

Generative AI – an overview

Research Briefing

July 2024



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 2024**

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

July 2024

Author:

Amandine Debus

Senedd Research acknowledges the parliamentary fellowship provided to Amandine Debus by the Natural Environment Research Council, which enabled this Research Briefing to be completed.



Contents

Summary	1
1. What is generative AI and what are the opportunities for the public sector?.....	3
Definition	3
Relevance for the public sector	5
Economic opportunities.....	6
Specific opportunities in Wales.....	6
2. What are the risks and challenges?.....	8
Technical risks and limitations of AI.....	8
Theoretical technical limitations.....	8
‘Real-world’ examples	9
Risks beyond technical limitations: stakeholder concerns	9
Existing guidance to mitigate risks for ethical, fair, and safe use	14

3. What is the current regulation of generative AI in Wales and other countries?16

Regulations and competency in Wales 16

 Existing legislation for Wales.....16

 Competency for further legislation18

UK policy and legislation 19

 AI-specific strategy.....19

 Implications of the AI regulation white paper for Wales.....21

EU policy.....22

Other countries24

4. What are the challenges to designing and implementing policy?.....27

Fast-evolving27

Defining harm and assessing damage.....27

Need for international approaches27

Acronyms.....29

Summary

What are the opportunities and risks of generative AI in the public sector?

Generative artificial intelligence (or GenAI) is the use of AI for the creation of new content.

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

However, generative AI outputs depend on the **quality of the data** used to train models which can lead to bias, discrimination, mistakes, or disinformation. It can also limit AI capabilities if there is less data available to train models in some languages **such as Welsh**.

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

What is the current approach to regulating generative AI in Wales?

There is currently **no specific holistic approach** to legislating for generative AI (and AI in general) in Wales in terms of its development, deployment or use. Legislative tools rely on an existing regulatory framework, including domain-specific legislation or legislation that applies to multiple domains.

Protection of personal information, intellectual property, and internet services **are reserved matters**, meaning that the Welsh Government and the Senedd do not have the power to make law in relation to 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 also limited scope for the Senedd to make law relating to **equal opportunities to address** faulty AI outputs.

The UK Government published its **AI regulation white paper** in March 2023, which states AI should be done within the existing powers of regulators and be devolved to sectors. It is based on a 'pro-innovation' approach and does not include a plan to bring forward specific new regulations or regulatory bodies. The **UK Government response** in February 2024 highlighted support from stakeholders and emphasised that regulators were already taking action to follow the proposed framework.

Other countries have adopted a different approach based on prescriptive legislative frameworks. These include **Brazil** and **China**. The **EU** has also been at the forefront of AI legislation with the development of an **AI Act**.

What are the challenges of regulating AI?

Since the generative AI environment evolves rapidly, regulation **needs to be flexible** to account for new innovations.

In addition, the impacts of generative AI **are not always easy** to define, which means it can be difficult to define 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 AI Safety Summit in November 2023, the strategies planned by different countries vary.

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

Definition

Generative artificial intelligence, also sometimes described as GenAI, is the use of AI for the creation of new content. This includes text, images, audio, videos, or music.. The **Turing Institute** defines Generative AI as referring to:

a type of artificial intelligence that involves creating new and original data or content. Unlike traditional AI models that rely on large datasets and algorithms to classify or predict outcomes, generative AI models are designed to learn the underlying patterns and structure of the data and generate novel outputs that mimic human creativity.

Generative AI learns patterns in **“human-created content” such as webpages, social media, and other online content**, and often uses **large language models (LLMs)**. **LLMs** learn from text and can perform a variety of Natural Language Processing (NLP) tasks by mimicking human language. **NLP** describes the branch of computer science which focuses on giving computers the ability to understand and generate speech. **LLMs** learn hundreds of millions to hundreds of billions of **parameters** (i.e. the values used to formulate the model) during training and are therefore computationally intensive, can be expensive and can take a long time to be trained.

In recent years, improvements in computer memory, dataset size, processing power, and long text sequences modelling techniques led to the development of powerful LLMs with greater capabilities and size, such as OpenAI’s **ChatGPT**. LLMs often use **transformers**, which apply a mechanism called **“self-attention”** to find relationships and patterns in sequential data.

Artificial intelligence (AI) can be defined as **“the science and engineering of making intelligent machines”**. In other words, AI describes the ability of machines to **replicate the capabilities of the human brain** for problem-solving and decision-making. Within AI, **machine learning (ML)** uses data and algorithms to build systems with learning capabilities. A sub-branch of ML is **deep learning (DL)**, which is characterised by the use of neural networks (NN) with layers that gradually learn, mimicking how neurons work.

