

Generative AI – an overview

Research Briefing

September 2026



The Welsh Parliament is the democratically elected body that represents the interests of Wales and its people.

Commonly known as the Senedd, it makes laws for Wales, agrees Welsh taxes and holds the Welsh Government to account.

An electronic copy of this document can be found on the Senedd's website:
research.senedd.wales

Copies of this document can also be obtained in accessible formats including Braille, large print, audio or hard copy from:

Welsh Parliament
Tŷ Hywel
Cardiff Bay
CF99 1SN

X: **[@SeneddResearch](https://twitter.com/SeneddResearch)**
Senedd Research: **research.senedd.wales**
Subscription: **[Newsletter](#)**

© **Senedd Commission Copyright 2026**

The text of this document may be reproduced free of charge in any format or medium providing that it is reproduced accurately and not used in a misleading or derogatory context. The material must be acknowledged as copyright of the Senedd Commission and the title of the document specified.

Generative AI – an overview

Research Briefing

September 2026

Authors:

Amandine Debus and Jessica Woodgate

Senedd Research acknowledges the parliamentary fellowships provided to Amandine Debus by the Natural Environment Research Council and Jessica Woodgate by PolicyBristol and the Research England Policy Support Fund, which enabled this Research Briefing to be completed.

This Research Briefing outlines generative AI and its relevance to the Welsh public sector, including opportunities for use and relevant risks. It explains current regulation of AI in Wales, the UK and international approaches.



Contents

Summary.....	6
1. What is generative AI and what are the opportunities for the public sector?	8
Definition	8
AI agents.....	10
Wearable AI	10
Relevance for the public sector	11
Economic opportunities	12
Specific opportunities in Wales.....	14
2. What are the risks and challenges?	16
Technical risks and limitations of generative AI	16
Theoretical technical limitations.....	16
‘Real-world’ examples.....	18
Risks beyond technical limitations: stakeholder concerns.....	18
Existing guidance to mitigate risks for ethical and safe use	26
3. What is the current regulation of generative AI in Wales and other countries?	28
Regulations and competency in Wales.....	28
Welsh Government AI policy.....	28
Existing legislation for Wales	29
Competency for further legislation in Wales.....	31
UK policy and legislation	33
AI-specific strategy	33
Implications of UK AI policy for Wales.....	34
EU policy.....	35
Other countries	38
4. What are the challenges to designing and implementing policy?	42
Fast-evolving.....	42

Defining harm and assessing damage 42

Need for international approaches..... 43

Acronyms.....45

Summary

What are the opportunities and risks of generative artificial intelligence (AI) in the public sector?

AI can be defined as the science and engineering of machines to perform tasks typically associated with human intelligence, such as learning, reasoning and planning.

Generative AI is a type of AI to produce content like text, images, audio, or video. Generative AI is increasingly being applied to make technologies like **AI agents** and **AI glasses** interactive and autonomous.

Generative AI offers **opportunities** to improve the efficiency, quality and accessibility of services provided by the public sector. Generative AI also offers **specific opportunities** to promote and support the use of Welsh language.

However, generative AI outputs depend on the **quality of the data** used to train models, as poor quality inputs can lead to bias, mistakes, or misinformation. If there are limited resources dedicated to training models in some languages, **such as Welsh**, it can also mean that models don't perform well for these languages.

Beyond technical limitations, other risks concerning stakeholders include the current **lack of regulation, impacts on jobs** and **human behaviour, privacy, data protection** and **copyright issues**.

What is the current approach to regulating AI in Wales?

There is currently **no specific holistic approach** to legislating for AI (of which generative AI is a subfield) in Wales. In the UK, AI is regulated **in the context it is used** through **existing regulatory frameworks**, for example, **financial services, copyright** and **healthcare** legislation.

The previous Welsh Government outlined its **AI Plan for Wales**, which does not include additional legislation, in November 2025. The Plan establishes the **Office for AI and the Strategic AI Advisory Group**. It outlines the Government's approach to use AI to drive economic growth and enhance public services, emphasising collaborative decision-making and equity.

The UK Government published its **AI regulation white paper** in March 2023, which states AI regulation should be done within the existing powers of regulators and devolved to sectors. It is based on a ‘pro-innovation’ approach, with **policies mainly focusing on** boosting the AI sector to drive growth and productivity across the UK economy.

In its January 2025 **AI Opportunities Action Plan**, the UK Government committed to developing **AI Growth Zones (AIGZs)** to accelerate AI infrastructure development. Two AIGZs are in Wales: one in **North Wales**; and one along the M4 corridor in **South Wales**.

Protection of personal information, intellectual property and internet services **are reserved matters**, meaning that the Welsh Government and the Senedd do not have power to make law in these areas, so Wales is subject to the same law as the UK. However, the wider governance of data protection regulations **does not always involve a straightforward distinction between ‘reserved’ and ‘devolved’**.

There is limited scope for the Senedd to make law relating to **equal opportunities to address discriminatory use of AI**. The **Well-being of Future Generations Act (Wales) 2015** means that decisions on AI adoption by public bodies must consider the impact on future generations, sustainability and long-term prosperity, as well as immediate benefits.

What are the challenges of regulating AI?

Since the field of AI evolves rapidly, regulation **needs to account for new innovations**.

In addition, the impacts of AI **are not always easy** to define, which means it can be difficult to specify who has been harmed and when.

Finally, **it is recognised that** international collaboration is needed to avoid displacing problems. Despite **efforts** to collaborate, for example at the **India AI Impact Summit** in February 2026, strategies planned by different countries vary.

1. What is generative AI and what are the opportunities for the public sector?

This section overviews the field of AI, explains what generative AI is, and summarises its relevancy for the public sector.

Definition

AI studies the science and engineering of machines to perform tasks typically associated with human intelligence, including learning, reasoning and planning. An algorithm is a list of rules to complete a task, like a recipe.

Within AI, machine learning (ML) algorithms apply statistical analysis to identify patterns in data and make predictions. A sub-branch of ML is deep learning, characterised by the use of artificial neural networks (ANNs) which are inspired by how neurons in the brain work. ANNs are composed of multiple artificial neurons defined by numerical parameters. Parameters can be incrementally tweaked so that the network adapts to input data (also known as training data). An ANN gradually encodes patterns from training data to produce a model, which is a mathematical representation of patterns in data that can be applied to make predictions about new data. Various techniques can be used to train ML models, such as:

- Supervised learning, i.e. the model is trained on labelled data (i.e. data tagged with properties in the context of a specific task);
- Unsupervised learning, i.e. the training data is not labelled;
- Semi-supervised learning, i.e. a small part of the training data is labelled and a large part is unlabelled;
- Self-supervised learning, i.e. the training data is unlabelled and the model iteratively predicts pseudo-labels; and
- Reinforcement learning, i.e. the model learns by trial and error.

Generative AI is a subfield of AI that uses ML to produce content such as text, images, audio, videos, or music. Generative AI encodes patterns in data such as

webpages, social media and other online content. There are two main branches of generative AI: generating language; and generating images, audio and video.

Approaches to generate language apply **natural language processing (NLP)** techniques (the use of machines to handle language), in particular the **generative pre-trained transformer (GPT)** algorithm. A GPT can encode patterns in sequential data like text, by **detecting relationships** between parts of a sentence and using these patterns to predict the next likely word given an input. A **large language model (LLM)** does this on massive amounts of text data with millions or billions of parameters. **LLMs are computationally intensive and can take a long time to be trained.**

Approaches to generate images, audio and videos include **diffusion models** which start from “**noise**” (random irregularities in data) and gradually refine it into a meaningful output. Diffusion models have been used for applications like **deepfake technology**, defined in Note 1 below.

