nesc
Taighde, Idirphlé, Comhairle
Research, Dialogue, Advice

Highlights | April 2026

Artificial Intelligence in Service of Society:

Navigating our way Forward

This 'highlights document' summarises the key insights and priority areas identified in the NESC report **Artificial Intelligence in Service of Society: Navigating Our Way Forward.**

Key messages

- **AI is not purely a technical issue; it is a socio-technical transformation:** AI's impacts emerge from the interaction between technology and its social context. Outcomes depend not only on algorithms, but on governance, institutional capacity, labour market adaptation, and democratic oversight. The challenge is to shape AI intentionally in line with our values and priorities, rather than passively absorb the technology.
- **Responsible and strategic AI adoption:** AI adoption should be driven by clearly defined problems and public needs, with models chosen that are proportionate to task complexity and mindful of environmental impact. Sustainable value creation requires aligning deployment with organisational capacity, data quality and workforce capability, pursued alongside employee involvement in how AI is introduced.
- **Safe, ethical and trustworthy AI:** AI systems are probabilistic and imperfect. Meaningful human control is essential to prevent over-reliance, loss of judgement and accountability gaps. High-level ethical principles must be translated into concrete practices with individuals and institutions building genuine ethical capability to ensure AI operates safely, fairly and effectively.
- **Governance must keep pace with uncertainty:** The trajectory of AI capability remains genuinely uncertain. Regulatory approaches must be agile, anticipatory and continuously updated rather than static responses to already-embedded technologies.
- **AI literacy is national infrastructure:** The capacity to understand, critically evaluate and responsibly use AI is not a technical training issue but a foundational societal capability that underpins democratic oversight, workforce adaptation and public trust.
- **Social licence must be earned:** Legal compliance is insufficient. Public deliberation and genuine two-way engagement are essential to ensure AI deployment reflects democratic values and retains legitimacy over time.

Introduction

Artificial Intelligence (AI) has in recent years moved from specialist technology into the centre of economic, policy and societal debate. Since the widespread release of generative AI tools in late 2022, dramatic jumps in model capability, falling computation costs and deepening integration into digital services have all accelerated adoption of AI across public and private sectors. Governments and organisations are actively exploring how AI can improve productivity, enhance public services and drive innovation, while simultaneously grappling with significant and unresolved questions about reliability, safety, fairness and broader societal impact of AI. These are issues that require considered and sustained attention. Ireland has strong foundations from which to engage with the opportunities AI offers, including a vibrant technology ecosystem, a highly skilled workforce and a set of ambitious goals elaborated in the *National Digital & AI Strategy 2030*. The central challenge is how to think about AI's opportunities and risks and to ensure its development and use aligns with our values, priorities and aspirations for the future.

This document draws on the fuller NESC report Artificial Intelligence in Service of Society: Navigating Our Way Forward which sets out five priority areas and associated actions to guide the safe and responsible development and use of AI in Ireland. This highlights document surfaces key issues and challenges and offers high-level reflections on where further thinking and work on AI is needed. It is intended for policymakers, senior officials and organisational leaders who need an accessible but rigorous framework for engaging with AI and its implications.

The analysis considers both the technical and societal dimensions of AI. By adopting a socio-technical lens we can avoid falling into the trap of techno-solutionism and recognise that AI systems are not simply a technical phenomenon but also a social and institutional one. Rather than starting with AI in search of a problem, we need to start with the problem and ask whether AI has something useful to offer. How AI develops, who benefits, what harms arise and how effectively they are managed will depend not only on the technology but on the governance, institutional capacity and societal choices brought to bear on it. Rather than adopting AI in a passive manner, the goal should be to actively shape AI, so that its benefits can be realised responsibly, equitably, and sustainably.

Figure 1: Navigating the Future of AI through a Socio-Technical Lens

Source: NESC Secretariat.

The Evolution & Future Direction of AI

What Is Artificial Intelligence?
Artificial intelligence (AI) refers to the science of building machines that can perform tasks normally requiring human intelligence such as learning, reasoning, and decision-making. The field has evolved from early rule-based systems through machine learning, to today's deep learning models built on neural networks.

Neural networks are computational systems loosely inspired by the structure of the human brain. During training, the system processes vast amounts of data and adjusts the strength of connections between artificial neurons (known as weights) to minimise prediction errors. Through this process, the network gradually learns patterns in the data. Once trained, when given a prompt, information passes through the network layer by layer, with these learned weights shaping the output. Large language models (LLMs) are neural networks trained on hundreds of billions or even trillions of tokens (i.e. units of text such as words). They work by estimating the probability of possible next tokens (next words) in a sequence. This means that more likely words come up more often, but less likely ones can still be chosen. This process involves sophisticated pattern recognition. The system has learned what tends to follow based on its training data and doesn't understand the meaning of words in the sense that humans do. As randomness is built into every selection, the same prompt or question posed multiple times can produce different answers, meaning outputs cannot be taken at face value and require consistent human validation.

Generative (GenAI) and Agentic AI
Recent advances have been driven by GenAI systems capable of producing text, images, audio, video, and code that can be indistinguishable from work produced by humans. An important emerging development is agentic AI, which goes further still. Rather than responding to prompts, these systems plan and execute multi-step tasks with limited human intervention. This distinction matters for governance. GenAI primarily raises concerns around reliability and misinformation, while agentic AI raises particular challenges regarding accountability and oversight.

Uneven and Fragile Capabilities
Despite considerable progress, current AI systems exhibit what researchers call "jagged capabilities." They can produce answers comparable to those of a PhD-level specialist in fields such as physics, law, or mathematics, while failing unexpectedly on tasks humans find routine e.g. counting the number of objects in an image. AI system performance often becomes more fragile as task complexity increases. Systems that perform well in familiar settings may break down when facing novel combinations of factors or multi-step reasoning that falls outside their training data. This means strong benchmark performance under controlled testing conditions is not always a dependable guide to real-world reliability, underscoring the importance of ongoing evaluation of AI systems across their entire life cycle.

The Future of AI
Major technology companies and governments are investing heavily in AI in the expectation that future systems will become significantly more capable. Much of this investment is linked to the aspiration of developing Artificial General Intelligence (AGI), that would be capable of matching human cognitive performance across most

domains. Some go further, pursuing the more speculative goal of superintelligence; AI that would surpass human capabilities entirely. However, it remains unclear whether such systems are achievable, or even how these terms should be defined and measured in practice. Much of the recent improvement in AI systems has come from scaling by increasing the size of models, the amount of training data, and the computing power used to train them. While this has produced substantial gains, there are indications that this approach may be facing diminishing returns. In addition, current systems lack any genuine model of the world; recognising statistical patterns without understanding physical causality or how actions produce consequences.