Most of the recent generative AI technologies **rely on the use of foundation models**, which are large-scale ML models trained on a vast quantity of data. Foundation models are **not task-specific** and can be adapted to a variety of

domains. They are often called **general-purpose** models. However, **generative AI can** be designed for narrow specific tasks and use approaches that are not foundation models (e.g. **deepfake technology**, defined in Note 1 below, has been used since 2014 and uses another type of NNs).

Note 1

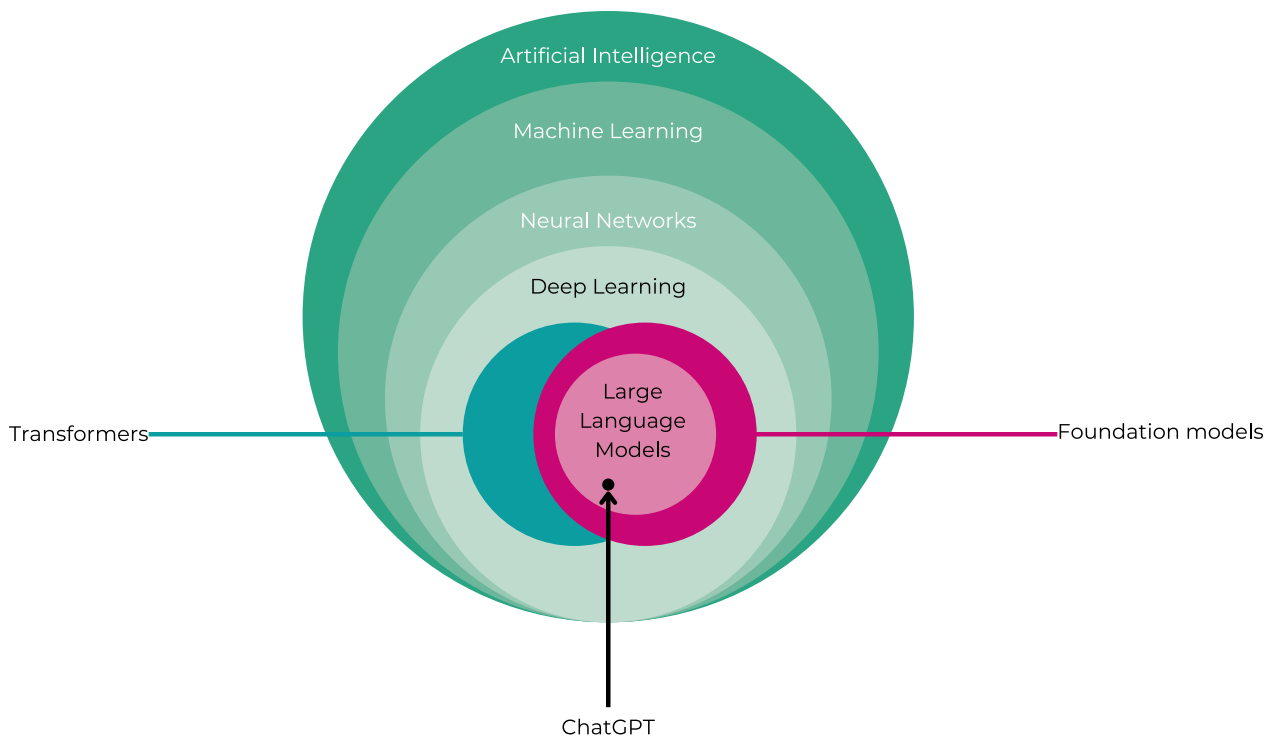
Deepfakes describe synthetic media “that replaces someone’s face or voice with that of someone else, in a way that appears real” (often videos or sound recordings, but it could also be pictures).

Various techniques can be used to train generative AI models, such as:

- **Supervised learning**, i.e. the model is trained with **labelled data** (i.e. data tagged with their properties in the context of the specific task performed by the AI model);
- **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 and learns from its mistakes.

Different types of learning can be **combined**. For example, generative models can be trained using unsupervised learning and then further trained using supervised learning for a specific task.

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



Other terms are sometimes used such as ‘frontier models’ to describe cutting-edge models, or ‘artificial general intelligence (AGI)’ and ‘strong AI’ to describe AI capable of any task a human could undertake. However, these terms are often contested because of inconsistent definitions. ‘Frontier models’ do not have a consistent way of measuring how ‘frontier’ it is, and AGI/‘strong’ AI do not describe current AI capabilities.

Relevance for the public sector

Multiple technology and consulting companies, such as **Microsoft**, **Deloitte**, and **Google**, have highlighted the opportunities offered by generative AI for the public sector. **Teachers** and several **public bodies** such as the **Financial Conduct Authority (FCA)**, the **Medicines and Healthcare products Regulatory Agency (MHRA)**, and **the Office of Communication (Ofcom)**, also emphasised the benefits of generative AI.

