada, 2023a), the Canadian government opted to regulate AI at the federal level, through the proposed *Artificial Intelligence and Data Act*, which is currently under legislative review (Government of Canada, 2023b). China's approach to AI governance and regulation is a hybrid one, sitting between the centralised, top-down approach of the EU and the decentralised, free-market approach in the US. China does not have a single comprehensive law on AI governance but has implemented industry-specific binding regulations and technical standards which often target AI outputs as distinct from AI systems (Chun, Schroeder & Elkins, 2024). For example, in March 2025, the Cyberspace Administration of China (2025) introduced rules requiring internet service providers to clearly label AI-generated content, using both explicit and implicit methods. China's AI governance principles emphasise human control, fairness and transparency and, interestingly, endorse the principle of open-source models. DeepSeek-R1, a Chinese

LLM optimised for reasoning, was launched in January 2025, while Deepseek V3.2, which incorporates a 'sparse attention' mechanism that reduces computational work to provide similar-quality outputs, was introduced in December 2025.

5.3.1 Ireland

The *EU AI Act* (discussed in detail below) provides the overarching legal framework for AI in Ireland. In February 2026, the *General Scheme of the Regulation of Artificial Intelligence Bill 2026* was published by the Department of Enterprise, Tourism and Employment (2026b) to give effect to the Regulation.

Ireland's first *National Artificial Intelligence Strategy, AI – Here for Good*, was published in 2021 (Department of Enterprise, Trade and Employment, 2021) and established an initial framework for the responsible development and adoption of AI. This was followed by a strategic refresh in 2024 (Department of Enterprise, Tourism and Employment, 2024) reflecting evolving technological, regulatory and economic contexts. The current *National Digital and AI Strategy 2030, Digital Ireland Connecting our People, Securing our Future*, builds on these earlier initiatives, maintaining a consistent emphasis on trust, governance, skills development and enterprise adoption across all three policy iterations (Department of the Taoiseach, 2026). The 2026 strategy articulates an integrated vision for positioning Ireland as a digitally enabled, AI-ready society and economy, and is structured around five strategic ambitions, 20 high-level strategic objectives, and supported by 90 key deliverables, designed to guide co-ordinated public, private and societal action.

Beyond these strategy-linked commitments, additional steps have been taken to enhance Ireland's AI governance architecture. The AI Advisory Council, established in January 2024, provides expert advice to government and engages with the public to build confidence in trustworthy AI. While the *National Digital and AI Strategy* is silent on the future of the AI Advisory Council, it commits to pooling and institutionalising expertise through the establishment of an AI Advisory Unit to support public bodies in the effective and responsible adoption of AI. In addition, a National AI Fellowship Programme is to be established by Research Ireland to embed advanced research expertise within the public service and strengthen evidence-based and ethical AI adoption, while strengthening of knowledge-sharing and co-ordination of regulatory-related matters will be done through the Digital Regulators Group. In addition, the Oireachtas established a Joint Committee on Artificial Intelligence, chaired by Malcolm Byrne TD, in May 2025 to examine and make recommendations on AI's development, deployment, regulation and ethical implications, ensuring that governance both supports innovation and safeguards societal interests. In December 2025, the committee published its first interim report in which it made 85 recommendations (Joint Committee on Artificial Intelligence, 2025).

5.4 AI Regulation in the EU

The *AI Continent Action Plan* is the European Commission's overarching policy, setting out 'a set of bold actions' to make the EU a leading AI continent, emphasising competitiveness, democratic values and cultural diversity. It highlights the need to invest in large-scale AI computing infrastructure, data, skills and innovation ecosystems, while ensuring human-centric and trustworthy AI (European Union, 2025). The 'AI Continent' ambition is operationalised through the AI Innovation Package launched on 24 January 2024. A central pillar of this is the GenAI4EU initiative, which aims to boost the uptake of generative AI in 14 strategic industrial ecosystems (e.g. robotics, health, manufacturing). The *Apply AI Strategy*, adopted in October 2025, builds on that foundation, but shifts the focus from supporting AI creation to promoting AI adoption across strategic sectors of the European economy and public sector.

5.4.1 EU AI Act

The *EU Artificial Intelligence Act* (AI Act) (Regulation {EU} 2024/1689 of the European Parliament and of the Council of 13 June 2024 on artificial intelligence) is a landmark, legally binding regulatory framework that officially became law on 1 August 2024, with the Act being implemented on a staggered basis. The general-purpose rules of the Act came into force on 2 August 2025. The regulation lays down harmonised rules for the placing on the market, putting into service and use of AI systems, with the twin aim of fostering the uptake of safe, trustworthy AI and protecting health, safety and fundamental rights across the EU. The Act adopts a risk-based approach that categorises AI systems into different levels of risk; there are stricter obligations for higher-risk uses, with some AI practices prohibited (e.g. some biometric uses), with narrowly defined exceptions (see further detail under Section 5.5.1). The Commission has begun issuing non-binding guidance to support early application of the Act. In February 2025, it published *Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)*, to help providers and other actors determine whether particular software falls under the legal definition of an AI system. The Commission has also issued *Guidelines on prohibited artificial intelligence (AI) practices*, explaining which AI practices are considered unacceptable and providing examples to support compliance. Further, the *EU AI Act* contains dedicated provisions on AI regulatory sandboxes (Articles 57–59), designed as controlled environments where competent authorities can support the development, testing and validation of innovative AI systems, including in real-world conditions. Each member state must ensure that their competent authorities establish at least one AI regulatory sandbox at national level.

EU AI Office

To implement and enforce the *AI Act*, the Commission has created a multi-level governance framework centred on the European AI Office, national competent authorities and EU-level advisory bodies. The AI Office, established within the Commission and operational since 2024, plays a key role in implementing the AI Act, especially for general-purpose AI models (European Commission, 2024). Its tasks include supporting coherent application of the Act across member states, developing tools and benchmarks for evaluating general-purpose AI, drafting codes of practice, preparing guidance and investigating possible infringements. It also advances policies for trustworthy AI (including AI sandboxes and real-world testing), co-ordinates with the European Artificial Intelligence Board (AI Board), the AI Advisory Forum and a Scientific Panel, and promotes the EU's approach internationally.

Member state obligations

Article 70 of the *EU AI Act* mandates that each member state designate national competent authorities and a single point of contact for the application and implementation of the Act. Article 28 requires each member state to designate at least one notifying authority responsible for assessing, designating and monitoring conformity-assessment bodies (notified bodies), and Article 74 requires the designation of market surveillance authorities. Article 77 requires that member states identify national public authorities which supervise or enforce the respect of obligations protecting fundamental rights.

The Irish Government has opted for a distributed regulatory model to implement the Act and has designated 15 public bodies as national competent authorities within their respective sectors,[6] and a further nine bodies as fundamental rights authorities for the Act.[7] A distributed model was chosen as it allows for existing regulatory experience to be leveraged. A distributed model makes sense given the wide range of fields in which AI will be deployed, each with its own regulatory particularities; however, it also carries the risk of producing a fragmented approach if co-ordination is not carefully maintained.

In that context, Ireland has signalled its intention to establish an AI Office of Ireland (AIOI) as the central, co-ordinating authority for implementing the *EU Artificial Intelligence Act*. The AIOI will serve as the 'single point of contact' to co-ordinate the activities of the sectoral competent authorities. Responsibility for its establishment lies with the Department of Enterprise, Tourism and Employment (DETE), perhaps reflecting the Government's view that AI oversight should align with enterprise, innovation and economic policy. The AIOI's core tasks will include co-ordinating the work of the designated competent authorities for consistent nationally coherent application of the *EU AI Act*; acting as Ireland's single national contact point under the Act; providing centralised access to technical expertise for regulators; and hosting a national regulatory sandbox to support innovation and safe deployment of AI systems.

Moreover, the National Standards Authority of Ireland (NSAI) acts as the State's primary body for developing, co-ordinating, and contributing to technical standards that support compliance with the *EU AI Act*. Because the Act relies heavily on harmonised European standards, which are being developed through the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC), NSAI's role is to represent Ireland in these committees, ensure Irish interests are reflected, and facilitate the adoption of these standards nationally.

6 Competent authorities currently designated under the EU AI Act are: Central Bank of Ireland; Coimisiún na Meán; Commission for Communications Regulation; Commission for Railway Regulation; Commission for Regulation of Utilities; Competition and Consumer Protection Commission; Data Protection Commission; Health and Safety Authority; Health Products Regulatory Authority; Health Services Executive; Marine Survey Office of the Department of Transport; Minister for Enterprise, Tourism and Employment; Minister for Transport; National Transport Authority; Workplace Relations Commission.

7 An Coimisiún Toghcháin; Coimisiún na Meán; Data Protection Commission; Environmental Protection Authority; Financial Services and Pensions Ombudsman; Irish Human Rights and Equality Commission; Ombudsman; Ombudsman for Children's Office; Ombudsman for the Defence Forces.

In addition, NSAI has a parallel international role as Ireland's national member of ISO/IEC, where global AI standards are being developed – especially in ISO/IEC JTC 1/SC 42 on Artificial Intelligence. Through this channel, NSAI participates in drafting and refining international standards on AI terminology, governance, risk management, bias mitigation, quality, robustness and lifecycle processes. These ISO standards often feed into or are aligned with the European standardisation process. It is worth noting that Ireland holds the Convenorship and Secretariat of the ISO Working Group 3 on AI Trustworthiness, positioning the country to play a leading role in shaping international standards for safe, ethical and reliable AI and to influence how core principles of trustworthiness are operationalised globally.

Digital Omnibus

Europe was a 'first mover' in the context of AI regulation, which both offers opportunities and poses challenges. There has been some discussion as to whether the *EU AI Act* will generate a 'Brussels Effect', the phenomenon whereby EU regulation becomes a de facto global standard as firms and other jurisdictions adapt to EU rules. However, observers are sceptical that the Act will be emulated in the same way as the GDPR, noting that AI is not a single, uniform policy problem but a diverse set of technologies and domain-specific risks, making wholesale regulatory convergence far less likely (Ebers, 2024). The recently proposed Digital Omnibus on AI (European Commission, 2025e), which seeks to amend and fine-tune the *EU Artificial Intelligence Act*, is a critical inflection point, potentially reshaping how and when the regulatory ambitions contained in the Act will crystallise, casting further doubt on whether the 'Brussels Effect' for AI will materialise.

The Digital Omnibus proposals introduce a significant recalibration of the *EU AI Act* by modifying the timelines for compliance and shifting several obligations to a more conditional, standards-based schedule. Instead of fixed dates (the original requirement that most high-risk AI obligations apply by August 2026 or, at the latest, August 2027), the Omnibus package links the entry into application of many provisions to the availability of harmonised standards or common specifications, with 'long-stop' deadlines that may extend into late 2027 or even 2028. The Digital Omnibus proposals are currently the subject of a public consultation process running until March 2026, after which the proposals will enter the EU's trilogue process involving the European Parliament, the Council and the Commission before any measures can be adopted.

The European Commission argues that these adjustments are necessary to ensure legal certainty, reduce administrative burdens, and allow businesses and regulators to prepare effectively, given that the required technical standards and EU-level support tools are still under development. Indeed, many member states have found it challenging to meet the original timelines laid down in the Act, raising concerns that the race to transpose and operationalise complex requirements could result in rushed national legislation and uneven implementation, each carrying risks of inconsistency, legal uncertainty and diminished regulatory effectiveness.

However, the proposals have also sparked critical commentary (European Civic Forum, 2025). Several observers have noted that major technology firms lobbied intensively for these delays, framing compliance as impracticable without extended timelines. This raises concerns about the potential influence of powerful industry actors on the EU's regulatory trajectory and whether such revisions could dilute the original political commitment to strong, timely safeguards for fundamental rights and societal oversight in the deployment of advanced AI systems.

The *National Digital and AI Strategy* commits Ireland to working with EU partner states to advance an ambitious digital simplification agenda and has prioritised this issue for Ireland's 2026 EU Presidency (Department of the Taoiseach, 2026, p.56). At national level, this commitment is reflected in a streamlined regulatory approach focused on reducing administrative burden through single reporting mechanisms and enhanced co-ordination via the Digital Regulators Group.

Governance v innovation

The Digital Omnibus initiative is at least in part motivated by the narrative that the EU's extensive regulatory approach to digital technologies, including AI, is causing Europe to fall behind in the 'AI race'. Proponents of this view argue that regulation raises costs, diverts resources and slows innovation. This is particularly relevant for SMEs that may be forced out of the market or discouraged from entering by the regulatory burden imposed.

Others dispute this trade-off logic, arguing that regulation is essential for consumer trust, provides predictable legal frameworks that reduce uncertainty, thereby promoting investment, and can even stimulate innovation by pushing firms toward more efficient, socially beneficial technological solutions (Porter, 1991). It has also been argued that Europe's innovation deficit is driven less by regulation and more by structural factors such as fragmented capital markets, lower risk-tolerant investment, weaker scaling ecosystems, and under-investment in digital infrastructure (Bradford, 2024). Allen (2025) argues that policymakers may be overestimating the competitiveness gains from reducing the regulatory burden, while underestimating the unintended harms of such action. Rather than seeing regulation as a constraint, in the European context it could be seen as a positive differentiator enabling trust, adoption and scale in sensitive, high-value use cases (Tournesac *et al.*, 2025).

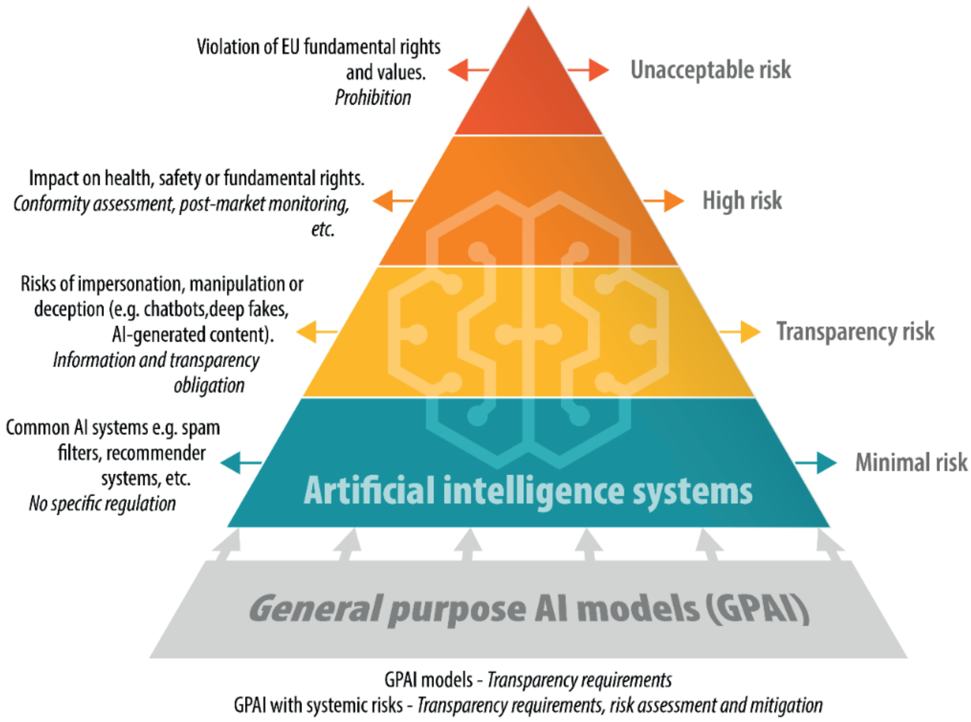
5.5 Common Threads in AI Governance

While approaches vary across jurisdictions, several common themes in relation to AI governance have emerged, including the adoption of a risk-based approach and an emphasis on trustworthiness and ethical principles, as well as the necessity for human agency and oversight.

5.5.1 Risk

The adoption of a risk-based approach involves classifying AI systems into categories with varying regulatory burdens associated with each. The *EU AI Act* classifies AI systems into four categories: *unacceptable risk systems* which are strictly prohibited – e.g. social scoring, manipulative subliminal techniques or real-time biometric identification (with limited law enforcement exceptions); *high risk systems* – e.g. critical infrastructure, healthcare, justice, which must undergo rigorous risk impact assessment; *limited-risk systems* such as chatbots which carry transparency requirements so that users know they are interacting with AI; and *minimal or no systems* such as translation tools which are largely unregulated but encourage adherence to voluntary codes of conduct.

Figure 5.2: EU AI Risk-based Approach



Source: European Parliament, 2021.

While this framework provides legal clarity and aims at proportionality, a clear, detailed methodology for assessing AI risks in concrete situations is lacking (Novelli *et al.*, 2024). The notion of risk within the Act is often vaguely articulated, leaving key definitions and thresholds open to interpretation (e.g. what constitutes 'high risk'?). Another challenge with the risk-based framing is that classical conceptions of risk typically rely on quantifiable probabilities and measurable harms, but AI often introduces deep uncertainty and 'known unknowns' (Ebers, 2025). As previously discussed, frontier AI systems can demonstrate unpredictable emergent behaviours, complex interactions with social systems, and harms that may not be foreseeable at design time. As a result, a purely probabilistic, quantification-based regulatory lens may systematically underestimate or even miss serious but non-quantifiable harms.

The Act does not call for a risk–benefit analysis, even though ethical evaluation typically requires weighing potential harms against potential societal gains rather than considering risks in isolation. Instead, the Act focuses almost exclusively on mitigating risks, with little consideration of the potential social, economic or scientific benefits of AI deployment. As pointed out by Ebers (2025), the lack of such a risk-benefit analysis may lead to opportunity costs as there is no balanced appraisal of what might be lost.

Many of the harms associated with AI – especially those affecting EU fundamental human rights, as protected under the *Charter of Fundamental Rights of the EU* and explicitly referenced throughout the *EU AI Act* – are poorly suited to a standard risk-based framing. These rights are not marginal trade-offs but, in many cases, represent non-negotiable guarantees for individuals. Applying tools such as quantification or acceptable risk thresholds runs the risk of obscuring or normalising rights violations. International bodies have increasingly embedded human rights at the centre of AI governance. The OECD AI Principles (2019) explicitly anchors this approach in *Principle 1*, which calls for AI systems to 'respect the rule of law, human rights, democratic values and diversity'. UNESCO's *Recommendation on the Ethics of Artificial Intelligence* (2021) similarly grounds AI governance in dignity, fairness and human rights protections.

Most prominently, the Council of Europe (CoE), consistent with its foundational pillars of human rights, democracy and the rule of law, has adopted a rights-first regulatory approach, which places the safeguarding of fundamental rights at the core of all stages of AI design, development and deployment. The Commissioner for Human Rights in the CoE highlights that AI technologies are not only sources of risk but also hold significant potential to promote and strengthen human rights – e.g. AI could help identify where individuals are entitled to public benefits. This would require AI to be approached through a holistic, human-rights-centred lens, rather than one focused narrowly on productivity gains or securitisation (Commissioner for Human Rights, 2025). The Ombudsman for Children (2025) has recommended adopting a rights-based approach to AI, to ensure that AI systems are designed and governed in ways that safeguard the best interests, privacy, dignity and developmental needs of children.

5.5.2 Trustworthy AI

Ethical principles play a foundational role in the governance of artificial intelligence, providing a normative framework to guide its design, deployment and oversight. By incorporating ethical principles into the fabric of AI governance, the ambition is to achieve technologically advanced systems that are aligned with democratic values, fundamental rights and the public good. Within the EU, seven principles (human agency and oversight, technical robustness and safety, privacy and data governance, transparency and explainability, diversity/non-discrimination, fairness, and societal and environmental wellbeing) have been formally consolidated into a foundational concept of trustworthy AI, which serves as the premise for the *EU AI Act*. Trustworthy AI is conceived as AI that is lawful (complies with existing laws), ethical (upholds fundamental values and rights) and robust (secure and reliable in practice). This framing has become a reference point for global AI governance.

The OECD's *Framework for Trustworthy AI in Government* provides a structure for how governments can ensure their use of AI is trustworthy by focusing on three essential pillars: *Enablers, Guardrails and Engagement* (OECD, 2025a). Key enablers include strong data foundations, digital infrastructure, skills, governance and purposeful investment and procurement. In relation to guardrails, the OECD stresses the importance of promoting transparency and explainability, as well as empowering oversight bodies and having the appropriate policy levers in place. Engagement is crucial, with both citizens and social partners and users being involved in AI development. There is also an emphasis on collaborating across borders.

However, trustworthy AI remains a difficult ideal to achieve and there is little concrete guidance for how to go about that goal (Laux, Wachter & Mittelstadt, 2023). As pointed out by Ballot Jones, Thornton and De Silva (2025), there is a danger that trustworthy AI becomes, in effect, a 'regulatory visibility tactic', a symbolic label rather than a guarantee of substantive safety, fairness and accountability.

5.5.3 Human Oversight

Human agency and oversight are consistently emphasised as core principles in global AI governance instruments, reflecting the desire that AI should augment rather than replace human decision-making. The regulatory frameworks developed by the OECD, UNESCO and EU all stress the importance of mechanisms such as human-in-the-loop or human in command to ensure that people retain meaningful control over AI systems, particularly in high-stake or sensitive contexts.

Human oversight is a key requirement of the *EU AI Act*, which mandates that high-risk AI systems must be designed and developed so they can be effectively overseen by humans during their operation (Article 14). This obligation is grounded in the Act's overarching goals of safeguarding health and safety, ensuring system reliability, and protecting fundamental rights. Yet the rationale for human oversight extends well beyond these regulatory imperatives. Across domains, oversight serves indispensable governance functions by introducing moral judgment, contextual sensitivity and empathy into decision-making processes which would otherwise be governed by opaque outputs. It helps to some extent to counteract algorithmic bias and anchor accountability in identifiable human or institutional actors, and provides a mechanism for aligning AI behaviour with societal values and ethical norms.

