Policy briefing

July 2023

Regulating AI in the UK

This briefing examines the UK's current plans for artificial intelligence (AI) regulation as set out in the March 2023 white paper 'A pro-innovation approach to regulating AI'. It sets out 18 recommendations for the Government and the Foundation Model Taskforce that, if acted on, will help to strengthen the proposed regulatory framework.

It is accompanied by a longer report – *Regulating AI in the UK* – which further contextualises and summarises the Government's proposals.[1]

The Ada Lovelace Institute (Ada) is an independent research institute with a mission to make data and AI work for people and society. This means making sure that the opportunities, benefits and privileges generated by data and AI are justly and equitably distributed.

1 Ada Lovelace Institute, *Regulating AI in the UK* (2023), https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk

Ada Lovelace Institute

For more information about the Ada Lovelace Institute or to discuss this policy briefing, contact Matt Davies: mdavies@adalovelaceinstitute.org

Key considerations for UK AI regulation

The UK Government has laid out its ambition to make the UK an 'AI superpower', leveraging the development and proliferation of AI technologies to benefit the UK's society and economy, and hosting a global summit in autumn 2023.

This ambition will only materialise with effective domestic regulation, which will provide the platform for the UK's future AI economy.

The Ada Lovelace Institute welcomes the allocation of significant Government resource and attention to the challenge of AI governance, and its commitment to driving AI safety forward at a global level. We contend that:

- Regulation will need to ensure that AI systems are trustworthy, that AI risks are mitigated, and that those developing, deploying and using AI technologies can be held accountable when things go wrong – a key ask of the British public in relation to AI regulation.[2]
- The definition of 'safety' adopted by the Government must be an expansive one, reflecting the wide variety of harms arising as AI systems become more capable and embedded in society. The solutions to well-documented AI harms on the one hand, and putative 'existential' risks on the other, are likely to stem from the same institutional capabilities.
- It is unlikely that international agreements will be effective in making AI safer and preventing harm, unless they are underpinned by robust domestic regulatory frameworks that can shape corporate incentives and developer behaviour in particular. The credibility of the UK's AI leadership aspirations therefore rests on getting the domestic regime right.

Regulating AI in the UK: our recommendations

This briefing sets out 18 recommendations for the Government and the Foundation Model Taskforce. Our recommendations fall into three categories, reflecting our three tests for effective AI regulation in the UK: coverage, capability and urgency.

2 Ada Lovelace Institute and The Alan Turing Institute, *How do people feel about AI? A nationally representative survey of public attitudes to artificial intelligence in Britain* (2023) https://www.adalovelaceinstitute.org/report/public-attitudes-ai/

Coverage

AI is being deployed and used in every sector, but the UK's diffuse legal and regulatory network for AI currently has significant gaps. Clearer rights and new institutions are needed to ensure that safeguards extend across the economy.

Challenge	Recommendation
Legal rights and protections New legal analysis shows safeguards for AI-assisted decision-making don't properly protect people.	**Recommendation 1:** Rethink the elements of the Data Protection and Digital Information Bill that are likely to undermine the safe development, deployment and use of AI, such as changes to the accountability framework. **Recommendation 2:** Review the rights and protections provided by existing legislation such as the UK General Data Protection Regulation (GDPR) and the Equality Act 2010 and – where necessary – legislate to introduce new rights and protections for people and groups affected by AI to ensure people can achieve adequate redress. **Recommendation 3:** Publish a consolidated statement of the rights and protections that people can expect when interacting with AI-based products and services, and organisations providing them.
Routes to redress Even when legal safeguards are in place, accessing redress can be costly and unrealistic for many affected people.	**Recommendation 4:** Explore the value of establishing an 'AI ombudsman' to support people affected by AI and increase regulators' visibility of AI harms as they occur.
Regulatory gaps The Government hasn't addressed how its proposed AI principles will apply in many sectors.	**Recommendation 5:** Set out how the five AI principles will be implemented in domains where there is no specific regulator and/or 'diffuse' regulation, and also across the public sector.

Capability

Regulating AI is resource-intensive and highly technical. Regulators, civil society organisations and other actors need new capabilities to properly carry out their duties.