Generative AI offers opportunities to:

- **Increase efficiency;**
- **Improve the satisfaction of employees** by reducing mundane tasks;
- **Enhance citizen services (e.g. through the use of chatbots);**

- **Provide creative aid;**
- **Analyse deep data** for comprehensive outputs; and
- **Improve accessibility and inclusivity.**

Many specific **opportunities for parliaments** also exist, for example for recommendations on legislation based on identified gaps, consultation summaries, or horizon scanning for scientific research. This is a non-exhaustive list and **specificities** of each parliamentary setting and their digital strategies should be taken into consideration to assess relevance and priority.

Generative AI has already been used in the public sector across the world. For instance, it has been used in the **education sector** 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, to explain concepts or find information, and to produce content for presentations. It has also been **used by councils** to build chatbots to respond to enquiries. Pilots **have been carried out** for legislative drafting in Brazil, Argentina, Finland and Italy. In **Brazil**, the public can ask about parliamentary outputs (e.g. bills) and generative AI is used to pull information from those.

In Wales, generative AI **has been used** by both academic staff and students, and the University of South Wales published **guidelines** for appropriate use. ChatGPT has also been tested to write a **speech** in the Welsh Parliament.

Economic opportunities

McKinsey evaluates that “generative AI’s impact on productivity could add trillions of dollars in value to the global economy”. **KPMG** estimates that generative AI could add “£31 billion of GDP to the UK economy in the next decade”. AI in general could lead to an increase of £7.9 billion (9.8% growth in GDP) in Wales by 2030, according to a 2017 report by **PwC**.

However, there is still significant **uncertainty** regarding the global economic impacts of generative AI.

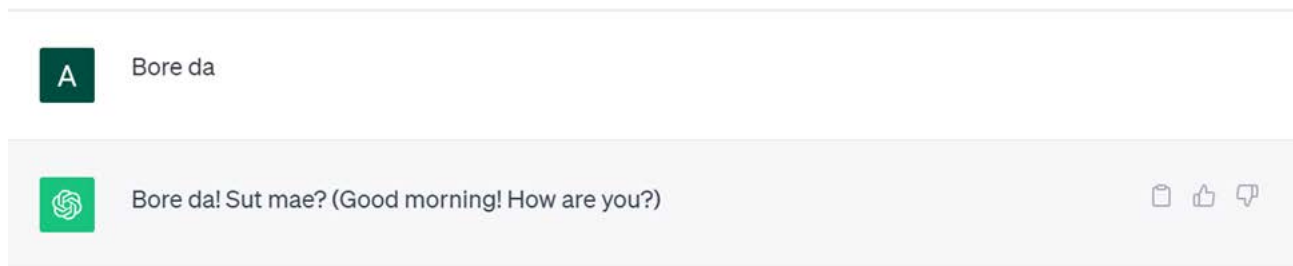
Specific opportunities in Wales

As a bilingual country, generative AI offers opportunities to promote and support the use of the Welsh language. Much of the translation sector in Wales, including the Welsh Parliament, **is already using** a neural machine translation model for Welsh, which **predicts** the probability of a sequence of words in Welsh, given

a sequence of words in English. “Traditional AI” (i.e. non-generative AI) Welsh automatic speech recognition tools, such as the **voice assistant Maccsen**, have also been developed. These could be improved with generative AI in the future, following **the example of Alexa**.

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. This Action Plan includes a work package specifically dedicated to Welsh-language machine learning and conversational AI.

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



In late 2023, the **Centre for Digital Public Services (CDPS)** launched a survey to get more information about how the Welsh public sector is using AI and conducted interviews with public services providers. **CDPS published a report** in early 2024 recommending **an automation and AI community of practice** for all Welsh public sector organisations.

2. What are the risks and challenges?

Technical risks and limitations of AI

Theoretical technical limitations

Generative AI outputs depend on the **quality of the training data** that is fed into the algorithms. An algorithm **corresponds** to the instructions a machine receives to find answers to a question or problem. 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 2 below);
- **Disinformation;**
- **Mistakes, creation of made-up knowledge** and **misinformation;** and
- **Discrimination.**

Note 2

Generative AI outputs will **reflect the values** of the creators of the training data, or the developers who selected the training datasets to use. As a result, marginalised voices will often not be represented in generative AI outputs.

These can lead to **wider effects** on society as a whole. For example, they can lead to an increasing erosion of trust in information, political and societal influence, and cascading failures if used in large-scale organisations.

This problem **may also worsen** with the development of future models, which are likely to use the outputs of previous models as training data and therefore be compromised because of this feedback loop.

In addition, it is often difficult to know why an error happened and to provide a human feedback loop due to a lack of oversight of the way the algorithm works. That is why many AI systems are called **'black boxes'**: the developer has limited understanding of internal mechanisms. The **Alan Turing Institute** highlighted these interpretability issues, and **academics** have called for interpretable models.