However, achieving meaningful oversight presents substantial challenges, many of which stem from the very characteristics that make AI powerful. The opacity and scale of complex machine-learning models can make real-time monitoring or comprehension impracticable in many situations. At the same time, human cognitive limitations – including automation bias, vigilance decline (difficulty of maintaining attention over time) and reduced moral agency (tendency of humans to relinquish their sense of responsibility when interacting with technology) – all undermine the assumption that a human ;'in the loop' will necessarily detect or correct errors (Holzinger, Zatloukal & Müller, 2024). Moreover, organisational constraints, such as insufficient training and inadequate time for review, can further erode operators' ability or willingness to intervene. Meaningful human oversight is neither automatic nor guaranteed and, therefore, must be deliberately designed and institutionally supported to function as an effective governance mechanism.

5.6 Governance in Practice

Despite a convergence around trustworthy AI and the ethical principles which underpin it, it is less clear how it can be operationalised in practice. Realising abstract values such as fairness, accountability and transparency into measurable, verifiable criteria that can withstand regulatory and public scrutiny is very challenging. This 'principle to practice gap' is a major area of focus as adoption of AI tools increases.

A range of modalities are beginning to emerge to assure governance across the AI life-cycle. Regulatory frameworks adopting a risk-based framework require impact assessments depending on the potential harms of an AI system, while conformity assessment procedures – e.g. the University of Oxford *capAI* protocol (Floridi *et al.*, 2022) and independent-based ethics auditing – provide structured means of validating compliance and ethical alignment with principles. The Council of Europe's HUDERIA methodology (Council of Europe, 2024b) offers a structured framework for assessing how AI systems may affect human rights and democracy and offers practical tools to identify and address harms.

Regulatory sandboxes can allow authorities to engage firms to test AI tools that challenge existing legal frameworks in a supervised setting (OECD, 2023d). Private and public-sector organisations are also employing internal controls tools such as datasheets and scorecards to improve accountability and transparency. Ethics committees, the creation of AI accountability roles and staff training further reinforce responsible practices. Thus, a multi-layered strategy is being adopted, but the effectiveness of such measures will depend on continuous monitoring and adaption.

What is clear is that practical, accessible tools are essential to help practitioners bridge the 'principle to practice' gap. The UK has developed a series of eight practice-based workbooks offering end-to-end guidance on applying ethical principles in public-sector AI projects, covering issues from problem formulation and data use to safety, accountability and deployment (Alan Turing Institute, 2023). These kinds of grounded, operational tools are critical for enabling organisations to move beyond well-intentioned principles and embed ethical and safe AI practices in everyday decision-making.

5.7 New Forms of Governance

Innovation is difficult to govern because it creates novelty and surprise. The implementation of technology into society is a complex and unpredictable endeavour. By the time the full extent of risks and unintended consequences of a given innovation is fully appreciated, it has usually become embedded in social infrastructures, and at that stage it can be exceptionally difficult to change course (O'Sullivan, 2020).

The development of social media provides an illustrative case in point. Early policy assumptions framed social media platforms as neutral intermediaries rather than as powerful socio-technical systems capable of reshaping behaviours, markets, information ecosystems and democratic processes in systemic ways. Arguably, meaningful regulation arrived only after mass adoption, which meant that governance became reactive and path-dependent. Regulators were forced to manage an entrenched status quo shaped by dominant business models, technical architectures and user lock-in, rather than to shape the role of platforms in society *ex ante*.

The rapid evolution of AI technology presents a significant challenge for effective governance, as legal and regulatory frameworks often struggle to keep pace with technological innovation. This 'law lag' creates a gap in which AI systems may be deployed before adequate safeguards are in place, which increases the risk of unintended and/or unexpected consequences at both individual and societal levels. In response, academics and policymakers have called for new forms of governance, including anticipatory innovation governance and experimental governance as a future-oriented approach to navigating uncertainty.

Experimental governance is an adaptive approach which emphasises iteration, evidence gathering and participation to address complex and uncertain challenges (Sabel & Zeitlin, 2012). Rather than relying on fixed rules, experimental governance is open to revision and responsive to emerging data and stakeholder feedback. In the AI context, one could argue that regulatory sandboxes and algorithmic impact assessments are a form of experimental governance, as these tools allow governments to trial regulatory approaches, generate evidence on risk and impacts, and adjust frameworks as the technology evolves.

Building on the logic of experimental governance, anticipatory governance is an even more developed approach in the AI context, a framing most notably advanced by the OECD. It emphasises the need for public institutions to proactively explore emerging futures, identify potential risks and opportunities, and adapt policy frameworks before problems fully materialise.

5.7.1 Governance in Situations of High Uncertainty

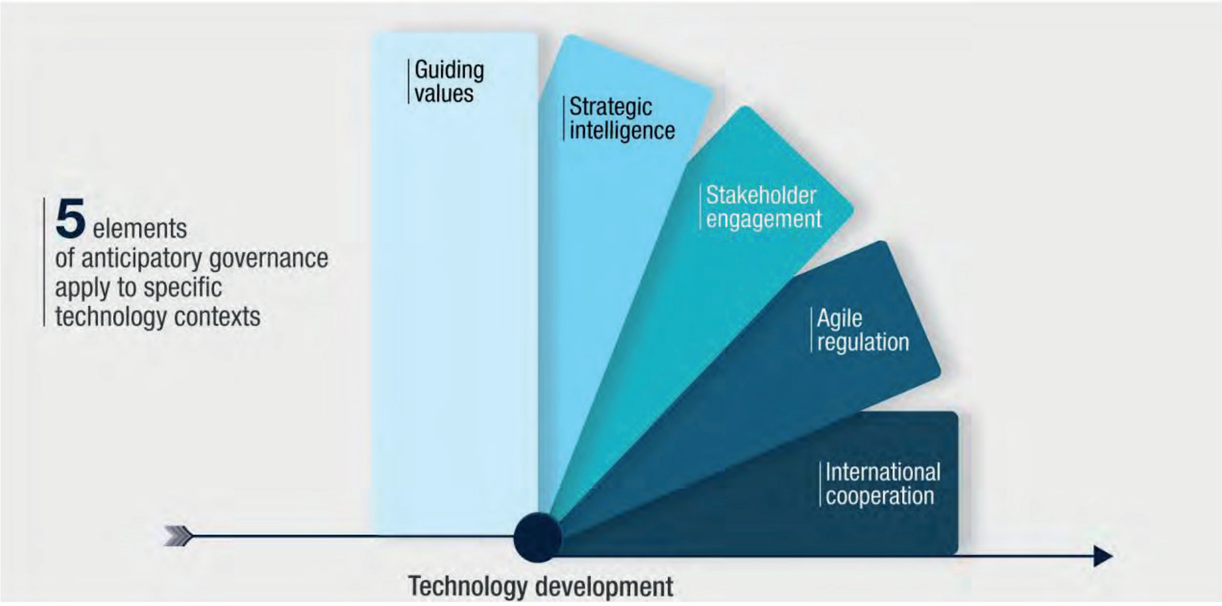
Anticipatory governance (AG) is specifically designed for high-uncertainty environments where the timeline, pathways and ultimate societal impacts are difficult to predict. By combining foresight, flexible policy design and iterative learning, AG provides institutions with the capacity to prepare for, rather than merely react to, rapidly evolving socio-technical landscapes. Anticipatory governance addresses uncertainty by stressing the need for a mix of problem-solving and problem-finding approaches, which involves an active and systematic search for potential future problems that the technology may raise.

Anticipatory governance provides a flexible scaffolding for navigating the unknowns inherent in AI. By integrating foresight, broad engagement, and continuous learning, it can help policymakers prepare for diverse and evolving futures, rather than being constrained by narrow predictions or reactive responses. This adaptability makes AG especially well-suited to AI where technological trajectories are open-ended and their societal consequences not yet fully visible.

5.7.2 Anticipatory Governance for AI

The recent *Steering AI's Future* report from the OECD focuses on five interdependent elements of the OECD *Framework for Anticipatory Governance of Emerging Technologies*: guiding values, strategic intelligence, stakeholder engagement, agile regulation and international co-operation. The five elements function collectively, with each reinforcing the others.

Figure 5.3: Five Elements of Anticipatory Governance



Source: OECD, 2024b.

Values

A robust anticipatory governance strategy begins with a shared set of guiding values that intentionally shape AI development and deployment. As discussed in earlier sections, there is growing international convergence around values frameworks, with alignment across the OECD *AI Principles*, the EU's AI governance instruments (including the *Seven Requirements for Trustworthy AI*), the Council of Europe's *Framework Convention on AI*, and UNESCO's *Recommendation on the Ethics of AI*. This convergence provides a crucial foundation for global interoperability, reducing fragmentation and enabling coherent cross-border governance.

While the OECD framework operates at a broad policy level and the EU requirements focus on operational, implementation-level guidance, both share a commitment to fairness, transparency, accountability and human-centred design. An interesting divergence is that the OECD explicitly incorporates sustainability, emphasising environmental, social and economic well-being as a core objective, whereas the EU principles do not feature sustainability as a standalone requirement, instead addressing it only indirectly through risk and impact considerations.

Several types of tools and processes are being developed in an effort to embed guiding values throughout the AI system lifecycle. The *OECD.AI Catalogue of Tools & Metrics* enables practitioners to identify and compare techniques to operationalise fairness, explainability, robustness and other principles, while deliberative processes, including public dialogues and multi-stakeholder roundtables, can help to elucidate societal values, identify red lines and surface concerns about emerging AI capabilities.

Strategic intelligence

Strategic intelligence provides the 'early warning system' necessary for anticipatory governance. Because AI evolves rapidly and in unpredictable ways, governments require mechanisms that capture weak signals, synthesise expert insight and illuminate plausible medium- and long-term trajectories of the technology.

Foresight methodologies such as scenario building, horizon scanning, Delphi surveys and backcasting[8] enable policymakers to explore divergent futures, challenge prevailing assumptions, and prepare for high-impact uncertainties. The OECD.AI Expert Group on AI Futures has used foresight methods to map possible AI trajectories, identifying numerous benefits, risks and policy options relevant to governments. Akin to the public health approach to infectious disease, sentinel and real-time monitoring of AI can identify weak signals such as patterns of misuse, failure modes and systemic vulnerabilities, which may indicate future governance challenges.

Stakeholder engagement

Stakeholder engagement is indispensable for anticipatory governance because AI systems affect diverse communities, rely on public trust, and raise normative questions that cannot be resolved by experts alone. Engagement processes broaden understanding, surface blindspots and can promote legitimacy. A comprehensive approach to engagement involves civil society organisation, industry and technical experts, academia, public-sector actors and general publics, whose perspectives are essential for shaping values and expectations.

Several forms of engagement are being used in AG, including informative engagement – for example, explainers and transparent communication of risks and system behaviours. Consultative engagement includes surveys, targeted interviews and public consultations, which can be useful in collecting views on proposed regulation. Collaborative engagement, the most demanding and potentially most rewarding, is where stakeholders co-design governance tools, participate in deliberative assemblies or citizens' juries, and contribute to community red teaming[9] or participatory audits. As previously discussed, AI itself can be leveraged to enhance engagement by enabling citizen participation in policymaking and processing consultation data.

Participation-washing, where the appearance of engagement can mask predetermined agendas and sideline community interests, poses a risk in public discussions on AI. An analysis of national AI strategies reveals a persistent gap between governments' rhetoric of public involvement and the absence of concrete mechanisms to secure meaningful input (Wilson, 2021). As Wilson (2021) argues, private-sector values like efficiency and competitiveness often eclipse democratic commitments to equity, deliberation and accountability. Governance frameworks should embed genuine, inclusive participation and ensure that AI policy development is grounded in public interest values rather than performative consultation.

8 Backcasting is a strategic foresight method which starts with a desired future outcome and works backward to identify the steps, decisions and interventions needed to reach that future from the present.

9 Red teaming is a structured, adversarial testing exercise designed to identify vulnerabilities, potential harms and failure modes in an AI system before it is widely deployed.

Agile governance

Given AI's rapid evolution, governance systems must remain adaptable, iterative and capable of learning through experimentation. Agile governance complements anticipatory governance by enabling policy innovation alongside technological innovation. Agile governance also requires integrating good practice 'by design', such as safety-by-design, privacy-by-design and ethics-by-design. Standards and shared risk-management frameworks provide predictable structures that promote interoperability while supporting rapid adaptation.

Table 5.1: Anticipatory Innovation in Policymaking

	Traditional policymaking	Anticipatory innovation governance
Evaluation approach	Evaluation as the last stage in an often multi-year policy cycle	Continuous evaluation and assessment; exploring future effects (e.g. changes in public values, ethics, intergenerational fairness)
Policy cycle	Long research and drafting cycles, with policy implemented accordingly	Recognition that cause-effect relationships are impossible to know in advance, and that the policy implementation itself changes the problem space
Research and analysis approach	Exploring the problem space through research and analysis	Exploring the problem space through small-scale real-world experiments and innovation
Research and analysis focus	Research and analysis focused on what has happened	Research and model development focused on a range of possible futures
Participation	Policy domain experts and primary affected population	System of related policy areas and affected populations, which changes over time

Source: OECD, 2024c.

Regulatory sandboxes allow developers and regulators to test innovations in controlled environments. They provide temporary adjustments or exemptions from certain rules, enabling regulators to observe real-world risks and gather evidence for longer-term policymaking. Norway's Regulatory Sandbox for Responsible Artificial Intelligence and privacy has enabled firms to experiment with privacy-preserving machine learning systems while regulators observe risks and identify areas requiring legal clarification or policy reform. The sandbox has produced actionable insights on data minimisation, transparency practices and novel approaches to safeguarding rights.

The *Digital & AI Strategy 2030* positions Ireland as a trusted, agile and forward-looking digital regulatory hub, and has committed to the establishment of a national AI regulatory sandbox by the AI Office in 2026 (Department of the Taoiseach, 2026).

Table 5.2: Benefits and Challenges of Regulatory Sandboxes

	Benefits	Challenges
To regulators	• Inform long-term policy making through learning and experimentation • Signal commitment to innovation and learning • Engage and communicate collaboratively with market participants • Adjust regulations that may restrain innovation or result in safety risks	• Insufficient technical expertise within regulatory bodies • Multi-disciplinary scope of AI products calls for collaboration of multiple stakeholders • Defining the evaluation methods to determine the eligible participating firms
To firms	• Reduce time to market by streamlining the approval process • Reduce regulatory uncertainty by providing clarity on prohibited technologies • Gather feedback on regulatory requirements or risks • Improve access to capital • Facilitate market entry for companies, especially SMEs and start-ups, by providing accessible information about legal frameworks	• Difficulty in understanding the eligibility and selection criteria to participate • Capacity constraints may limit the number of firms that can participate
To consumers	• Promote introduction of innovative and potentially safer products • Increase access to AI products and services	• Inefficient implementation of regulatory sandboxes and their associated safeguards can lead to negative impacts on consumers

Source: OECD, 2025d.

International co-operation

International co-operation is fundamental to anticipatory governance, enabling interoperability, pooling of expertise and co-ordinated responses to shared risks. The transboundary nature of AI means that no single country can govern it effectively alone. While Ireland operates within a wider European regulatory framework, it is also essential to remain cognisant of and work collaboratively with other countries and their distinct regulatory systems.

International co-operation avoids regulatory fragmentation, as without alignment developers or deployers of the technology could engage in 'ethics shopping', by choosing the least restrictive jurisdiction in which to operate. It also recognises that issues such as cyber-security vulnerabilities or harms arising from global deployment require co-ordinated solutions. Moreover, by involving countries with diverse capacities, it prevents governance architectures from being shaped solely by technologically dominant actors.

Effective co-operation requires a multilayered approach, and Ireland is well positioned in this regard thanks to its expert and engaged participation in working groups, standard-setting processes and wider AI initiatives across the European Commission, Council of Europe, OECD and ISO, while also ensuring it continues to build capacity in strategic intelligence and related capabilities.

5.7.3 Importance of Monitoring and Evaluation

Monitoring and evaluation must be embedded throughout the entire AI lifecycle rather than be treated as activities that begin and end at the point of deployment. Because AI systems are dynamic, context-dependent and capable of behaving unpredictably in real-world environments, ongoing assessment is essential to ensure safety, effectiveness and alignment with societal values.

Traditional evaluation models are insufficient for fast-moving technologies whose impacts unfold over time. Developmental and real-time evaluation support iterative learning and allow policymakers to revisit assumptions, adjust strategies and refine interventions as conditions change. Rather than relying on retrospective, end-stage assessments, anticipatory governance requires continuous feedback loops across development, testing, deployment and operation. Such feedback loops ensure that real-world evidence informs the evolution of policies, system design and implementation strategies. Multidimensional evaluation spanning social (e.g. access to services and inclusion, impacts on employment), environmental (e.g. energy consumption, water and land use) and economic impacts (e.g. productivity gains, impacts on regional and sectoral development) ensures that governance systems capture the full range of outcomes rather than relying solely on technical metrics such as accuracy, speed and cost-efficiency. By integrating these dimensions into monitoring and evaluation frameworks, public bodies can better understand how AI systems affect society as a whole, not just how well they function technically.

As previously described, AI systems frequently behave differently in controlled testing environments compared with real-world settings, where data quality, user behaviour, operational pressures and contextual variation introduce complexities that cannot be fully simulated in advance. This makes ongoing monitoring essential to detect performance degradation, biases, emergent risks and unintended consequences. The Epic Sepsis Prediction tool serves to illustrates this point (Patient-Safety-Learning, 2024). Although it demonstrated strong performance and high accuracy during internal testing, real-world deployment revealed a significant gap; the tool failed to identify two-thirds of sepsis cases when first implemented in a hospital setting. A recently published randomised study found that, although LLMs performed very well when tested on complete clinical cases (correctly identifying relevant conditions in ~95 per cent of cases), lay users interacting with the same models identified relevant conditions in fewer than 35 per cent of cases (Bean *et al.*, 2026). People using LLMs were no better than those relying on standard internet searches at identifying important conditions or judging how urgently care was needed. The performance drop was largely driven by communication failures, as users often provided incomplete information, misunderstood or ignored advice from the LLM, or struggled to interpret mixed or inconsistent suggestions from the LLM. This mismatch between laboratory performance and 'in the wild' behaviour, sometimes referred to as the 'evaluation gap', highlights the critical need for continuous monitoring, post-deployment evaluation and system recalibration to ensure clinical safety and reliability. Similar patterns have emerged in sectors such as education, where optimistic performance claims of AI systems have not yet translated into consistent improvements in student learning outcomes at scale (Fengchun *et al.*, 2021; Bauer, 2025). This reinforces why early-stage and ongoing evaluation should be considered foundational to responsible anticipatory AI governance.

Chapter 6: AI Literacy

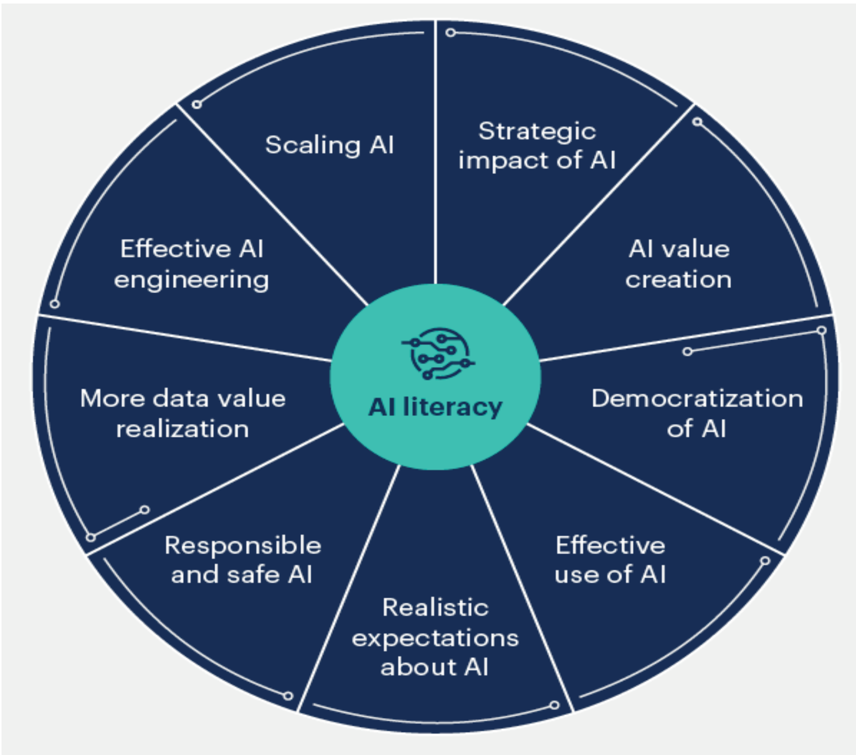
6.1 Introduction

This chapter examines the growing importance of AI literacy as a foundational competency for participating in an increasingly AI-mediated society. It explores how AI literacy has evolved from a niche technical skill to a civic, educational and organisational necessity. It outlines the key components of AI literacy, traces its development across established conceptual frameworks, and surveys how governments, educational institutions, businesses and the wider public are cultivating the knowledge, skills and critical capacities needed to engage with AI responsibly and effectively.

6.2 The Imperative for AI Literacy

The increasing integration of artificial intelligence across economic, social and civic domains has rendered AI literacy an increasingly indispensable competency for meaningful participation in contemporary society. As AI systems increasingly mediate decisions in healthcare, finance, education and the public sector, AI literacy has become a civic, strategic and economic necessity. The capacity of individuals and institutions to understand, use and evaluate AI has become key to realising the opportunities the technology offers.