Challenge	Recommendation
Scope and powers Regulator mandates and powers vary greatly, and many will be unable to force AI users and developers to comply with all the principles.	**Recommendation 6:** Introduce a statutory duty for legislators to have regard to the principles, including strict transparency and accountability obligations. **Recommendation 7:** Explore the introduction of a common set of powers for regulators and *ex ante*, developer-focused regulatory capability. **Recommendation 8:** Clarify the law around AI liability, to ensure that legal and financial liability for AI risk is distributed proportionately along AI value chains.
Resourcing AI is increasingly a core part of our digital infrastructure, and regulators need significantly more resourcing to address it.	**Recommendation 9:** Significantly increase the amount of funding available to regulators for responding to AI-related harms, in line with other safety-case based regulatory domains.
The regulatory ecosystem Other actors such as consumer groups, trade unions, charities and assurance providers will need to play a central role in holding AI accountable.	**Recommendation 10:** Create formal channels to allow civil society organisations, particularly those representing vulnerable groups, to meaningfully feed into future regulatory processes, the work of the Foundation Model Taskforce and the AI Safety Summit. **Recommendation 11:** Establish funds and pooled support to enable civil society organisations like consumer groups, trade unions and advisory organisations to hold those deploying and using AI accountable. **Recommendation 12:** Support the development of non-regulatory tools such as standards and assurance.

Urgency

The widespread availability of foundation models such as GPT-4 is accelerating AI adoption and risks scaling up existing harms. The Government, regulators and the Foundation Model Taskforce need to take urgent action.

Challenge	Recommendation
Legislation and enforcement New legislation, and more robust enforcement of existing laws, will be necessary to ensure foundation models are safe.	**Recommendation 13:** Allocate significant resource and future parliamentary time to enable a robust, legislatively supported approach to foundation model governance as soon as possible **Recommendation 14:** Review opportunities for and barriers to the enforcement of existing laws – particularly the UK GDPR and the intellectual property (IP) regime – in relation to foundation models and applications built on top of them.
Transparency and monitoring Too often, foundation models are opaque 'black boxes', with limited information available to the Government and regulators.	**Recommendation 15:** Invest in pilot projects to improve Government understanding of trends in AI research, development and deployment. **Recommendation 16:** Introduce mandatory reporting requirements for developers of foundation models operating in the UK or selling to UK organisations.
Leadership Priorities for AI development are currently set by a relatively small group of large industry players.	**Recommendation 17:** Ensure the AI Safety Summit reflects diverse voices and an expansive definition of 'AI safety'. **Recommendation 18:** Consider public investment in, and development of, AI capabilities to steer applications towards generating long-term public benefit.

Detailed recommendations

Coverage – protections that extend across the economy

AI harms can occur across the economy, and the mitigations afforded by the AI principles should extend across the whole economy too. We are concerned, however, that the coverage of the regulatory system proposed by the Government will be uneven.

Recommendations to improve coverage

Recommendation 1: Rethink the elements of the Data Protection and Digital Information Bill that are likely to undermine the safe development, deployment and use of AI, such as changes to the accountability framework.

We are concerned to see the Government proceed with plans to deregulate the use of data in the UK through the Data Protection and Digital Information Bill. Against an already poor landscape of redress and accountability in cases of AI harms, the Bill's changes will further erode the safeguards provided by underlying regulation.

Among other changes, most elements of the existing accountability framework for personal data use will be required only for 'high-risk processing', the approach to automated decision-making will be more permissive, and the ability of the Information Commissioner's Office (ICO) to issue guidance independently of the Government will be curtailed.

At a time when cross-economy access to powerful commoditised AI systems is growing, altering these legal protections is a serious misstep that risks undermining the Government's vision for AI safety and therefore the UK's credibility as an AI leader. The Government should reconsider the elements of the Data Protection and Digital Information Bill that will make AI less safe for affected people in light of increased AI adoption by businesses and individuals, and the outcomes of the review in Recommendation 2.

Recommendation 2: Review the rights and protections provided by existing legislation and where necessary, legislate to introduce new rights and protections.

Legal analysis commissioned by the Ada Lovelace Institute finds people affected by AI-informed decisions lack sufficient protection from harm or ability to get redress when things go wrong under existing legislation.