For Wales specifically, one limitation, at least currently, could also be the **lack of training data in Welsh**, limiting the ability of generative AI to provide satisfactory outputs. More broadly, generative AI is **often trained** on internet data, which is

mostly in a few data-rich languages (e.g. English, Spanish, Mandarin), creating language gaps for any non-data rich language. The outputs of a GenAI tool produced for an English prompt and a Welsh prompt could vary greatly. There could also be implications in terms of bias when using different languages.

Finally, **a study** showed that the behaviour of generative AI systems can change relatively quickly over time, by showing a decrease in ChatGPT's ability to follow user instructions in the span of three months, showing 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.

Generative AI has also **repeatedly** given outputs exacerbating harmful stereotypes, or made sexist and racist statements, **even with filters put in place to remove prejudicial comments**.

AI 'real-world' applications in general have had issues with bias in the public sector. For example, facial recognition used by South Wales Police was **declared unlawful** as it breached 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 in the case, 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 divorces. 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 the 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, and among other areas, clarification around liability

specifically. A **survey** conducted by the Ada Lovelace Institute highlighted that 62% of the people in Britain “would like to see laws and regulations guiding the use of AI technologies”. The consumer advocacy group BEUC **called for more regulation**, mostly due to privacy concerns (more details about privacy concerns in part 5 below). Trades Union Congress (TUC) also **called** for legislation to protect workers’ rights and launched an AI taskforce. The taskforce published a **draft AI and Employment Bill** in April 2024 and **lobby** for its incorporation into UK law.

- 2. Impacts on jobs and potential job losses:** **KPMG** estimates that 40% of UK jobs will be impacted by generative AI. **Creative workers** have raised concerns. In addition to writers, authors, and translators, **KPMG** also predicts impacts on computer programmers and software development professionals due to the capability of generative AI to write code. The UK Department for Education produced a **report on the impact of AI on UK jobs and training**, which highlighted high impacts of LLMs 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. The Institute for the Future of Work **produced a paper** based on a survey of over a thousand UK firms to understand AI adoption and potential impacts on jobs. They concluded that the outcomes are uncertain, but urgent action is required to avoid the aggravation of existing regional and demographic inequalities. The **Workforce Partnership Council** agreed to publish a paper on **Artificial Intelligence - Opportunities and Threats for the Public Sector**, which should provide more information about the impacts on jobs specifically in the public sector.

In August 2023, Wales TUC, with Kings College London and Connected by Data, **started an investigation** on the response of workers to the use of AI across sectors. This investigation has highlighted that workers worry that AI, and digitalisation in general, could worsen the workplace. For example, this investigation showed concerns about **surveillance by management**, **redundancies and lack of consultation with unions**, and **scepticism** about AI capabilities compared with human skills. TUC **produced a report** revealing notable increases in worker surveillance, and **guides** for unions on AI and the workplace. In addition, TUC highlighted the risk of bias and discrimination **starting at recruitment**. The House of Commons Library published a paper on **Artificial intelligence and employment law** which highlighted similar risks of using AI for recruitment, line management, and surveillance.

- 3. 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.

Figure 3: AI-generated image with the prompt ‘Wales’ on Pixlr



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

5. 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**. Privacy concerns have led to a temporary **ban of ChatGPT in Italy in 2023**. ChatGPT has been **reinstated** a month later "with enhanced transparency and rights for European users", including an option to refuse the use of conversations for ChatGPT training algorithms, age verification, and a notice explaining that ChatGPT may provide inaccurate information.

6. 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**. The **UK Intellectual Property Office** recommended clarifications on intellectual property around generative AI, and is working on a code of practice on copyright and AI. Lawsuits due to copyright issues with code and visual art generated from generative AI **have been filed in the US**. The **Alan Turing Institute** highlights issues with plagiarism in the context of education and assessments.

7. 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 ChatGPT RLHF reviewers were mostly from the global south and paid low incomes (e.g. less than \$3 per hour in Kenya) while being faced with content with detrimental effects on their wellbeing.

8. Environmental impact: Training generative AI, especially general purpose ones, **requires power and has an important carbon footprint**. **University College London reports** that:

(...) it is estimated that the training of GPT3 (the GPT used by the first version of ChatGPT made available to the public) consumed 1,287 megawatt hours of electricity and generated 552 tons of carbon dioxide, the equivalent of 123 cars driven for one year.