Figure 6.1: Key Benefits of AI Literacy



Source: Gartner, 2025c.

Ireland's *National Digital and AI Strategy* (2026) positions artificial intelligence as a central enabler of digital transformation, public service reform and sustainable economic competitiveness within an integrated national digital policy framework. Yet as AI technologies evolve at remarkable speed, the gap in understanding among professionals and the public risks widening, threatening both engagement and responsible adoption. This concern is echoed at the European level through the *EU AI Act* (European Union, 2024), which makes explicit in Article 4 of the Regulation the requirement for a 'sufficient level of AI literacy' among all staff involved in providing or deploying AI systems. The Act recognises that ethical and safe implementation of AI cannot occur without the human capacity to interpret, challenge and govern these systems responsibly.

Global economic and workforce trends also underscore the urgency of fostering AI literacy. The World Economic Forum's *Future of Jobs* report (2025a) anticipates that 44 per cent of workers' core skills will be disrupted by technological change by 2030, with AI playing a leading role. In this context, AI literacy is not 'a nice-to-have' but rather should be considered a foundational skill to navigate the digital world, access opportunity and participate in the shaping of the future of AI.

6.3 What is AI Literacy?

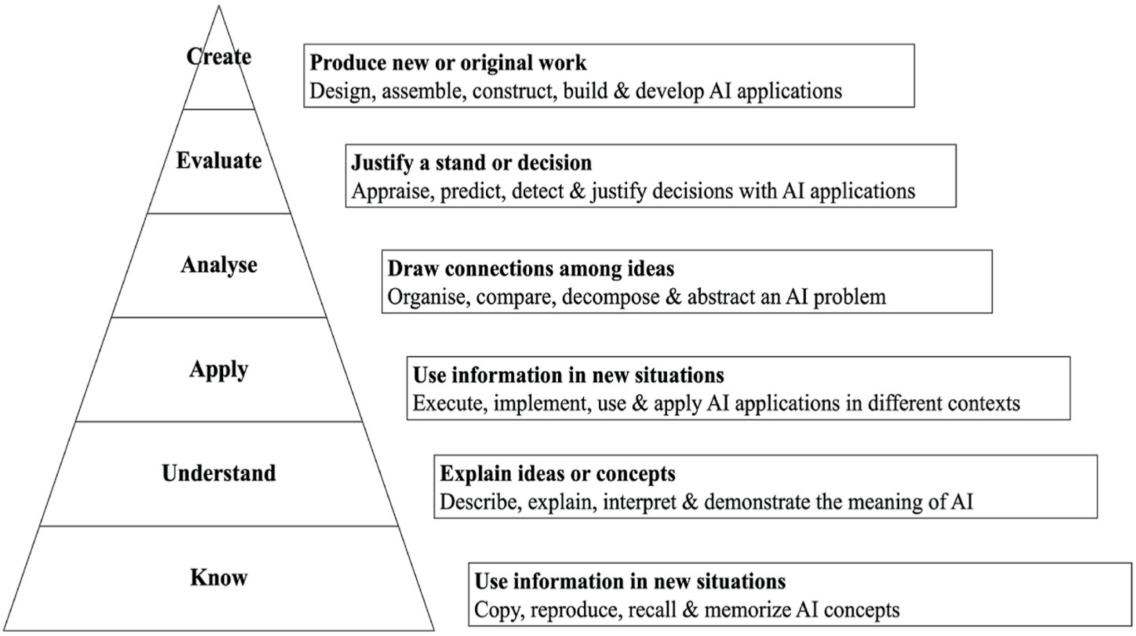
AI literacy refers to the foundational knowledge, skills and dispositions required to understand, interact with, evaluate and use AI systems responsibly and effectively. The concept builds on earlier literacies, particularly data literacy and digital literacy, yet extends beyond them in both scope and purpose. While related literacies form the foundation for AI literacy, data literacy fosters the ability to interpret and reason with data. It enables individuals to use computational devices, but AI literacy emphasises a functional and critical understanding of AI's mechanisms and implications (Chiu, 2025). This involves knowing how AI works, what it can and cannot do, and how to use it responsibly.

Kandlhofer and colleagues (2016), were among the first to formalise the term, defining AI literacy as a set of competencies that enable individuals to know, understand and use AI technologies. Long and Magerko (2020) later expanded this definition, framing AI literacy as 'a set of competencies that enables individuals to evaluate AI technologies critically; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace'.

Long and Magerko's (2020) framework remains one of the most comprehensive early conceptualisations of AI literacy. The authors identify 17 specific competencies necessary for AI literacy by reference to five guiding questions: What is AI? What can AI do? How does AI work? How should AI be used? How do people perceive AI? The questions serve as a thematic framework for exploring what individuals need to know, be able to do, and critically reflect upon to participate meaningfully in an AI-driven world. The 17 competencies span technical, conceptual, social and ethical dimensions, from recognising and understanding AI systems to appreciating their social implications. The study positions AI literacy as a multidimensional construct that integrates technical understanding with ethical reasoning and social awareness.

Ng, Leung, Chu and Shen (2021) expand on this conceptual groundwork by proposing a structured framework for AI literacy that links cognitive development with ethical understanding. They identify four dimensions: *know and understand*, *use and apply*, *evaluate and create*, and *ethical issues*. The authors explicitly align these with Bloom's Taxonomy (Bloom *et al.*, 1956, pp.1103–1133), a cognitive model of learning that describes progression from foundational to higher order thinking.

Figure 6.2: Bloom's Taxonomy and AI Literacy

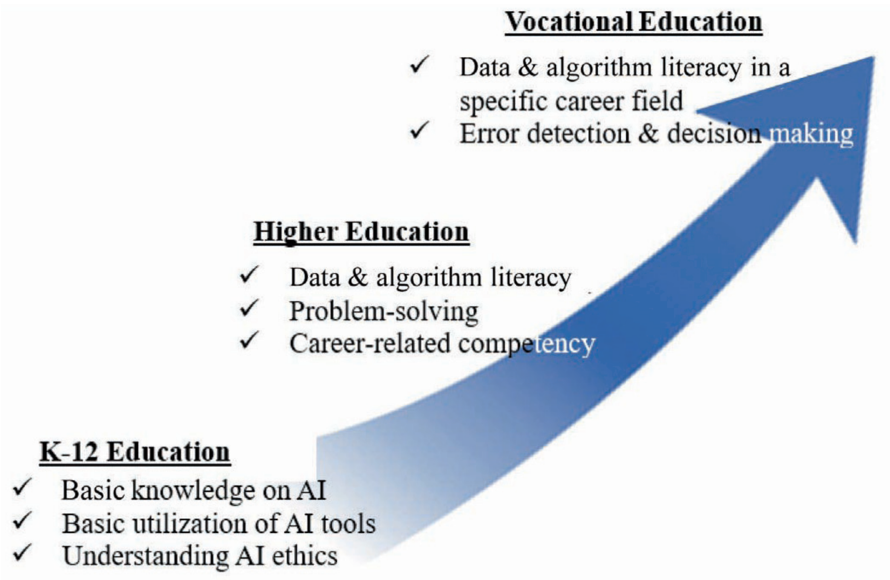


Source: Ng *et al.*, 2021.

Their framework illustrates how AI literacy involves not only the acquisition of knowledge but also the capacity to analyse, evaluate and act responsibly in relation to AI technologies. The authors further proposed three inter-related components – *conceptual*, *practical* and *ethical* – that provide a basis for curriculum design and policy development.

Extending the focus beyond formal education, Chee, Ahn and Lee (2024) frame AI literacy as a lifelong and cross-sectoral capability. They argue that AI literacy must be understood as a competence relevant to all groups in society, each requiring different levels of engagement and cognitive complexity. Education may focus on awareness and responsible use, while professional and policy domains require more advanced analytical, evaluative and ethical capacities.

Figure 6.3: Pathway for Educating Competencies for AI Literacy



Source: Chee, Ahn and Lee, 2024.

These frameworks collectively reflect a growing global consensus. AI literacy is more than technical fluency; it is a structured, developmental capability that moves from knowledge to application to critique and creative engagement.

6.4 AI Literacy Across the Life Course

AI literacy is not a monolithic competency but a differentiated set of capabilities that spans a continuum from foundational awareness to advanced technical proficiency, calibrated to the specific requirements of diverse audiences and contexts. It is also a lifelong learning activity, requiring continuous opportunities for people to develop and update their understanding so that they can engage constructively and ethically with AI as the technology evolves. The *Digital & AI Strategy* 2030 frames AI literacy as a form of critical, ethical and interpretive competence for citizens, learners and businesses, and contains actions to support AI literacy through targeted awareness campaigns for SMEs, curriculum and teacher guidance in education, and national initiatives to strengthen basic digital, media and AI literacy across the life course (Department of the Taoiseach, 2026).

6.4.1 Primary, Secondary & Tertiary Education

Children and adolescents are growing up immersed in AI-mediated environments, often interacting with recommendation algorithms, chat bots or generative AI systems long before they understand how these tools work. The OECD (2026) estimates that student use of generative AI ranges from about 8 per cent in primary education to 70–90 per cent in upper secondary and over 86 per cent in higher education, while around 36 per cent of lower secondary teachers on average report using AI tools, mainly for lesson planning, assessment support and resource design. Embedding AI literacy into primary and secondary education is therefore vital, not only to equip students for future work but also to enable them to become informed, ethical digital citizens. Higher education institutions play a dual role in AI literacy, preparing students for AI-driven careers and equipping them to think critically about AI societal impacts. Students are preparing for a rapidly evolving labour market shaped by automation, algorithmic decision making and digital transformation.

The OECD *Digital Education Outlook 2026* highlights that generative AI offers substantial benefits for personalised learning, teaching productivity and system efficiency, but it also warns that poorly implemented systems can amplify inequities, weaken pedagogy and undermine professional judgment (OECD, 2026). An important finding is that learning gains from generative AI are not evenly distributed; large-scale trials show stronger effects for students with higher prior attainment and higher socio-economic status, indicating that without careful design and targeted support, generative AI risks widening rather than narrowing existing educational gaps. The report indicates that many students are using chatbots to generate complete answers, which can shortcut cognitive effort and reduce deep learning, increasing the likelihood of surface-level engagement rather than conceptual understanding. In contrast, the clear educational advantage of fine-tuned, purpose-built systems co-created with teachers and students – which can be aligned to curricula, restrict direct answer-giving and embed scaffolding and socratic questioning – is highlighted. On that basis, the OECD recommends a shift away from general-purpose chatbots toward rigorously governed, pedagogy-first generative AI tools, strengthened AI literacy for teachers and learners, and robust public oversight. This closely aligns with the Irish Children's Rights Alliance's (2025) call for Government to systematically review and monitor EdTech applications for compliance with children's safety, learning and wellbeing across all educational environments.

International frameworks

UNESCO's *Guidance for generative AI in education and research* (UNESCO, 2023a) sets out a policy framework for the ethical and responsible integration of AI technologies into teaching, learning and academic inquiry. It emphasises that generative AI should enhance human creativity and critical thinking rather than replace them, and it calls on governments to develop national regulations, teacher training programmes and institutional policies to ensure safe and equitable use of AI. Notably, UNESCO recommends a minimum age threshold of 13 years for the independent use of generative AI tools by students, aligning with international standards for digital consent and data protection. This safeguard, the organisation argues, is essential to protect learners' rights, privacy and cognitive development in the face of rapidly evolving AI systems.

In 2024, UNESCO introduced the *AI Competency Framework for Students* (UNESCO, 2024a), a global initiative designed to equip learners to be both responsible users and active co-creators of AI. The framework provides a human-centred, ethics-first roadmap structured around four competency aspects (Table 6.1). Together these competency blocks outline a comprehensive model for cultivating not only technical proficiency, but also for ethical critical and reflective capacities needed to shape AI for the public good.

Table 6.1: AI Competency Framework for Students

Competency aspects	Progression levels		
	Understand	**Apply**	**Create**
• Human-centred mindset	• Human agency	• Human accountability	• Citizenship in the era of AI
• Ethics of AI	• Embodied ethics	• Safe and responsible use	• Ethics by design
• AI techniques and applications	• AI foundations	• Application skills	• Creating AI tools
• AI system design	• Problem scoping	• Architecture design	• Iteration and feedback loops

Source: UNESCO, 2024a.

Complementing the student framework, UNESCO's *AI Competency Framework for Teachers* (UNESCO, 2024b) outlines the knowledge, pedagogical strategies and ethical principles teachers require to integrate AI safety and meaningfully into classrooms. The framework emphasises three core dimensions: fostering teachers' AI literacy and critical understanding of generative tools, equipping them to guide students' responsible engagement with AI, and enabling them to use AI to enhance inclusion, assessment and creativity in teaching practice.

In May 2025 the OECD (2025e) in conjunction with the European Commission published a draft *AI Literacy Framework for Primary and Secondary Education* (*AILit Framework*) for public consultation. The finalised framework will be published in 2026. It emphasises that AI literacy is not solely technical but civic, ethical and creative. It recommends that AI literacy become a foundational competence in primary and secondary curricula, calls for the development of teacher training and professional learning pathways in AI pedagogy, and encourages and investment in high-quality, age-appropriate resources and open learning materials. The framework further recommends national co-ordination mechanisms to ensure coherence between education, technology and data governance policies, and calls for the involvement of students, teachers and wider communities in co-designing AI learning experiences that are inclusive, equitable and relevant.

Building on earlier international efforts, including the UNECSO work, the *AILit Framework* identifies four interrelated domains (*engaging with AI, creating with AI, managing AI, and designing AI*) that describe the diverse ways learners engage with AI, encompassing 22 competencies in total. It recognises that learners may develop proficiency across these domains to varying degrees without necessarily achieving full mastery in any single one. Within the framework, *knowledge, skills, and attitudes* operate as the core building blocks that structure each competence. They ensure that learning addresses conceptual understanding, practical capability and ethical awareness in equal measure. Together, these elements enable learners to engage with AI confidently and responsibly as technologies and contexts evolve, as they invariably will.

Figure 6.4: Dimensions of AI Literacy

Knowledge

The knowledge statements in the framework focus on conceptual knowledge, outlining the technical and societal understandings that learners need to apply and engage with AI systems. These concepts include how AI processes data, how AI differs from human thinking, and how bias can emerge in AI systems.

Skills

The skills demonstrate how fundamental abilities, such as critical thinking, creativity, and computational thinking, apply in an AI context. They guide learners in using AI effectively and ethically, ensuring that learners actively shape how AI fits into their lives.

Attitudes

The attitudes reflect mindsets and dispositions that prepare learners to engage with AI, not only with technical skills, but also with an awareness of AI's impact on themselves and others. These include a sense of curiosity and adaptability in using AI systems, as well as a readiness to question outputs and a commitment to using AI responsibly.

Source: OECD, 2025e.

The EU is advancing a comprehensive agenda on AI and education, with a strong emphasis on AI literacy as a cornerstone of digital readiness. The *Digital Education Action Plan 2021–2027* highlights the need for both learners and educators to develop critical digital and AI competences, supported by investments in infrastructure, teacher training and research (European Commission , 2020a). The EU-funded Artificial Intelligence for and by Teachers (AI4T) project aims to strengthen AI literacy among secondary school teachers by helping them understand core AI concepts, ethical considerations and practical classroom applications. Central to the project is a 'Massive Open Online Course' and an open textbook that provide accessible training on both teaching about AI and teaching with AI (Ai4t.eu, 2023). The project also includes school-based experimentation and evaluation to understand how teachers engage with and apply AI tools in practice. Ireland was one of the five participating countries, contributing to the project's cross-national piloting and insights on effective teacher learning. The Commission's *2030 Roadmap on the Future of Digital Education and Skills*, expected in 2026, is set to strengthen efforts to ensure equal access to AI-enhanced learning and to embed AI literacy across education systems.

National initiatives

At the national level, many governments are taking steps to incorporate AI literacy into formal education frameworks. UNESCO's 2022 global survey on K-12 AI curricula (UNESCO, 2022) found that only 11 countries had developed and officially endorsed AI programmes for primary and secondary education, with a further four in development. The report concluded that, while global awareness of the importance of AI literacy was growing, formal curriculum integration remained limited and uneven. More recent research (Yeter, Yang & Sturgess, 2024; Edwards, 2025) shows that the integration of AI literacy into primary and secondary education is gaining momentum but remains uneven across countries. China and the United Arab Emirates have made AI a mandatory component of their national computing curricula from early grades onward, while Portugal, Singapore and New Zealand have integrated computational thinking, robotics and AI fundamentals across primary and secondary education. South Korea has introduced basic AI principles and ethics into primary school curricula and elective courses at second level.

Universities are developing programmes to support faculty, staff and students, often with an interdisciplinary focus. Many institutions are experimenting with AI-across-the-curriculum approaches, where students in non-technical disciplines learn to use AI tools critically for analysis, drafting and design, while technical students are exposed to ethical, legal and social implications. A 2024 study of university students in the US, UK and Germany identified three distinct groups based on their AI-related cognitive and behavioural traits. These were AI advocates (exhibiting a high level of AI literacy, interest and positive attitudes to the technology), cautious critics (low levels of AI literacy coupled with negative attitudes towards AI, and pragmatic observers (representing an intermediate group with moderate AI literacy and agnostic views towards the technology (Bewersdorff *et al.*, 2024). This suggests that educational strategies need to go beyond teaching technical concepts and need to foster AI literacy and interest to build students' confidence. This is especially important from a labour market perspective as AI literacy needs to be understood as part of a broader digital skills portfolio, thereby future-proofing graduate careers.

National approaches vary but share an emphasis on making complex AI concepts accessible through hands-on, engaging, and ethical learning experiences. Teachers have been identified as the key agents of change in developing AI literacy across educational systems. Thus, engendering understanding, confidence and pedagogical capacity to integrate AI meaningfully are pivotal to ensuring equitable and ethical student engagement with AI (UNESCO, 2024b; OECD, 2025e). In that context it is worth noting that the speed of adoption of AI in the education sector has outpaced the upskilling of educators, many of whom report low AI literacy and uncertainty about how to apply these tools ethically and effectively (UNESCO, 2023a). A persistent weakness in many initiatives is the lack of rigorous assessment frameworks. Few programmes systematically measure what students learn about AI, making it difficult to evaluate the depth of understanding or the long-term impact of AI literacy interventions (Lorena Casal Otero *et al.*, 2023).

Ireland

In Ireland, substantial work is underway to shape policy and practice around the integration of AI in education. Efforts span national strategy, sectoral guidance and institutional initiatives. The publication of the Department of Education and Youth's (2025) *Guidance on Artificial Intelligence in Schools* is an important step within this international landscape. The guidance emphasises safe, ethical and appropriate AI use that supports rather than replaces teachers, prioritising student wellbeing and learning outcomes. Importantly, it situates AI literacy not as a standalone subject but as a transversal competency to be embedded across curricula. The Irish AI Advisory Council statement on education reinforces this approach, characterising AI literacy as a civic skill like reading or critical thinking rather than purely technical competency (AI Advisory Council, 2025a).

This framing situates AI literacy within broader educational objectives of fostering informed, engaged citizenship. The pilot phase of the ADAPT Centre's *AI Literacy in the Classroom* initiative, supported by Google and launched in 2024, involved over 340 teachers. Evaluation data from ADAPT itself shows that 96 per cent of teachers reported improved ability to explain AI concepts, while 92 per cent felt more confident discussing AI with students. Building on this, the programme plans to expand and aims to train a further 500 teachers, with targeted pilots in DEIS schools (Irish Tech News, 2025). This effort is supported by a wider ecosystem of resources, included those curated by Oide, the support service for teachers funded by the Department of Education and Youth.

The Higher Education Authority (HEA) has developed a sector-wide resource portal titled *Artificial Intelligence in Irish Higher Education*, which offers institutional guidelines, open educational resources and policy materials to help staff and students build foundational AI literacy, critically covering both ethical/critical awareness and practical engagement with AI tools. The National Forum for the Enhancement of Teaching and Learning in Higher Education has issued *Ten Considerations for Generative Artificial Intelligence Adoption in Irish Higher Education*, offering practical and ethical guidance for institutions (Higher Education Authority, 2025b). Research commissioned by the HEA stresses that strengthening AI literacy across the sector is essential for a coherent and ethical response to AI. It recommends equipping both students and staff with not only practical skills for using AI tools, but also the critical capacity to understand their limits, risks and implications for academic integrity and learning. The report highlights the need for professional development, updated assessment practices and sector-wide co-ordination to ensure that AI literacy becomes a foundational competence within Irish higher education (O'Sullivan *et al.*, 2025). Irish higher education institutions (HEIs) have also developed academic supports as well as modules designed to integrate AI literacy across diverse disciplines.

In February 2026, a new suite of Further Education and Training (FET) micro-qualifications in AI, developed with Microsoft, were launched to help upskill citizens and businesses in emerging AI technologies, covering topics such as machine learning basics, ethical AI and data analysis. These accredited short courses will be delivered through the network of 16 Education & Training Boards nationwide to help address critical AI skills gaps and strengthen digital capability across the workforce.

While these initiatives are welcome and valuable, they tend to focus on specific skills and use cases, rather than adopting a holistic AI literacy framework that recognises the need for a differentiated range of capabilities, spanning a continuum from foundational awareness to critical understanding and engagement.

6.4.2 Employees & Organisations

In the corporate sphere, AI literacy has transitioned from a niche technical skill to a core business competency. This shift is being driven by a desire to leverage AI for a competitive advantage as well as the legal obligation to ensure its responsible deployment. AI literacy is increasingly being recognised as a requirement across all levels of an organisation to adopt workflows, use AI tools effectively, interpret AI outputs and maintain critical oversight. Roles such as AI champions, AI governance and AI risk functions are becoming more common in organisations in order to lead adoption, tailor training and ensure compliance.