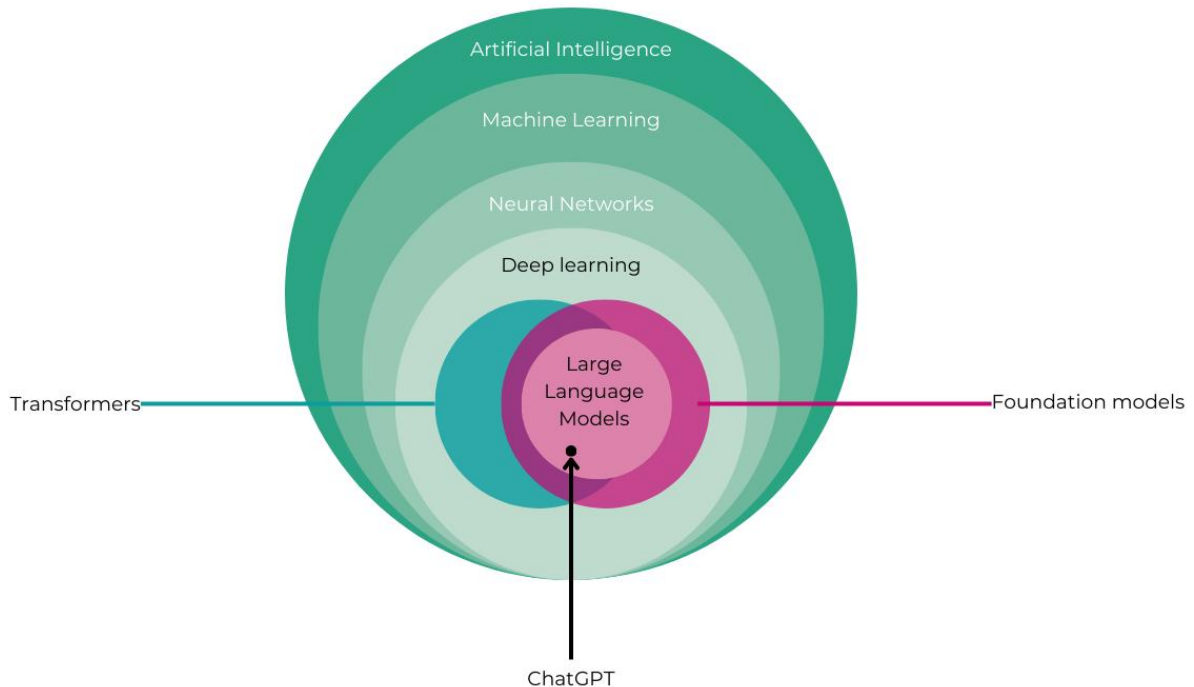
Note 1

Deepfakes describe synthetic media, often videos, sound recordings, or pictures, that can be used to “replace someone's face or voice with that of someone else, in a way that appears real”.

Most of the popular generative AI technologies **rely on the use of foundation models**, which are large-scale ML models trained on vast quantities of data. Foundation models are **not task-specific** and are often called **general-purpose** models. However, **generative AI can** also be designed for specific tasks and use approaches that are not foundation models (e.g. small generative AI models **for creative work** like making music).

Other terms are sometimes used such as “**frontier models**” to describe cutting-edge models, or “**artificial general intelligence (AGI)**” and “**strong AI**” to describe highly advanced AI. However, these terms are inconsistently defined.

Figure 1: Venn diagram describing the common distinction between types of AI



AI agents

An AI agent is a computational system implementing AI technologies, including generative AI, to **act autonomously and achieve goals**. Rather than using a single model or piece of software, an agent is a design architecture that draws on multiple computational tools. Agents can be used to automate procedures and processes, like workflows or repetitive tasks. An agent's goals are defined by a user, e.g. asking an agent to find the best meeting slot on their team's calendars, or pre-defined by the software system owner, e.g. an agent that is designed to research. A user can have an ongoing interaction with an agent and the agent can **autonomously select which tools are best suited to achieve the objective**. Many agents can search the web, query databases and implement various software programs.

Wearable AI

Wearable technologies are electronic or mobile devices, or computers with wireless communication, that are incorporated into gadgets, accessories, or clothes. AI and generative AI in particular, **is increasingly being used** in wearable technologies, such as **healthcare monitoring, augmented and**

virtual reality, sports and **entertainment**. Using generative AI in wearable technologies can make them more adaptive, personalised and interactive. In healthcare, for example, generative AI can monitor people in real time and adapt to individuals as their needs evolve. An increasingly popular application of wearable AI are **AI glasses**, defined in Note 2 below.

Note 2

AI glasses are wearable technologies that are connected to remote devices or AI. AI glasses have sensors, like **tiny speakers, microphones and cameras**, to collect data about the user and their environment, and can be used to feed back information to the user.

Relevance for the public sector

Multiple technology and consulting companies, such as **Microsoft, Deloitte** and **Google**, have highlighted opportunities offered by generative AI for the public sector. Generative AI is already being used in parts of the UK public sector. A 2025 survey found that **22% of respondents were using generative AI** for tasks such as drafting documents, summarising lengthy emails, updating records, preparing reports and analysis. Similar uses were identified in a **Workforce Partnership Council (WPC) report** and a **UK Government trial of 20,000 civil servants**.

Generative AI offers opportunities to:

- **Increase efficiency;**
- **Improve accessibility and inclusivity;**
- **Improve the satisfaction of employees** by reducing mundane tasks;
- **Enhance citizen services** (e.g. through the use of **chatbots**);
- **Provide creative aid;** and
- **Analyse large amounts of data** for comprehensive outputs.

Many specific **opportunities for parliaments** exist, for example, using generative AI to analyse legislation and identify gaps, summarise consultations, or gather information for research. AI agents can help **automate routine processes** like information gathering, administrative workflows, or improving service delivery. These are non-exhaustive lists and **specificities** of each

parliamentary setting and their digital strategies should be taken into consideration to assess relevance and priority.

£1 million was allocated by the Welsh Government in October 2025 to **support** the adoption of AI across Welsh public services. The **Office for AI and Strategic AI Advisory Group** have been working on applying AI to enhance public service delivery across Wales. **Projects are underway** to evaluate AI scribe tools, deploy a platform for the public sector to test and use AI tools, develop chatbots and deliver guidance and training for public sector workers.

“Humphrey”, a UK Government-developed AI toolkit, has been used to **analyse public responses** to a Scottish Government consultation; **speed up admin** in planning and social care; improve **GOV.UK Chat**; and **take notes** in a meeting chaired by the UK Prime Minister. The Welsh Government **is working with** the UK Government to assess how the Humphrey toolkit can be made available and used in Wales.

The **education sector** is also increasingly adopting generative AI. For instance, generative AI has been used by teachers to generate educational resources tailored for individual students or to generate feedback on students’ work. Students have used it, for example, to help them complete assignments, investigate concepts or find information and structure presentations.

In Wales, generative AI **has been used** by both academic staff and students. The University of South Wales published **guidelines** for appropriate use. A **review of generative AI in schools** published in October 2025 by Estyn, the education and training inspectorate for Wales, found many schools in Wales to be in the early stages of exploring AI.

Economic opportunities

The **2026 AI Index Report** from Stanford University finds that productivity gains from implementing AI are measurable within narrow tasks. AI can improve productivity in tasks that are structured, language heavy, or have clear feedback loops. One study of workers using generative AI based customer support chatbots found **15% more issues were solved per hour**, although this varied substantially across workers.

At a macro level, the AI Index Report says that evidence “remains early and mixed”, that there are indicators that AI is starting to impact aggregate productivity, although substantial productivity gains may take time to materialise.

A 2026 study of over 12,000 European firms found that AI adoption increases labour productivity levels by **4% on average**, but with uneven benefits depending on the size and capacities of the firm. AI adoption is projected to contribute to a **2.98% overall increase, equivalent to £79.3 billion, in the UK's GDP by 2035**. A 2026 UK Government report found that jobs directly involving AI activities could rise from **158,000 in 2024 to 3.9 million by 2035**.

Economic opportunities for Wales

Wales contributes to the global AI infrastructure market through its **semiconductor** industry, which produces essential parts of computer chips, and **data centres**, which host computational equipment.

South Wales is home to the **UK compound semiconductor cluster**, a centre of excellence in compound semiconductor research, innovation and manufacturing. The cluster accelerates growth in areas including AI research. It supports **£436 million in Welsh Gross Value Added (GVA), 3,140 jobs across Wales and £480 million in exports representing more than 90% of output**. A 2026 Department for Science, Innovation and Technology (DSIT) report found that Wales accounted for **11% of employment and 24% of revenues for the UK's semiconductor industry**, despite being home to only 7% of dedicated UK headquartered companies.

Cardiff also hosts **one of Europe's largest hyper-scale data centre campuses**. The South Wales AIGZ is **projected to bring around £10 billion of investment** by the early 2030s.

A May 2025 **report** from the Sixth Senedd's Economy, Trade and Rural Affairs Committee recommended that, to fully realise the benefits of AI, the Welsh Government should examine how it can support Welsh businesses adapt to and develop AI, analyse where job displacement will be most significant, and assess which skills the Welsh workforce requires to maximise benefits from AI.

In **response**, the previous Welsh Government welcomed guidance published on ethical AI. It also conducted a **review** into improving productivity of small medium enterprises (SMEs), recommending helping SMEs realise the benefits of AI through awareness campaigns and enhanced digital support. The Plaid Cymru Manifesto 2026 states that a new **National Development Agency for Wales** would support Welsh businesses to **"respond to new challenges and opportunities – including... digital and artificial intelligence"**.

Welsh language opportunities

The 2025 **AI Plan for Wales** sets out the previous Welsh Government's vision for AI:

To harness the power of AI responsibly to deliver economic growth, enhance public services in English and Welsh, and equip people in Wales with the skills to thrive in a world shaped by intelligent technologies whilst ensuring equity for all.