9. Loss of control and dependence: Loss of control over data fed into models is a risk that was mentioned by **PwC**. Over-reliance on generative AI can also lead to passiveness in learning, lack of engagement, and impact the development of

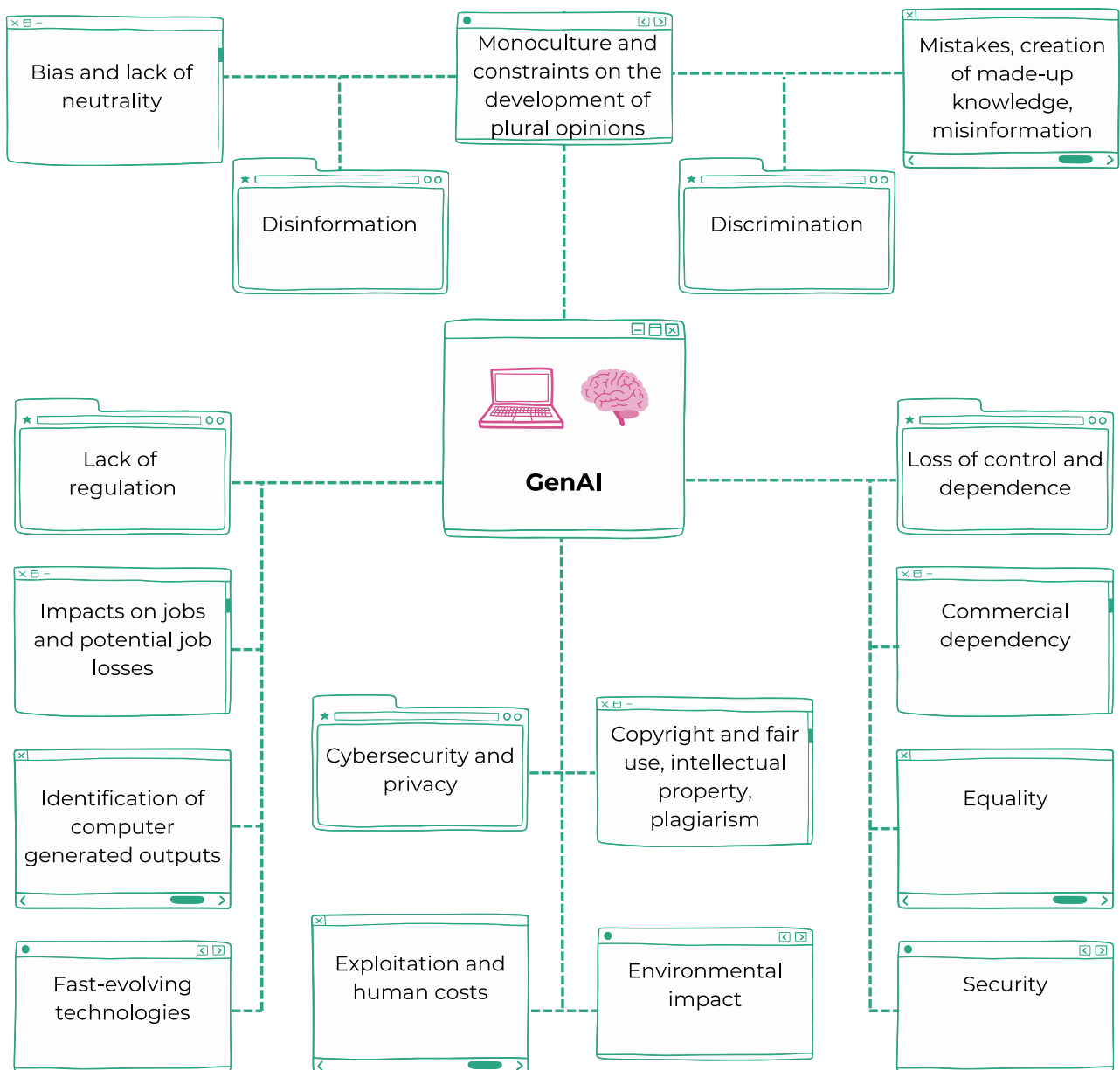
skills such as teamwork, communication, and critical thinking, **according to the Alan Turing Institute**.

- 10. 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 private actors do not share their technology due to commercial sensitivity.
- 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 about it. The EHRC emphasised that there is currently 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 believed 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).
- 12. Security:** 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 also mentions 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.