To support appropriate coverage of the AI principles across all sectors in which AI is likely to be deployed, we urge the Government to review the protections afforded by the UK GDPR and the Equality Act to people and groups affected by AI.

Where necessary, these existing rights may need to be strengthened to ensure an appropriate baseline of protection is available even in unregulated and partially regulated sectors. We have highlighted areas where this is the case, such as the regulation of biometric technologies.

The Ada Lovelace Institute is continuing to investigate how particular areas of law – such as the automated decision-making provisions contained in the UK GDPR and modified by the Data Protection and Digital Information Bill – could be updated for the era of widespread AI deployment. We expect to publish further information on this work later in the year.

Recommendation 3: Produce a consolidated statement of the rights and protections people can expect when interacting with (organisations using) AI.

The Government should take steps to provide a clear and consolidated statement of AI rights and protections, ensuring that members of the public have a clear understanding of the level of transparency and protection they should expect when using or interacting with AI systems.

The Government has said that it envisages regulators issuing joint guidance – albeit with the primary function of providing clarity to businesses, not individuals – and this could also be an appropriate mechanism for communicating with the public.

Another model could be the White House Office of Science and Technology's 'AI Bill of Rights'.[3] This document does not in itself have any legal standing but acts as a clear signal of the US Government's intent to act in certain ways when deploying or using AI, and makes explicit some protections that are provided under the US Constitution and existing laws.

Recommendation 4: Explore the establishment of an 'AI ombudsman' to support people affected by AI.

3 'Relationship to Existing Law and Policy | OSTP' (The White House) https://www.whitehouse.gov/ostp/ai-bill-of-rights/relationship-to-existing-law-and-policy/

There is a need for the Government to provide some sort of redress or dispute resolution mechanism for individuals affected by AI in sectors where no formal mechanisms currently exist.

Adopting an ombudsman-style model could act as a complement to other central functions the Government has set out. It could support individuals in resolving their complaints, direct them to appropriate regulators where this is not possible, and provide the Government and regulators with important insights into the sorts of AI harms people are experiencing, and whether they are effectively securing redress.

Ombudsmen have worked well in other areas such as financial services and maladministration. Their advantage in the context of AI would be providing a single point of contact, with a mandate to represent the individual in their capacity as citizen, consumer or worker, and covering a range of legal angles. This is in contrast with regulators who are mandated to balance different interests and often take a particular view on questions of law or policy. To operate effectively, an AI ombudsman would require access to sector-specific expertise and would therefore need to work closely with sector-specific regulators and ombudsmen.

Where businesses trying to embed AI principles in their products and services will have access to the ICO/Digital Regulation Cooperation Forum (DRCF) Multi-Agency Advisory Service pilot[4] and the proposed AI Sandbox,[5] there is at present no proposed equivalent for citizens trying to understand how to seek redress when they have suffered harm.

We propose an ombudsman pilot, which would represent a relatively modest investment from the Government, but – if successful – could dramatically improve redress for AI harms and the functionality of the framework as a whole.

Recommendation 5: Set out how the five AI principles will be implemented in domains where there is no specific regulator, where there is 'diffuse' regulation and across the public sector.

If implemented, Recommendations 1–3 should help to improve the legal safeguards available to people affected by AI, even in unregulated sectors. An AI ombudsman, as set out in Recommendation 4, would also ensure that there are meaningful routes to redress and contestability available to people affected by AI in these sectors. Taken together, Recommendations 1–4 offer a route to achieving the minimum viable standard of protection in instances of AI harm.

4 'Projects Selected for the Regulators' Pioneer Fund (2022)' (GOV.UK) https://www.gov.uk/government/publications/projects-selected-for-the-regulators-pioneer-fund/projects-selected-for-the-regulators-pioneer-fund-2022

5 Department for Science, Innovation & Technology and Office for Artificial Intelligence, *A pro-innovation approach to AI regulation* (2023) https://www.gov.-approach

However, the vision articulated by the Government's AI principles sets a higher bar than this. There will also be a need for the Government to clearly articulate how the principles will apply and be implemented, in scenarios where there is no regulator with obvious current responsibility for doing so. These sectors include:

- sensitive practices such as recruitment and employment, which are not comprehensively monitored by regulators, even within regulated sectors
- public-sector services such as education and policing, which are monitored and enforced by an uneven network of regulators
- activities carried out by central government departments, which are often not directly regulated, such as benefits administration or tax fraud detection
- unregulated parts of the private sector, such as retail.