As a bilingual country, generative AI **offers opportunities** to promote and support the use of the Welsh language, making it easier for more people to use the Welsh language in everyday life. The Welsh Government's **2018 Welsh Language Technology Action Plan** highlights the importance of Welsh chatbots for citizens to make enquiries to their local councils in Welsh. The Action Plan includes a work package specifically dedicated to Welsh-language machine learning and conversational AI.

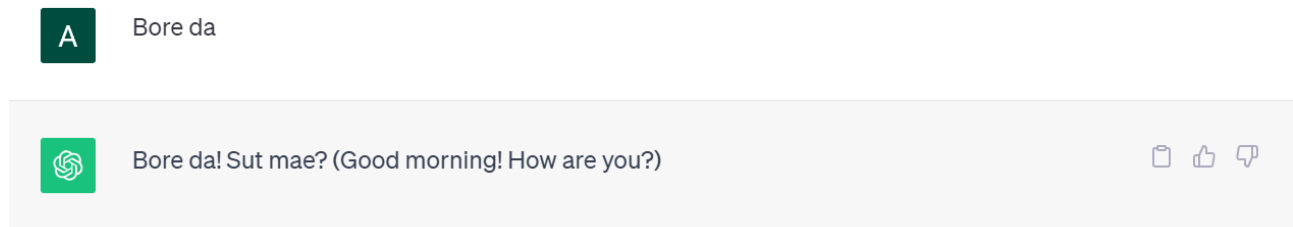
As quality data is vital for building tools that can support Welsh language use, the **AI Plan for Wales** emphasises the need for **Open Data** to make Welsh language data, and data about the uniqueness of life in Wales available. Resources such as **Helo Blod** are widening access to Welsh language technology. In 2025, **UK Research and Innovation (UKRI) confirmed £26 million investment** to support data use for research in Wales. **Projects are underway** to build resources of Welsh text and audio data. Data about the Welsh language **has been shared** with AI projects like **AINA**, led by the Government of Catalonia, and **OpenAI**.

Leveraging the benefits of open-source techniques and datasets, University College London, NVIDIA and Bangor university have **developed an LLM** to enable AI reasoning for Welsh and other UK languages. The model can be used to support the delivery of public services, such as health care, education and legal resources, in Welsh. The Welsh Government has worked with industry leaders to improve the quality of Welsh in popular AI tools such as **Microsoft's Copilot** and **OpenAI's ChatGPT**.

AI techniques other than generative AI are also being used to improve access to the Welsh language. For example, much of the translation sector in Wales, including the Welsh Parliament, **uses a neural machine translation model** which **predicts** the probability of a sequence of words in Welsh, given a sequence of words in English. Welsh automatic speech recognition tools

include the **voice assistant Macsen**, as well as text-to-speech models **Whisper** and **Trawsgrifiwr**.

Figure 2: Start of a conversation in Welsh on ChatGPT-3.5



2. What are the risks and challenges?

This section outlines risks associated with generative AI and points to relevant guidelines for fair and responsible use.

Technical risks and limitations of generative AI

Theoretical technical limitations

Generative AI outputs are bounded by the **quality of the training data** that a model is trained on. As a result, potential risks in outputs are:

- **Bias and lack of neutrality;**
- **Monoculture** and **constraints on the development of plural opinions** (more explanations in Note 3 below);
- **Mistakes, creation of made-up knowledge**, and **misinformation;**
- **Disinformation;** and
- **Discrimination.**

Note 3

Generative AI outputs **reflect the values** of the creators of the training data and the developers who select which training datasets to use. As a result, marginalised voices are often not represented in generative AI outputs.

AI agents raise **additional fairness concerns** because they automate choices about what goals to prioritise. These choices may favour some people or groups more than others, even if it isn't obvious to the user.

Errors or bias in generative AI outputs can have **wider effects** on society. For example, they can erode trust in information, political and societal influence, and cause cascading failures if used incorrectly in large-scale organisations.

This problem **may also worsen** as future generative AI models are trained on data taken from the internet, which is **increasingly filled with** the outputs of previous models. These feedback loops **contaminate training data and deteriorate model performance.**

Even with filters put in place to reduce prejudicial outputs, there is an inherent risk of generative AI producing harmful content. Large generative AI models have been trained on data that contains **illegal and toxic content**, such as **harmful stereotypes** or **sexist** and **racist** statements, which means that there is a chance this data can be **reproduced in outputs**.

Research has also proven it to be **very difficult to stop “jailbreaking” attacks**, which exploit vulnerabilities in generative AI models to get them to behave in undesirable ways. A **“prompt injection”** is a similar type of attack, where malicious instructions are hidden within other input data. When a prompt injection attack happens to an agent, the agent can be manipulated to take actions on behalf of a malicious actor.

Prompt injection is especially problematic for AI agents, because agents automate decisions away from the user, which makes it harder for the user to identify when an attack is happening. Agents can also be given authority to take consequential actions like data sharing or deletion, which means an attack can be highly impactful, **raising significant security and privacy concerns**.

Issues are made worse because many AI technologies, like generative AI, rely on complex internal decision-making process, making it difficult to understand how an output is reached from a given input. That is why many AI systems are called **“black boxes”**, because, metaphorically, you can see what goes into the box and what comes out, but not what happens inside it.

AI agents automate decisions away from the user, which makes it **even more challenging** to understand how the technology is working and **who is accountable**. Making AI interpretable is **an active research direction** and researchers, policymakers and industry standards **stress the need for meaningful human oversight of AI**.

For Wales specifically, a **lack of quality training data in Welsh** could reduce the standard of Welsh outputs made using generative AI. More broadly, training data for generative AI is **largely gathered from** the internet, which is mostly in a few data-rich languages (e.g. English, Spanish, Mandarin), creating gaps for any non-data rich language.

Efforts to improve data availability are laid out in the **AI Plan for Wales**, including the use of **Open Data** to widen access to Welsh language data and data about what makes life in Wales unique. **Progress** includes Bangor University’s **dataset of sentences from Welsh Government websites** and Cardiff University’s **SENTimental** tool to create Welsh training data.

Improvements to the quality of Welsh training data **does not eliminate risk of bias**, however. Generative AI tools **exhibit biases towards the values of certain countries even when using other languages**. In addition, because generative AI is probabilistic, outputs vary even for the same input. This means that using generative AI for translation will not be consistent.

Finally, the behaviour of generative AI systems can change relatively quickly over time. **One study** found a decrease in ChatGPT's ability to follow user instructions in the span of three months, underlining the need for continuous monitoring.

'Real-world' examples

Various **experimental tests** have shown the technical limitations of generative AI for 'real world' applications, even for basic tasks that could be solved by non-expert humans.

Issues with bias in AI algorithms other than generative AI have been seen in the public sector. For example, facial recognition used by South Wales Police was **declared unlawful** and a breach of human rights due to: a lack of guidance on its use; defective consideration of the data protection impact; and a failure to adequately address risks of racial and gender bias. However, if these issues had not been present, the Court of Appeal stated that use of facial recognition would have been lawful.

Other examples outside of the UK include a **fraud prediction algorithm** used by Dutch Tax Authorities which was found to be guilty of **indirect discrimination on the basis of race and ethnicity**. Parents were **wrongly accused of fraud**, leading to **unemployment, bankruptcies and parents losing custody of their children**. The Government ended up **resigning** and 5 billion euros were distributed in **compensation**.

Risks beyond technical limitations: stakeholder concerns

Beyond technical limitations of generative AI due to input data, other risks concern stakeholders including:

1. Lack of regulation: Generative AI experts, **such as the OpenAI co-founder**, have urged for legislation to regulate the use and development of AI. The **Ada Lovelace Institute recommends** a more holistic and comprehensive regulatory framework for AI in the UK as a way to both manage risks and support innovation. Without clear regulation, issues arise with how to allocate

accountability when problems occur. A **survey** published by the Ada Lovelace Institute in December 2025 reported that 72% of respondents supported AI legislation, increasing from 62% in 2023, and 89% supported an independent AI regulator.

Concerns have been raised that current legislation is inadequate for **regulating AI agents**. As the level of automation in AI agents increases, the scope and impact of errors or misuse also **increases**, which raises **questions around accountability**. Researchers have **highlighted** that existing regulators will face challenges in scaling enforcement to keep up with the destabilising effects of increasing AI agent use.

The UK Government advises that AI agent implementation must **comply with consumer law**. The Competition and Markets Authority (CMA) has published a report finding that AI agents could “**materially change how people engage with markets and how value is created**”. In Estonia, steps have been taken to regulate AI agents by assigning each agent a “**personal identification code**” to facilitate tracking agent behaviour and the users behind them.