Future AI may therefore look quite different from today's large general-purpose models, potentially involving collections of specialised systems working together in hybrid architectures. What is clear is that the trajectory of AI capability remains uncertain, meaning regulatory and oversight approaches will need to be agile enough to accommodate a wide range of possible technological developments.

Risks, Harms and the Ethical Imperative

As AI systems become embedded into our social and economic infrastructure, ensuring they are both technically safe and ethically robust becomes a central public policy challenge. Safety and ethics are not peripheral concerns, rather they are the preconditions for trust, adoption and long-term legitimacy of AI.

Reliability & Safety

Current frontier AI systems demonstrate impressive capabilities, but they remain imperfect. LLM's and GenAI systems can produce outputs that appear convincing but are factually incorrect, a phenomenon commonly referred to as hallucinations or confabulations. As previously discussed, these errors arise from the probabilistic nature of such systems, which generate likely responses rather than verified facts. While mitigation strategies and refinement of systems can reduce the occurrence of hallucinations (there was a 26 per cent reduction in hallucinations in GPT5 compared to GPT4o), given they are inherent to the ways AI systems operate, they cannot be fully eradicated. Reliance on AI-generated outputs without appropriate verification may therefore create risks of physical, psychological, reputational, or financial harm for individuals and organisations, and may undermine trust in AI systems. Evidence suggests that such reliance is not uncommon. A 2025 global study on attitudes toward artificial intelligence found that 42 per cent of employees report relying on AI outputs without evaluating the information, at least sometimes (Gillespie et al., 2025). This indicates that inaccurate or fabricated AI-generated content may influence real-world decisions when outputs are not critically assessed.

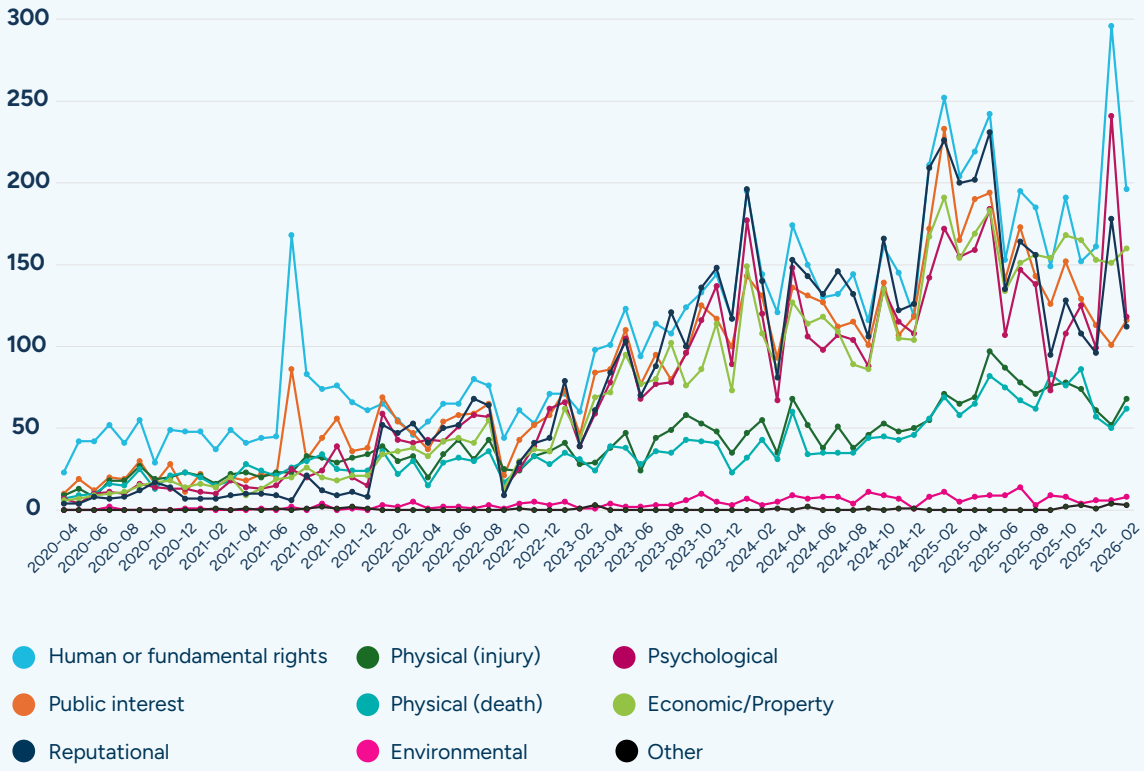
Malicious Use and Misuse

Alongside technical limitations, AI systems introduce new opportunities for misuse and deliberate harm. Generative AI tools can now produce highly convincing text, images, audio and video at minimal cost. While these capabilities enable valuable applications, they also lower the barriers for activities such as fraud, impersonation and disinformation. Deepfakes provide a clear illustration of this risk. A 2019 report found that 96 per cent of all deepfake video online were pornographic with almost all content targeting women (Ajder et al., 2019). Reports of AI-generated child sexual abuse material have also surged during recent years. In 2025, the UK's Internet Watchdog Foundation documented a 26,362 per cent rise in photo-realistic AI videos of child sexual abuse, compared to 2024 (Internet Watch Foundation, 2026). Similar deepfake techniques have been used in financial scams, including voice-cloning attacks in which criminals impersonate senior executives to authorise fraudulent transfers. Deepfakes also pose risks to democracy, journalism, and public discourse by enabling the creation of fabricated speeches, news footage, and other misleading content. In several recent elections, including in Slovakia, Taiwan, and Romania, AI-generated political content circulated online in attempts to manipulate public debate and undermine trust in institutions. Even when such material is identified and removed, its rapid dissemination can still shape public perception.

Cybersecurity is another domain where AI creates a dual-use dynamic. AI can improve defensive capabilities by automating threat detection and analysing network behaviour. However, the same technologies can be used by attackers to automate malware development, generate convincing phishing messages or identify system vulnerabilities at scale. As a result, the widespread availability of AI tools can reduce the technical expertise required to conduct sophisticated cyberattacks.