Figure 4: Summary of the risks of generative AI



Existing guidance to mitigate risks for ethical, fair, 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 Alan Turing Institute **Guide for the responsible design and implementation of AI systems in the public sector** (2019): it covers all types of AI, not only generative AI;
- The CDDO and Office for Artificial Intelligence **Guide to using artificial intelligence in the public sector** (2019) includes **guidance for ethics and safety**: it covers all types of AI;

- The Central Digital and Data Office (CDDO) **Data Ethics Framework** (2020): this is not specific to generative AI or AI;
- The EHRC **Public Sector Equality Duty (PSED) Artificial Intelligence Guidance** (2022): it is worth noting that, even though this guidance applies to Wales, the **PSED deep dive into AI carried out by EHRC did not include Wales**;
- The CDDO and Centre for Data Ethics and Innovation Algorithmic report on **Transparency Recording Standard - Guidance for Public Sector Bodies** (January 2023): it applies to any algorithmic tool, not only generative AI;
- The Cabinet Office and CDDO **Guidance to civil servants on use of generative AI** (September 2023);
- The Department for Education policy paper on **Generative artificial intelligence (AI) in education** (October 2023); and
- The National Cyber Security Centre (NCSC) **Guidelines for secure AI system development** (November 2023).

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

Regulations and competency in Wales

Existing legislation for Wales

There is currently **no specific holistic approach** to legislating for generative AI (and AI in general) in Wales in terms of development, deployment or use. Legislative tools rely on an existing regulatory framework, including domain-specific legislation or legislation that applies to multiple domains.

For example, existing legislation that could be used to legislate against misuse of generative AI in Wales include:

- The **Equality Act 2010**, which protects people against discrimination and supports equal employment opportunities;
- The **Data Protection Act 2018**, which implements the European Union's General Data Protection Regulation (GDPR) in the UK; and
- The **Online Safety Act 2023**, which includes criminalisation of the non-consensual sharing of explicit deepfakes.

However, the law-making powers of the Welsh Government and the Senedd are limited in the above areas by **the Government of Wales Act 2006**. Protection of personal information is reserved to UK Parliament, but there is limited scope for the Senedd to make law relating to equal opportunities.

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. However, **the EHRC commissioned research** in 2022 showed that no public bodies in Wales referred to AI or digital matters in their published strategic equality plans. **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.

The **Socio-Economic Duty** came into force in Wales in March 2021. The duty requires public bodies, when making strategic decisions about how to exercise its functions, to have due regard to the need to reduce the inequalities of outcome which result from socio-economic disadvantage. Public bodies therefore need to consider whether the use of generative AI increases or reduces socio-economic inequalities.

Competency for further legislation

The 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.

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

In addition to the guidance for the use of AI cited above, the UK Government:

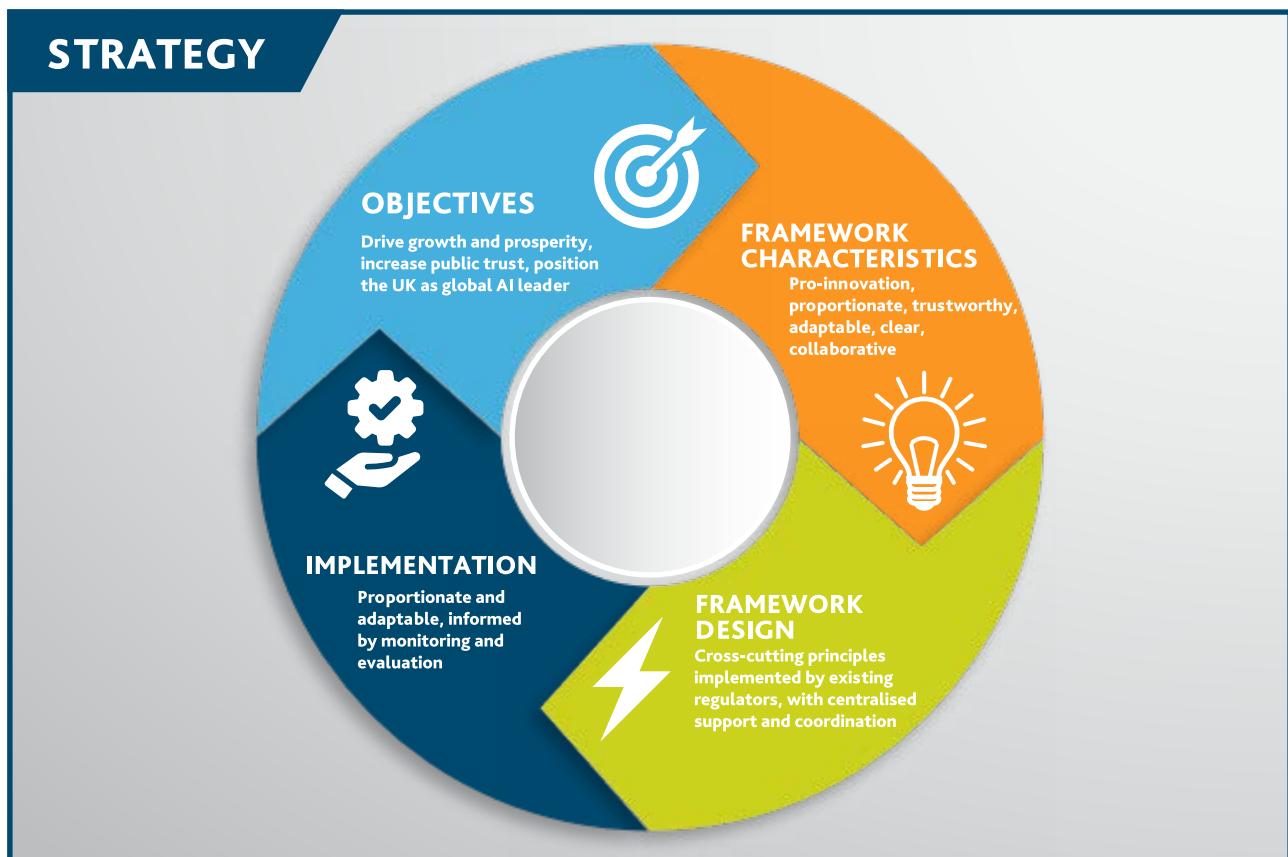
- Published its **National AI strategy** in September 2021 (updated in December 2022), with an accompanying **AI Action Plan** published in July 2022;
- Published a policy paper **Establishing a pro-innovation approach to regulating AI** in July 2022, which describes the high level principles to be applied by sector, and states that the next steps do not include new legislation: the policy paper explains that regulating AI should be done within the existing powers of regulators and be devolved to sectors, with cross-sectoral principles being initially non-statutory and regulators being encouraged to provide guidance and voluntary measures in the first instance;
- Created a new department, the **Department for Science, Innovation and Technology (DSIT)**, in **February 2023, which** brings together the relevant parts of the former Department for Business, Energy and Industrial Strategy and the former Department for Digital, Culture, Media and Sport, with the goal to drive "innovation that will deliver improved public services, create new better-paid jobs and grow the economy";
- Published its **AI regulation white paper** in March 2023, which is based on the 'pro-innovation' approach described in the 2022 policy paper mentioned above. 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.