There are a number of ways that the Government could do this. It could for example expand the remit and functionalities of existing regulators, to ensure that sectors are adequately covered. It could also consider introducing a 'backstop regulator' linked to the AI ombudsman, to implement and enforce the AI principles in contexts and sectors that are not comprehensively regulated at present.

Capability – an empowered and well-resourced regulatory ecosystem

Regulating AI effectively is a resource-intensive technical challenge. A second key test for UK AI regulation will be whether regulators and other actors involved in making AI accountable – such as civil society organisations and third-party providers of AI assurance services – have the necessary capabilities to discharge their functions.

Recommendations to improve capability

Recommendation 6: Introduce a statutory duty for regulators to have regard to the principles, including strict transparency and accountability obligations.

It will be important for a statutory duty to be introduced, mandating regulators to implement the AI principles. However, a statutory duty to merely 'have regard' to the principles could be discharged by regulators by simply stating to the Government, or providing minimal evidence, of their consideration in strategy setting.

To be effective in providing regulator engagement with the principles and accountability for their progress in implementing them, a statutory duty would also need to be supported with robust transparency and accountability obligations. These would include a requirement to report progress

to Parliament against specified KPIs, and to publish open data that supports monitoring and evaluation of the entire framework in its effectiveness at mitigating AI risks identified by the central risk function.

Where regulators or the central functions identify AI risks that are poorly mitigated or unmanaged by existing regulation, a policy response will be required from the Government. The process for the reporting of these risks, and the Government's consideration and response to them, should be formalised in a notification and reporting process, ideally with some level of public transparency to ensure accountability for responding.

Recommendation 7: Explore the introduction of a common set of powers for regulators, including an *ex ante*, developer-focused regulatory capability.

The Government should consider the case for legislation that would equip regulators with a common set of AI powers that would put them on an even footing in addressing AI. We are aware of ongoing research at The Alan Turing Institute which seeks to map existing regulator powers and identify gaps. This could complement additional work by the Government or the Foundation Model Taskforce to identify gaps in relation to foundation models specifically, as set out in Recommendation 14.

One area that should be considered in this regard is the introduction of greater powers to request information of companies developing, deploying or using AI systems, and to compel those organisations to make that information available more widely when appropriate.

Another major gap in the regulatory toolkit is the lack of powers to ensure that organisations developing or selling AI tools adhere to safety requirements. Regulators could in theory bring in *ex ante* product safety requirements but it is doubtful whether this is currently feasible in practice. Many existing regulators focus on outcomes, meaning – in practice – that they are only equipped to look at technology at the point of use or commercialisation.

AI, and foundation models (like GPT-4) in particular, confound this model of regulation: they are often the building blocks behind specific technological products that are available to the public (like Bing) and sit upstream of complex value chains.[6] This means that regulators may struggle to reach the companies or other actors most able to address AI-related harms, with the potential consequence that responsibility for addressing risks and liability will accrue to the organisation using an AI tool.

6 Ian Brown, *Allocating Accountability in AI Supply Chains* (Ada Lovelace Institute 2023) https://www.adalovelaceinstitute.org/resource/ai-supply-chains/

Recommendation 8: Clarify the law around AI liability.

A further area where statutory interventions would be useful concerns the potential value for the law governing legal and financial liability. This could ensure that actors within the AI lifecycle who are in the best position to mitigate given AI risks are appropriately incentivised to address them.

While theoretically it may be possible to address this through contract, market dynamics may result in legal and financial risk being passed towards smaller actors who tend to sit at the end of AI value chains ('AI users', which can include organisations and members of the public).

From a UK perspective, this dynamic could be a particularly undesirable one: the UK's comparative strengths in AI tend towards products and services further down the AI value chain (such as in services associated with the deployment and implementation of AI), rather than in upstream activities. There may therefore be an important role for legislation in clarifying the law around AI liability and potentially redistributing it.

Recommendation 9: Significantly increase the amount of funding available to regulators for responding to AI-related harms.

To ensure that the UK's regulatory ecosystem has the necessary capabilities to implement the AI principles, the Government should introduce funding for cross-cutting regulators such as the EHRC and ICO to scale up monitoring and enforcement.