EU AI Act

As previously, mentioned, Article 4 of the *EU AI Act* mandates that providers and deployers of AI systems ensure their staff have a 'sufficient level of AI literacy'. Under Article 4, this obligation falls on providers and developers of AI systems, while in the proposed Digital Omnibus on AI, it is the responsibility of member states to 'encourage' providers and deployers of AI systems to provide AI literacy (European Commission, 2025e).
The specific level and nature of literacy required are not prescribed, leaving flexibility for organisations to tailor their approach based on staff knowledge and on the specific application of the AI system in question. To support organisations in meeting their obligations under Article 4, the EU AI office has established a living repository of best practices in AI literacy. It is notable that most examples are drawn from larger organisations and relatively few examples from small-to-medium enterprises, potentially reflecting the lower levels of adoption of AI in this sector.

Training

Internationally, a diverse market of AI training has emerged to meet the demands of businesses and public bodies. Offerings range from compliance focused e-learning modules to strategic, non-technical diplomas for business leaders and specialised workshops for senior public servants, ensuring that the current workforce can navigate AI's operational, ethical and legal dimensions.

In Ireland, a diverse ecosystem of corporate training has also emerged. This includes CeADAR's free *AI for You: An introduction to AI and the EU AI Act* course, developed in partnership with the Department of Enterprise, Trade and Employment to demystify the regulation for Irish SMEs. The UCD Professional Academy offers a Diploma in AI and Business to prepare organisations to integrate AI technology, taking account of people and processes. For the public sector, the Institution of Public Administration offers a one-day AI masterclass for senior leaders and a practical workshop on implementing the AI guidelines for operational staff.

Senior leaders

Recent literature characterises AI literacy as an increasingly important competence for senior organisational leaders. Studies note that, as AI systems become embedded in core business processes, senior executives are more frequently required to engage with decisions that involve algorithmic outputs, data-driven insights and automated processes. As AI-related decisions can influence areas such as operational performance, compliance and reputational resilience, senior leaders are often expected to understand the strategic implications of AI, its potential contributions to growth and efficiency, and the limitations that may affect its reliability or suitability for specific applications. While detailed technical expertise is not necessarily required, insufficient executive understanding can contribute to fragmented AI initiatives, misaligned investments and governance gaps. Crucially, senior leaders also play an important role in shaping cultural norms, setting expectations around responsible AI use, and communicating strategic priorities. Their engagement is associated with clearer decision-making processes, improved alignment across business units, and more consistent application of safeguards during AI deployment.

The OECD stresses the necessity for leadership knowledge about data inputs, model behaviour and system reliability (OECD, 2025a). The governance domain incorporates executive responsibility for risk management, regulatory compliance, ethical standards and accountability. Likewise, the European Commission guidance emphasises the need for leaders to ensure systems are transparent, traceable and deployed in accordance with legal and organisational requirements (European Commission, 2019).

Many governments and organisations have implemented structured initiatives to strengthen executive AI literacy. Singapore mandates AI literacy training for all civil servants, with dedicated executive-level modules developed by the Smart Nation and Digital Government Group. These modules focus on strategic, governance and assurance aspects of AI, reflecting the country's public-sector governance framework (Smart Nation Singapore, 2020). Telefónica has introduced a *Responsible AI Culture Plan* that incorporates role-specific AI governance training for board members and establishes a Responsible AI Champions Network to promote governance consistency and strategic alignment across leadership levels (UNESCO, 2024c).

6.4.3 Public

Artificial intelligence is transforming public life. Algorithmic systems now mediate access to credit, welfare, information and even justice. For this reason, AI literacy is now a civic necessity, empowering individuals to understand and question the technologies that shape their lives. In October 2025, OpenAI CEO Sam Altman announced that ChatGPT had 800 million weekly active users, although only a small fraction (around 5%) are paid subscribers. Interestingly, most AI interactions today are personal rather than professional; about 70 per cent of ChatGPT use focuses on non-work activities such as advice-seeking, entertainment and self-reflection. Across studies, six broad use categories have emerged: content creation and editing; technical assistance; personal and professional support; learning and education; creativity and recreation; and research and decision-making. Notably, therapeutic, companionship and life-organisation uses are rapidly becoming prominent, indicating a shift from productivity-oriented applications toward emotional and existential support (Chatterji *et al.*, 2025). Evidence on the psychological effects of companion chatbots is mixed; however, some early research suggests they may contribute to loneliness and reduced social interaction for frequent users (Bengio *et al.*, 2026).

Despite the extraordinary number of people using AI tools, recent surveys reveal substantial gaps in public AI understanding, alongside complex, sometimes contradictory attitudes. According to the IPSOS AI Monitor 2025 survey, 65 per cent of Irish participants said they had a good understanding of what AI is (slightly below the 30-country average of 67%). However, when asked whether they knew which products and services use AI, only 43 per cent of Irish respondents said yes, compared with a 30-country average of 52 per cent (Carmichael, 2025). Similar findings emerge in the *Attitudes and Use of Artificial Intelligence: A Global Study 2025* (Gillespie, *et al.*, 2025). The survey reports that, while 52 per cent of Irish respondents feel confident using AI tools effectively, a much smaller share (38%) believe they have the skills and knowledge to use AI appropriately, highlighting a gap between perceived ease of use and deeper understanding. This gap is reinforced by the fact that only 32 per cent have received any form of AI-related training, whether formal or informal. The report concludes that, globally, high levels of adoption are coupled with low levels of AI training and literacy, and that while people may find AI intuitive to use, this does not necessarily translate into knowledge about where and how AI systems are being deployed.

Survey data also reveal additional dimensions of public understanding and acceptance, including notable gender divides in AI attitudes and usage patterns. Research demonstrates that males report higher AI usage, more positive attitudes and less concern about AI chat bots compared to females, who have more concerns regarding transparency and fairness in relation to the technology; these disparities have implications for inclusive AI literacy programme design.

A recent study involving Irish young people aged 13–17 examined their understanding, use and confidence in engaging with AI, with the aim of informing education and policy on AI literacy (Ombudsman for Children Office, 2025). Young people reported using AI regularly for schoolwork, fact-checking, creative projects, entertainment and, in some cases, advice on health and wellbeing. Although they expressed confidence in using AI, they also highlighted risks, including misinformation, bias, over-reliance and inadequate safeguards for younger users. The recommendations made by the Ombudsman Office emphasise the need for structured AI and digital-health literacy education in schools, clearer guidance for safe and age-appropriate use, better support for parents and educators, and stronger transparency and safety measures when AI provides health-related information.

The rationale for public AI literacy operates at multiple levels. At the individual level, AI literacy enables informed decision-making on AI mediated services, products and interactions. At the civic level, AI literacy facilitates meaningful participation in policy debates, regulatory processes and value alignment discussions about AI development and deployment. At the societal level, widespread AI literacy represents a precondition for democratic governance of AI technologies. Simply raising AI literacy does not necessarily guarantee higher adoption or receptivity to the technology. A recent multi-study investigation challenges the common assumption that increasing AI literacy will naturally enhance public receptivity to AI technologies (Tully, Chiara Longoni & Appel, 2025). The authors found that individuals with lower AI literacy consistently report higher openness, usage and positive attitudes towards AI. The research suggests that lower-literacy users may rely on 'magical' or overly optimistic perceptions of AI, whereas higher-literacy individuals tend to hold more calibrated, and sometimes more cautious, views of AI's capabilities and limitations. These findings indicate that, while AI literacy remains essential for informed engagement, it should not be treated as a straightforward lever for increasing adoption. Effective public AI literacy must address knowledge gaps but also perceptions, expectations and trust in AI-enabled systems.

Within the EU, the *Digital Competence Framework for Citizens* (DigComp 3.0), while not strictly an AI-literacy programme, provides foundational digital and data literacy competencies that underpin citizens' ability to critically engage with AI-enabled technologies by systematic and transversal integration of AI across the framework (Cosgrove & Cachia, 2025) The University of Helsinki's *Elements of AI* programme represents a pioneering initiative in mass public AI education. Launched in Finland, the programme has been translated into over thirty languages and has reached around 1 per cent of European Union citizens through a free, accessible online course designed for non-technical audiences. The programme's success demonstrates the feasibility of large-scale public education and provides a model for similar initiatives. Australia's *AI for All* government initiative explicitly targets the general public, recognising that AI literacy should not remain confined to professional or educational contexts. The programme emphasises accessible, practical understanding tailored to everyday AI encounters in consumer products, public services and media.

While Ireland has developed a rich ecosystem of AI literacy initiatives aimed at organisations and the workforce, a comparable set of programmes specifically designed to build AI literacy among the general public is notably lacking. The *National Digital and AI Strategy* contains a commitment 'to ensuring that all learners acquire the basic digital skills, digital literacy skills, and media literacy skills needed to thrive in an AI-driven world' (Department of the Taoiseach, 2026, p.69). The AI Advisory Council has called for a co-ordinated national approach to public AI literacy, positioning it as essential to democratic debate and ethical innovation (AI Advisory Council, 2025b).

Box 6.1: AI in Transport and Logistics

Modern transportation and logistics systems are under immense pressure from rapid urbanisation, population growth and increasing motorisation. Ireland illustrates this strain; in 2025 Dublin was ranked the 11th most congested city globally, with drivers losing approx. 95 hours annually to delays (INRIX, 2025). Artificial intelligence holds the promise of providing data-driven solutions to persistent challenges of congestion, safety, inefficiency and sustainability (World Economic Forum, 2025b).

In real-time traffic management, AI can combine CCTV, roadside sensors and connected vehicle and navigation data to detect incidents earlier and optimise network response. On Ireland's motorway network, Transport Infrastructure Ireland's use of intelligent transportations systems using AI reported incident detection up to 25 minutes earlier on the M1 and 35 minutes earlier on the M6, supporting faster intervention and reduced secondary disruption (Valerann, 2025). In public transport, AI supports demand forecasting, dynamic scheduling and predictive maintenance (e.g. using sensor data to anticipate failures and minimise service disruption) (Son *et al.*, 2025). In logistics, AI improves route planning, fleet maintenance and warehouse operations, reducing empty miles and emissions. Autonomous vehicles (AVs) go further, using AI for perception, prediction and planning, but raise more acute concerns around safety assurance, cybersecurity (evasion/poisoning attacks) and transparency. Under the *EU AI Act*, AI used as a vehicle 'safety component' is classified as high-risk, triggering requirements for risk management, robustness, and human oversight (Fernández Llorca *et al.*, 2025).

Widespread adoption will depend on high-quality interoperable data, resilient connectivity (IoT/5G), rigorous assurance and cybersecurity engineering, clear operational accountability, and harmonised regulation that enables innovation while safeguarding public trust.

Chapter 7: Strategic Reflections & Priority Actions for Navigating AI

The debate around AI is often framed in extremes, as a revolutionary cure-all or as an existential threat to humanity. Neither of these framings is likely to be true and conceptualising the debate in such terms can be unhelpful. It diminishes the role of human agency and risks crowding out the more important discussion about the need to intentionally shape AI in line with our goals and values, and what that requires of policymakers, institutions and society. The impacts of AI will vary across domains and unfold over time in ways that are difficult to predict.

The diffusion and embedding of AI into everyday life, workplaces and the broader economy will take time, creating a critical window in which Ireland can act deliberately rather than reactively. This period should be used to clarify where AI can generate meaningful value, identify the tasks to which it is best suited, and establish agile risk-informed and proportionate governance frameworks that can guide its responsible development and deployment. It also provides time to design mitigation strategies that address emerging risks and unintended consequences, expand AI literacy and technical skills, and support the adaptation of labour markets as roles evolve.

NESC seeks to broaden the debate on AI by emphasising that it should not be seen merely as another tool in the digital toolbox. Instead, AI should be understood as a socio-technical system whose design, application and impacts are shaped by human decisions, institutional and economic incentives, and social norms. The effects of AI are therefore neither automatic nor inevitable; they reflect the priorities embedded within systems, the quality of governance, and the contexts in which AI is applied. Approaching AI in this way brings questions of responsibility, power, equity and accountability to the forefront, and highlights the need for intentional stewardship to ensure that technological advancement aligns with societal values. This framing provides an important foundation for considering how Ireland can guide the development and use of AI in a manner that is both strategic, safe, rights-respecting and aligned with the public interest.

The Council offers five interconnected reflections that can help Ireland pursue a responsible, rights-respecting and inclusive approach to developing and using AI, one which supports productivity, economic prosperity, better public services and wider societal benefits. These reflections establish a strategic framework from which a set of priority actions is identified. While not exhaustive, the actions highlighted here focus on areas where deliberate and timely intervention is likely to be most impactful, building on the imperative for proactive stewardship outlined above. Taken together, the reflections and associated priorities are intended to help the Irish AI ecosystem translate broad ambition into co-ordinated, practical progress.

Reflections are centred on five main themes: **Responsible and Strategic Adoption of AI**; **Safe, Ethical and Trustworthy AI**; **Anticipatory Governance and Institutional Readiness**; **AI Literacy as National Infrastructure**; **Public Deliberation, Legitimacy and Social Licence**. Ireland already has expertise in a number of these areas, providing a good foundation for future policy and implementation efforts. The planned AI Advisory Unit will also play an important role in this regard. These capacities can be strategically leveraged to support the delivery of priority actions, strengthen institutional co-ordination, and accelerate progress towards inclusive, trustworthy and sustainable AI adoption.

Reflection 1: Responsible and Strategic Adoption of AI

A first reflection concerns the need for responsible, strategic and problem-led adoption of AI. Ireland's ambition cannot be realised through a technology-first mindset or by pursuing AI adoption for its own sake. Too often, enthusiasm outpaces organisational readiness, leading to fragmented pilots, wasted investment and erosion of public trust. A sustainable path requires beginning with clearly defined problems and societal needs, and then determining whether AI provides a safe, effective and rights-respecting solution. This approach helps avoid techno-solutionism, opportunity costs and the risk of introducing AI into domains where the conditions for success – including high-quality, curated data and a supportive socio-technical environment – do not exist.

Strategic adoption also requires attention to the type of AI model deployed. Responsible practice involves matching model complexity to problem complexity, selecting energy-efficient tools, and ensuring transparency around environmental impacts. Equally, strategic adoption means focusing on transformation rather than incremental automation. Ireland should be ambitious in its use of AI and think beyond simply streamlining existing processes. We need to rethink how public services and organisational systems could be reorganised to enhance value, inclusion and efficiency using AI tools.

Ultimately, a responsible adoption strategy demands socio-technical integration: investment in data governance, digital infrastructure, strengthened workforce capability, participatory design, and early engagement with employees and affected communities. Without these foundations, AI is unlikely to deliver sustained productivity or public benefit, and risks deepening distrust or embedding inequities into decision-making systems.

Priority Actions

1. *Establish a Problem-First Adoption Framework*

   Work with public and private sector stakeholders to develop a national decision framework enabling organisations, particularly in the public sector, to clearly define the problem to be solved before pursuing AI solutions. This should include structured needs assessments, options analysis (including non-AI alternatives) and explicit tests of public value in the case of public-sector adoption. Embedding a 'problem first' approach can reduce fragmented experimentation and direct investment toward high-impact use cases.

2. *Implement 'Right-Sized' Model Protocols*
   Create a Model Selection Matrix that guides organisations to match model complexity to the scale and sensitivity of the task at hand. This should encourage reflection on environmental impacts and sustainability considerations.

3. *Incentivise Transformational Rather Than Only Incremental Uses*
   Design public funding mechanisms and innovation programmes that reward projects capable of redesigning services or organisational processes, rather than only automating existing workflows. Encouraging system-level redesign can unlock greater long-term value and help Ireland avoid sub-optimal productivity gains.

Reflection 2: Safe, Ethical and Trustworthy AI

A second reflection centres on the imperative for safe, ethical and trustworthy AI. It is important to avoid undertones of techno-solutionism when we speak of ethical AI, as if AI itself had some inherent capability to be ethical. Rather, we need to focus on the integration of human ethical deliberation into AI policy discussions, as well as adoption and oversight of AI systems.

Ensuring trustworthiness requires moving beyond high-level principles toward practical mechanisms such as algorithmic audits, impact assessments, structured documentation and transparent reporting in relation to firms' safety testing procedures and results, and the training data used in model development (e.g. public registry of AI systems). These mechanisms make fairness, accountability and transparency meaningful in practice. They also support the detection of bias – not only algorithmic bias, but also the systemic biases embedded in historical data and human decision-making. Importantly, the goal is not to eliminate bias entirely, which is neither realistic nor a standard met by human systems, but to minimise harm, enhance scrutiny and ensure proportionate and equitable outcomes.

To bridge the principle-to-practice gap, Ireland needs concrete guidance such as, for example, sector-specific playbooks that translate high-level ethical principles into actionable steps for real-world settings. However, such tools alone will not be sufficient. To ensure that ethical AI is not merely performative, organisations must also invest in building ethical capability, both among individual practitioners and within institutions, so that trustworthy AI becomes embedded in everyday decision-making rather than remaining an aspirational ideal.

Trustworthy AI also depends on clear human oversight. Yet oversight cannot be assumed; it requires ensuring that systems are explainable enough for humans to interrogate, and that organisational conditions do not incentivise blind acceptance of AI outputs. There is evidence that uncritical reliance on AI can erode human proficiency, diminish skill over time and weaken epistemic capability. Addressing this requires careful design, training and culture-building, and it demands clarity on who is accountable when AI is used in decision-making. A key distinction must be maintained between trust in AI and trustworthy AI. The policy goal is not to persuade the public to trust AI systems, but to ensure that systems, and the institutions deploying them, are genuinely worthy of trust through verifiable, transparent and responsible practices.

Priority Actions

1. *Build Ethics Capability through Sector-Specific Guidance and Institutional Capacity*

   Work with public and private stakeholders to translate high-level ethical principles into sector-specific playbooks that provide practical, context-sensitive guidance for real-world AI use. Playbooks could include concrete decision tools, escalation pathways, and minimum documentation standards, in line with the *EU AI Act*, to support consistent and defensible practice. To ensure that ethical AI is not merely procedural or performative, organisations must also invest in ethical capability. This could include developing multidisciplinary ethics governance structures, embedding responsible-AI roles within teams, and providing professional training that equips practitioners and leaders to identify trade-offs, interrogate system behaviour and exercise informed judgment.

2. *Embed Human Oversight and Accountability in AI-Assisted Decision-Making*

   In line with EU AI Act requirements, establish clear lines of responsibility so that accountability remains traceable and with identifiable human decision-makers. This should include explicit guidance on when human review is mandatory, who holds final decision authority, and how affected individuals can challenge or seek redress for AI-influenced outcomes. In addition, minimum standards for explainability and interpretability in high-stakes applications should be set in line with EU AI Act requirements, so that human reviewers can meaningfully scrutinise and question system outputs rather than simply endorse them. Oversight frameworks should be supported by appropriate training and workflow design to mitigate automation bias and preserve human judgment and expertise.

3. *Integrate Safe and Ethical AI into Procurement and Funding Criteria*

   Leverage public procurement to promote the development and adoption of trustworthy AI by embedding expectations for safety, transparency and ethical governance within purchasing frameworks. This could be facilitated through a central procurement arrangement, as posited in the National Digital & AI Strategy 2030.

Reflection 3: Anticipatory Governance and Institutional Readiness

A third reflection concerns the need for anticipatory and adaptive governance. The rapid pace, unpredictability and heterogeneous nature of AI technologies mean that governance must be capable of learning, adjusting and responding to emerging risks and opportunities. Ireland is already embedded within the regulatory structure of the *EU AI Act*, which provides a strong baseline for trustworthy AI. The purpose of anticipatory governance is not to 'gold-plate' this regulatory effort, but to complement it with a broader, future-oriented perspective that strengthens institutional resilience and prepares the State for uncertain technological trajectories. Anticipatory governance processes can assess the costs of delayed adoption alongside potential harms and allow precaution to be proportionally balanced with strategic ambition, ensuring Ireland can leverage beneficial innovation opportunities.

Anticipatory governance involves integrating strategic foresight, horizon scanning and scenario planning into policy cycles, enabling policymakers to identify weak signals of change and respond proactively rather than reactively. It also requires institutionalising monitoring and evaluation so that real-world evidence continuously informs decision-making. While AI systems may demonstrate impressive performance under controlled conditions, their behaviour can degrade in dynamic, real-world contexts where variables cannot be easily constrained. For this reason, rigorous piloting, careful evaluation and continuous monitoring are essential to understand how systems operate over time and across diverse populations. Such monitoring cannot be episodic but should be embedded throughout the entire lifecycle of AI systems. This supports early detection of harm, helps scale successful innovations and prevents policy or technological lock-in. It is important that the metrics chosen for evaluation are suitably broad, capturing social and economic impacts as well as technical performance. Ongoing oversight should be matched by systematic and regular sharing of information across organisations and sectors, enabling the development of best practices and helping to operationalise core principles of transparency and accountability.

Governance must also be a whole-of-government endeavour, with clear lines of responsibility and strong co-ordination across departments, regulators and public bodies. While it is understandable and appropriate that much attention has focused on the risks of AI, this should not blind us to the opportunity costs of inaction. Anticipatory governance offers a way to stay agile, avoid technological and policy lock-in, and take advantage, where appropriate, of innovative new AI tools or novel applications of existing tools across different domains. Regulatory sandboxes and testbeds, as provided for in the *EU AI Act*, can support trustworthy AI and regulatory innovation while maintaining safeguards, while modular, adaptive governance frameworks can reduce the risk of rigidity in the face of rapid technological evolution.