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 and hazards captured by AIM.

Figure 2: Evolution of AI Incidents and Hazards by Harm Type



Source: OECD Incidents Monitor.

Fairness, Bias and Structural Inequality

AI systems can reproduce and amplify existing social inequalities through several mechanisms. First, bias can arise from unrepresentative training data. Facial recognition systems, for example, have shown lower accuracy for women and people with darker skin tones because training datasets were dominated by lighter-skinned male images. Second, bias can be historically embedded in otherwise representative data. AI systems used in hiring or mortgage assessments have reproduced disparities in employment (less women hired) and lending (black applicants less likely to be approved) because the data reflects discriminatory patterns present when it was generated. Bias can also arise from the lack of diversity in AI development teams. Globally, women make up roughly 30 per cent of the AI workforce while more than half of data scientists and machine learning experts are aged 25-34 (Pal, Marino Lazzaroni and Mendoza, 2024; OECD, 2025). When development teams are relatively homogeneous in terms of gender, geography or cultural background, their assumptions and perspectives can inadvertently shape system design. As a result, systems may fail to anticipate the needs or experiences of more diverse populations. When these factors combine, AI systems can scale inequality across large numbers of decisions. While human judgement can also be biased, automated systems can apply such bias at far greater scale, turning it from an individual fairness issue into a systemic risk.

These concerns about bias within individual systems sit within a larger structural problem: who gets to build AI, and on whose terms. Advanced AI development is concentrated in a small number of companies and countries, primarily in North America, Europe and China. This concentration of influence could enable private actors to shape the trajectory of AI in ways in which risks are widely distributed but benefits remains narrowly concentrated. Currently the distribution of AI benefits is very uneven with adoption of AI in the Global North roughly twice that of the Global South. The concentration of AI development also shapes which languages are supported by AI systems, with dominant global languages receiving far greater representation in training data, tools and services than smaller or low-resource languages, including Irish. These structural inequalities are also evident within societies, where disparities in digital access, skills and AI literacy risk excluding certain population groups from the benefits of AI. In Ireland, digital exclusion is particularly pronounced among older adults, low-income households and rural communities, with younger adults almost nine times more likely to use AI frequently than those aged 55–64 (Eurofound, 2025). Addressing this divide will require targeted policies that expand digital access, strengthen AI literacy and ensure that emerging AI systems are designed with the needs of diverse populations in mind.

Privacy, Transparency and the Black Box Problem

AI systems also raise significant challenges for privacy and accountability. Many modern AI models rely on enormous datasets collected from the internet or other sources, which may contain personal or sensitive information. This raises questions about consent, data protection and the appropriate use of personal data in AI training.

Many advanced AI systems operate as opaque "black boxes." Even developers may not fully understand how complex neural networks arrive at particular outputs. This lack of interpretability can make it difficult to explain or challenge decisions influenced by AI systems. Opacity becomes particularly problematic when AI systems affect individuals' rights or opportunities, for example in hiring, lending, policing or welfare decisions. Closely related is what philosophers of technology call the "problem of many hands", where AI systems are developed by large teams, integrated by third parties and used across diverse contexts. This can often make it very difficult to attribute responsibility when something goes wrong.

Mitigation of AI Risks

Mitigating the risks associated with AI requires a layered approach combining technical safeguards, organisational practices and regulatory oversight.

Technical measures such as improving the representativeness of training data, applying bias detection tools, and using privacy-preserving methods like differential privacy or federated learning can help reduce harmful outcomes. Organisational practices also play an important role, including diverse development teams, robust testing procedures and continuous monitoring of systems after deployment. At the governance level, transparency tools, independent audits and regulatory frameworks such as the EU AI Act help establish accountability and common safety standards. While these measures cannot eliminate all risks, together they can significantly reduce the likelihood and impact of harmful outcomes while supporting responsible innovation.

AI and Society: Impacts on Work, Economy and Environment

Artificial intelligence is increasingly shaping economic activity, labour markets and everyday life. Understanding these dynamics requires a socio-technical perspective, recognising that AI systems do not operate in isolation but interact with the economic structures, institutional arrangements, workforce capabilities and social norms within which they are deployed. The opportunities and risks associated with AI therefore arise not only from the technical properties of the systems themselves but from how they are implemented, governed and used in practice.

The Pace and Pattern of AI Adoption

AI adoption is advancing rapidly but unevenly across sectors, organisations and economies. Within the European Union, approximately 13.5 per cent of enterprises reported using AI technologies in 2024, although adoption rates vary widely across countries and industries. Ireland has experienced relatively rapid growth in enterprise adoption, with the share of firms using AI increasing from 8 per cent in 2023 to 14.9 per cent in 2024 (Eurostat, 2025). Adoption also varies significantly by firm size, with larger and more productive firms are better positioned to invest in the infrastructure, data capabilities and specialist expertise required to deploy AI systems effectively. In Ireland, 51.2 per cent of large enterprises reported using AI technologies in 2024, compared with 25.1 per cent of medium-sized firms and 12 % of small enterprises (Central Statistics Office, 2025). At the level of individuals, diffusion has been particularly rapid. By the end of 2025, an estimated 44.6 per cent of Ireland's working-age population were using generative AI tools, placing Ireland among the fastest adopters globally (Microsoft AI Economy Institute, 2026).

Nonetheless, many organisations encounter difficulties moving from initial pilots to widespread operational use. This dynamic is sometimes described as the 'AI adoption paradox': while experimentation with AI tools is widespread, integrating these systems into core workflows at scale often proves far more challenging. Successful deployment depends on foundational digital capabilities, including robust digital infrastructure, high-quality and interoperable data, and effective governance frameworks. Consequently, AI adoption is not simply a technical upgrade but a broader organisational transformation. Firms must invest not only in AI tools themselves but also in the underlying digital architecture, data governance frameworks and workforce capabilities needed to support their effective use.

This gap between formal adoption and operational reality is further complicated by the rise of 'shadow AI'. Employees increasingly use publicly available generative AI tools without formal organisational approval, exposing a gap between official adoption policy and workplace reality. Worker sentiment adds a further dimension to this picture. Many employees recognise the potential of AI tools to improve productivity and reduce repetitive tasks, but concerns about job security, workplace surveillance and the erosion of professional autonomy remain widespread. These attitudes are not simply a barrier to uptake; they reflect legitimate questions about how the gains and risks of AI adoption are distributed within organisations. Addressing them

requires meaningful employee involvement in decisions about how AI technologies are introduced and used, not as an afterthought to deployment, but as a condition of it.