AI is a general-purpose technology with significant safety implications, which will increasingly form part of the UK's digital infrastructure. In other domains where safety and public trust are paramount and where underlying technologies form important parts of national infrastructure – such as civil nuclear, civil aviation, medicines, road and rail – annual regulatory funding is in the region of tens of millions of pounds, if not higher.

Regardless of whether AI regulation is delivered on a centralised or distributed basis, or of the funding model, we contend that the challenge of governing a general-purpose technology like AI effectively will be on a similar scale, and the Government should consider models for providing resourcing accordingly – both for regulators, but also for central Government policy capacity.

We anticipate the needs of digital regulators (such as those that are members of the DRCF) will be different to less-digitally mature regulators, which will have smaller, less-specialist teams of AI-focused experts (if any) and which would benefit more from centralised capacity in the absence of increased ring-fenced AI funding. Digital regulators could play a significant role

in upskilling and sharing learning across the wider regulatory ecosystem through the central functions, as well as building on existing coordination mechanisms such as the DRCF and Regulators' AI Working Group.

Recommendation 10: Create formal, funded channels to involve civil society organisations, particularly those representing vulnerable groups, meaningfully in the regulatory process, the work of the Foundation Model Taskforce and the AI Safety Summit.

One activity that the Government could consider is the provision of formal channels to involve civil society organisations – such as consumer groups, trade unions, and groups representing underrepresented and vulnerable people – in the work of regulators and the central functions. This should include opportunities to meaningfully participate in the work of the Taskforce and the AI Safety Summit. Strategic partnership arrangements between the Government and civil society organisations, which have led to significant policy improvements (e.g. in health),[7] could serve as a useful model to follow.

This work would need to be appropriately funded: many civil society organisations are under-resourced, particularly those that provide frontline services or that work with vulnerable communities, and failing to fund participation would risk excluding these perspectives. Conversely, some civil society organisations are wholly or mostly funded by large private-sector organisations, and in these cases their research or policy positions may not be independent. Providing an even playing field for an array of civil society voices will be crucial to ensuring that civil society participation in the AI governance system is of genuine value rather than becoming an opportunity for regulatory capture. The Government could seek to draw lessons from the experience of previous initiatives such as Open Banking, which has been praised for having civil society appointees on expert groups but criticised for failing to fund participation.[8]

Recommendation 11: Establish funds and pooled support for civil society participation in all levels of the regulatory process.

In addition to formal input at a national level from civil society organisations, the Government should also explore how civil society organisations at a local level can be supported to engage with the regulatory system.

7 'NHS England » Voluntary, Community and Social Enterprise (VCSE) Health and Wellbeing Alliance' https://www.england.nhs.uk/hwalliance/

8 'FCA Open Finance Call for Input - Lab Response' (Finance Innovation Lab) https://financeinnovationlab.org/insights/open-finance-response/

Trade union branches, consumer groups, local community organisations and organisations representing people with protected characteristics are in close contact with those who are likely to be the most affected by AI technologies. As such these organisations will need to play an important role in the regulatory ecosystem: holding organisations deploying or using AI to account, and supporting individuals to navigate redress mechanisms and report incidences of harm to regulators and the AI ombudsman.

As AI systems continue to be integrated into our everyday lives – from schools and workplaces to shops and public spaces – these organisations will require funding and expertise to ensure they can continue to effectively serve their communities. As such, we contend that the Government should consider introducing ringfenced funding and pooled support to help upskill a diverse range of civil society organisations in AI and resource their meaningful engagement with regulators, the central functions, and the AI ombudsman.

Recommendation 12: Support the development of non-regulatory tools for trustworthy AI.

The Government also expects non-regulatory tools such as standards and assurance to play a role alongside regulation in improving AI outcomes. It has committed to collaboration with partners such as the UK AI Standards Hub[9] to develop these tools and support responsible innovation.

We believe that supporting the flourishing of an 'ecosystem of assessment, assurance and audit' can help to mitigate AI harms. However, we are concerned that this could become a point of failure within the regulatory system if policymakers overestimate the capability of a still-nascent AI assurance market to catch certain risks and drive up standards.