- Provided **£100m funding to an AI Foundation Model Taskforce** to “lead vital AI safety research as part of driving forward the safe and reliable development of Foundation Models while seizing the extraordinary opportunities they present”, according to the press release from June 2023;
- Led the **International Summit on AI Safety** in November 2023;
- Announced the launch of the **AI Safety Institute** (ASI) in November 2023 and provided **more information** about its functions and approaches in February 2024;
- Put forward the **Data Protection and Digital Information (No.2) Bill**, which was being considered in the UK Parliament, and was expected to impact AI regulation as it made provision for the regulation of personal data processing. This Bill will **not progress further** as the 2023-24 session of the UK Parliament has prorogued;
- Recognised the need for “legislative action in every country once understanding of risk has matured” in its **response** to the AI regulation white paper from February 2024;
- **Announced** a £80 million investment to launch nine new research hubs across the UK to boost AI research in February 2024; and
- Provided **initial AI guidance** to regulators in February 2024 and **asked** key regulators to provide updates on their strategic approach to AI, which were all **published** by the end of April 2024.

The EHRC **highlighted** 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.

The Ada Lovelace Institute detailed **18 recommendations** for regulating AI in the UK following the white paper. These recommendations were published in July 2023. In July 2023, the House of Commons Science, Innovation and Technology Committee (SITC) **published** their interim report on the governance of AI. In this report, the SITC drew attention to the risks of the approach proposed in the white paper and provided 22 recommendations. The SITC highlighted the risk of not being able to keep up with the fast development of AI and the frameworks from other nations. They also recommended a “tightly-focussed AI Bill” to help the UK establish itself as an AI governance leader. The UK Government **replied** to this report in November 2023, restating their will to avoid “rushing to legislate” and highlighting that new developments in the AI regulatory framework would be addressed in the white paper consultation response.

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



(Source: [UK Government](#))

Implications of the AI regulation white paper for Wales

The AI white paper **explicitly states** that it applies to the whole of the UK as AI impacts on 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 relation to 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.

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.

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 Direct (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 800 AI policy initiatives** around the world.

As of May 2023, the US had the highest number of national-level AI policies. The US **National Artificial Intelligence Initiative Act of 2020** aimed to accelerate AI research and application and established the National AI Initiative Office, the National AI Advisory Committee, and the National AI Research Resource Task Force. The White House published a **Blueprint for an AI Bill of Rights** in October 2022, which constitutes non-binding guidelines for a safe and effective use of AI, with algorithmic discrimination protections and data privacy considerations.

The US National Institute of Standards and Technology (NIST) released an **AI Risk Management Framework** in January 2023 and launched the **Trustworthy and Responsible AI Resource Center** in March 2023. In July 2023, the White House **secured voluntary commitments** from leading AI companies to manage AI risks.