Trust & Social License

Public attitudes will play a critical role in shaping the trajectory of AI adoption. While many people recognise the potential of AI to improve productivity, public services and economic growth, concerns remain about privacy, misinformation, bias and the concentration of technological power.

Evidence suggests that Ireland exhibits relatively cautious public attitudes toward AI compared with many other countries. A recent international survey found that only 38 per cent of Irish respondents reported trusting AI systems, compared with a 47-country average of 46 per cent (Gillespie, et al., 2025). Trust also varies across institutions. Irish respondents expressed higher levels of confidence in AI systems developed by universities, research organisations and healthcare institutions than in those deployed by governments or private technology companies. These findings highlight that trust in AI is shaped not only by the technology itself but also by the credibility of the institutions responsible for its development and deployment. Perceptions of fairness, transparency and accountability play an important role in determining whether individuals are willing to accept the use of AI systems in areas that affect their lives.

In this context, the concept of social licence becomes increasingly important. Legal compliance alone may not be sufficient to secure public legitimacy for AI deployment, particularly in high-impact areas such as healthcare, policing or public administration. Public deliberation in the form of genuine two-way dialogue can give citizens a role in determining where AI should and should not be used, what boundaries should be set, and what trade-offs are acceptable.

By embedding deliberation into governance cycles, Ireland can ensure that AI development remains aligned with democratic values and gives citizens genuine agency in shaping technological futures. Without meaningful public engagement, AI systems risk rejection or loss of legitimacy regardless of their technical performance.

Labour Market Impacts

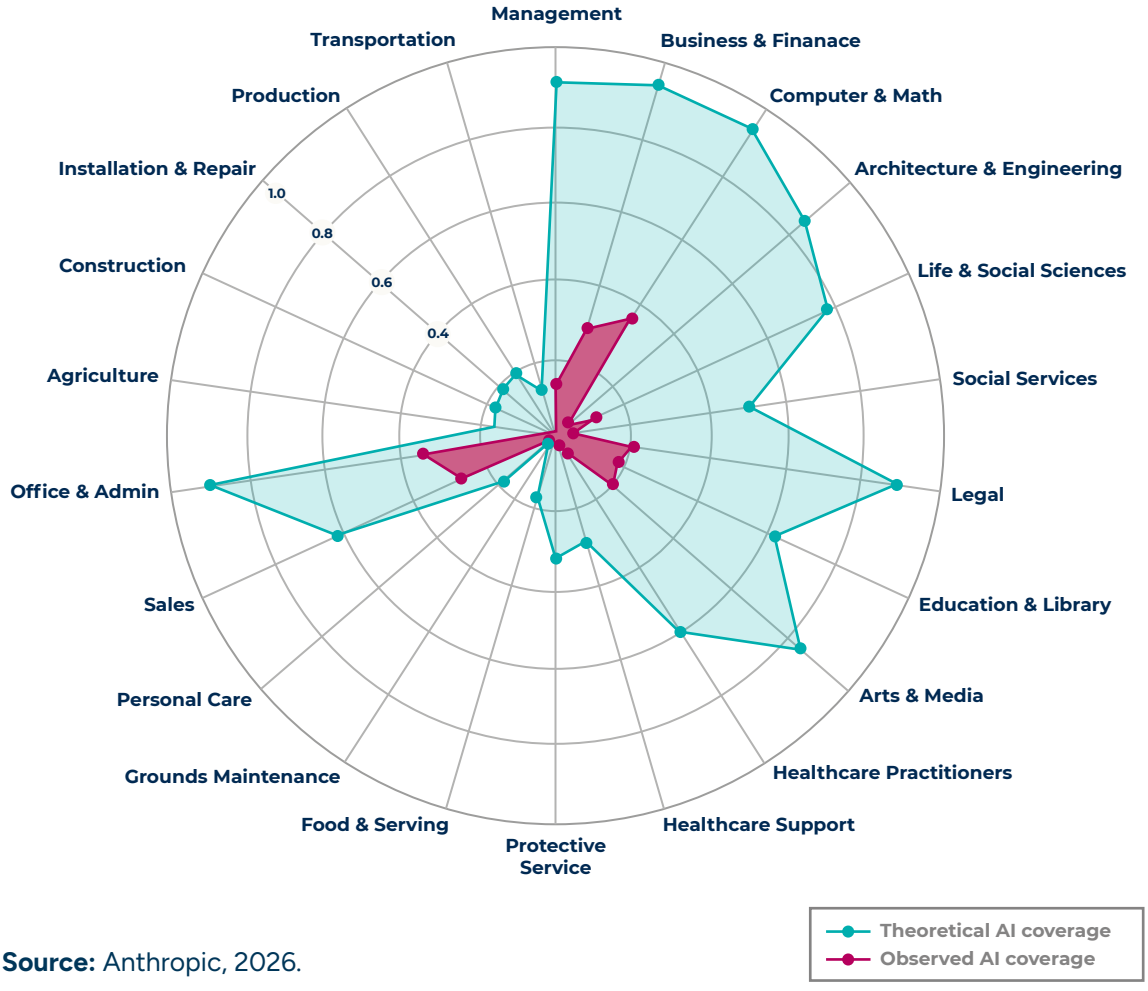
AI represents not merely a technological shift but a reconfiguration of work capable of reshaping labour markets and redistributing skills and responsibilities. The technology has the potential to both substitute for and complement human labour, and the balance between these dynamics will largely determine its social impact. Where complementarity dominates, workers benefit by focusing on higher-value and interpersonal tasks; where substitution dominates, risks of displacement, deskilling and unemployment are more acute.

Current evidence suggests the principal near-term effect will be reallocation of tasks within jobs rather than elimination of occupations. A recent Anthropic study found no systematic increase in unemployment in highly exposed occupations since late 2022, consistent with broader international evidence (Anthropic, 2026). Displacement risk is nonetheless concentrated in identifiable areas including roles involving information processing, routine drafting, summarisation and standardised interaction, with clerical, administrative and sales support functions most vulnerable. Professional roles in accountancy, legal services and

software development present a more mixed picture, with outcomes depending heavily on how organisations redesign work. The Anthropic study illustrates the scale of latent exposure showing that within computer and mathematical occupations, AI could theoretically handle 94 per cent of tasks, though observed usage currently covers around 33 per cent. For now, there exists a substantial gap between what AI could theoretically automate across occupational categories and how much it is being used in practice.