The Government is already supporting the development and adoption of assessment, assurance and audit mechanisms through vehicles like case studies and its support for the AI Standards Hub. We think there are a number of other ways that the Government could support the creation of an effective assessment ecosystem:

- Create incentives for companies, drawing on external expertise and certification where appropriate, to assess risks from AI systems, e.g. mandated algorithmic impact assessments in particular sectors, or introducing requirements as part of data-access processes and procurement requirements in the public sector.

- Introduce domain or sector-specific guidance on societal risks (perhaps produced by regulators) that could support the development of AI risk and impact assessment methods tailored to specific sectors.

9 'AI Standards Hub – The New Home of the AI Standards Community' (*AI Standards Hub*) https://aistandardshub.org/

- Developing the skills base. The technology sector will need teams, roles and staff with the skills to conduct risk and impact assessments. In particular, many methods involve identifying and coordinating diverse stakeholders, and the use of participatory or deliberative methods that are not currently widespread in the technology sector, but are more established in other domains such as participatory research, policy, design, academic sociology and anthropology.
- Resourcing and empowering organisations to assess risks and impact. Many of the most well-known and significant AI risk assessments to date have been conducted by civil society groups, academics and companies that evaluate a system's impacts without the permission of the company. However, these organisations often lack access or information about emerging AI systems, and may not be well resourced to conduct these kinds of assessments.

For more information on these activities, and how the Government and regulators can facilitate them, read the Ada Lovelace Institute's recent research paper.[10]

Urgency – taking action before it's too late

The third factor is sufficient urgency on current and emerging risks. The Government envisions a timeline of at least a year before the first iteration of the new AI framework is implemented, with further time needed to evaluate its effectiveness and address any emerging limitations.

Under ordinary circumstances, that would be considered a reasonable schedule for establishing a long-term framework for governing an economically and societally cross-cutting technology. But there are significant harms associated with AI use today, many of which are felt disproportionately by the most marginalised members of society. In particular, the pace at which foundation models are being integrated into the economy and our everyday lives means that they risk scaling up and exacerbating these harms.

Recommendation 13: Immediately allocate significant resource and future Parliamentary time to enable a robust, legislatively supported approach to foundation model governance.

Foundation models are being integrated into the practices of organisations across the economy. The major factor determining the trustworthiness of foundation models developed or deployed in the UK will be the presence of a strong domestic regulatory framework that can effectively shape incentives and developer behaviour.

10 Ada Lovelace Institute, *AI assurance? Assessing and mitigating risks across the AI lifecycle* (2023) https://www.adalovelaceinstitute.org/report/risks-ai-systems/

We therefore contend that it will be important for the Government to allocate significant resource and future Parliamentary time to enable the creation of such a framework. The announcement £100m for the Foundation Model Taskforce chaired by Ian Hogarth is a welcome acknowledgement of this urgency, and Recommendations 14 and 15 make a number of suggestions for how this resource could be fruitfully spent.

It is likely however that certain parts of the solution to foundation model governance will require new primary legislation: for example the introduction of new *ex ante* powers for regulators (Recommendation 7), clarification to liability rules (Recommendation 8), and mandatory transparency requirements for developers (Recommendation 16). Parliamentary time is a scarce resource, and the Government should act now to ensure that legislation can be passed as swiftly as possible.

Recommendation 14: Review opportunities for and barriers to the enforcement of existing law.

We also contend that there is a need to review the opportunities for more proactive enforcement of existing UK law and regulation that addresses the risks of foundation models (notably the UK GDPR, the Equality Act 2010 and the intellectual property regime). At present, the compliance of many widely available foundation models with these legal regimes is questionable.

However, cross-cutting UK regulators are constrained in the powers, resources and the sources of information available to them, as well as cultural barriers to enforcement. There are also particular challenges associated with enforcing the law in relation to foundation models, chiefly among them the opacity of many widely used models and the datasets used to train them.

The Government – and the Foundation Model Taskforce – could play a constructive role in reviewing these opportunities for, and barriers to, the enforcement of existing law in relation to foundation models. This would strengthen Government understanding of where legislative change might be necessary (Recommendation 13) and what sort of transparency requirements might need to be imposed on developers to facilitate effective regulatory action (Recommendation 16).

Recommendation 15: Invest in pilot projects to improve Government understanding of trends in research, development and deployment.