Crucially, anticipatory governance expands the view beyond risk mitigation alone. It is concerned with steering AI development toward public benefit, enabling re-imagining of systems and ensuring Ireland can respond effectively to multiple possible futures.

Priority Actions

1. *Institute Strategic Foresight into National AI Governance*

   Establish a dedicated and coherent national AI foresight function with responsibility for horizon scanning, scenario development and long-range analysis of technological, societal and economic impacts. This capability could be integrated into decisions on AI policy and investment through scenario testing, stress-testing against plausible technological trajectories, and explicit assessment of opportunity costs as well as risks. Embedding foresight into routine decision-making would shift AI governance from reactive responses to anticipatory, strategically informed action at the cabinet level.

2. *Institutionalise Life-cycle Monitoring of AI Systems*

   Move beyond point-in-time approval models by requiring continuous, proportionate evaluation of AI systems once deployed. Monitoring frameworks should track not only technical performance but also social outcomes, distributional effects and unintended consequences. This supports early harm detection, enables timely recalibration and reduces the risk of technological or policy lock-in. Obligations under the EU AI Act relating to post-market monitoring systems can be leveraged to support and institutionalise these continuous evaluation practices.

3. *Establish a National AI Evaluation and Learning Framework*
   Develop and publish a cross-sector national framework for evaluating AI deployments that defines shared metrics and methodologies for assessing public value, equity, safety, environmental impacts and economic effects, alongside technical performance. This framework should be supported by systematic knowledge-sharing mechanisms that enable regular exchange of evaluation results, operational lessons and incident reports across departments, regulators and sectors.

Reflection 4: AI Literacy as National Infrastructure

A fourth reflection highlights AI literacy as a form of national digital infrastructure, essential for responsible innovation, democratic engagement and organisational readiness. Seen through this lens, AI literacy initiatives should be grounded in a clear public service mandate, designed by independent expertise, adapted to local needs, and subject to strong public accountability. AI literacy is not simply knowledge of tools or technical concepts; it is a socio-technical capability that enables individuals to interpret outputs critically, understand system limitations, identify opportunities, recognise ethical implications and participate effectively in decisions about AI procurement and deployment.

Ireland has a growing ecosystem of AI literacy initiatives across education, enterprise and civil society, but they remain fragmented. A co-ordinated national approach is needed to embed AI literacy across all levels of education, professional training and public engagement in line with the European Union's *Digital Decade* framework, under which member states have committed to ensuring that at least 80 per cent of adults possess basic digital skills by 2030. A national approach should include age-appropriate curricula in schools; accredited programmes and continuing professional development for educators; expanded AI-related training across disciplines such as law, health, humanities and public administration; and, critically, sustained AI literacy initiatives for the general public.

Leadership literacy is particularly important. Executives and senior public-sector leaders shape organisational culture and determine how AI is procured, governed and used. Without AI-literate leadership, organisations risk adopting systems they cannot adequately assess, oversee or evaluate. Embedding AI literacy within risk management, audit processes and governance frameworks is therefore integral to ensuring responsible deployment.

A national commitment to AI literacy would empower citizens to critically evaluate AI, to be appropriately trusting or distrustful where warranted, and to take an active role in shaping Ireland's AI future rather than being passive recipients of technological change.

Priority Actions

1. *Implement a Comprehensive National AI Literacy Strategy*

   Adopt a whole-of-society national AI literacy strategy that defines core competencies, sets measurable objectives and aligns efforts across education systems, workforce development, public services and civic engagement. The strategy should be delivered through sustained public AI literacy initiatives that provide accessible learning resources, community-based programmes and trusted information campaigns aimed at enabling informed, critical engagement with AI. To ensure coherence and quality, establish a national AI literacy hub to leverage existing initiatives in the first instance, to curate high-quality materials, share best practices and co-ordinate initiatives across government, business, academia and civil society. Treat AI literacy as long-term national infrastructure by introducing periodic assessments to track literacy levels, identify demographic and regional gaps, and guide targeted interventions. Throughout, prioritise inclusion to prevent a new digital divide, ensuring that AI understanding and capability are equitably distributed across age groups, regions and socio-economic backgrounds.

2. *Embed AI Literacy as a Core Expectation for Senior Leadership and Governance*

   Foster AI literacy as a standard component of effective leadership and governance for senior public-sector leaders, board members of state bodies and executives in regulated sectors. This should be reflected in leadership development pathways and board education, with a focus on strategic judgment, procurement scrutiny, opportunity and risk evaluation, and governance, rather than only on the technical capabilities of AI systems. Organisations should incorporate AI literacy into routine governance and risk practices, including audit committees, risk frameworks and assurance processes, so that senior decision-makers are equipped to interrogate AI-enabled systems, avoid uncritical adoption or vendor over-reliance, and exercise informed oversight and accountability.

Reflection 5: Public Deliberation, Legitimacy and Social Licence

The final reflection concerns public deliberation and the broader question of social licence. Artificial intelligence has the potential to reshape society in ways that are distributed unevenly, creating different opportunities and risks across communities. What counts as AI for the public good cannot be determined solely by experts, industry or government; it must be shaped through sustained, inclusive engagement with the public. In this regard, the Council welcomes the commitment in the *National Digital & AI Strategy 2030* to launch a National Conversation on AI to ensure societal values and concerns can directly inform the adoption of AI technologies.

Public deliberation must be more than awareness-raising or consultation. It requires meaningful two-way dialogue that recognises diverse values, lived experiences and perspectives, and, in the context of national policymaking, may take different forms and employ different methodologies depending on the issue, scale and level of public impact involved. Citizens should have an informed role in determining where AI should or should not be used, what boundaries should be set, and what trade-offs – be they ethical, social or economic – are acceptable. Without this engagement, AI systems risk rejection, resistance or loss of legitimacy, regardless of their technical performance.

Deliberation is also essential for navigating contested issues such as the balance between innovation and rights protection, concerns about surveillance or misinformation, the impact on labour markets, and questions of environmental sustainability. By embedding public deliberation into governance cycles, Ireland can ensure that AI development is aligned with democratic values, strengthens institutional trust and gives citizens agency in shaping technological futures.

Priority Actions

1. *Integrate Inclusive Public Deliberation in AI Governance*

   Integrate structured public deliberation into AI policy, regulatory and high-risk public-sector deployment cycles, positioning engagement upstream at defined stages of decision-making rather than after choices have been made. In a sociotechnical framing of AI, where impacts emerge from the interaction between technology, institutions and society, ongoing public deliberation is a necessary condition for legitimate and effective governance. Engagement processes should prioritise inclusion and representativeness, and be treated as a continuous democratic practice, with sustained channels for dialogue that evolve alongside AI systems and reinforce public trust over time.

2. *Engage Workers and Communities Affected by AI in Deliberative Dialogue*

   Prioritise early and ongoing dialogue with workers and communities likely to experience the direct impacts of AI deployment, particularly in sectors and local settings where changes to roles, services and decision-making will be most tangible. Supporting deliberation at workplace, sectoral and community levels can highlight lived experience, practical concerns and context-specific opportunities and risks that are often missed by national processes. Building trust and mutual understanding through sustained discussion is critical to securing co-operation and ensuring that AI adoption is socially legitimate and operationally effective.

Concluding Remarks

The path forward is not about allowing AI to determine our future but about defining our future with AI. The progress of this technology is non-linear, and history should make us cautious about predicting what it will or will not achieve. Rather than treating the complexity of AI as a source of apprehension, we should recognise it as a marker of opportunity. The question is not whether AI will match or surpass human intelligence, especially when such comparisons often rely on opaque or narrow benchmarks, but rather how we understand the different forms of intelligence involved. It is important that we do not conflate intelligence (either human or machine) with wisdom, which remains a uniquely human trait. By focusing on how human and AI capabilities can be harnessed together in safe, ethical and purposeful ways, we can ensure that AI becomes a tool for human flourishing – advancing social wellbeing, economic prosperity and democratic values.

Bibliography

Acemoglu, D. (2024) The simple macroeconomics of AI. Massachusetts Institute of Technology. Available at: https://economics.mit.edu/sites/default/files/2024-04/The%20Simple%20Macroeconomics%20of%20AI.pdf (Accessed: 7 January 2026).

Acemoglu, D. and Restrepo, P. (2019) Robots and jobs: Evidence from US labor markets. National Bureau of Economic Research Working Paper No. 23285. Cambridge, MA: NBER. Available at: https://www.nber.org/papers/w23285 (Accessed: 30 August 2025).

African Union (2024) Continental artificial intelligence strategy: Harnessing AI for Africa's development and prosperity. Available at: https://au.int/sites/default/files/documents/44004-doc-EN-_Continental_AI_Strategy_July_2024.pdf (Accessed: 20 August 2025).

Aghion, P. and Bunel, S. (2024) AI and growth: Where do we stand? Federal Reserve Bank of San Francisco. Available at: https://www.frbsf.org/wp-content/uploads/AI-and-Growth-Aghion-Bunel.pdf (Accessed: 12 January 2026).

Agüera y Arcas, B. and Norvig, P. (2023) 'Artificial general intelligence is already here', Noema, 10 October. Available at: https://www.noemamag.com/artificial-general-intelligence-is-already-here/ (Accessed: 12 August 2025).

AI Advisory Council (2025a) AI and education. Department of Enterprise, Tourism and Employment. Available at: https://enterprise.gov.ie/en/publications/publication-files/ai-advisory-council-ai-on-education-paper.pdf (Accessed: 21 August 2025).

AI Advisory Council (2025b) Ireland's AI Advisory Council recommendations: Helping to shape Ireland's AI future. Dublin: Department of Enterprise, Trade and Employment. Available at: https://assets.gov.ie/static/documents/Irelands_AI_Advisory_Council_Recommendations_Helping_to_Shape_Irelands_AI_Future.pdf (Accessed: 19 September 2025).

AI Advisory Council (2026) AI Advisory Council recommendations to Government regarding alleged creation and public dissemination of AI-generated non-consensual intimate images, including child sexual abuse material ('CSAM'). Government of Ireland. Available at: https://assets.gov.ie/static/documents/8a7c0d52/AI_Advisory_Council_Paper_January_2026.pdf (Accessed: 20 January 2026).

AI4T (2023) 'Resources – AI4T project'. Available at: https://www.ai4t.eu/resources/ (Accessed: 14 October 2025).

Aijaz, N., Lan, H., Raza, T., Yaqub, M., Iqbal, R. and Pathan, M.S. (2025) 'Artificial intelligence in agriculture: Advancing crop productivity and sustainability', Journal of Agriculture and Food Research. doi: 10.1002/fer3.59.

Ajder, H., Patrini, G., Cavalli, F. and Cullen, L. (2019) The state of deepfakes: Landscape, threats, and impact. Available at: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf (Accessed: 27 August 2025).

Allen, S. (2025) The misaligned compass: How Europe's quest for digital competitiveness and sovereignty is going off track. Institute of International and European Affairs. Available at: https://www.iiea.com/publications/the-misaligned-compass-how-europes-quest-for-digital-competitiveness-and-sovereignty-is-going-off-track (Accessed: 18 December 2025).

Alowais, S.A., Alghamdi, S.S., Alsuhebany, N. et al. (2023) 'Revolutionizing healthcare: The role of artificial intelligence in clinical practice', BMC Medical Education, 23. Available at: https://doi.org/10.1186/s12909-023-04698-z.

Anastasiou, E., Fountas, S., Voulgaraki, M. et al. (2023) 'Precision farming technologies for crop protection: A meta-analysis', Smart Agricultural Technology, 5, p. 100323. Available at: https://doi.org/10.1016/j.atech.2023.100323.

Anthropic (2026) 'Claude's constitution'. Available at: https://www.anthropic.com/constitution (Accessed: 28 January 2026).

Association for the Advancement of Artificial Intelligence (2025) AAAI 2025 Presidential Panel on the future of AI research. Available at: https://aaai.org/wp-content/uploads/2025/03/AAAI-2025-PresPanel-Report-Digital-3.7.25.pdf (Accessed: 22 September 2025).

Australian Government (2024) Voluntary AI safety standard. Available at: https://www.industry.gov.au/sites/default/files/2024-09/voluntary-ai-safety-standard.pdf (Accessed: 20 August 2025).

Autor, D.H. and Thompson, N. (2025) 'Expertise', Journal of the European Economic Association, 23(4), pp. 1203–1271. Available at: https://doi.org/10.1093/jeea/jvaf023.

Bainbridge, L. (1983) 'Ironies of automation', Automatica, 19(6), pp. 775–779. Available at: https://doi.org/10.1016/0005-1098(83)90046-8.

Bajwa, J., Munir, U., Nori, A. and Williams, B. (2021) 'Artificial intelligence in healthcare: Transforming the practice of medicine', Future Healthcare Journal, 8(2), pp. 188–194. Available at: https://doi.org/10.7861/fhj.2021-0095.

Ballot Jones, L., Thornton, J. and De Silva, D. (2025) 'Limitations of risk-based artificial intelligence regulation: A structuration theory approach', Discover Artificial Intelligence, 5(14). Available at: https://doi.org/10.1007/s44163-025-00233-9.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö. and Mariman, R. (2024) Generative AI can harm learning. SSRN. Available at: https://doi.org/10.2139/ssrn.4895486 (Accessed: 26 February 2026).

Bauer, E., Greiff, S., Graesser, A.C., Scheiter, K. and Sailer, M. (2025) 'Looking beyond the hype: Understanding the effects of AI on learning', Educational Psychology Review, 37(2). Available at: https://doi.org/10.1007/s10648-025-10020-8.

Bean, A.M., Kearns, R.O., Romanou, A. et al. (2025) 'Measuring what matters: Construct validity in large language model benchmarks', arXiv. Available at: https://arxiv.org/abs/2511.04703v1 (Accessed: 10 January 2026).

Bean, A.M., Payne, R.E., Parsons, G., Kirk, H.R., Ciro, J., Mosquera-Gómez, R., Sara, H.M., Ekanayaka, A.S., Tarassenko, L., Rocher, L. and Mahdi, A. (2026) 'Reliability of LLMs as medical assistants for the general public: A randomized preregistered study', Nature Medicine. Available at: https://doi.org/10.1038/s41591-025-04074-y.

Belcak, P., Heinrich, G., Fu, Y., Dong, X., Muralidharan, S., Lin, Y.C. and Molchanov, P. (2025) 'Small language models are the future of agentic AI', arXiv. Available at: https://arxiv.org/abs/2506.02153 (Accessed: 11 November 2025).

Bengio, Y. (2025) 'The first international AI safety report', SuperIntelligence – Robotics – Safety and Alignment, 2(2). Available at: https://doi.org/10.70777/si.v2i2.14755.

Bengio, Y., Clare, S., Prunkl, C. et al. (2026) International AI Safety Report 2026. DSIT 2026/001. Available at: https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026 (Accessed: 5 February 2026).

Bewersdorff, A., Hornberger, M., Nerdel, C. and Schiff, D. (2024) 'AI advocates and cautious critics: How AI attitudes, AI interest, use of AI, and AI literacy build university students' AI self-efficacy', Computers and Education: Artificial Intelligence. Available at: https://doi.org/10.1016/j.caeai.2024.100340.

Bhuiyan, J. (2025) 'Character.AI bans users under 18 after being sued over child's suicide', The Guardian, 29 October. Available at: https://www.theguardian.com/technology/2025/oct/29/character-ai-suicide-children-ban (Accessed: 2 November 2025).

Blanco-González, A., Cabezón, A., Seco-González, A., Conde-Torres, D., Antelo-Riveiro, P., Piñeiro, Á. and Garcia-Fandino, R. (2023) 'The role of AI in drug discovery: Challenges, opportunities, and strategies', Pharmaceuticals, 16(6), p. 891. Available at: https://doi.org/10.3390/ph16060891.

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H. and Krathwohl, D.R. (1956) Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. New York: Longman.

Booth, R. (2025) 'ChatGPT offered bomb recipes and hacking tips during safety tests', The Guardian, 28 August. Available at: https://www.theguardian.com/technology/2025/aug/28/chatgpt-offered-bomb-recipes-and-hacking-tips-during-safety-tests (Accessed: 23 September 2025).

Bostrom, N. (2014) Superintelligence: Paths, dangers, strategies. Oxford: Oxford University Press.

Bowen, D.E. III, Price, S.M., Stein, L.C.D. and Yang, K. (2025) 'Measuring and mitigating racial bias in large language model mortgage underwriting', in 31st Annual European Real Estate Society Conference. Available at: https://eres.architexturez.net/doc/oai-eres-id-eres2025-75 (Accessed: 4 September 2025).

Bradford, A. (2024) 'The false choice between digital regulation and innovation', Northwestern University Law Review, 119(2). Available at: https://scholarlycommons.law.northwestern.edu/nulr/vol119/iss2/3/ (Accessed: 22 November 2025).

Bria, F., Timmers, P. and Gernone, F. (2025) EuroStack – A European alternative for digital sovereignty. Available at: https://eurostack.eu/eurostack-a-european-alternative-for-digital-sovereignty-francesca-bria-bertelsmann-stiftung/ (Accessed: 18 January 2026).

Brollo, F., Dabla-Norris, E., de Mooij, R., Garcia-Macia, D., Hanappi, T., Liu, L. and Nguyen, A.D.M. (2024) Broadening the gains from generative AI: The role of fiscal policies. Washington, DC: International Monetary Fund. Available at: https://www.imf.org/-/media/Files/Publications/SDN/2024/English/SDNEA2024002.ashx (Accessed: 29 August 2025).

Brynjolfsson, E., Rock, D. and Syverson, C. (2021) 'The productivity J-curve: How intangibles complement general purpose technologies', American Economic Journal: Macroeconomics, 13(1), pp. 333–372. Available at: https://doi.org/10.1257/mac.20180386.

Brynjolfsson, E., Chandar, B., Chen, R. et al. (2025a) Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence. Available at: https://digitaleconomy.stanford.edu/wp-content/uploads/2025/08/Canaries_BrynjolfssonChandarChen.pdf (Accessed: 22 August 2025).

Brynjolfsson, E., Li, D. and Raymond, L. (2025b) 'Generative AI at work', The Quarterly Journal of Economics, 140(2). Available at: https://doi.org/10.1093/qje/qjae044.

Bubeck, S., Chandrasekaran, V., Eldan, R. et al. (2023) 'Sparks of artificial general intelligence: Early experiments with GPT-4', arXiv. Available at: https://arxiv.org/abs/2303.12712 (Accessed: 12 August 2025).

Budzyń, K., Romańczyk, M., Kitala, D. et al. (2025) 'Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: A multicentre observational study', The Lancet Gastroenterology and Hepatology, 10(10). Available at: https://doi.org/10.1016/S2468-1253(25)00133-5.

Business Outstanders (2025) 'Infrastructure of the future: Chaslau Koniukh tells how Europe is building "sovereign AI"', Business Outstanders. Available at: https://businessoutstanders.com/artificial-intelligence/europe-ai-infrastructure-and-regulation-2025 (Accessed: 12 September 2025).

Capgemini Research Institute (2025) Harnessing the value of AI: Unlocking scalable advantage. Available at: https://www.capgemini.com/wp-content/uploads/2025/09/Final-Web-version-Report-Gen-AI-in-Organizations.pdf (Accessed: 3 September 2025).

Cardona, M., Rodríguez, R. and Ishmael, K. (2023) Artificial intelligence and the future of teaching and learning: Insights and recommendations. U.S. Department of Education, Office of Educational Technology. Available at: https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf (Accessed: 13 August 2025).

Carmichael, M. (2025) The Ipsos AI monitor 2025: A 30-country Ipsos Global Advisor survey. Available at: https://resources.ipsos.com/rs/297-CXJ-795/images/Ipsos-AI-Monitor-2025.pdf (Accessed: 28 October 2025).

Cazzaniga, M., Jaumotte, F., Li, L. et al. (2024) Gen-AI: Artificial intelligence and the future of work. IMF Staff Discussion Note No. SDN/2024/001. Washington, DC: International Monetary Fund. Available at: https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379 (Accessed: 26 February 2026).

Center for Countering Digital Hate (2026) 'Grok floods X with sexualized images of women and children', Center for Countering Digital Hate. Available at: https://counterhate.com/research/grok-floods-x-with-sexualized-images/ (Accessed: 28 January 2026).

Central Statistics Office (2025a) Key findings: Data centres metered electricity consumption 2024. Available at: https://www.cso.ie/en/releasesandpublications/ep/p-dcmec/datacentresmeteredelectricityconsumption2024/keyfindings/ (Accessed: 25 August 2025).

Central Statistics Office (2025b) Artificial intelligence: Information society statistics – enterprises 2024. Available at: https://www.cso.ie/en/releasesandpublications/ep/p-isse/informationsocietystatistics-enterprises2024/artificialintelligence/ (Accessed: 16 August 2025).

Challapally, A., Pease, C., Raskar, R. and Chari, P. (2025) The GenAI divide: State of AI in business 2025. MIT NANDA. Available at: https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf (Accessed: 30 October 2025).

Chatterji, A., Cunningham, T., Deming, D. et al. (2025) How people use ChatGPT. NBER Working Paper No. 34255. Available at: https://doi.org/10.3386/w34255 (Accessed: 26 February 2026).

Chebrolu, K., Ressler, D. and Varia, H. (2020) 'Smart use of artificial intelligence in health care', Deloitte Insights. Available at: https://www.deloitte.com/us/en/insights/industry/health-care/artificial-intelligence-in-health-care.html (Accessed: 12 August 2025).

Chee, H., Ahn, S. and Lee, J. (2024) 'A competency framework for AI literacy: Variations by different learner groups and an implied learning pathway', British Journal of Educational Technology. Available at: https://doi.org/10.1111/bjet.13556.