Ireland is marginally more exposed than the advanced economy average, with 63 per cent of Irish employment in highly AI-exposed occupations compared with an advanced-economy benchmark of approximately 60 per cent (Department of Finance and Department of Enterprise, Trade and Employment, 2024). The costs of adapting to these changes are unlikely to be shared equally. Women are disproportionately represented in higher-risk administrative roles, and effects on younger workers are an emerging concern. Employment among 15–29-year-olds in AI-exposed sectors in Ireland fell between 2023 and 2025 despite overall sectoral

Figure 3: Theoretical capability and observed usage by occupational category



Source: Anthropic, 2026.

growth (Department of Finance, 2026). This is a pattern replicated internationally, with hiring among workers aged 22–25 slowing in AI-exposed occupations, suggesting a contraction in entry-level roles (Brynjolfsson et al., 2025). Longer-term impacts remain harder to predict, and will depend on the pace of adoption, organisational responses and the effectiveness of policy in supporting workforce adjustment.

Skills, Reskilling and the Risk of De-skilling

As AI reshapes the tasks performed within occupations, it is also altering the skills required across the workforce. Technical capabilities such as data analysis, programming and AI literacy are becoming increasingly valuable, while human-centred skills such as critical thinking, creativity, problem-solving and ethical judgement, are likely to grow in importance as workers collaborate more closely with AI systems. Ensuring workers can adapt to these changes will be a central determinant of whether AI adoption generates inclusive productivity gains or widens existing inequalities.

Individuals need opportunities for lifelong learning as job roles evolve; organisations need to invest in workforce training to support effective AI integration; and governments need to strengthen education and training systems that enable workers to transition between roles as technology changes.

Ireland enters this transition with a comparatively strong foundation. An IMF analysis suggests Ireland is among the countries best positioned to meet future AI-related skills needs, reflecting its higher education system and established technology sector (Jaumotte et al., 2026). However, demand for AI and data science expertise is growing rapidly across advanced economies, and competition to attract and retain skilled workers is intensifying. Maintaining and expanding Ireland's talent base will require sustained investment in education, training and retention policies.

A less discussed but important risk is de-skilling. While AI tools can enhance productivity by automating routine elements of tasks, excessive reliance on AI-generated outputs may over time erode the underlying competencies they are designed to support. This 'de-skilling paradox' suggests that realising the full potential of AI as a complement to human capability requires careful attention to how tools are integrated into workflows: preserving opportunities for workers to exercise and develop core professional skills, maintaining meaningful human oversight, and ensuring workers understand how AI systems generate their outputs.

Productivity Potential

Artificial intelligence has the potential to deliver substantial productivity and economic benefits by automating routine tasks, supporting decision-making and accelerating research and development. These capabilities can lower costs, increase efficiency and stimulate innovation. Early empirical evidence suggests that AI tools can deliver measurable productivity improvements in specific contexts. Studies examining the use of generative AI in workplace settings have found that workers using these systems can complete certain tasks more quickly, particularly in areas such as software development, customer service and writing-intensive work. A recent meta-analysis of research in this area estimated average productivity improvements of around 17 per cent for specific tasks when generative AI tools are used effectively. Other studies indicate that workers may save several hours

per week when AI systems are integrated into routine workflows, although these gains depend heavily on training, task suitability and organisational readiness. This highlights the risks associated with inefficient forms of automation, i.e. those that can lead to increased costs or the need for additional work, sometimes referred to as 'so-so automation'.

Beyond efficiency improvements, AI may also support new forms of higher-value work. By automating routine elements of tasks, AI systems have the potential to enable workers to concentrate on activities that generate greater economic and social value, including innovation, complex problem-solving and relationship-based work. Realising these benefits, however, will depend on how organisations redesign roles, workflows and management practices to integrate AI effectively.

Productivity benefits typically lag behind technological implementation, and AI's impact remains modest and difficult to detect in national productivity statistics. This pattern is consistent with the 'productivity J-curve' hypothesis, which holds that gains are initially suppressed by the need for complementary investment in data infrastructure, workforce training and workflow redesign before they materialise at scale (Brynjolfsson et al., 2021).

Environmental Impacts

The environmental footprint of AI is substantial and operates across multiple dimensions. Direct impacts arise from the energy and water required to train and operate large-scale AI systems. Data centres are significant energy consumers: in Ireland, they accounted for approximately 22 per cent of metered electricity consumption in 2024, up from 5 per cent in 2015, with contracted demand projected to reach ≥30 per cent of Ireland's supply by 2030. Each query to a large language model is estimated to consume considerably more energy than a conventional internet search. Early estimates suggested as much as ten times more (de Vries, 2023), though more recent data indicate the gap may be narrowing as model efficiency improves.

The wider structural consequences of AI's expansion raise additional concerns. Where AI delivers efficiency gains e.g. in logistics, energy use or industrial processes, these can bring environmental benefits but rebound effects may erode them. When a process becomes cheaper or faster, overall activity often increases, consuming more resources than the efficiency saved. Over the longer term, risks include lock-in to energy-intensive digital infrastructures, growing demand for critical minerals used in advanced computing hardware, and increasing pressure on environmental governance frameworks as digital infrastructure expands.

At the same time, AI has meaningful potential as a tool for environmental benefit by supporting climate modelling, optimising renewable energy integration, improving demand forecasting in electricity grids, and enhancing efficiency in agriculture, logistics and building management. Realising this potential while managing the sector's own footprint will require coordinated action across technology developers, infrastructure providers, energy planners and policymakers.

Governing AI: From Compliance to Anticipatory Leadership

The International Governance Landscape

The governance of AI is increasingly shaped by a growing ecosystem of international principles, frameworks and cooperative initiatives.

Key Legal AI Governance Instruments

- OECD - Recommendation on AI **(2019, updated 2023).**
  Sets global principles and policy guidance for trustworthy, human-centred AI.
- UNESCO - Recommendation on the Ethics of AI **(2021).**
  Establishes universal ethical and human rights standards for AI governance.
- CoE Framework Convention on AI **(2024).**
  Binding pan-European legal instrument to govern AI based on human rights.
- EU AI Act **(2024).**
  Legally regulates AI using a risk-based framework, high-risk attracts additional legal obligations.