There are a number of pilot projects that could be carried out – probably by the Foundation Model Taskforce – to improve Government understanding of trends in AI research, development and deployment.

At present, the Government is largely reliant on external expertise from industry for these insights. While collaboration with industry will continue to be an important component of effective AI governance, there are inherent risks in over-optimising regulation to the needs and perspectives of incumbent industry corporations and companies.

We contend that the Government understanding of the sector, and of necessary governance interventions, would be strengthened by conducting systematic in-house analysis.

In the longer term, the horizon scanning and cross-sectoral risk assessment functions envisaged by the Government will be important vehicles for this. However, we propose that the timeline of 12 or more months for their establishment, coupled with the current fast pace of AI development and uptake, means that there is a strong case for action sooner.

We propose that the Foundation Model Taskforce should look to invest immediately in small pilot projects that could begin to build this in-house expertise and infrastructure and which – if successful – could be continued as part of the central functions.

Immediate pilot projects could include:

- establishing a national-level public repository of the harms, failures and unintended negative consequences of AI systems deployed in the real-world and potential future harms of in-development applications, building on the work of the Responsible AI Collaborative's AI Incident Database[11]
- developing benchmarks and evaluations to test for the potential harms and risks foundation models may raise in deployment
- beginning to regularly monitor, aggregate (and potentially publish) data on compute use and demand trends, building on the work of the Future of Compute review.[12]

For more information on potential monitoring activities, read Ada's recent report *Keeping an eye on AI*.[13]

11 'Welcome to the Artificial Intelligence Incident Database' https://incidentdatabase.ai/

12 'Independent Review of the Future of Compute' (*GOV.UK*, 6 March 2023) https://www.gov.uk/government/publications/future-of-compute-review

13 Ada Lovelace Institute, *Keeping an eye on AI: Approaches to government monitoring of the AI landscape* (2023) https://www.adalovelaceinstitute.org/report/keeping-an-eye-on-ai/

Recommendation 16: Begin to introduce mandatory reporting requirements for developers of foundation models operating in the UK.

To facilitate the monitoring and analysis activities detailed in Recommendation 15 – and the growth of an 'ecosystem of assessment' around foundation models as discussed in Recommendation 13 – the Government should also consider introducing mandatory reporting and transparency requirements for developers of foundation models operating in the UK. This would give the Government and regulators greater visibility and understanding of AI development and uptake, and could therefore help to alleviate some of the barriers to the enforcement of existing law in relation to foundation models as discussed in Recommendation 14.

Working with industry, the Taskforce could play a useful role in developing and piloting these requirements. These could be introduced on an initially voluntary or contractual basis with developers, building on welcome recent commitments from leading foundation model developers Google DeepMind, OpenAI and Anthropic, to give early or priority access to models for research and safety purposes.

In time, these requirements will need to be further specified and made mandatory. This could be done in the first instance through contract-based agreements with developers, which could be secured quickly in anticipation of formal enforceable legislative provision.

One way to extend and build on these relationships would be to require notification when these organisations (and similar labs) begin large-scale training runs of new models. This would provide the Government with an early warning of advancements in AI capabilities, allowing policymakers and regulators to prepare for the impact of these developments, rather than being caught unaware.

We contend that it will be important that reporting requirements are appropriately scoped: the Government should consider how mandatory requirements can ensure transparency not only of new or 'frontier' models – a term which is difficult to define or measure, and which is likely to change over time – but of all powerful foundation models made available (whether through application or API access) in the UK.

The Government and regulators will require access to a variety of different types of information on these models to appropriately tackle the spectrum of AI harms. As such, reporting requirements should also include information such as access to the data used to train models, results from in-house audits, and supply chain data. We contend that reporting requirements are a good example of how solutions to different AI harms are often complementary, stemming from a common set of institutional mechanisms.

Recommendation 17: Ensure the AI Safety Summit reflects a diverse range of voices.

As a UK-based research institute whose mission is to ensure that data and AI work for people and society, we have welcomed the commitment of significant Government resource and attention to these important issues as represented by the announcement of the AI Safety Summit.