Wearable AI has also drawn the attention of regulators. Concerns have been raised about AI glasses **infringing privacy**. **Calls have also been made** for regulation to address the possibility of biases and safety risks in AI as a medical device, which includes wearable technologies.

2. Impacts on jobs and potential job losses: **Data suggests** that UK hiring is falling faster in occupations that are more exposed to AI, and the UK workforce **faces high AI exposure** relative to other advanced economies. **Analysis** found that between 2022 and 2025, job adverts fell by 38% for high AI exposed occupations compared to 21% for low exposed roles. **Labour market costs** could fall **disproportionately** on junior and entry-level workers. In Wales, overall unemployment has risen from **3.0% in 2022** to **4.5% in 2025**, whilst unemployment in people aged 16 to 24 has increased from 7.8% to 15.4%.

Creative workers have raised concerns and **computer programmers and software development professionals** are likely to be significantly impacted by the effectiveness of using generative AI to write code. A UK Department for Education **report on the impact of AI on UK jobs and training** highlights impacts of generative AI on professional occupations normally requiring a degree, especially those involving clerical work in finance, law and business management roles. However, this report also highlighted that Wales is one of the UK regions where it is expected that workers will be the least impacted by AI overall, as the occupational structure of Wales has a lower level of AI

exposure. The Institute for the Future of Work **surveyed** 1,000 UK firms about AI's impact on jobs, concluding that outcomes are uncertain but urgent action is required to avoid aggravating **existing regional and demographic inequalities**.

In August 2023, Wales TUC, with Kings College London and Connected by Data, **launched an investigation** into worker attitudes on the use of AI across sectors. Findings suggested workers worry that AI, and digitalisation in general, could worsen the workplace. Concerns included **surveillance by management, redundancies and lack of consultation with unions** and **scepticism** about AI capabilities compared with human skills. A TUC **report** revealed notable increases in worker surveillance and highlighted risk of bias and discrimination **starting at recruitment**. The House of Commons Library's briefing on **Artificial intelligence and employment law** outlines risks of using AI for recruitment, line management and surveillance.

WPC guidelines recommend that decisions about algorithmic systems should put human oversight and interaction first, protect jobs, create jobs and invest in the workforce. TUC have **guides** for unions on AI and the workplace.

3. Overreliance and lack of AI literacy: Dependence on generative AI can increase **social isolation**, risk of **emotional manipulation** and impact **mental health**. It can also **reduce critical thinking skills** and prevent users from learning by encouraging excessive **cognitive offloading**. Studies have **found** that generative AI use can impair cognitive performance, decrease the quality of attention and increase the likelihood of memory loss.

Risks of overreliance are **decreased by improving AI literacy**. As with any technology, concerns predominantly arise not from AI itself but **from the ways that it is used by people**. Many risks related to AI come from misunderstandings about the how the technology works and what its limitations are. A **2024 survey into automation and AI** in the Welsh public sector published by the **Centre for Digital Public Services (CDPS)** emphasises the importance of being user centred and problem led, rather than seeking out problems to apply AI to.

4. Loss of control and accountability: As generative AI relies on probability, it is difficult to control what outputs are produced and what happens to data inputted into models, risking **loss of control over data**. Data entered into generative AI systems might be **re-used in training** and **reproduced elsewhere**. Difficulties in understanding what happens to data and why certain outputs are made can create **accountability gaps**, which **can result in serious**

consequences including unaddressed harm to people and services, regulatory action and reputational damage.

AI agents further reduce human control by increasing the level of automation. Allowing an agent to make decisions on a user's behalf **narrows the user's choices** and can reduce their autonomy. **Overreliance** on an AI agent can be problematic if **software updates change an agent's behaviour**, an agent starts taking **unpredictable actions**, or the technology provider removes certain functionalities. As agents are delegated **more autonomy** and operate across increasingly **disperse technologies and platforms**, accountability becomes **fragmented and concealed**.

5. Commercial dependency: **The Alan Turing Institute warned** against dependency on commercial generative AI, which could result in lack of transparency and undefined liability. The **Equality and Human Rights Commission (EHRC) also stressed** that several AI companies do not share their technology due to commercial sensitivity. The UK Parliament is **concerned** about the **dependence** of the UK public sector on a small number of US-based providers.

6. Identification of computer generated outputs: The UK Government highlighted that **synthetic media generated by AI can be difficult to identify as such**, and that authentication solutions are unreliable for now. This issue was also identified by the **Alan Turing Institute**. The **Institute for Government emphasised** the issue of 'deepfakes' and trust in interactions that could lead to government decisions.

As generative AI outputs are difficult to identify and can be quickly produced at scale, it increases the availability, ubiquity and realism of **misinformation** (unintentional spreading of false information) and **disinformation** (deliberate creation and dissemination of false or misleading information). Fake content **can be used** to manipulate narratives, distort historical facts, discount cultural nuances, polarise discussions and erode trust in genuine information. Generative AI accelerates the spreading of fake information and encourages **perceptions of consensus** through the sheer volume of content. One example of this is "**political astroturfing**" where a political actor uses technologies (including **generative AI** and **AI agents**) to forge the appearance of real citizen behaviour.

Figure 3: AI-generated image with the prompt ‘Wales’ on [Pixlr](#)



7. Cybersecurity and privacy: The [UK Government Safety and Security Risks of Generative Artificial Intelligence to 2025 Report](#) says that generative AI can help automate cyber-attacks and increase digital vulnerabilities by:

(...) corrupting training data ('data poisoning'), hijacking model output ('prompt injection'), extracting sensitive training data ('model inversion'), misclassifying information ('perturbation') and targeting computing power.

Queries stored online could also make it easier for hackers to **access potentially sensitive data**. AI agents pose additional risks of misuse such as **data disclosure** and increasing the scope of **disinformation campaigns**. Agents **threaten privacy and information governance** as they continuously collect and combine data across contexts that would usually be kept separate, and have unpredictable variations in performance across domains.

Additional privacy, cybersecurity and surveillance threats arise from wearable AI. The Information Commissioner's Office (ICO) **contacted** AI glasses manufacturer Meta in March 2026 regarding a report that found videos of users, including in highly personal situations like using the toilet, were reviewed by international subcontractors. Women have been **non-consensually filmed** by users of wearable AI. European lawmakers are also **concerned about** whether the technology breaches regulation such as GDPR and have issued **warnings** that wearable AI normalises surveillance and increases risk of abuse.

8. Copyright and fair use, intellectual property, plagiarism: Generative AI draws data from various sources, and **summarises or imitates existing content often without the permission of owners**. Generative AI has been found to reproduce **exact copies** of its training data. **The Alan Turing Institute highlights** issues with plagiarism in the context of education and assessments.

Lawsuits due to copyright issues with training and using generative AI have been filed in the UK, for example, **Getty vs Stability AI**. The Judge ruled that there was not an indefensible infringement of copyright in training, developing and deploying generative AI, however, there was an infringement on the reproduction of Getty trademarks by generative AI.

Following a **consultation** from December 2024 to February 2025, the UK Government published a **report** on **copyright and artificial intelligence**. The report recommended no immediate change to copyright law, but recognised the need for broader transparency and technical mechanisms to support copyright enforcement.

9. Exploitation and human costs: In the generative AI development process, there are risks of **exploiting workers**. Models are often checked through a process known as **Reinforcement Learning from Human Feedback (RLHF)**, which involves human reviewers. **University College London reported** that RLHF reviewers for OpenAI's ChatGPT tool were mostly from the global south and paid low incomes (e.g. less than \$3 per hour in Kenya). Reviewers had to work with toxic content that had **detrimental effects** on their wellbeing.

10. Environmental impact: Training and using generative AI, especially large-scale models, is energy intensive, relying on electricity generation and water consumption. Whilst **reports** from 2026 suggest significant efficiency gains are being made in hardware and software used for AI, improvements are outweighed by rapid increases in model size and popular use. Training large-scale AI technologies often **requires weeks of computation** with high levels of emissions. For example, emissions from training the Grok 4 model reached an **estimated 72,816 tons** of CO2 equivalent. This is magnitudes more than lifetime emissions of an average petrol car of **54.1 tons**.

Power consumed each time a generative AI model is used varies depending on the model and task. Simple text queries now **typically consume** less electricity than running a television over the same period of time, but are still about **40% more expensive** than a Google search. Queries that use images, documents, videos, audio, or tasks involving AI agents, can consume **hundreds or thousands of times more energy** per query than simple text generation.

In addition to electricity, **water use associated with AI** happens directly, through liquid cooling in AI data centres, and indirectly, through electricity generation or supply chain manufacturing. **2026 estimates** of annual cumulative water consumption from users interacting with OpenAI's GPT-4o model range between 1.3 to 1.6 million kilolitres which, at the upper end, exceeds the average annual drinking water needs of 1.2 million people.