Chen, W.X., Srinivasan, S. and Zakerinia, S. (2024) Displacement or complementarity? The labour market impact of generative AI. Harvard Business School Working Paper No. 25-039. Available at: https://www.hbs.edu/ris/Publication%20Files/25-039_05fbec84-1f23-459b-8410-e3cd7ab6c88a.pdf (Accessed: 31 August 2025).

Children's Rights Alliance (2025) Online safety monitor 2025. Available at: https://childrensrights.ie/publications/online-safety-monitor/ (Accessed: 14 November 2025).

Chiu, T.K.F. (2025) 'AI literacy and competency: Definitions, frameworks, development and future research directions', Interactive Learning Environments, 33(5), pp. 3225–3229. Available at: https://doi.org/10.1080/10494820.2025.2514372.

Choi, R.Y., Coyner, A.S., Kalpathy-Cramer, J., Chiang, M.F. and Campbell, J.P. (2020) 'Introduction to machine learning, neural networks, and deep learning', Translational Vision Science and Technology, 9(2), p. 14. Available at: https://doi.org/10.1167/tvst.9.2.14.

Chun, J., Schroeder, C. and Elkins, K. (2024) 'Comparative global AI regulation: Policy perspectives from the EU, China, and the US', arXiv. Available at: https://arxiv.org/abs/2410.21279 (Accessed: 21 August 2025).

Cloud, A., Le, M., Chua, J. et al. (2025) 'Subliminal learning: Language models transmit behavioral traits via hidden signals in data', arXiv. Available at: https://arxiv.org/abs/2507.14805 (Accessed: 22 August 2025).

Commission for Regulation of Utilities (2025a) Large energy users connection policy. Available at: https://cruie-live-96ca64acab2247eca8a850a7e54b-5b34f62.divio-media.com/documents/CRU2025236_Large_Energy_User_connection_policy_decision_paper.pdf (Accessed: 10 January 2026).

Commission for Regulation of Utilities (2025b) 'CRU approves record investment in Ireland's electricity grid and network', CRU. Available at: https://www.cru.ie/about-us/news/cru-approves-record-investment-in-irelands-electricity-grid-and-network/ (Accessed: 24 February 2026).

Commissioner for Human Rights (2025) The human line: Safeguarding rights and democracy in the AI era: Meeting report. Available at: https://rm.coe.int/meeting-report-navigating-the-future-human-rights-in-the-face-of-emerg/488028f7a0 (Accessed: 10 November 2025).

Cosgrove, J. and Cachia, R. (2025). DigComp 3.0: European Digital Competence Framework. JRC Publications Repository. [online] doi: https://doi.org/10.2760/0001149 (Accessed: 11 January 2026).

Cotton, S. (2024) 'Why must workers be included in decision-making', World Economic Forum. Available at: https://www.weforum.org/stories/2024/01/rebuilding-trust-workers-included-decision-making/ (Accessed: 1 September 2025).

Council of Europe (2024a) Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Strasbourg: Council of Europe. Available at: https://rm.coe.int/1680afae3c (Accessed: 23 August 2025).

Council of Europe (2024b) 'HUDERIA: New tool to assess the impact of AI systems on human rights', Council of Europe portal. Available at: https://www.coe.int/en/web/portal/-/huderia-new-tool-to-assess-the-impact-of-ai-systems-on-human-rights (Accessed: 10 September 2025).

Council of Europe (2024c) Report on the application of artificial intelligence in healthcare and its impact on the doctor–patient relationship. Available at: https://edoc.coe.int/en/health-care/12135-report-on-the-application-of-artificial-intelligence-in-healthcare-and-its-impact-on-the-patient-doctor-relationship.html (Accessed: 25 August 2025).

Coupé, T. and Wu, W. (2025) The impact of generative AI on productivity: Results of an early meta-analysis. University of Canterbury. Available at: https://repec.canterbury.ac.nz/cbt/econwp/2509.pdf (Accessed: 23 October 2025).

Cyberspace Administration of China (2025) 'Notice on printing and distributing the "Measures for the identification of artificial intelligence generated synthetic content"'. Available at: https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm (Accessed: 7 January 2026).

Dalal, M. and Mittal, P. (2025) 'A systematic review of deep learning-based object detection in agriculture: Methods, challenges, and future directions', Computers, Materials and Continua, 84(1), pp. 57–91. Available at: https://doi.org/10.32604/cmc.2025.066056.

Danton, A., Roux, J.-C., Dance, B., Cariou, C. and Lenain, R. (2020) 'Development of a spraying robot for precision agriculture: An edge following approach', in 2020 IEEE Conference on Control Technology and Applications (CCTA), pp. 267–272. Available at: https://doi.org/10.1109/CCTA41146.2020.9206304.

Data Protection Commission (2025) Public attitudes survey. Available at: https://www.dataprotection.ie/sites/default/files/uploads/2025-06/DPC_Public_Attitudes_Report_May_2025_updated.pdf (Accessed: 8 September 2025).

Davies, C. and Jung-a, S. (2024) 'South Korea's plan for AI textbooks hit by backlash from parents', Financial Times. Available at: https://www.ft.com/content/1f5c5377-5e85-4174-a54f-adc8f19fa5cb (Accessed: 19 August 2024).

de Vries, A. (2023) 'The growing energy footprint of artificial intelligence', Joule, 7(10). Available at: https://doi.org/10.1016/j.joule.2023.09.004.

Dell'Acqua, F., Saran, A., McFowland, R. et al. (2023) Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Working Paper No. 24-013. Available at: https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf (Accessed: 22 September 2025).

Dellibarda Varela, I., Romero-Sorozabal, A., Rocon, E. and Cebrian, M. (2025) 'Rethinking the illusion of thinking', arXiv. Available at: https://arxiv.org/abs/2507.01231 (Accessed: 25 October 2025).

Department for Business and Trade (2025) The evaluation of the M365 Copilot pilot in the Department for Business and Trade. Available at: https://assets.publishing.service.gov.uk/media/68adbe409e1cebdd2c96a19d/dbt-microsoft-365-copilot-evaluation.pdf (Accessed: 20 September 2025).

Department for Science, Technology and Innovation (2023) A pro-innovation approach to AI regulation. GOV.UK. Available at: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper (Accessed: 26 November 2025).

Department of Education and Youth (2025) Guidance on artificial intelligence in schools. Available at: https://assets.gov.ie/static/documents/dee23cad/Guidance_on_Artificial_Intelligence_in_Schools_2025.pdf (Accessed: 1 November 2025).

Department of Enterprise, Trade and Employment (2021) AI – here for good: A national artificial intelligence strategy for Ireland. Available at: https://enterprise.gov.ie/en/Publications/Publication-files/National-AI-Strategy.pdf (Accessed: 26 February 2026).

Department of Enterprise, Tourism and Employment (2024) Ireland's national AI strategy: AI – here for good (refresh 2024). Available at: https://enterprise.gov.ie/en/publications/publication-files/national-ai-strategy-refresh-2024.pdf (Accessed: 24 August 2025).

Department of Enterprise, Trade and Employment (2026a) LEAP – large energy user action plan. Available at: https://enterprise.gov.ie/en/publications/leap.html (Accessed: 18 January 2026).

Department of Enterprise, Tourism and Employment (2026b) General scheme of the Regulation of Artificial Intelligence Bill 2026. Available at: https://enterprise.gov.ie/en/legislation/general-scheme-of-the-regulation-of-artificial-intelligence-bill-2026.html (Accessed: 6 February 2026).

Department of Finance and Department of Enterprise, Trade and Employment (2024) Artificial intelligence: Friend or foe? Summary and public policy considerations. Available at: https://assets.gov.ie/static/documents/artificial-intelligence-friend-or-foe-an-analysis-of-how-ai-could-impact-irelands-labo.pdf (Accessed: 27 August 2025).

Department of Finance (2026) Economic insights: Volume 1 2026. Government of Ireland. Available at: https://assets.gov.ie/static/documents/391b8952/Economic_Insights_Volume_1_2026.pdf (Accessed: 20 February 2026).

Department of Health (2024) Digital for care: A digital health framework for Ireland 2024–2030. Available at: https://www.gov.ie/en/department-of-health/publications/digital-for-care-a-digital-health-framework-for-ireland-2024-2030/ (Accessed: 13 August 2025).

Department of Health (2026). AI for Care - The Artificial Intelligence (AI) Strategy for Healthcare in Ireland 2026 - 2030. [online] gov.ie. Available at: https://www.gov.ie/en/department-of-health/publications/ai-for-care-the-artificial-intelligence-ai-strategy-for-healthcare-in-ireland-2026-2030 (Accessed 12 March 2026).

Department of Public Expenditure, Infrastructure, Public Service Reform and Digitalisation (2023) Digital for good: Ireland's digital inclusion roadmap. Available at: https://www.gov.ie/en/department-of-public-expenditure-infrastructure-public-service-reform-and-digitalisation/publications/digital-for-good-irelands-digital-inclusion-roadmap/ (Accessed: 2 September 2025).

Department of Public Expenditure, National Development Plan Delivery and Reform (2025) Guidelines for the responsible use of artificial intelligence in the public service. Available at: https://assets.gov.ie/static/documents/09fe3ad4/Guidelines_for_the_Responsible_Use_of_AI_in_the_Public_Service_20250918.pdf (Accessed: 1 September 2025).

Department of Rural and Community Development and the Gaeltacht (2025) 'Minister Calleary announces €4.9m in funding for Irish language digital projects in Dublin City University', Press release. Available at: https://www.gov.ie/en/department-of-rural-and-community-development-and-the-gaeltacht/press-releases/minister-calleary-announces-49m-in-funding-for-irish-language-digital-projects-in-dublin-city-university/ (Accessed: 26 November 2025).

Department of the Taoiseach (2022) Harnessing digital: The digital Ireland framework. Available at: https://assets.gov.ie/static/documents/harnessing-digital-the-digital-ireland-framework.pdf (Accessed: 20 August 2025).

Department of the Taoiseach (2025) Three pillars of policy development. Available at: https://www.gov.ie/en/department-of-the-taoiseach/publications/three-pillars-of-policy-development/ (Accessed: 23 July 2025).

Department of the Taoiseach (2026) Digital Ireland – connecting our people, securing our future. Government of Ireland. Available at: https://www.gov.ie/en/department-of-the-taoiseach/campaigns/digital-ireland-connecting-our-people-securing-our-future/ (Accessed: 19 February 2026).

Dillon, E., Donnellan, T., Moran, B. and Lennon, J. (2025) Preliminary results: Teagasc National Farm Survey 2024. Teagasc. Available at: https://teagasc.ie/wp-content/uploads/uploads/NFS-Preliminary-Report-2024.pdf (Accessed: 1 September 2025).

Draghi, M. (2024) The future of European competitiveness: Part A – a competitiveness strategy for Europe. European Union. Available at: https://commission.europa.eu/topics/competitiveness/draghi-report_en (Accessed: 9 September 2025).

Eastwood, B. (2025) '10 AI healthcare trends to watch in 2025 and beyond', Health IT and EHR. Available at: https://www.techtarget.com/searchhealthit/feature/AI-healthcare-trends-to-watch (Accessed: 26 February 2026).

Ebers, M. (2024) 'Truly risk-based regulation of artificial intelligence: How to implement the EU's AI Act', European Journal of Risk Regulation, 16(2), pp. 1–20. Available at: https://doi.org/10.1017/err.2024.78.

Edwards, R. (2025) Global integration of AI education. Available at: https://iorma.com/wp-content/uploads/2025/06/Global-Integration-of-AI-Education-4.pdf (Accessed: 7 November 2025).

eHealth Ireland (2023) 'How AI is changing patient care: A conversation with Professor Peter MacMahon'. Available at: https://www.ehealthireland.ie/news-media/news/2025/how-ai-is-changing-patient-care-a-conversation-with-professor-peter-macmahon/ (Accessed: 22 September 2025).

Elsevier (2024) Insights 2024: Attitudes toward AI. Available at: https://assets.ctfassets.net/o78em1y1w4i4/6BWRibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (Accessed: 20 August 2025).

ESOFT (2024) 'Ireland's digital transformation in agriculture'. Available at: https://esoftskills.ie/irelands-digital-transformation-in-agriculture/ (Accessed: 15 August 2025).

Eurofound (2025) Narrowing the digital divide: Economic and social convergence in Europe's digital transformation. Luxembourg: Publications Office of the European Union. Available at: https://www.eurofound.europa.eu/en/publications/all/narrowing-digital-divide-economic-and-social-convergence-europes-digital (Accessed: 1 September 2025).

European Central Bank (2025) Financial stability review, November 2025. Frankfurt: European Central Bank. Available at: https://www.ecb.europa.eu/press/financial-stability-publications/fsr/html/ecb.fsr202511~263b5810d4.en.html (Accessed: 28 January 2026).

European Civic Forum (2025) 'Joint letter: The EU must uphold hard-won protections for digital human rights'. Available at: https://civic-forum.eu/publications/open-letter/joint-letter-the-eu-must-uphold-hard-won-protections-for-digital-human-rights (Accessed: 28 November 2025).

European Commission (2018) Artificial intelligence for Europe, COM(2018) 237 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN (Accessed: 8 August 2025).

European Commission (2020a) Digital education action plan 2021–2027: Resetting education and training for the digital age, COM(2020) 624 final. Luxembourg: Publications Office of the European Union. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0624 (Accessed: 29 October 2025).

European Commission (2020b) European enterprise survey on the use of technologies based on artificial intelligence. Luxembourg: Publications Office of the European Union. Available at: https://digital-strategy.ec.europa.eu/en/library/european-enterprise-survey-use-technologies-based-artificial-intelligence (Accessed: 3 September 2025).

European Commission (2021) Ethics by design and ethics of use approaches for artificial intelligence. Brussels: European Commission. Available at: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf (Accessed: 27 August 2025).

European Commission (2024) 'European AI Office'. Available at: https://digital-strategy.ec.europa.eu/en/policies/ai-office (Accessed: 22 November 2025).

European Commission (2025a) European democracy shield: Empowering strong and resilient democracies. Luxembourg: Publications Office of the European Union. Available at: https://commission.europa.eu/document/download/2539eb53-9485-4199-bfdc-97166893ff45_en (Accessed: 15 November 2025).

European Commission (2025b) 'EU launches InvestAI initiative to mobilise €200 billion investment in artificial intelligence'. Available at: https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence (Accessed: 2 August 2025).

European Commission (2025c) Special Eurobarometer 566: The digital decade 2025. Brussels: European Union. Available at: https://europa.eu/eurobarometer/surveys/detail/3362 (Accessed: 26 August 2025).

European Commission (2025d) Special Eurobarometer 554: Artificial intelligence and the future of work. Brussels: European Union. Available at: https://europa.eu/eurobarometer/surveys/detail/3222 (Accessed: 23 November 2025).

European Commission (2025e) Digital omnibus on AI regulation proposal. Available at: https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-ai-regulation-proposal (Accessed: 21 November 2025).

European Commission (2025f) Study on the deployment of AI in healthcare. Luxembourg: Publications Office of the European Union. Available at: https://op.europa.eu/en/publication-detail/-/publication/9ddf7bf8-62bf-11f0-bf4e-01aa75ed71a1 (Accessed: 26 February 2026).

European Union (2024) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (Accessed: 27 August 2025).

European Union (2025) AI continent action plan, COM(2025) 165 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52025DC0165 (Accessed: 28 October 2025).

Eurostat (2025) 'Use of artificial intelligence in enterprises'. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises (Accessed: 19 August 2025).

Expert Group on Future Skills Needs (2025) Skills insights note 2025-2: How AI is transforming the Irish labour market. Dublin: Department of Enterprise, Tourism and Employment. Available at: https://enterprise.gov.ie/en/publications/publication-files/skills-insights-note-2025-2-how-ai-is-transforming-the-irish-labour-market.pdf (Accessed: 23 November 2025).

Faiyazuddin, M., Rahman, S.J.Q., Anand, G. et al. (2025) 'The impact of artificial intelligence on healthcare: A comprehensive review of advancements in diagnostics, treatment, and operational efficiency', Health Science Reports, 8(1). Available at: https://doi.org/10.1002/hsr2.70312.

Fattorini, L., Maslej, N., Perrault, R. et al. (2024) The global AI vibrancy tool: November 2024. Stanford Institute for Human-Centered AI. Available at: https://hai.stanford.edu/assets/files/global_ai_vibrancy_tool_paper_november2024.pdf (Accessed: 22 September 2025).

Feldstein, S. (2022) AI & Big Data Global Surveillance Index (2022 updated). Mendeley Data, V3. Available at: https://doi.org/10.17632/gjhf5y4xjp.3.

Fengchun, M., Holmes, W., Huang, R. and Zhang, H. (2021) AI and education: Guidance for policy-makers. Paris: UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000376709 (Accessed: 12 August 2025).

Fernández Llorca, D., Hamon, R., Junklewitz, H. et al. (2025) 'Testing autonomous vehicles and AI: Perspectives and challenges from cybersecurity, transparency, robustness and fairness', European Transport Research Review, 17. Available at: https://doi.org/10.1186/s12544-025-00732-x.

Filippucci, F., Gal, P., Jona-Lasinio, C., Leandro, A. and Nicoletti, G. (2024) The impact of artificial intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges. Paris: OECD Publishing. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/04/the-impact-of-artificial-intelligence-on-productivity-distribution-and-growth_d54e2842/8d900037-en.pdf (Accessed: 25 August 2025).

Filippucci, F., Gal, P., Laengle, K. and Schief, M. (2025) Macroeconomic productivity gains from artificial intelligence in G7 economies. Paris: OECD Publishing. Available at: https://www.oecd.org/en/publications/macroeconomic-productivity-gains-from-artificial-intelligence-in-g7-economies_a5319ab5-en.html (Accessed: 28 October 2025).

Financial Services Ireland and IBEC (2025) People at the helm: Harnessing the benefits of AI in Irish financial services. Available at: https://www.ibec.ie/connect-and-learn/insights/insights/2025/06/23/fsi-harnessing-the-benefits-of-ai (Accessed: 4 January 2026).

Floridi, L., Holweg, M., Taddeo, M., Amaya Silva, J., Mökander, J. and Wen, Y. (2022) 'capAI: A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act', SSRN. Available at: https://doi.org/10.2139/ssrn.4064091.

Fortune Business Insights (2025) Artificial intelligence in healthcare market size 2029. Available at: https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-in-healthcare-market-100534 (Accessed: 19 September 2025).

Future of Life Institute (2025a) Statement on superintelligence. Available at: https://superintelligence-statement.org/ (Accessed: 31 October 2025).

Future of Life Institute (2025b) AI safety index: Winter 2025. Available at: https://futureoflife.org/ai-safety-index-winter-2025/ (Accessed: 9 January 2026).

G7 Hiroshima Conference (2023) Hiroshima Process international guiding principles for advanced AI systems. Available at: https://www.mofa.go.jp/files/100573471.pdf (Accessed: 26 August 2025).

Galaz, V., Schewenius, M., Donges, J.F. et al. (2025) 'AI for a planet under pressure', arXiv. Available at: https://arxiv.org/abs/2510.24373 (Accessed: 28 November 2025).

Gartner (2025a) 'The 2025 hype cycle for artificial intelligence goes beyond GenAI'. Available at: https://www.gartner.com/en/articles/hype-cycle-for-artificial-intelligence (Accessed: 26 February 2026).

Gartner (2025b) 'Gartner predicts over 40% of agentic AI projects will be canceled by end of 2027'. Available at: https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027 (Accessed: 22 September 2025).

Gartner (2025c) 'AI literacy: Why and how business leaders must build it'. Available at: https://www.gartner.com/en/articles/ai-literacy (Accessed: 14 November 2025).

Gathmann, C.E.W., Grimm, F. and Winkler, E. (2024) AI, task changes in jobs, and worker reallocation. IZA Discussion Paper No. 17554. Bonn: Institute of Labor Economics. Available at: https://docs.iza.org/dp17554.pdf (Accessed: 31 August 2025).

Gillespie, N., Lockey, S., Ward, T., Macdade, A. and Hassed, G. (2025) Trust, attitudes and use of artificial intelligence: A global study 2025. University of Melbourne and KPMG International. Available at: https://mbs.edu/-/media/PDF/Research/Trust_in_AI_Report.pdf (Accessed: 21 August 2025).

Gmyrek, P., Berg, J., Kamiński, K. et al. (2025) Generative AI and jobs: A refined global index of occupational exposure. ILO Working Paper No. 140. Geneva: International Labour Organization. Available at: https://www.ilo.org/sites/default/files/2025-05/WP140_web.pdf (Accessed: 9 September 2025).

Goel, M. and Pandey, M. (2024) 'Crop yield prediction using AI: A review', in 2024 2nd International Conference on Disruptive Technologies (ICDT). Greater Noida, India, pp. 1547–1553. Available at: https://ieeexplore.ieee.org/document/10489432 (Accessed: 18 August 2025).

Goldman Sachs (2025) 'Why AI companies may invest more than $500 billion in 2026'. Available at: https://www.goldmansachs.com/insights/articles/why-ai-companies-may-invest-more-than-500-billion-in-2026 (Accessed: 29 January 2026).

Gonzales, S. (2024) 'AI literacy and the new digital divide: A global call for action'. Available at: https://www.unesco.org/ethics-ai/en/articles/ai-literacy-and-new-digital-divide-global-call-action (Accessed: 15 November 2025).

Government Digital Service (2023) Algorithmic transparency recording standard hub. GOV.UK. Available at: https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub (Accessed: 9 September 2025).