Source: NESC Secretariat.

Within this landscape, the EU AI Act represents the most comprehensive binding regulatory framework currently governing artificial intelligence. It adopts a risk-based model, categorising AI systems according to the level of risk they pose. Systems presenting unacceptable risk, such as manipulative or exploitative AI and social scoring are prohibited outright. High-risk systems, including those used in employment, education, healthcare, law enforcement and critical infrastructure, are subject to extensive obligations covering risk management, data quality, transparency, documentation, human oversight and post-market monitoring, and must undergo conformity assessment before entering the European market. Limited-risk systems such as chatbots must comply with transparency requirements ensuring users are aware they are interacting with AI. Most other systems fall into a minimal-risk category where regulatory intervention remains limited. Complementing the EU AI Act, the European Commission's proposed Digital Omnibus initiative seeks to streamline elements of the EU's expanding digital regulatory framework and reduce unnecessary administrative complexity.

It is worth noting that regulation need not function solely as a constraint on technological development. In emerging technology sectors, clear and credible governance frameworks can actively support innovation by providing legal certainty, establishing common standards and strengthening public trust.

Governance of AI in Ireland

Ireland's national approach to AI governance is shaped by both European regulatory obligations and domestic strategic priorities. The Government's Digital and AI Strategy 2030 sets out an ambition for Ireland to become a leader in the development and deployment of responsible and trustworthy AI, emphasising regulatory frameworks that are proportionate yet robust. National legislation is currently being developed to give effect to the EU AI Act, with the *General Scheme of the Regulation of Artificial Intelligence Bill 2026* setting out the proposed governance arrangements.

Ireland has opted for a distributed regulatory model, under which existing sectoral regulators spanning finance, communications, consumer protection, transport and employment, have been designated as national competent authorities responsible for supervising AI systems within their respective domains. Several public bodies have also been designated as fundamental rights authorities tasked with overseeing compliance with rights protections under the Act. This structure leverages existing regulatory expertise and reflects the cross-sectoral nature of AI deployment, though it necessitates strong coordination mechanisms to ensure consistency in the application of the Regulation.

To support this coordination, the Government has signalled its intention to establish an AI Office of Ireland as the central authority responsible for implementing the Act and serving as the State's single national point of contact. The Office will coordinate sectoral competent authorities, provide access to specialised technical expertise, and oversee initiatives such as national regulatory sandboxes for testing innovative AI systems. These arrangements are complemented by a range of advisory and capacity-building mechanisms, set out in the National Digital and AI Strategy 2030.

Trustworthy AI and Human Oversight

Across emerging governance frameworks there is increasing convergence around trustworthy AI as a guiding principle,

encompassing values such as safety, fairness, transparency, accountability and respect for fundamental rights. It is considerably less clear how these principles can be operationalised in practice. Translating abstract values into measurable, verifiable criteria that can withstand regulatory and public scrutiny is challenging, and this 'principle-to-practice gap' has become a major area of focus as AI adoption accelerates.

Practical, accessible tools such as sectoral practice-based workbooks offering end-to-end guidance on applying ethical principles are critical for enabling organisations to move beyond well-intentioned principles and embed ethical and safe AI practices in everyday decision-making.

Closing the principle-to-practice gap will also require organisations to build genuine internal ethical capability thereby equipping practitioners with the knowledge, confidence and contextual judgement to interpret and apply principles in the specific circumstances they encounter.

Human oversight is a central component of trustworthy AI, and while it was historically framed as an ethical expectation, it is increasingly embedded as a regulatory requirement, particularly within the EU AI Act for high-risk systems. The objective is to ensure that humans retain meaningful control over consequential decisions and that responsibility for outcomes remains clearly assigned.

Meaningful human oversight requires robust accountability structures, transparency regarding how systems function, clearly defined decision-making responsibilities, and institutional processes that allow human decision-makers to intervene effectively when required.

Governing Under Uncertainty: Anticipatory Governance

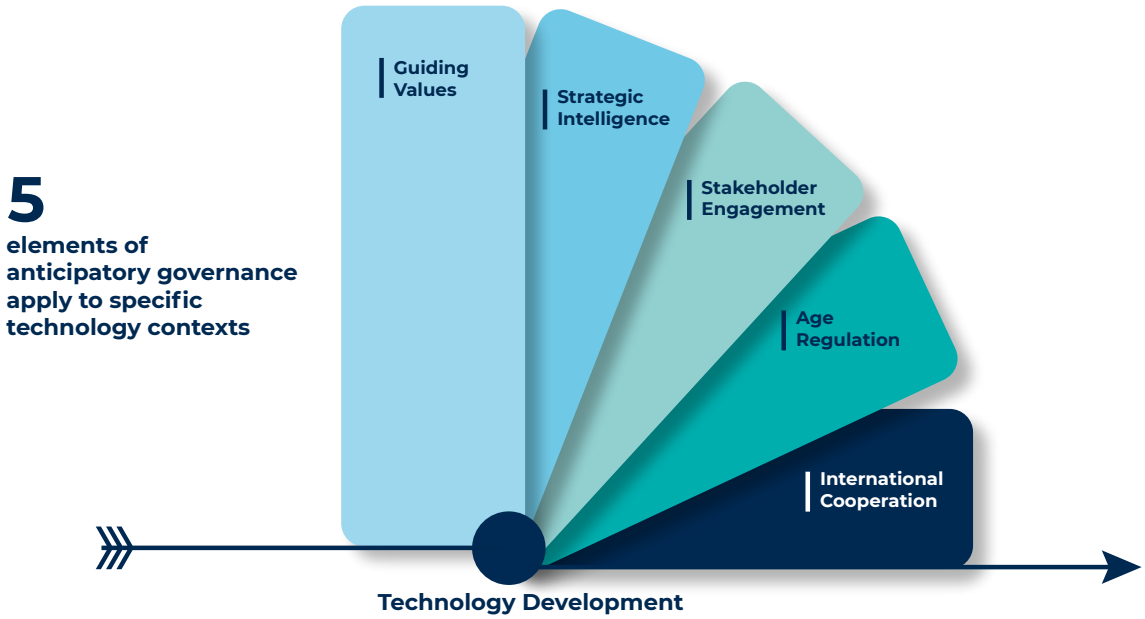
AI presents policymakers with a fundamental governance challenge: regulating a technology whose capabilities, applications and societal impacts remain highly uncertain. Traditional regulatory approaches are often poorly suited to this environment, as they typically assume relatively stable technologies and clearly identifiable risks. In contrast, AI contrast, evolves rapidly, diffuses across sectors and frequently demonstrates uneven or jagged performance in real-world settings.