11. Equality: **The EHRC highlighted** that most local authorities do not appear to have built equality considerations into their procurement processes, and that some do not recognise that they are using generative AI or do not communicate it. The EHRC emphasised that there is confusion between ethics and equality and cited the **Data Protection Impact Assessment (DPIA)** as an example of this confusion. A DPIA is a process to help identify and minimise data protection risks in a project. It is compulsory for any processing “likely to result in a high risk to individuals” and is considered best practice for any project handling personal data.

The EHRC said that local authorities often believe they addressed equality issues with a DPIA, but it was often not the case and more was needed. The EHRC also said that the **AI White Paper** published by the UK Government (more details in part 3) did not emphasise the equality and human rights framework enough. The **Alan Turing Institute also warns** against a potential widening of the digital divide and breach of children's rights (e.g. commercial exploitation). **Estyn raised concerns** that there is a risk of AI exacerbating a

digital divide in Welsh schools as pupils who can afford paid-for AI tools may gain advantages that others cannot.

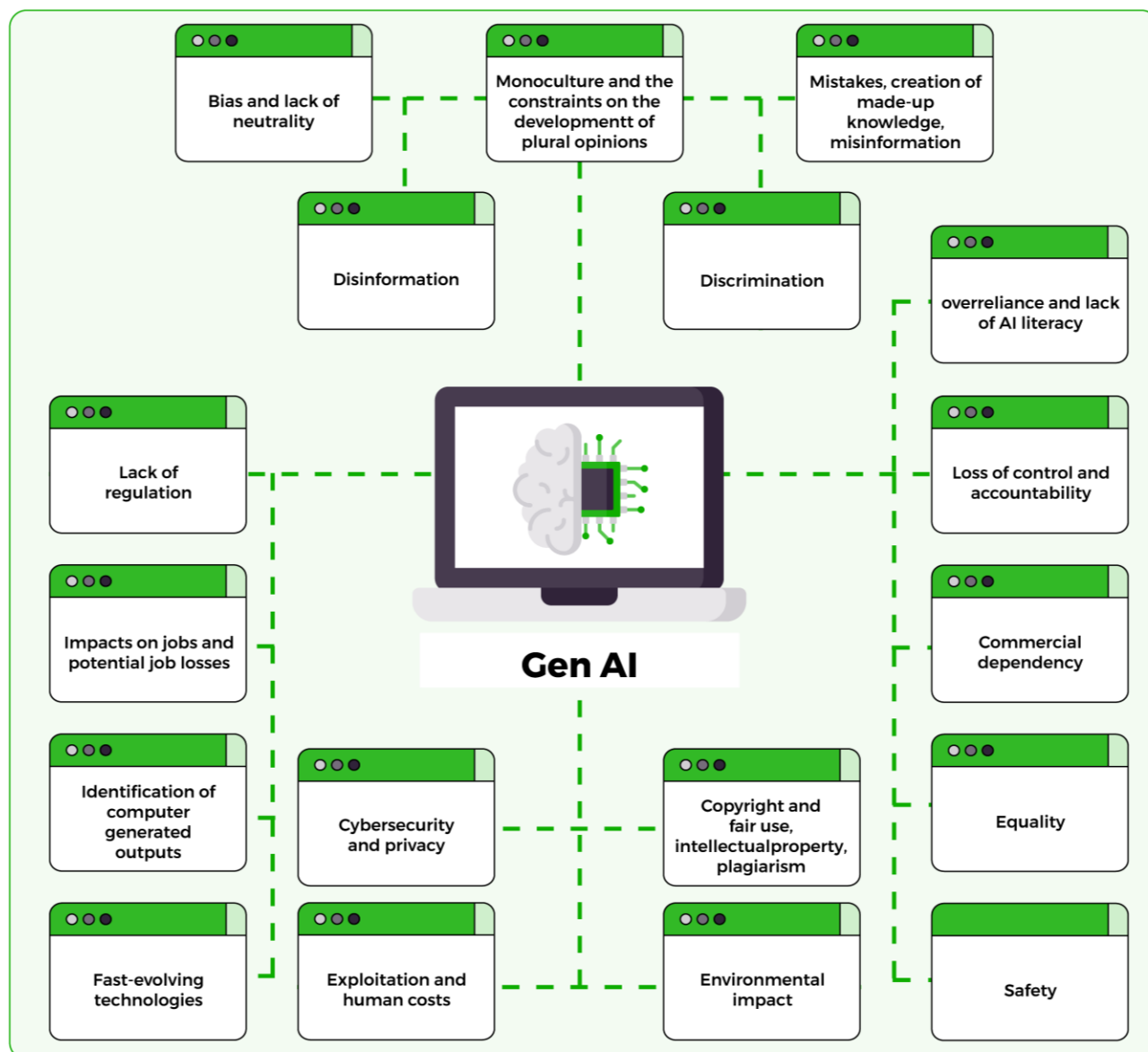
The **AI Plan for Wales** emphasises equitable delivery and ensuring that AI benefits all of Wales, developing Welsh public services aligned with the **Strategic Equality and Human Rights Plan 2025-2029**, the **Anti-Racist Wales Action Plan**, the **LGBTQ+ Action Plan for Wales** and the **Disabled People's Rights Plan for Wales**. In 2025 the Welsh Government **received more than £440k of funding from the UK Government's Digital Inclusion Innovation Fund** to support work on the concept of a **minimum digital living standard (MDLS)**. The MDLS covers minimum requirements for digital inclusion such as access to devices, connectivity and basic digital skills.

12. Safety: The **UK Government Safety and Security Risks of Generative Artificial Intelligence to 2025 Report** states:

Generative AI can be used to assemble knowledge on physical attacks by non-state violent actors, including for chemical, biological and radiological weapons.

The report says that safeguards are not always effective, and that generative AI will help accelerate the removal of current barriers such as access to components or tacit knowledge. Other **safety concerns** include misusing AI to produce inappropriate or toxic material, **phishing**, or **social engineering** through **highly personalised but fake content** which is difficult to detect.

13. Fast-evolving technologies: **The Alan Turing Institute emphasises** that generative AI is evolving at an 'extremely fast-moving' pace, meaning that new risks can emerge rapidly.

Figure 4: Summary of the risks of generative AI

Existing guidance to mitigate risks for ethical and safe use

Various non-statutory guidance exists to mitigate the risks of generative AI listed above. Guidance relevant for the public sector in Wales are:

- The UK Government Digital Service (GDS) **Algorithmic Transparency Recording Standard (ATRS)** (January 2023): helps public sector organisations provide clear information about how and why they are using algorithmic tools. This is useful in case of audits where technology, including AI, is used to structure talks with unions about buying or building new systems;

- The National Cyber Security Centre (NCSC) **Guidelines for secure AI system development** (November 2023);
- The Inter-Parliamentary Union (IPU) **Guidelines for AI in Parliaments** which follow the IPU **resolution on the impact of AI on democracy, human rights and the rule of law** (October 2024);
- The Organisation for Economic Co-operation and Development (OECD) **G7 Toolkit for Artificial Intelligence in the Public Sector** (October 2024);
- The **ICO's UK GDPR guidance on Artificial intelligence** (2025): it focuses especially on data protection and explaining decisions made using AI;
- DSIT's **AI Cyber Security Code of Practice** (January 2025);
- The **AI Playbook for the UK Government** (February 2025) and **AI resources for the public sector** (updated June 2025) published by the GDS; and
- The WPC's **ethical guidance to using algorithmic management systems in the Welsh public sector** (2025): it covers all types of AI and algorithmic systems.

3. What is the current regulation of generative AI in Wales and other countries?

This section examines AI policy and legislation in Wales, relevant policy in the UK and other international approaches.

Regulations and competency in Wales

Welsh Government AI policy

In the 2025 [**AI Plan for Wales**](#), the previous Welsh Government emphasises that opportunities presented by AI, such as driving economic growth and enhancing public services in both Welsh and English, must be matched by a commitment to responsible use. It proposes a four pillar framework – Economic Growth, Educating Wales, Equitable Delivery, Excellence & Trust – grounded in values of social partnership, collaborative decision-making, fairness, transparency, equality and equity. First Minister Rhun ap Iorwerth MS has said that the current Welsh Government is “[**supporting the safe and responsible adoption of artificial intelligence in the health service**](#)”.

As part of the AI Plan for Wales, the Welsh Government established an [**Office for AI and Strategic AI Advisory Group**](#). The Office for AI facilitates informed policymaking, internal capability and collaboration with stakeholders across Wales. The Strategic AI Advisory Group provides expert advice on AI adoption across Welsh public services. The [**Digital Service Standards for Wales**](#) and the [**Digital and Data Standards Board for Wales**](#) define what good digital public services look like. The [**AI Commission for Health and Social Care**](#) inform the safe and responsible use of AI in health and social care.