Government of Canada (2023a) Voluntary code of conduct on the responsible development and management of advanced generative AI systems. Available at: https://ised-isde.canada.ca/site/ised/en/voluntary-code-conduct-responsible-development-and-management-advanced-generative-ai-systems (Accessed: 20 August 2025).

Government of Canada (2023b) The Artificial Intelligence and Data Act (AIDA): Companion document. Available at: https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document (Accessed: 20 August 2025).

Grace, K., Stewart, H., Sandkühler, J.F. et al. (2024) 'Thousands of AI authors on the future of AI', arXiv. Available at: https://arxiv.org/abs/2401.02843 (Accessed: 10 August 2025).

Grand View Research (2024) AI in education market size and share report, 2022–2030. Available at: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-education-market-report (Accessed: 13 August 2025).

Green, A. (2024) Artificial intelligence and the changing demand for skills in the labour market. Paris: OECD Publishing. Available at: https://www.oecd.org/en/publications/artificial-intelligence-and-the-changing-demand-for-skills-in-the-labour-market_88684e36-en.html (Accessed: 11 November 2025).

Gu, X., Zheng, X., Pang, T. et al. (2024) 'Agent Smith: A single image can jailbreak one million multimodal LLM agents exponentially fast', arXiv. Available at: https://arxiv.org/abs/2402.08567 (Accessed: 22 November 2025).

Haenlein, M. and Kaplan, A. (2019) 'A brief history of artificial intelligence: On the past, present, and future of artificial intelligence', California Management Review, 61(4), pp. 5–14. Available at: https://doi.org/10.1177/0008125619864925.

Hagendorff, T. (2024) 'Mapping the ethics of generative AI: A comprehensive scoping review', Minds and Machines, 34. Available at: https://doi.org/10.1007/s11023-024-09694-w.

Hammond, G. and Kinder, T. (2025) 'OpenAI's computing deals top $1tn', Financial Times. Available at: https://www.ft.com/content/5f6f78af-aed9-43a5-8e31-2df7851ceb67 (Accessed: 28 October 2025).

Hendrycks, D., Song, D., Szegedy, C. et al. (2025) 'A definition of AGI', arXiv. Available at: https://arxiv.org/abs/2510.18212 (Accessed: 30 October 2025).

Higher Education Authority (2025a) Key facts and figures for Ireland's publicly-funded higher education institutions. Available at: https://hea.ie/statistics/data-for-download-and-visualisations/key-facts-figures-report/ (Accessed: 20 November 2025).

Higher Education Authority (2025b) 'Ten considerations for generative artificial intelligence adoption in Irish higher education', National Resource Hub. Available at: https://hub.teachingandlearning.ie/genai/ten-considerations-for-generative-artificial-intelligence-adoption-in-irish-higher-education/ (Accessed: 22 August 2025).

HM Government (2021) National AI strategy. London: HM Government. Available at: https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf (Accessed: 20 August 2025).

Hobbs, H., Docherty, D., Aranda, L. et al. (2026) Exploring possible AI trajectories through 2030. OECD Artificial Intelligence Papers No. 55. Paris: OECD Publishing. Available at: https://www.oecd.org/en/publications/exploring-possible-ai-trajectories-through-2030_cb41117a-en.html (Accessed: 10 February 2026).

Holzinger, A., Zatloukal, K. and Müller, H. (2024) 'Is human oversight to AI systems still possible?', New Biotechnology, 85, pp. 59–62. Available at: https://doi.org/10.1016/j.nbt.2024.12.003.

Hughes, S. and Bae, M. (2023) Vectara hallucination leaderboard. GitHub. Available at: https://github.com/vectara/hallucination-leaderboard/ (Accessed: 5 September 2025).

IAPP (2025) US state AI governance legislation tracker. Available at: https://iapp.org/resources/article/us-state-ai-governance-legislation-tracker/ (Accessed: 28 November 2025).

IAPP Research and Insights (2025) Global AI law and policy tracker. IAPP. Available at: https://iapp.org/media/pdf/resource_center/global_ai_law_policy_tracker.pdf (Accessed: 26 August 2025).

Incident Database (n.d.) Welcome to the Artificial Intelligence Incident Database. Available at: https://incidentdatabase.ai/ (Accessed: 26 February 2026).

INRIX (2025) 2025 INRIX global traffic scorecard. Available at: https://inrix.com/scorecard/ (Accessed: 5 January 2026).

International Energy Agency (2024) Electricity 2024: Analysis and forecast to 2026. Paris: International Energy Agency. Available at: https://iea.blob.core.windows.net/assets/18f3ed24-4b26-4c83-a3d2-8a1be51c8cc8/Electricity2024-Analysisandforecastto2026.pdf (Accessed: 25 August 2025).

International Energy Agency (2025) Energy and AI: World energy outlook special report. Paris: International Energy Agency. Available at: https://iea.blob.core.windows.net/assets/601eaec9-ba91-4623-819b-4ded331ec9e8/EnergyandAI.pdf (Accessed: 25 August 2025).

International Monetary Fund (2024) World economic outlook: Steady but slow – resilience amid divergence. Washington, DC: International Monetary Fund. Available at: https://digitallibrary.un.org/record/4065012?v=pdf#files (Accessed: 2 September 2025).

International Monetary Fund (2025) World economic outlook (April 2025): Commodity special feature – Annex 1.1. Washington, DC: International Monetary Fund. Available at: https://www.imf.org/-/media/Files/Publications/WEO/2025/April/English/ch1onlineannex.ashx (Accessed: 2 September 2025).

International Organization for Standardization (2025) ISO policy brief: Harnessing international standards for responsible AI development and governance. Geneva: ISO Central Secretariat. Available at: https://www.iso.org/publication/PUB100498.html (Accessed: 23 November 2025).

Irish Tech News (2025) 'ADAPT's "AI literacy" programme scales up to reach hundreds more teachers across Ireland', Irish Tech News. Available at: https://irishtechnews.ie/adapts-ai-literacy-programme-scales-up/ (Accessed: 7 November 2025).

Jaumotte, F., Kim, J., Koll, D., Li, E.Z., Li, L., Melina, G., Song, A. and Tavares, M.M. (2026) Bridging skill gaps for the future: New jobs creation in the AI age. IMF Staff Discussion Note SDN/2026/001. Washington, DC: International Monetary Fund.

Joint Committee on Artificial Intelligence (2025) First interim report. Houses of the Oireachtas. Available at: https://www.oireachtas.ie/en/committees/34/artificial-intelligence/ (Accessed: 17 December 2025).

Jolles, D. and Lordan, G. (2025) Bridging the generational AI gap: Unlocking productivity for all generations. The Inclusion Initiative at LSE. Available at: https://www.protiviti.com/sites/default/files/2025-11/lse-generational-surveyreport-bklt-1025-iz-en.pdf (Accessed: 25 November 2025).

Josten, C. and Lordan, G. (2019) Robots at work: Automatable and non-automatable jobs. IZA Discussion Paper No. 12520. Bonn: Institute of Labor Economics. Available at: https://docs.iza.org/dp12520.pdf (Accessed: 2 September 2025).

Kalai, A.T., Nachum, O., Vempala, S.S. and Zhang, E. (2025) 'Why language models hallucinate', arXiv. Available at: https://arxiv.org/abs/2509.04664 (Accessed: 5 September 2025).

Kandlhofer, M., Steinbauer, G., Hirschmugl-Gaisch, S. and Huber, P. (2016) 'Artificial intelligence and computer science in education: From kindergarten to university', in 2016 IEEE Frontiers in Education Conference (FIE). Available at: https://doi.org/10.1109/FIE.2016.7757570 (Accessed: 26 February 2026).

Kong, S.-C., Cheung, W.M.-Y. and Zhang, G. (2021) 'Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds', Computers and Education: Artificial Intelligence, 2. Available at: https://doi.org/10.1016/j.caeai.2021.100026.

Kong, S.-C., Cheung, W.M.-Y. and Tsang, O. (2024) 'Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach', Computers and Education: Artificial Intelligence, 6, p. 100214. Available at: https://doi.org/10.1016/j.caeai.2024.100214.

Korst, J., Puntoni, S. and Tambe, P. (2025) Accountable acceleration: GenAI fast-tracks into enterprise. Available at: https://ai.wharton.upenn.edu/wp-content/uploads/2025/10/2025-Wharton-GBK-AI-Adoption-Report_Executive-Summary.pdf (Accessed: 22 January 2026).

Kumar Jha, A. and Danks, N. (2025) The AI economy in Ireland 2025: Trends, impact and opportunity. Available at: https://www.tcd.ie/media/tcd/business/pdfs/research/Microsoft-Report.pdf (Accessed: 28 August 2025).

Lane, M. and Saint-Martin, A. (2021) The impact of artificial intelligence on the labour market: What do we know so far? OECD Social, Employment and Migration Working Papers No. 256. Paris: OECD Publishing. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/01/the-impact-of-artificial-intelligence-on-the-labour-market_a4b9cac2/7c895724-en.pdf (Accessed: 11 September 2025).

Laux, J., Wachter, S. and Mittelstadt, B. (2023) 'Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk', Regulation and Governance, 18, pp. 3–32. Available at: https://doi.org/10.1111/rego.12512.

Lawder, D. (2025) 'AI investment boom may lead to bust, but not likely systemic crisis, IMF chief economist says', Reuters, 14 October. Available at: https://www.reuters.com/legal/transactional/ai-investment-boom-may-lead-bust-not-likely-systemic-crisis-imf-chief-economist-2025-10-14/ (Accessed: 10 January 2026).

Logg, J.M., Minson, J.A. and Moore, D.A. (2019) 'Algorithm appreciation: People prefer algorithmic to human judgment', Organizational Behavior and Human Decision Processes, 151, pp. 90–103. Available at: https://doi.org/10.1016/j.obhdp.2018.12.005.

Long, D. and Magerko, B. (2020) 'What is AI literacy? Competencies and design considerations', in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–16. Available at: https://doi.org/10.1145/3313831.3376727.

Lorena Casal Otero, Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro López, B. and Barro, S. (2023) 'AI literacy in K-12: A systematic literature review', International Journal of STEM Education, 10. Available at: https://doi.org/10.1186/s40594-023-00418-7.

Lu, Z., Afridi, I., Kang, H.J., Ruchkin, I. and Zheng, X. (2024) 'Surveying neuro-symbolic approaches for reliable artificial intelligence of things', Journal of Reliable Intelligent Environments, 10, pp. 257–279. Available at: https://doi.org/10.1007/s40860-024-00231-1.

Lyu, J. and Tang, S. (2025) 'Power hungry data centers are driving green energy demand', BloombergNEF. Available at: https://about.bnef.com/insights/clean-energy/power-hungry-data-centers-are-driving-green-energy-demand/ (Accessed: 27 November 2025).

Mackowski, M.J., Maschek, W.A., Goldstein, B.L., Jacobson, J.B., Friel, A.L. and Kirk, M. (2025) 'Key insights on President Trump's new AI executive order and policy and regulatory implications', The National Law Review. Available at: https://natlawreview.com/article/key-insights-president-trumps-new-ai-executive-order-and-policy-regulatory (Accessed: 26 February 2026).

Maphoto, K.B., Sevnarayan, K., Mohale, N.E., Suliman, Z., Ntsopi, T.J. and Mokoena, D. (2024) 'Advancing students' academic excellence in distance education: Exploring the potential of generative AI integration to improve academic writing skills', Open Praxis, 16(2), pp. 142–159. Available at: https://doi.org/10.55982/openpraxis.16.2.649.

Maple, C., Szpruch, L., Epiphaniou, G., Staykova, K., Singh, S., Penwarden, W., Wen, Y., Wang, Z., Hariharan, J. and Avramovic, P. (2023) 'The AI revolution: Opportunities and challenges for the finance sector', arXiv. Available at: https://arxiv.org/abs/2308.16538 (Accessed: 12 January 2026).

Marcus, G. (2025) 'Game over. AGI is not imminent, and LLMs are not the royal road to getting there', Substack. Available at: https://garymarcus.substack.com/p/the-last-few-months-have-been-devastating (Accessed: 25 October 2025).

Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J.C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L. and Betts Lotufo, J. (2025) The AI Index 2025 annual report. Stanford: Institute for Human-Centered Artificial Intelligence (HAI). Available at: https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf (Accessed: 28 October 2025).

McCann, D. (2025) 'Louth shellfish processor is pioneering AI technology', All Ireland Sustainability. Available at: https://www.allirelandsustainability.com/louth-shellfish-processor-is-pioneering-ai-technology/ (Accessed: 15 August 2025).

McCarthy, J. (1955) 'A proposal for the Dartmouth summer research project on artificial intelligence'. Available at: https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html (Accessed: 26 February 2026).

McKinsey & Company (2024) 'What is AI (artificial intelligence)?'. Available at: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai (Accessed: 26 February 2026).

Meng, Y., Bing, Z., Yao, X., Chen, K., Huang, K., Gao, Y., Sun, F. and Knoll, A. (2025) 'Preserving and combining knowledge in robotic lifelong reinforcement learning', Nature Machine Intelligence, 7(2), pp. 256–269. Available at: https://doi.org/10.1038/s42256-025-00983-2.

Merino-Campos, C. (2025) 'The impact of artificial intelligence on personalized learning in higher education: A systematic review', Trends in Higher Education, 4(2), p. 17. Available at: https://doi.org/10.3390/higheredu4020017.

Microsoft AI Economy Institute (2026) Global AI adoption in 2025: A widening digital divide. Available at: https://www.microsoft.com/en-us/research/wp-content/uploads/2026/01/Microsoft-AI-Diffusion-Report-2025-H2.pdf (Accessed: 18 January 2026).

Milanez, A. and Bratta, B. (2019) Taxation and the future of work. OECD Taxation Working Papers. Paris: OECD Publishing. Available at: https://doi.org/10.1787/20f7164a-en.

Misuraca, G. and van Noordt, C. (2020) Overview of the use and impact of AI in public services in the EU. Luxembourg: Publications Office of the European Union. Available at: https://ai-watch.ec.europa.eu/publications/ai-watch-artificial-intelligence-public-services_en (Accessed: 19 August 2025).

Moix, A., Ledebev, K. and Klein, J. (2025) Threat intelligence report: August 2025. Anthropic. Available at: https://www-cdn.anthropic.com/b2a76c6f6992465c09a6f2fce282f6c0cea8c200.pdf (Accessed: 27 August 2025).

Mon-Williams, R., Li, G., Long, R., Du, W. and Lucas, C.G. (2025) 'Embodied large language models enable robots to complete complex tasks in unpredictable environments', Nature Machine Intelligence, 7. Available at: https://doi.org/10.1038/s42256-025-01005-x.

Moravec, H. (1988) Mind children: The future of robot and human intelligence. Cambridge, MA: Harvard University Press.

Morozov, E. (2013) To save everything, click here: Technology, solutionism and the urge to fix problems that don't exist. New York: PublicAffairs.

Mucci, T. (2024) 'AI generated content', IBM Think. Available at: https://www.ibm.com/think/insights/ai-generated-content (Accessed: 26 February 2026).

Najem, R., Bahnasse, A., Fakhouri Amr, M. and Talea, M. (2025) 'Advanced AI and big data techniques in e-finance: A comprehensive survey', Discover Artificial Intelligence, 5(1). Available at: https://doi.org/10.1007/s44163-025-00365-y.

Narayanan, A. and Kapoor, S. (2025) AI as normal technology: An alternative to the vision of AI as a potential superintelligence. Available at: https://kfai-documents.s3.amazonaws.com/documents/0ee1da899a/AI-as-Normal-Technology---Narayanan---Kapoor-Final.pdf (Accessed: 22 September 2025).

National AI Leadership Forum (2025) Recommendations for action developed through the National AI Leadership Forum. Available at: https://www.adaptcentre.ie/wp-content/uploads/2025/10/NationalAILeadershipForumReport22Oct2025.pdf (Accessed: 2 November 2025).

National Broadband Ireland (2025) 'Two-thirds of rural homes and businesses can now connect to high-speed broadband'. Available at: https://nbi.ie/news/updates/2025/06/03/two-thirds-of-rural-homes-and-businesses-can-now-connect-to-high-speed-broadband/ (Accessed: 18 August 2025).

National Economic and Social Council (2024) Towards a national better work strategy. Report No. 165. Dublin: National Economic and Social Council. Available at: https://www.nesc.ie/app/uploads/2024/07/165_towards_a_national_better_work_strategy.pdf (Accessed: 10 September 2025).

National Economic and Social Council (2025) Ireland's future power system and economic resilience. Dublin: National Economic and Social Council. Available at: https://s3.eu-west-1.amazonaws.com/files.nesc.ie/nesc_reports/en/167_energy_resilience.pdf (Accessed: 27 September 2025).

National Standards Authority of Ireland (2023) AI standards and assurance roadmap: Action under 'AI – Here for good', the national artificial intelligence strategy for Ireland. Dublin: National Standards Authority of Ireland. Available at: https://www.nsai.ie/images/uploads/general/NSAI_AI_report_digital.pdf (Accessed: 21 August 2025).

Ng, D.T.K., Leung, J.K.L., Chu, S.K.W. and Shen, M.Q. (2021) 'Conceptualizing AI literacy: An exploratory review', Computers and Education: Artificial Intelligence, 2(1). Available at: https://doi.org/10.1016/j.caeai.2021.100041.

Niederhoffer, K., Rosen Kellerman, G., Lee, A., Liebscher, A., Rapuano, K. and Hancock, J.T. (2025) 'AI-generated "workslop" is destroying productivity', Harvard Business Review. Available at: https://hbr.org/2025/09/ai-generated-workslop-is-destroying-productivity (Accessed: 26 February 2026).

NIST (2024) Artificial intelligence risk management framework: Generative artificial intelligence profile. Available at: https://doi.org/10.6028/nist.ai.600-1.

Novelli, C., Casolari, F., Rotolo, A., Taddeo, M. and Floridi, L. (2024) 'AI risk assessment: A scenario-based, proportional methodology for the AI Act', DISO, 3(13). Available at: https://doi.org/10.1007/s44206-024-00095-1.

Noy, S. and Zhang, W. (2023) 'Experimental evidence on the productivity effects of generative artificial intelligence', Science, 381(6654), pp. 187–192. Available at: https://doi.org/10.1126/science.adh2586.

O'Sullivan, S. (2020) 'Models of governance for innovation in medicine and health research', European Journal of Health Law, 27(3), pp. 324–334. Available at: https://doi.org/10.2307/48712707.

O'Sullivan, J., Lowry, C., Woods, R., Marrinan, B. and Hutchinson, C. (2025) Generative AI in higher education teaching and learning: Sectoral perspectives. Dublin: Higher Education Authority. Available at: https://hea.ie/assets/uploads/2025/09/Gen-AI-in-Higher-Education-Teaching-and-Learning-Sectoral-Perspectives.pdf (Accessed: 14 November 2025).

OECD (2019) The OECD artificial intelligence (AI) principles. Available at: https://oecd.ai/en/ai-principles (Accessed: 20 August 2025).

OECD (2023a) Recommendation of the Council on artificial intelligence. Available at: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449#mainText (Accessed: 26 February 2026).

OECD (2023b) OECD employment outlook 2023: Artificial intelligence and the labour market. Paris: OECD Publishing. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/07/oecd-employment-outlook-2023_904bcef3/08785bba-en.pdf (Accessed: 23 August 2025).

OECD (2023c) Generative artificial intelligence in finance. OECD Artificial Intelligence Papers No. 9. Paris: OECD Publishing. Available at: https://www.oecd.org/en/publications/generative-artificial-intelligence-in-finance_ac7149cc-en.html (Accessed: 4 January 2026).

OECD (2023d) 'Regulatory sandboxes in artificial intelligence'. OECD Digital Economy Papers No. 356. Paris: OECD Publishing. Available at: https://www.oecd.org/en/publications/regulatory-sandboxes-in-artificial-intelligence_8f80a0e6-en.html (Accessed: 3 September 2025).

OECD (2024a) The impact of artificial intelligence on productivity, distribution and growth. Paris: OECD Publishing. Available at: https://www.oecd.org/en/publications/the-impact-of-artificial-intelligence-on-productivity-distribution-and-growth_8d900037-en.html (Accessed: 25 November 2025).

OECD (2024b) Framework for anticipatory governance of emerging technologies. Paris: OECD Publishing. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/04/framework-for-anticipatory-governance-of-emerging-technologies_14bf0402/0248ead5-en.pdf (Accessed: 12 August 2024).

OECD (2024c) 'Anticipatory governance'. Available at: https://www.oecd.org/en/topics/anticipatory-governance.html (Accessed: 26 February 2026).

OECD (2024d) Regulatory approaches to artificial intelligence in finance. Paris: OECD Publishing. Available at: https://www.oecd.org/en/publications/regulatory-approaches-to-artificial-intelligence-in-finance_f1498c02-en.html (Accessed: 4 January 2026).

OECD (2025a) Governing with artificial intelligence: The state of play and way forward in core government functions. Paris: OECD Publishing. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/06/governing-with-artificial-intelligence_398fa287/795de142-en.pdf (Accessed: 20 September 2025).

OECD (2025b) Live data from OECD.AI. Available at: https://oecd.ai/en/data?selectedArea=ai-demographics&selectedVisualization=ai-demographics-by-age (Accessed: 4 September 2025).

OECD (2025c) AI and the global productivity divide: Fuel for the fast or a lift for the laggards? Paris: OECD Publishing. Available at: www.oecd.org/content/dam/oecd/en/publications/reports/2025/12/ai-and-the-global-productivity-divide_f47026c5/c315ea90-en.pdf (Accessed: 10 January 2026).

OECD (2025d) Regulatory sandbox toolkit: A comprehensive guide for regulators to establish and manage regulatory sandboxes effectively. Paris: OECD Publishing. Available at: https://doi.org/10.1787/de36fa62-en.