In October 2023, the President of the United States of America signed an **Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence**. In April 2024, federal agencies **announced** they completed the actions in the Executive Order.

The US has also been collaborating with the EU, for example with the **EU-US Trade and Technology Council (TTC)** producing a **joint AI Roadmap**. The UK and the US agreed on a AI safety **memorandum of understanding** in April 2024. **Roberts et al.** mentioned the potentiality of a ‘Transatlantic effect’ instead of a ‘Brussels Effect’ on the UK.

However, the US has not introduced any nationwide regulation. States and city authorities **are developing** their own regulations. For example, New York City published its **Artificial Intelligence Action Plan** in October 2023.

Other countries have been developing AI regulations:

- Brazil published a **National Strategy for AI** and **has put forward draft AI regulation**, which **promotes** a similar approach to the EU Act. This regulation **requests** AI providers to provide information to users, gives users power to request more information or contest an AI decision, requires developers to conduct risk assessments, and makes them liable for the damages caused by their systems;
- China published **Interim Measures for the Management of Generative Artificial Intelligence Services** which came into effect in August 2023. These state that the producers of the content “bear responsibility” for the content generated. They also specify that providers need to make sure the sources used to train generative AI are lawful, and that there are restrictions on the nature of these sources, which must not be infringing intellectual property rights. These measures also require security assessments to be carried out before generative AI products are used to provide services to the public. Finally, it says that the providers should “employ effective measures to increase transparency in generative AI services and to increase the accuracy and reliability of generated content”. China is also working on a **general “AI Law”** with a draft expected in 2024; and
- Canada proposed an **Artificial Intelligence and Data Act**, which **follows** a similar approach to the EU. The Act would require developers to have workflows to minimise risks, improve transparency, ensure the respect of anti-discrimination laws, as well as clear decision-making processes. As of May 2024, the Bill is **under consideration** in the House of Commons.

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**, but this strategy does not lay plans for regulations besides reviewing “national approaches to issues such as data management, ethics, and cybersecurity” and “the latest international best practices in legislation and global risks from AI”;
- Japan **has no regulations** specific to AI and adjacent laws (e.g. for data protection) are 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 however **announced** a voluntary international framework to regulate generative AI called the Hiroshima AI Process Friends Group;
- India **has developed** a National Strategy on AI, published guidance for responsible AI, launched a taskforce for India’s economic transformation, launched a national program on AI (India AI), and recently **announced** a requirement for “significant” tech companies to obtain government approval before the release of new models, but this is not legally binding and India **has not** introduced regulations as of May 2024; and
- Australia **has promoted** eight AI Ethics Principles and a Voluntary Framework to “ensure AI is safe, secure and reliable”, but **no regulation**.

These four countries signed the **Bletchley Declaration** in November 2023 and therefore acknowledge an agenda to implement “risk-based policies” (see below for more details).

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

Fast-evolving

Since the generative AI environment evolves rapidly, regulations **need to be flexible** to account for new innovations. Academics and regulators therefore **suggest** technology-neutral laws with a focus on specific high risk applications instead.

While being flexible, regulations **also need** to provide sufficient legal clarity for them to be effective. In addition, they **need** to provide coverage to avoid legal gaps and need rapid adoption.

Regulators, civil society, and organisations therefore **need** 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, as highlighted in Harvard Business Review. It can be difficult to define who has been harmed and when, for example with inaccurate outputs. In addition, a small mistake could seem to have minimal repercussions but with proliferation, can lead to detrimental societal consequences. It leads to difficulties in identifying which kind of harm is 'enough' to be deemed unlawful.

As a result, assessing the consequences and the appropriate sanction can also prove to be challenging. It can be difficult to determine the penalties to impose and when they should be given. Similar issues have been seen with privacy issues.

In addition, if generative AI is viewed as 'speech', it can lead to additional challenges linked to difficulties in regulating speech. 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”. The **most recent AI Safety Summit** took place in May 2024 in Seoul, Republic of Korea.

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.

An international expert advisory panel has **been tasked** with producing an International Scientific Report on AI Safety to address “where and why there is disagreement in the wider expert community, and present the debate in an objective manner.in the expert community”. The full report is expected towards the end of 2024, with a first publication in Spring 2024.

Acronyms

AI	Artificial Intelligence
CDDO	Central Digital and Data Office
CDPS	Centre for Digital Public Services
DL	Deep learning
DPIA	Data Protection Impact Assessment
DSIT	Department for Science, Innovation and Technology
GenAI	Generative Artificial Intelligence
LLM	Large Language Model
ML	Machine Learning
NCSC	National Cyber Security Centre
NLP	Natural Language Processing
NN	Neural Network
PSED	Public Sector Equality Duty
RLHF	Reinforcement Learning from Human Feedback
SED	Socio-economic Duty