In response, policymakers are increasingly exploring anticipatory governance, an approach designed to embed foresight, adaptability and continuous learning into regulatory systems. Rather than reacting to technological developments after harms occur, anticipatory governance seeks to identify emerging risks and opportunities early and incorporate them into policymaking before systems become deeply embedded in society. It operates through a set of mutually reinforcing practices: strategic foresight to explore alternative technological futures and their societal implications; horizon scanning to systematically monitor emerging trends, technologies and risks; and broad stakeholder engagement that brings together policymakers, researchers, industry and civil society to ensure regulatory responses are grounded in diverse perspectives and practical experience.

Anticipatory governance also emphasises experimentation and learning, often through mechanisms such as pilot initiatives or regulatory sandboxes, which allow policymakers to test new technologies and governance approaches in controlled environments. What ultimately distinguishes it from traditional regulatory models, however, is the emphasis on continuous monitoring, evaluation and policy iteration.

As previously discussed, because AI systems often perform unevenly across different tasks and environments, governance frameworks must remain adaptable and capable of incorporating real-world evidence over time. In this way, anticipatory governance complements formal regulation such as the EU AI Act by ensuring that policy frameworks remain responsive to both emerging risks and evolving technological capabilities.

Figure 5: Framework for Anticipatory Governance of Emerging Technologies



Source: OECD, 2024.

AI Literacy

As AI becomes embedded across public services, workplaces and everyday life, the capacity of individuals and institutions to understand, use and evaluate AI is key to realising the opportunities the technology offers.

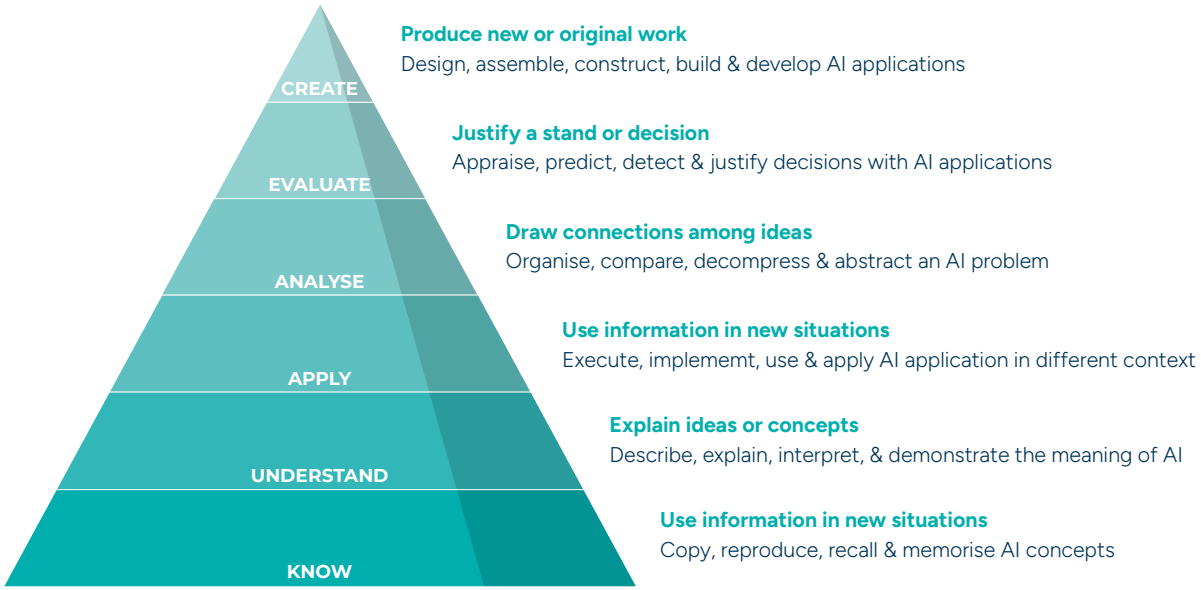
AI literacy should therefore be understood not as a technical training issue or a nice-to-have, but as a form of national capability infrastructure, that enables the transition from passive technological adoption to informed and responsible engagement with AI.

The Imperative for AI Literacy

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. Ng et al. (2021) suggest a model of categorising AI literacy; the six segments represent increased literacy levels from the most basic, 'knowing' AI, to the most advanced, 'creating'. This framework suggests that the 'know' and 'understand' levels of should be widely attainable and measurable.

The benefits of literacy operate at multiple levels. At the individual level, it enables people to interpret AI-generated outputs critically, recognise potential biases or errors, and use AI tools effectively and responsibly. At the organisational level, AI-literate workforces are better positioned to identify appropriate use cases, manage risks and integrate AI into workflows in ways that enhance productivity rather than create new vulnerabilities. At the societal level, widespread literacy supports democratic oversight by enabling citizens to participate in discussions about the acceptable uses of AI, the trade-offs between innovation and regulation, and the values that should shape the development of emerging technologies.

Figure 6: Categorising AI literacy



Source: Ng *et al.*, 2021.

Life-Course Approach to AI Literacy

Developing AI literacy requires a life-course approach, recognising that different groups encounter AI in different contexts and require different forms of knowledge and skills.

Education:

Schools and higher education institutions are central to building long-term societal capacity. Embedding AI literacy across curricula can help ensure that future workers, citizens and leaders develop a baseline understanding of AI technologies and their impacts. This should encompass not only a technical understanding of how AI systems operate but critical skills such as interpreting algorithmic outputs, understanding data biases, and recognising the broader societal implications of automated decision-making. Embedding AI literacy across curricula can help ensure that future workers, citizens and leaders develop a baseline understanding of AI technologies and their impacts.