OECD (2025e) Empowering learners for the age of AI: An AI literacy framework for primary and secondary education (review draft, May 2025). Paris: OECD. Available at: https://ailiteracyframework.org/wp-content/uploads/2025/05/AILitFramework_ReviewDraft.pdf (Accessed: 1 November 2025).

OECD (2026) OECD digital education outlook 2026. Paris: OECD Publishing. Available at: https://doi.org/10.1787/062a7394-en.

OECD and UNESCO (2024) G7 toolkit for artificial intelligence in the public sector. Paris: OECD/UNESCO. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/10/g7-toolkit-for-artificial-intelligence-in-the-public-sector_f93fb9fb/421c1244-en.pdf (Accessed: 16 August 2025).

OECD, BCG and INSEAD (2025) The adoption of artificial intelligence in firms: New evidence for policymaking. Paris: OECD Publishing. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/05/the-adoption-of-artificial-intelligence-in-firms_8fab986b/f9ef33c3-en.pdf (Accessed: 3 September 2025).

Ofcom (2024) Deepfake defences: Mitigating the harms of deceptive deepfakes (discussion paper). London: Ofcom. Available at: https://www.ofcom.org.uk/siteassets/resources/documents/consultations/discussion-papers/deepfake-defences/deepfake-defences.pdf?v=370754 (Accessed: 28 August 2025).

Office of the Government Chief Information Officer (2025) Learnings from three large language model proofs of concept. Available at: https://prod-g2g-assets.s3.amazonaws.com/documents/AI_POC_Learnings_Note_2024.pdf (Accessed: 19 October 2025).

Ombudsman for Children's Office (2025) AI and us: Young people's views and understanding of artificial intelligence. Available at: https://www.oco.ie/app/uploads/2025/09/OCO-AI-and-Us-Young-peoples-views-and-understanding-of-Artificial-Intelligence.pdf (Accessed: 11 November 2025).

OpenAI (2025) 'Introducing GPT-5'. Available at: https://openai.com/index/introducing-gpt-5/ (Accessed: 12 September 2025).

Pal, S., Marino Lazzaroni, R. and Mendoza, P. (2024) AI's missing link: The gender gap in the talent pool. Berlin: interface. Available at: https://www.interface-eu.org/publications/ai-gender-gap (Accessed: 4 September 2025).

Partel, V., Nunes, L., Stansly, P. and Ampatzidis, Y. (2019) 'Automated vision-based system for monitoring Asian citrus psyllid in orchards utilizing artificial intelligence', Computers and Electronics in Agriculture, 162, pp. 328–336. Available at: https://doi.org/10.1016/j.compag.2019.04.022.

Pati, A.K. (2025) 'Agentic AI: A comprehensive survey of technologies, applications, and societal implications', IEEE Access, 40(2), pp. 8–14. Available at: https://doi.org/10.1109/ACCESS.2025.3585609.

Patient Safety Learning (2024) 'Epic's overhaul of a flawed algorithm shows why AI oversight is a life-or-death issue (24 October 2022)'. Available at: https://www.pslhub.org/learn/commissioning-service-provision-and-innovation-in-health-and-care/digital-health-and-care-service-provision/288_artificial-intelligence/epic%E2%80%99s-overhaul-of-a-flawed-algorithm-shows-why-ai-oversight-is-a-life-or-death-issue-24-october-2022-r11327/ (Accessed: 22 September 2025).

Peng, S., Kalliamvakou, E., Cihon, P. and Demirer, M. (2023) 'The impact of AI on developer productivity: Evidence from GitHub Copilot', arXiv. Available at: https://arxiv.org/abs/2302.06590 (Accessed: 12 January 2026).

People's Republic of China (2025) Action plan for global governance of artificial intelligence. Available at: https://www.gov.cn/yaowen/liebiao/202507/content_7033929.htm (Accessed: 26 August 2025).

Perez, C. (2002) Technological revolutions and financial capital. Cheltenham: Edward Elgar.

Pizzinelli, C., Panton, A., Tavares, M.M., Cazzaniga, M. and Li, L. (2023) Labor market exposure to AI: Cross-country differences and distributional implications. Washington, DC: International Monetary Fund. Available at: https://www.imf.org/en/Publications/WP/Issues/2023/10/04/Labor-Market-Exposure-to-AI-Cross-country-Differences-and-Distributional-Implications-539656 (Accessed: 29 August 2025).

Porter, M.E. (1991) 'America's green strategy', Scientific American, 264(4), p. 168.
Raman, R., Kowalski, R., Achuthan, K., Iyer, A. and Nedungadi, P. (2025) 'Navigating artificial general intelligence development: Societal, technological, ethical, and brain-inspired pathways', Scientific Reports, 15(1). Available at: https://doi.org/10.1038/s41598-025-92190-7.

Randstad (2026) Work monitor 2026: The great workforce adaptation. Available at: https://www.randstad.ch/en/workmonitor-2026/ (Accessed: 30 January 2026).

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W.J., Sun, M., Day, I., Rather, R.A. and Heathcote, L. (2023) 'The role of ChatGPT in higher education: Benefits, challenges, and future research directions', JALT, 6(1). Available at: https://doi.org/10.37074/jalt.2023.6.1.29.

Raushan, P. (2023) AI in agriculture market by technology, offering, application: COVID-19 impact analysis. MarketsandMarkets. Available at: https://www.marketsandmarkets.com/Market-Reports/ai-in-agriculture-market-159957009.html (Accessed: 26 February 2026).

Reuters (2025) 'Britain boosts computing power in $1.3 billion AI drive', Reuters, 17 July. Available at: https://www.reuters.com/world/uk/britain-boosts-computing-power-13-billion-ai-drive-2025-07-17/ (Accessed: 3 September 2025).

Rodriguez, S.C. (2023) 'Consensus building in Taiwan, the poster child of digital democracy', Democracy Technologies. Available at: https://democracy-technologies.org/participation/consensus-building-in-taiwan/ (Accessed: 26 November 2025).

Rössler, B. (2005) The value of privacy. Cambridge: Polity.
Roy, N., Posner, I., Barfoot, T. et al. (2021) 'From machine learning to robotics: Challenges and opportunities for embodied intelligence', arXiv. Available at: https://arxiv.org/abs/2110.15245 (Accessed: 31 October 2025).

Russell, S. and Norvig, P. (2021) Artificial intelligence: A modern approach. 4th edn. Harlow: Pearson.

Ryan, Ó. (2025) 'Deepfake AI video depicting Catherine Connolly quitting presidential race removed by Meta', The Irish Times. Available at: https://www.irishtimes.com/politics/2025/10/22/meta-removes-ai-video-purporting-to-show-catherine-connolly-quitting-presidential-race/ (Accessed: 22 October 2025).

Sabel, C.F. and Zeitlin, J. (2012) 'Experimentalist governance', in Levi-Faur, D. (ed.) The Oxford handbook of governance. Oxford: Oxford University Press, pp. 169–184. Available at: https://academic.oup.com/edited-volume/34384/chapter-abstract/291590330 (Accessed: 3 September 2025).

Sagona, M., Dai, T., Macis, M. and Darden, M. (2025) 'Trust in AI-assisted health systems and AI's trust in humans', npj Health Systems, 2(10). Available at: https://doi.org/10.1038/s44401-025-00016-5.

Sahni, N., Stein, G., Zemmel, R. and Cutler, D.M. (2023) The potential impact of artificial intelligence on healthcare spending. NBER Working Paper No. 30857. Cambridge, MA: National Bureau of Economic Research. Available at: https://www.nber.org/system/files/working_papers/w30857/w30857.pdf (Accessed: 12 September 2025).

Santino, S. (2024) 'Artificial intelligence colonialism: Environmental damage, labor exploitation, and human rights crises in the Global South', SAIS Review of International Affairs, 44(2), pp. 75–92. Available at: https://doi.org/10.1353/sais.2024.a950958.

Sartori, L. and Theodorou, A. (2022) 'A sociotechnical perspective for the future of AI: Narratives, inequalities, and human control', Ethics and Information Technology, 24(4). Available at: https://doi.org/10.1007/s10676-022-09624-3.

Schneider, I., Xu, H., Benecke, S., Patterson, D., Huang, K., Ranganathan, P. and Elsworth, C. (2025) 'Life-cycle emissions of AI hardware: A cradle-to-grave approach and generational trends', arXiv. Available at: https://arxiv.org/abs/2502.01671 (Accessed: 10 January 2026).

Searle, J.R. (1980) 'Minds, brains, and programs', Behavioral and Brain Sciences, 3(3), pp. 417–424. Available at: https://doi.org/10.1017/S0140525X00005756.

Seger, E., Pearson, G., Avin, S., Briers, M., Ó Heigeartaigh, S. and Bacon, H. (2020) Tackling threats to informed decision-making in democratic societies: Promoting epistemic security in a technologically-advanced world. London: The Alan Turing Institute. Available at: https://www.turing.ac.uk/sites/default/files/2020-10/epistemic-security-report_final.pdf (Accessed: 27 August 2025).

Shan, R. (2024) 'Language artificial intelligence at a crossroads: Deciphering the future of small and large language models', Computer, 57(8), pp. 26–35. Available at: https://www.computer.org/csdl/magazine/co/2024/08/10632605/1ZgYa6HUNDq (Accessed: 31 October 2025).

Shojaee, P., Mirzadeh, I., Alizadeh, K., Horton, M., Bengio, S. and Farajtabar, M. (2025) 'The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity', Apple Machine Learning Research. Available at: https://machinelearning.apple.com/research/illusion-of-thinking (Accessed: 26 February 2026).

Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N. and Anderson, R. (2023) 'The curse of recursion: Training on generated data makes models forget', arXiv. Available at: https://arxiv.org/abs/2305.17493 (Accessed: 12 November 2025).

Son, H., Jang, J., Park, J., Balog, A., Ballantyne, P., Kwon, H.R., Singleton, A. and Hwang, J. (2025) 'Leveraging advanced technologies for (smart) transportation planning: A systematic review', Sustainability, 17(5), pp. 2245–2245. Available at: https://doi.org/10.3390/su17052245.

Sparkes, M. (2025) 'The AI bubble is heading towards a burst but it won't be the end of AI', New Scientist. Available at: https://www.newscientist.com/article/2499738-the-ai-bubble-is-heading-towards-a-burst-but-it-wont-be-the-end-of-ai/ (Accessed: 28 October 2025).

Sparrow, R., Howard, M. and Degeling, C. (2021) 'Managing the risks of artificial intelligence in agriculture', NJAS: Impact in Agricultural and Life Sciences, 93(1), pp. 172–196. Available at: https://doi.org/10.1080/27685241.2021.2008777.

Sproule, L. (2025) 'Will we control AI, or will it control us? Top researchers weigh in', CBC. Available at: https://www.cbc.ca/news/science/artificial-intelligence-predictions-1.7427024 (Accessed: 13 January 2025).

Sternfels, B. and Atsmon, Y. (2025) The learning organization: How to accelerate AI adoption. McKinsey & Company. Available at: https://www.mckinsey.com/~/media/mckinsey/business%20functions/strategy%20and%20corporate%20finance/our%20insights/the%20learning%20organization%20how%20to%20accelerate%20ai%20adoption/the-learning-organization-how-to-accelerate-ai-adoption_final2.pdf (Accessed: 10 October 2025).

Sutton, R. (2019) 'The bitter lesson'. Available at: https://www.cs.utexas.edu/~eunsol/courses/data/bitter_lesson.pdf (Accessed: 25 October 2025).

The Alan Turing Institute (2023) AI ethics and governance in practice. Available at: https://www.turing.ac.uk/research/research-projects/ai-ethics-and-governance-practice (Accessed: 1 November 2025).

The Economist (2025) 'What if the $3trn AI investment boom goes wrong?', The Economist. Available at: https://www.economist.com/leaders/2025/09/11/what-if-the-3trn-ai-investment-boom-goes-wrong (Accessed: 12 September 2025).

The Open Innovation Team and Department for Education (2024) Generative AI in education: Educator and expert views. London: Department for Education. Available at: https://assets.publishing.service.gov.uk/media/65b8cd41b5cb6e000d8bb74e/DfE_GenAI_in_education_-_Educator_and_expert_views_report.pdf (Accessed: 16 August 2025).

Thomas, B. (2025) 'The benefits of bubbles', Stratechery. Available at: https://stratechery.com/2025/the-benefits-of-bubbles/ (Accessed: 31 January 2026).

Thomasson, J.A., Ampatzidis, Y., Bhandari, M., Ferreyra, R., Gentimis, T., McReynolds, E., Murray, S.C., Peterson, M.B., Rodriguez Lopez, C.M., Strong, R.L., Tedeschi, L.O., Vitale, J. and Ye, X. (2025) AI in agriculture: Opportunities, challenges, and recommendations. CAST. Available at: https://cast-science.org/wp-content/uploads/2025/03/CAST_AI-in-Agriculture.pdf (Accessed: 19 August 2025).

Tierney, A.A., Gayre, G., Hoberman, B., Mattern, B., Ballesca, M., Wilson Hannay, S.B., Castilla, K., Lau, C.S., Kipnis, P., Liu, V. and Lee, K. (2025) 'Ambient artificial intelligence scribes: Learnings after 1 year and over 2.5 million uses', NEJM Catalyst, 6(5). Available at: https://doi.org/10.1056/cat.25.0040.

Times, T.K. (2023) 'AI digital textbooks to be introduced in schools from 2025', Korea Times. Available at: https://www.koreatimes.co.kr/southkorea/society/20230608/ai-digital-textbooks-to-be-introduced-in-schools-from-2025 (Accessed: 22 August 2025).

Tõnurist, P. and Hanson, A. (2020) Anticipatory innovation governance: Shaping the future through proactive policy making. OECD Working Papers on Public Governance No. 44. Paris: OECD Publishing. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/12/anticipatory-innovation-governance_d1aded4e/cce14d80-en.pdf (Accessed: 3 September 2025).

Tournesac, A., Hjartar, K., Krawina, M., Hillenbrand, P. and Olanrewaju, T. (2025) 'Accelerating Europe's AI adoption: The role of sovereign AI', McKinsey & Company. Available at: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/accelerating-europes-ai-adoption-the-role-of-sovereign-ai (Accessed: 6 January 2026).

Tuhin, M. (2025) 'The role of AI in tackling climate change: Harnessing technology for a sustainable future', Science News Today. Available at: https://www.sciencenewstoday.org/the-role-of-ai-in-tackling-climate-change-harnessing-technology-for-a-sustainable-future (Accessed: 24 September 2025).

Tully, S., Longoni, C. and Appel, G. (2025) 'Lower artificial intelligence literacy predicts greater AI receptivity', Journal of Marketing, 89(5). Available at: https://doi.org/10.1177/00222429251314491.

Turing, A. (1950) 'Computing machinery and intelligence', Mind, 59(236), pp. 433–460. Available at: https://doi.org/10.1093/mind/LIX.236.433.

UCL (2025) 'Practical changes could reduce AI energy demand by up to 90%', UCL News. Available at: https://www.ucl.ac.uk/news/2025/jul/practical-changes-could-reduce-ai-energy-demand-90 (Accessed: 22 November 2025).

UK Government (2023) The Bletchley Declaration by countries attending the AI Safety Summit, 1–2 November 2023. Available at: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 (Accessed: 1 September 2025).

UNESCO (2021) Recommendation on the ethics of artificial intelligence. Paris: UNESCO. Available at: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence (Accessed: 20 August 2025).

UNESCO (2022) K-12 AI curricula: A mapping of government-endorsed AI curricula. Paris: UNESCO. Available at: https://www.unesco.org/en/articles/k-12-ai-curricula-mapping-government-endorsed-ai-curricula (Accessed: 1 November 2025).

UNESCO (2023a) Guidance for generative AI in education and research. Paris: UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000386693 (Accessed: 21 August 2025).

UNESCO (2023b) 'Small language models (SLMs): A cheaper, greener route into AI'. Paris: UNESCO. Available at: https://www.unesco.org/en/articles/small-language-models-slms-cheaper-greener-route-ai (Accessed: 24 September 2025).

UNESCO (2024a) AI competency framework for students. Paris: UNESCO. Available at: https://www.unesco.org/en/articles/ai-competency-framework-students (Accessed: 1 November 2025).

UNESCO (2024b) AI competency framework for teachers. Paris: UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000391104 (Accessed: 1 November 2025).

UNESCO (2024c) 'Insights from practice: Telefónica's AI governance journey'. Available at: https://www.unesco.org/en/articles/insights-practice-telefonicas-ai-governance-journey (Accessed: 1 December 2025).

UNESCO (2025) AI and the future of education: Disruptions, dilemmas and directions. Paris: UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000395236 (Accessed: 1 November 2025).

UNICEF Innocenti – Global Office of Research and Foresight (2025) Guidance on AI and children: Updated guidance for governments and businesses to create AI policies and systems that uphold children's rights. Florence: UNICEF Innocenti. Available at: https://www.unicef.org/innocenti/reports/policy-guidance-ai-children (Accessed: 6 January 2026).

United Nations Environment Programme (2024) Artificial intelligence (AI) end-to-end: The environmental impact of the full AI lifecycle needs to be comprehensively assessed (issue note). Available at: https://wedocs.unep.org/rest/api/core/bitstreams/07b3c8fc-bd30-4b92-b5f4-d665e927b59d/content (Accessed: 8 October 2025).

Valerann (2025) 'Transforming Ireland's road network with AI-powered data fusion'. Available at: https://www.valerann.com/news/transforming-irelands-road-network-with-ai-powered-data-fusion (Accessed: 4 January 2026).

Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L. and Hobbhahn, M. (2024) 'Will we run out of data? Limits of LLM scaling based on human-generated data', arXiv. Available at: https://arxiv.org/abs/2211.04325 (Accessed: 31 October 2025).

Wen, D., Khan, S.M., Xu, A.J., Ibrahim, H., Smith, L., Caballero, J., Zepeda, L., Perez, C. de B., Denniston, A.K., Liu, X. and Matin, R.N. (2021) 'Characteristics of publicly available skin cancer image datasets: A systematic review', The Lancet Digital Health, 4(1). Available at: https://doi.org/10.1016/S2589-7500(21)00252-1.

Whiting, K. (2025) 'What is a small language model and should businesses invest in this AI tool?', World Economic Forum. Available at: https://www.weforum.org/stories/2025/01/ai-small-language-models/ (Accessed: 9 September 2025).

Wilson, C. (2021) 'Public engagement and AI: A values analysis of national strategies', Government Information Quarterly, 39(1), p. 101652. Available at: https://doi.org/10.1016/j.giq.2021.101652.

World Economic Forum (2024a) AI value alignment: Guiding artificial intelligence towards shared human goals (white paper). Geneva: World Economic Forum. Available at: https://www3.weforum.org/docs/WEF_AI_Value_Alignment_2024.pdf (Accessed: 27 August 2025).

World Economic Forum (2024b) Shaping the future of learning: The role of AI in education 4.0. Geneva: World Economic Forum. Available at: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Learning_2024.pdf (Accessed: 13 August 2025).

World Economic Forum (2025a) Future of jobs report 2025. Geneva: World Economic Forum. Available at: https://www.weforum.org/publications/the-future-of-jobs-report-2025/digest/ (Accessed: 10 October 2025).

World Economic Forum (2025b) Intelligent transport, greener future: AI as a catalyst to decarbonize global logistics. Geneva: World Economic Forum. Available at: https://reports.weforum.org/docs/WEF_Intelligent_Transport_Greener_Future_2025.pdf (Accessed: 5 January 2026).

World Economic Forum (2026a) Four futures for jobs in the new economy: AI and talent in 2030. Geneva: World Economic Forum. Available at: https://www.weforum.org/publications/four-futures-for-jobs-in-the-new-economy-ai-and-talent-in-2030/ (Accessed: 11 January 2026).

World Economic Forum (2026b) Chief economists' outlook: January 2026. Geneva: World Economic Forum. Available at: https://www.weforum.org/publications/chief-economists-outlook-january-2026/ (Accessed: 28 January 2026).

Worldwide Independent Network of Market Research (2025) WIN World AI index survey 2025. Available at: https://winmr.com/win-world-ai-index/ (Accessed: 19 September 2025).

Yeter, I.H., Yang, W. and Sturgess, J.B. (2024) 'Global initiatives and challenges in integrating artificial intelligence literacy in elementary education: Mapping policies and empirical literature', Future in Educational Research. Available at: https://doi.org/10.1002/fer3.59.

Yudkowsky, E. and Soares, N. (2025) If anyone builds it, everyone dies. London: Penguin Random House.

Zhao, C., Tan, Z., Ma, P., Li, D., Jiang, B., Wang, Y., Yang, Y. and Liu, H. (2025) 'Is chain-of-thought reasoning of LLMs a mirage? A data distribution lens', arXiv. Available at: https://arxiv.org/abs/2508.01191 (Accessed: 12 October 2025).

Zhou, Z., Ning, X., Hong, K., Fu, T., Xu, J., Li, S., Lou, Y., Wang, L., Yuan, Z., Li, X., Yan, S., Dai, G., Zhang, X.-P., Dong, Y. and Wang, Y. (2024) 'A survey on efficient inference for large language models', arXiv. Available at: https://arxiv.org/abs/2404.14294 (Accessed: 7 December 2025).

Zuckerberg, M. (2025) 'Personal superintelligence'. Available at: https://www.meta.com/superintelligence/ (Accessed: 26 February 2026).

nesc
Taighde, Idirphlé, Comhairle
Research, Dialogue, Advice

National Economic & Social Council
Parnell Square, Dublin 1, D01 E7C1
+353 1 814 6300 info@nesc.ie

www.nesc.ie