The previous Welsh Government’s [**Digital Strategy for Wales**](#), published in March 2021, highlights the importance of using AI “ethically and with integrity” to “ensure data ethics, transparency and trust”, and the importance it has for Wales’s economy. Building on this, the 2025 [**Digital inclusion strategy**](#) introduced the notion of a minimum digital living standard.

The **Socio-Economic Duty** came into force in Wales in March 2021. The statutory duty requires certain Welsh public bodies to consider how their strategic decisions can reduce inequalities caused by socio-economic disadvantage. Public bodies therefore need to consider whether the use of generative AI increases or reduces socio-economic inequalities. These considerations are also reflected in the **WPC guidance** to ensure socially responsible procurement and implementation of AI in line with the **Social Partnership and Public Procurement Act 2023**. The guidance also recommends that AI and algorithmic use in the public sector workplaces should prioritise: human oversight and interaction in decisions about algorithmic systems; fair work, which includes equality; building capability regarding technology use; and a commitment to protecting, creating and investing in the workforce.

Existing legislation for Wales

There is currently **no specific holistic approach** to legislating generative AI (and AI in general) in Wales in terms of development, deployment or use. In the UK, AI is **regulated in the context it is used** through **existing regulatory frameworks**, for example, in **financial services**, **copyright** and **healthcare**.

Existing legislation that could be used to address misuse of generative AI in Wales include:

- The **Equality Act 2010**, which protects people against discrimination and supports equal employment opportunities;
- The **Well-being of Future Generations Act (Wales) 2015** which requires decisions on AI adoption to consider impacts on future generations, and balance social, economic, environmental and cultural well-being;
- The **Data Protection Act 2018** and **UK General Data Protection Regulation (UK GDPR)** which controls **data protection**;
- The **Network and Information Systems (NIS) Regulations 2018**, to which the **Cyber Security and Resilience Bill** was added in November 2025 to bolster cybersecurity defences for essential and digital services;

- The **Social Partnership and Public Procurement Act (Wales) 2023** which covers socially responsible public service delivery and well-being, including the procurement and implementation of AI, in Wales;
- The **Online Safety Act 2023**, which empowers **Ofcom** to regulate online services and includes criminalisation of the **non-consensual sharing of explicit deepfakes** since January 2025;
- The **Digital Markets, Competition and Consumers Act 2024**, which protects consumer rights, including making sure that marketing materials are clear and provide accurate information. The UK Government recommends that **use of AI agents needs to comply with consumer law**;
- The **Cyber Security and Resilience Bill** which was added in November 2025 to the existing **Network and Information Systems (NIS) Regulations 2018**, proposes new laws to make essential and digital services more secure in the face of cyber criminals and disruptive state actors;
- The **Data (Use and Access) Act 2025**, under which **creating or requesting AI-generated intimate images without consent** is an offence from February 2026;
- The **Crime and Policing Act 2026** which combats illegal content related to AI chatbots; and
- The **Children’s Wellbeing and Schools Act 2026** which implements the results of the UK Government’s **consultation** on children’s lives online

Judicial review is a way to challenge the lawfulness of decisions or actions made by public bodies. If the use of generative AI by a public body results in that public body acting unlawfully, judicial review could be used to hold them to account.

Figure 5: A judicial review requires an application to the High Court of Justice, based in London



The **Public Sector Equality Duty (PSED)** is a legal obligation which requires public bodies to have due regard to the need to eliminate discrimination; advance equality of opportunity; and foster good relations between those with a protected characteristic and those who do not share it. **Wales specific duties** include carrying out steps to meet the above general equality duty and aid transparency; engage groups with protected characteristics; include equality requirements in procurement exercises; publish equality objectives; and assess equality impact.

Technical guidance on the PSED for Wales has been produced by the EHRC. **The EHRC also has powers** to conduct a 'Section 31 assessment', when it suspects an organisation is not compliant with the PSED, requiring evidence of their actions towards the PSED. If the EHRC finds the evidence to be non-satisfactory, it can request that the organisation enter into a 'Section 23 agreement' which requires **actions** (e.g. providing actions plan by a certain date, delivering learning material, publishing analyses, reviewing processes, engaging with specific stakeholders) to be taken. The EHRC also has investigation and enquiry powers around large themes (for example, it could be the use of AI in public bodies in Wales) or specific organisations.

Competency for further legislation in Wales

Data protection, as well as intellectual property and internet services, **are reserved matters**, meaning that the Welsh Government and the Senedd cannot make law in these areas.

However, the wider governance of data protection regulations **does not always involve a straightforward distinction between ‘reserved’ and ‘devolved’**. Protection of personal information is reserved to UK Parliament, but there is limited scope for the Senedd to make law relating to equal opportunities. Some aspects of data management are devolved (e.g. ‘information governance’ in healthcare), some responsibility can be shared by UK-wide and devolved regulators (e.g. equalities laws in Scotland), and former EU law that was applied UK-wide prior to Brexit has sometimes become the responsibility of devolved nations. **Roberts et al.** pointed out that:

(...) each of the devolved nations has different data strategies in place and different rules governing, for example, access to data for secondary purposes including the development of AI.

Since AI, and especially generative AI with general purpose applications, could be used in a variety of reserved (e.g. data protection) and devolved (e.g. healthcare, education) areas, **it is not clear** what devolved competence would look like for specific generative AI (and AI) legislation. **Roberts et al.** say it could lead to internal tensions due to devolved nations wanting to introduce AI legislation and the UK Parliament challenging it on the grounds that it relates to reserved matters or, conversely, due to the UK Parliament breaking the Sewel Convention by legislating on AI in areas that are within the competence of the Senedd without the Senedd’s consent:

The ambiguity surrounding how reserved and devolved powers relate to AI creates a real risk of regulatory divergence, which could lead to ineffective protections for citizens on the one hand, and a confusing regulatory environment for companies on the other, undermining a central aim of the UK’s overarching approach.

They highlighted that there are already some contradictions between the **‘pro-innovation’** strategy chosen by the UK Government and Scotland’s **‘ethical digital nation’** plan in their **AI Strategy**, published six months before the UK strategy.

Roberts et al. also highlighted that the AI White Paper published by the UK Government does not address the ambiguity surrounding reserved and devolved powers for AI and the potential coordination problems that may result from it (see below for more details).

UK policy and legislation

AI-specific strategy

The UK Government has said that “**most AI systems should be regulated at the point of use**”. AI is **overseen by existing regulators covering specific sectors**, such as **Ofcom** (the UK’s communications regulator), **Ofgem** (Great Britain’s energy regulator) and the **ICO** (the UK’s independent body for upholding information rights). Responsibility for AI policy and strategy was assigned to the **Minister of State (Minister for Artificial Intelligence)** Kanishka Narayan MP in July 2026, jointly in the Cabinet Office and the Department for Business, Innovation, Science and Trade. The House of Commons Library briefing “**AI regulation in the UK**”, published in June 2026, highlights how UK Government AI-related policies are mostly focused on boosting the AI sector to help drive growth and productivity across the UK economy, and comprehensively outlines how the UK’s approach to regulating AI has evolved.

UK Government policy that is relevant to Wales includes:

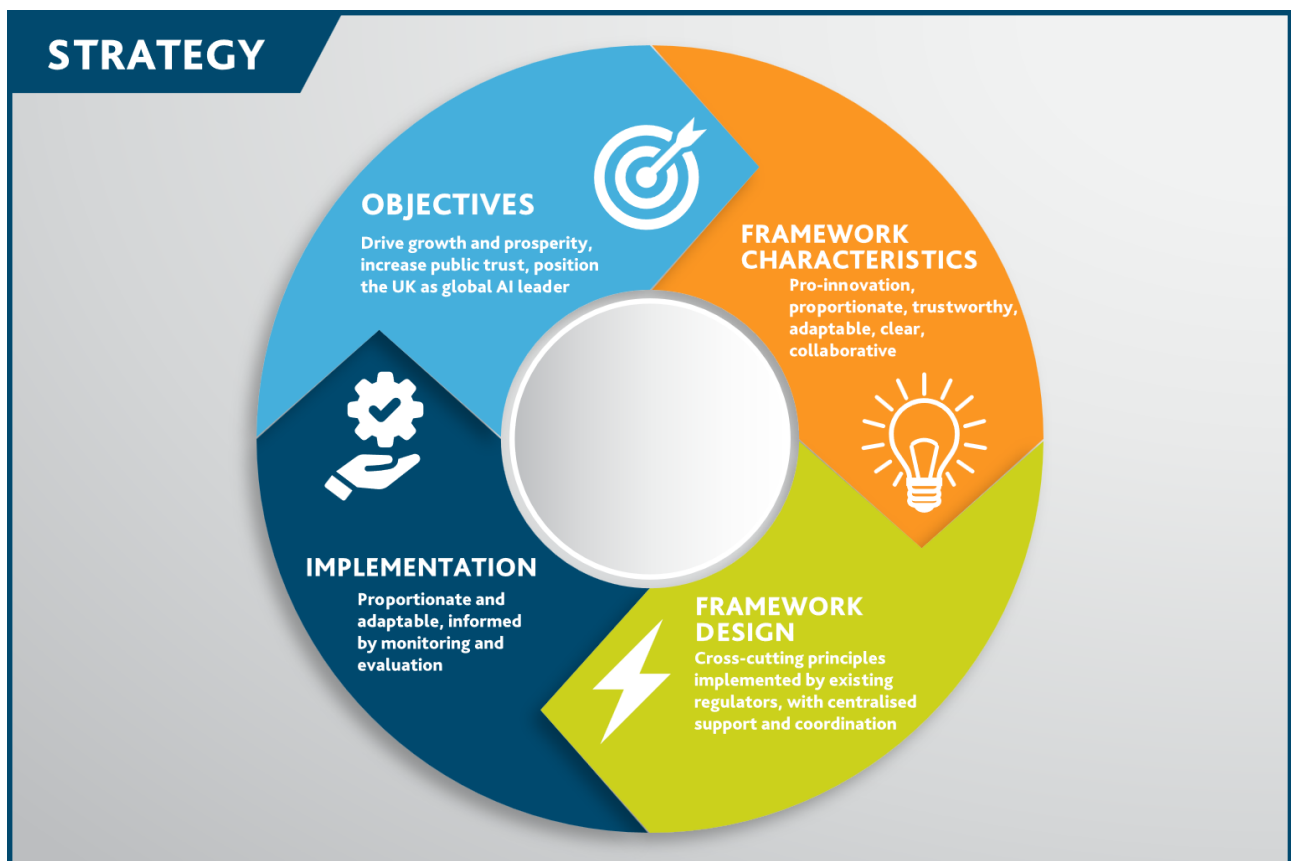
- The **AI regulation white paper** published in March 2023, which is based on the ‘pro-innovation’ approach described in the **2022 AI Action Plan** policy paper accompanying the **2021 National AI Strategy**. The white paper sets out the UK Government’s proposal for regulating AI but does not include a plan to bring specific new regulations or regulatory bodies. The consultation for this white paper was carried out from March until June 2023 and the UK Government published its **response** in February 2024. This response highlighted support from stakeholders and emphasised that regulators were already taking action to follow the proposed framework;
- Commitments made in **response to** the **AI Opportunities Action Plan** published by DSIT in January 2025, promising to develop AI infrastructure, create **AIGZs**, introduce an **AI Energy Council** and **National Data Library**, develop AI skills and talent, and set out an approach for AI

regulation. Two AIGZs are in Wales: one in **North Wales** and one in **South Wales**; and

- A commitment to **ban social media for under-16s** from spring 2027. A House of Commons Library briefing on “**Proposals to ban social media for children**” published in June 2026 discusses whether this will include AI services like AI chatbots. First Minister Rhun ap Iorwerth MS has said that action “**needs to be taken**” regarding under 16s use of social media.

The EHRC **highlights** that, with the pro-innovation cross-sectoral approach chosen by the UK Government, it might be challenging for regulators to have resources for these new responsibilities and for people to find redress due to the variety of regulators involved. Guidance for implementing the UK’s non-statutory principles was **published by DSIT in February 2024**.

Figure 6: Illustration of the UK Government strategy for regulating AI



(Source: **UK Government**)

Implications of UK AI policy for Wales

The AI white paper **explicitly states** that it applies to the whole of the UK as AI impacts a wide range of sectors and policy areas, some of which are reserved

and others devolved. It goes on to explain how the devolved administrations will be involved in the regulation of AI.

Roberts et al. highlighted that, while the white paper promotes flexibility in the fast-paced AI environment and acknowledges coordination challenges, it fails to address potential tensions, as well as capacity and funding challenges, with devolved governments. According to them, it also does not provide “any concrete next steps or timelines” for the UK Government to support devolved regulators. Roberts et al. suggested establishing an Interministerial Group on Digital Governance.

In June 2025, the UK Government opened applications for **AIGZs** to accelerate building infrastructure needed for AI growth through unlocking investment, improving access to power and providing planning support. It was announced in November 2025 that **Wales would host one AIGZ in North Wales and one in South Wales**. Infrastructure that will be built in AIGZ includes **AI data centres**, which house the computing equipment that is needed to train and use AI technologies.

Locations for AIGZs have been **selected** as places where **common barriers** to data centre development, including lengthy planning procedures and delays getting access to power, can be best addressed. Energy is a reserved power and developments in AIGZs will be given **priority access** to capacity on the energy grid. **Reforms** have been made to planning policy in the UK to explicitly incorporate data centres, however, planning is a devolved power and **planning policy in Wales does not currently mention data centres**.

EU policy

In April 2021, the **European Commission proposed** a legal framework on AI. This framework is based on different rules depending on risk levels: unacceptable risk, high risk, limited risk and minimal risk. The objective of the AI Regulation, known as the AI Act, is to **make sure** that:

(...) AI systems used in the EU are safe, transparent, traceable, non-discriminatory and environmentally friendly.

In June 2023, the European Parliament **adopted its negotiating position** on the **AI Act** with 499 votes in favour, 28 against and 93 abstentions. The **European Parliament Research Service (EPRS)** published **a briefing** in March 2024, stating that:

European Union lawmakers reached a political agreement on the draft artificial intelligence (AI) act in December 2023.

The **final text** was endorsed in February 2024, with 71 votes in favour, 8 votes against and 7 abstentions. On 21 May 2024 the European Council **formally adopted** the EU AI Act.

In terms of specific regulations applying to generative AI, **the AI Act includes:**

- The ban of “harmful AI practices that are considered to be a clear threat to people's safety, livelihoods and rights, because of the 'unacceptable risk' they create”, which include “AI systems that deploy harmful manipulative 'subliminal techniques'; AI systems that exploit specific vulnerable groups (physical or mental disability); AI systems used by public authorities, or on their behalf, for social scoring purposes; 'Real-time' remote biometric identification systems in publicly accessible spaces for law enforcement purposes, except in a limited number of cases”;
- Requirements on “risk management, testing, technical robustness, data training and data governance, transparency, human oversight and cybersecurity” for “high-risks AI systems” (e.g. systems used as safety components and in specific areas such as law enforcement);
- Transparency requirements for AI systems presenting “limited risk” (e.g. systems interacting with humans such as chatbots and deepfakes); and
- No legal obligations for AI systems with “low or minimal risk”.

The Act also proposes to task national supervisory authorities with supervising and implementing the regulations, national market surveillance authorities with assessing compliance, and a new European Artificial Intelligence Board with overseeing the regulations at EU level. The EU released a voluntary **General-Purpose AI Code of Practice** in July 2025 to help industry comply with the AI Act. Additional details about European policy can be found in the House of Commons Library briefing **“AI Regulation in the UK”**.

Figure 7: Thierry Breton, Commissioner for Internal Market of the EU, Plenary Session - AI Act, European Parliament in Strasbourg, 13 June 2023



(Source: [European Parliament](#))

The **EPRS briefing** emphasises regulatory divergences that arose and issues that were raised regarding the EU AI Act, including:

- The most adapted way to **define** general-purpose AI and AI systems;
- The potential **ineffectiveness** of the AI Act since the categorisation of the risk relies on self-assessment;
- The **need** to include more types of AI systems in the ban to avoid the use of any manipulative system (e.g. deepfakes) and the need for a procedure to add technologies in the ban;
- The need for a more **detailed** classification of risks;
- The **lack of mechanisms** available to users for complaints or judicial redress;
- The **incomplete alignment** of the AI Act with existing sectoral regulations, posing a challenge for implementation; and

- Questions around the specific rules for **military AI applications**.

The AI Act **could have influence** on policy adopted by other governments or companies, including in the UK, through the “Brussels Effect”. This describes the phenomenon whereby the capacity and market size of the EU influences companies to follow its standards and in turn, companies pressure other governments to align to the EU rules. This is what happened with the General Data Protection Regulation (GDPR).

In September 2022, the EU Commission also proposed **an AI Liability Directive (AILD)** and **a revision of the Product Liability Directive (PLD)**. Critics however highlighted that improvements are needed to provide harmonised regulation, include sustainability considerations and better define AI liability.

Other countries

There are **more than 1,000 AI policy initiatives across 72 countries** around the world.

National AI strategies were **expanding fastest** in 2026 among countries that had no formal AI policy five years previously. Global policies are increasingly emphasising **AI sovereignty**, which involves countries rethinking dependencies across the AI system, although definitions **vary significantly** between different regions. AI sovereignty was a key theme at the **India AI Impact Summit** in February 2026.