Employees and organisations:

In the workplace, AI literacy is increasingly a strategic capability. As AI systems become integrated into core organisational processes, employees need sufficient understanding of how these systems function, how their outputs should be interpreted and where their limitations lie. Leadership literacy is equally important. Senior executives and public sector leaders responsible for AI procurement, deployment and governance do not require deep technical expertise, but must understand the strategic implications of AI technologies, their potential contributions to productivity and innovation, and the risks associated with poorly governed deployment. Leadership engagement is critical in shaping organisational culture, establishing governance frameworks and ensuring regulatory compliance. Without it, organisations risk fragmented AI adoption, misaligned investment decisions and governance gaps that could expose them to operational or reputational risks.

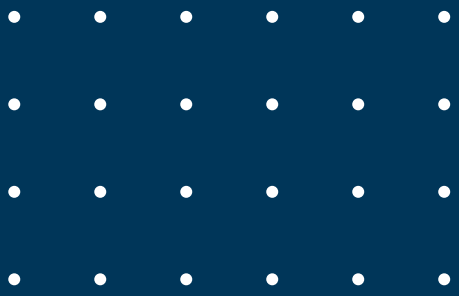
The public:

AI systems increasingly mediate access to information, financial services, employment opportunities and public services. As these technologies shape everyday experiences, citizens require sufficient understanding to recognise how AI influences decisions that affect their lives, evaluate AI-generated content, identify misinformation or manipulation, and engage in democratic discussions about where AI should and should not be used. Despite growing use of AI tools, surveys suggest that many people remain uncertain about how AI systems operate or where they are deployed in everyday products and services. In a 2025 global study, while 52 per cent of Irish respondents stated they felt confident using AI tools, only 38 per cent believed they had the skills and knowledge to use AI appropriately, highlighting a gap between perceived ease of use and deeper understanding (Gillespie, et al., 2025). Without a baseline level of public understanding, debates about AI risk being dominated by technical experts or industry actors, potentially undermining the legitimacy of governance decisions.

References

- Ajder, H., Patrini, G., Cavalli, F. and Cullen, L. (2019). The State of Deepfakes: Landscape, Threats, and Impact. [online] Available at: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf [Accessed 27 Aug. 2025].
- Anthropic (2026). Labor market impacts of AI: A new measure and early evidence. [online] Anthropic.com. Available at: https://www.anthropic.com/research/labor-market-impacts [Accessed 10 Mar. 2026].
- Azubuike, J. (2026). Page Restricted. [online] Linkedin.com. Available at: https://www.linkedin.com/search/results/all/?keywords=the%20global%20data%20and%20AI%20regulatory%20and%20privacy%20map&origin=GLOBAL_SEARCH_HEADER [Accessed 11 Mar. 2026]
- Brynjolfsson, E., Chandar, B., Chen, R., Bloom, N., Gans, J., Autor, D., Rock, D., Li, F.-F., Li, F., Langer, C., Bana, S., Cook, C., Forman, C., Wang, A., Ross, B., Maghzian, O., Halperin, B., Pei, J., Trammell, P. and Bergman, E. (2025). Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence. [online] Available at: https://digitaleconomy.stanford.edu/wp-content/uploads/2025/08/Canaries_BrynjolfssonChandarChen.pdf [Accessed 22 Aug. 2025].
- Central Statistics Office (2025). Artificial Intelligence Information Society Statistics - Enterprises 2024 - Central Statistics Office. [online] www.cso.ie. Available at: https://www.cso.ie/en/releasesandpublications/ep/p-isse/informationsocietystatistics-enterprises2024/ [Accessed 16 Aug. 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. doi: https://doi.org/10.1080/10494820.2025.2514372.
- de Vries, A. (2023). The growing energy footprint of artificial intelligence. Joule, 7(10). doi: https://doi.org/10.1016/j.joule.2023.09.004.
- Department of Finance (2026). Economic Insights- Volume 1 2026. [online] gov.ie. Government of Ireland. Available at: https://www.gov.ie/en/department-of-finance/publications/economic-insights-volume-1-2026/ [Accessed 20 Feb. 2026].
- Department of Finance & Department of Enterprise, Trade and Employment (2024). Artificial Intelligence: Friend or Foe? Summary and Public Policy Considerations . [online] Available at: https://www.gov.ie/en/department-of-finance/publications/artificial-intelligence-friend-or-foe/ [Accessed 27 Aug. 2025].
- Eurofound (2025). Narrowing the digital divide: Economic and social convergence in Europe's digital transformation. [online] 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 Sep. 2025].

- Eurostat (2025). Use of artificial intelligence in enterprises. [online] ec.europa.eu. Available at: https://ec.europa.eu/eurostat/statistics-explained/SEPDF/cache/106920.pdf [Accessed 19 Aug. 2025].
- Gillespie, N., Lockey, S., Ward, T., Macdade, A. and Hassed, G. (2025). Trust, attitudes and use of artificial intelligence: A global study 2025. [online] University of Melbourne, KPMG International. Available at: https://mbs.edu/faculty-and-research/trust-and-ai [Accessed 21 Aug. 2025].
- Internet Watch Foundation (2026). AI becoming 'child sexual abuse machine', warns IWF. [online] Iwf.org.uk. Available at: https://www.iwf.org.uk/news-media/news/ai-becoming-child-sexual-abuse-machine-adding-to-dangerous-record-levels-of-online-abuse-iwf-warns/ [Accessed 2 Mar. 2026].
- 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. Staff Discussion Note SDN2026/001. Washington, D.C. : International Monetary Fund.
- Microsoft AI Economy Institute (2026). Global AI Adoption in 2025 A Widening Digital Divide. [online] Available at: https://www.microsoft.com/en-us/research/wp-content/uploads/2026/01/Microsoft-AI-Diffusion-Report-2025-H2.pdf?msock-id=29b617f9e7d3673e14d101c8e639661b [Accessed 18 Jan. 2026].
- 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, [online] 2(1). doi: https://doi.org/10.1016/j.caeai.2021.100041.
- OECD (2024). Framework for Anticipatory Governance of Emerging Technologies. [online] 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 Aug. 2024].
- OECD (2025). Live data from OECD. AI. [online] Oecd.ai. Available at: https://oecd.ai/en/data?selectedArea=ai-demographics&selectedVisualization=ai-demographicsby-age [Accessed 4 Sep. 2025].
- Pal, S., Marino Lazzaroni , R. and Mendoza, P. (2024). AI's Missing Link: The Gender Gap in the Talent Pool. [online] Berlin: interface. Available at: https://www.interface-eu.org/publications/ai-gender-gap [Accessed 4 Sep. 2025].



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

www.nesc.ie