If it wants to secure international leadership on AI, the UK needs to have a credible domestic approach to trustworthy AI governance at home. All the recommendations in this report are relevant to this, and we would highlight the need for the Government to:

- address the gaps in its framework, including strengthening underlying regulation such as data protection law (as discussed in the section on Coverage)
- committing Parliamentary time to give regulators the right incentives, accountability and powers to deliver on the AI principles (as discussed in the section on Capacity)
- committing to ensuring AI regulation is properly resourced (as discussed in the section on Capacity).

As discussed above, it will be vital to ensure that the definition of 'AI safety' used by the summit is a broad one, providing a forum for both more proximate risks and larger but less knowable ones to be addressed. As part of this, it is important that voices representing those affected by AI are also heard at the summit, as well as the wider research community – and not solely governments or large industry players, who will have a particular perspective on risk.

Ultimately, the success of the AI Safety Summit will be determined by whether it can secure concrete commitments from international governments and industry that complement and build on existing work on AI governance in the UK and across the world.

Addressing AI safety will require legislative time and resource, and in the shorter term, the voluntary cooperation of industry. Achieving this will be more feasible if most major economies set the same expectations, and so reaching these agreements – for example on reporting requirements, as discussed in Recommendation 17 – should be a priority for the summit.

Recommendation 18: Consider public investment in, and development of AI, to unlock societal benefits.

The extent of market concentration in the digital economy raises serious questions around power and oversight. It also means that AI development is overwhelmingly focused on particular types of technologies (such as the recent spate of chatbot developments) with relatively narrow and commercial applications, rather than on technologies or use cases that centre individual, community and societal benefit.

The current 'AI moment' is a critical inflection point for these challenges: as AI uptake rapidly increases, societies risk unwittingly locking ourselves into a set of technologies, and economic dynamics, that are not necessarily optimal.

In other sectors, national and supra-national governments can rely on various tools to shape and 'direct' growth towards societal benefits. This is the rationale behind, for example, the Inflation Reduction Act in the USA,[14] and the Net-Zero Industry Act in the European Union,[15] which can be seen as legislative attempts to 'crowd in' private investment towards goals such as of tackling the climate crisis.

We propose that there might be a role for greater public investment in, and public development of, AI to rebalance existing concentrations of power, democratise the sector and direct AI towards better outcomes for people and society. This would potentially require new public capacities and institutions: for example, new public institutions for data and AI governance (as discussed in our reports, *Rethinking data and rebalancing digital power*[16] and *Legal mechanisms for data stewardship*);[17] new intelligence-gathering and market-shaping capabilities for regulators and government (as discussed in *Regulate to innovate*);[18] and new vehicles for the public sector to invest in or directly develop AI.

Any significant investment from the Government in public AI development would need to meet a high justificatory bar. We contend that it is unlikely, for instance, that model training looking to replicate or compete with the success of foundation models such as GPT-4 would unlock significant benefits for people and society at proportionate cost.

We do however propose that it would be valuable for the Government to explore how public support – whether through the Taskforce, the Advanced Research and Invention Agency (ARIA) or more established investment vehicles such as UK Research and Innovation (UKRI) – could facilitate the development of AI technologies and applications that are not currently well-served by market trends: public-service recommendation algorithms, for instance, or data analytics solutions optimised to the needs of local authorities.

14 'Inflation Reduction Act Guidebook | Clean Energy' (The White House) https://www.whitehouse.gov/cleanenergy/inflation-reduction-act-guidebook/

15 European Commission, 'The Net-Zero Industry Act' https://single-market-economy.ec.europa.eu/industry/sustainability/net-zero-industry-act_en

16 Ada Lovelace Institute, *Exploring Legal Mechanisms for Data Stewardship* (2021) https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/

17 Ibid.

18 Ada Lovelace Institute (n 1).

Regulating AI in the UK: full report

Read our report which further contextualises and summarises the UK's current plans for AI regulation.

Building on previous research by the Ada Lovelace Institute, its recommendations – and those in this briefing – are based on extensive desk research, two expert roundtables, independent legal analysis and the results of a nationally representative survey.

https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/

Ada Lovelace Institute
100 St John Street, London, WC1B 3JS
+44 (0) 20 7631 0566

Registered charity 206601

Website: adalovelaceinstitute.org
Twitter: @AdaLovelaceInst
Email: hello@adalovelaceinstitute.org