Regulation of AI in the US shifted significantly in 2025, transitioning from **previous safety focussed approaches** towards **accelerating AI adoption and innovation**. The **2026 National Policy Framework for Artificial Intelligence** emphasises AI infrastructure development and speeding up AI deployment. Federal agencies apply existing consumer protection, civil rights and financial laws related to AI. States have passed targeted AI legislation covering **algorithmic discrimination, deepfakes, automated decision-making, chatbots** and **social media**.

The UK and the US agreed on an AI safety **memorandum of understanding** in April 2024 and reached **further agreements** to cooperate over technology research in 2025. **Roberts et al.** mentioned the potentiality of a “Transatlantic effect” instead of a “Brussels Effect” on the UK.

The US has collaborated with the EU, for example by producing a **joint AI Roadmap** with the **EU-US Trade and Technology Council (TTC)**. However, cooperation has been waning. The US alongside the UK **refused to sign an**

international agreement on AI in February 2025. In the June 2026 G7 summit, EU Commission President Ursula von der Leyen **called for increased cooperation** between the US and EU on AI after the US administration **ordered** AI company Anthropic to block foreigners from two of its flagship models.

Other countries have been developing AI regulations, these include:

- Brazil published a **National Strategy for AI** and is constructing a **federal framework** for AI combining national policies, administrative coordination bodies and sector-specific regulations. Its **AI for the Good of All** strategy, published in 2025, promotes ethical and responsible use of AI. An AI **bill** creating rules on AI development and use following a risk-based model, was under review as of June 2026;
- China's **Interim Measures for the Management of Generative Artificial Intelligence Services** came into effect in August 2023. These state that: producers of content “bear responsibility” for generated outputs; providers must ensure sources used to train generative AI are lawful and do not infringe intellectual property rights; security assessments must be carried out before generative AI products are provided to the public; and providers must “employ effective measures to increase transparency in generative AI services and... the accuracy and reliability of generated content”. China introduced **mandatory labelling rules** for AI-generated and synthetic media and announced an **Action Plan for Global AI Governance** in July 2025, proposing a roadmap for global AI coordination. Further plans to accelerate legislation were announced in May 2026. A **comprehensive AI policy** to regulate the entire lifecycle of AI systems is also being drafted; and
- Japan enacted the **Act on Promotion of Research, Development and Utilisation of Artificial Intelligence-Related Technologies (the AI Promotion Act)** in 2025, forming an “**AI Basic Plan**” and establishing the **AI Strategic Headquarters**. The Act takes a risk based approach and aims to boost the national economy by promoting AI research. Adjacent laws (e.g. for data protection) are also used: its revised **Copyright Act** authorises the use of copyrighted data for data analysis, including as AI training data. In May 2024, the Japanese Prime Minister also **announced** a voluntary international framework to regulate generative AI called the Hiroshima AI Process Friends Group.

Figure 8: Someone consulting a document of the Cyberspace Administration of China on a screen



Some countries have devised strategies and issued guidelines, but do not have plans for regulations. For instance:

- The United Arab Emirates published a **National Strategy for Artificial Intelligence** in 2021 which outlines country's aim to become "a global hub for developing and adopting AI solutions and applications" by 2031. It is supplemented by the **Charter for the Development and Use of Artificial Intelligence**, published in June 2024, which outlines **principles** for compliance with existing legislation, rather than AI specific legislation
- India regulates AI based on the **India AI Governance Guidelines** published in 2025, which integrates AI governance into India's broader Digital Public Infrastructure (DPI) and combines broad principles with sectoral enforcement. In April 2026, the Government of India formed the **AI Governance and Economic Group** to coordinate the national approach to AI governance. India has also developed a **National Strategy on AI**, published **guidance for governing AI**, launched a **taskforce for India's economic transformation** and launched a national program on AI (**IndiaAI**);

- Australia have a risk-based approach framed in the **National AI Plan 2025**, which aims to grow the AI industry while spreading benefits and mitigating AI harms. The Plan will establish an **AI Safety Institute** to monitor, test and analyse advanced AI. Australia has no AI-specific legislation, instead leveraging existing laws to regulate AI and **promoting** eight AI Ethics Principles and a Voluntary Framework to “ensure AI is safe, secure and reliable”; and
- Canada has an **“AI for All” Strategy** published in June 2026 focusing on leveraging the benefits of AI across society. The Strategy emphasises data protection, equity and accessibility. An **Artificial Intelligence and Data Act**, which **followed** a similar approach to the EU, was **withdrawn** in January 2025 in favour of a lighted-handed approach to AI regulation focusing on federal level policies, sector-specific guidelines and a **Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems**.

These four countries signed the **Bletchley Declaration** in November 2023 and therefore acknowledge an agenda to implement “risk-based policies”.

4. What are the challenges to designing and implementing policy?

Key difficulties with policy for generative AI, include rapid change of technology, diffuse harms and international implications.

Fast-evolving

Since the generative AI environment evolves rapidly, regulations **need to account for new innovations**. Academics and regulators **suggest** technology-neutral laws with a focus on specific high risk applications instead. While being flexible, regulations **also need** to provide sufficient legal clarity to be effective and provide **coverage** to avoid legal gaps.

Regulators, civil society and organisations **require** appropriate powers and resources to be able to adapt to this fast-paced AI environment.

Defining harm and assessing damage

The impacts of generative AI are not always easy **to define**. It can be difficult to specify who has been harmed and when, for example, from inaccurate outputs. Small mistakes may appear to have minimal repercussions but with proliferation, can result in **detrimental societal consequences**.

Diffuse and cumulative impacts of generative AI make it difficult to identify **which kind of harm is 'enough' to be deemed unlawful**. As a result, assessing the consequences and appropriate sanctions, including what penalties to impose and when they should be given, can prove to be challenging. Similar concerns also arise with privacy issues.

If generative AI is viewed as 'speech', additional challenges linked to regulating speech arise. An article by the University of Manchester **explains**:

Regulating online content can be problematic due to challenges, such as defining the legally responsible actors for online hate speech and balancing a right to free speech with the desire to limit harmful content.

Need for international approaches

It is recognised that international collaboration is needed to avoid displacing problems. Social media is an example where no common strategy has been implemented to ban derogatory and discriminatory comments.

The Bletchley Declaration was signed by 28 countries and the European Union in November 2023. This declaration recognises that many risks are “inherently international in nature, and so are best addressed through international cooperation”. Signatories committed to cooperate on common principles and codes of conduct, for example through future international AI Safety Summits and building a “shared scientific and evidence-based understanding” of AI risks. However, they also recognise that their “approaches may differ based on national circumstances and applicable legal frameworks”.

In September 2024, the UK was one of the first countries to sign the Council of Europe’s first international treaty on AI. The UK and USA refused to sign an international agreement on AI at a global summit in Paris in February 2025 due to concerns about national security and “global governance”. In June 2025, G7 leaders signed a joint statement reaffirming cooperation on AI safety and risk management. Further efforts towards international collaboration brought G20 countries together at the India AI Impact Summit in February 2026.

Figure 9: Government representatives at the AI Safety Summit hosted by the UK in Bletchley Park



(Source: [Bletchley Park](#))

In March 2024, the Tech & Digital Ministers of the G7 signed a **declaration** setting out the development of an AI toolkit for the public sector. The toolkit was **published** in October 2024.

An international expert advisory panel was **tasked** with producing an **International AI Safety Report**, published in January 2025, to “help create a shared international understanding of risks from advanced AI and how they can be mitigated”.

The Welsh Government **acknowledges and supports** the **UN Global Dialogue on AI Governance** which encourages international cooperation and sharing of best practices. The Dialogue aims to ensure that governance of AI reflects the priorities of all nations and the benefits of AI are shared by all, encouraging international cooperation and sharing of best practices.

Acronyms

Acronym	Name
AI	Artificial Intelligence
AIGZ	Artificial Intelligence Growth Zone
ANN	Artificial Neural Network
CDDO	Central Digital and Data Office
CDPS	Centre for Digital Public Services
CMA	Competition and Markets Authority
DCMS	Department for Culture, Media & Sport
DPIA	Data Protection Impact Assessment
DSIT	Department for Science, Innovation and Technology
GDS	Government Digital Service
ICO	Information Commissioner's Office
IPU	Inter-Parliamentary Union
LLM	Large Language Model
ML	Machine Learning
MDLS	Minimum Digital Living Standard
NCSC	National Cyber Security Centre
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
PSED	Public Sector Equality Duty
RLHF	Reinforcement Learning from Human Feedback
SED	Socio-economic Duty

WPC	Workforce Partnership Council
-----	-------------------------------