Policy briefing

October 2023

# Mission critical

## Lessons from relevant sectors for AI safety

AI systems are rapidly becoming ubiquitous as they are integrated into almost every aspect of our lives: from schools to public services, and from our phones to the cars we drive. Organisations across all sectors of the global economy are looking to develop, deploy and make use of the potential of these technologies. At the same time, we are already seeing considerable harms caused by the use of AI systems: ranging from discrimination, misuse and system failure, to socioeconomic and environmental harms.

The UK Government's announcement of an 'AI Safety Summit' – taking place on 1 and 2 November 2023 – has led to a surge of public interest in these topics. As world leaders and leading figures from industry and civil society descend on Bletchley Park, the words 'AI safety' are likely to feature prominently – but this term is a contested one, with no consensus definition.

This policy briefing explores approaches to regulation and governance around other emerging, complex technologies and sets out a roadmap to securing governance that ensures AI works for people and society.

Ada Lovelace Institute

**For more information about the Ada Lovelace Institute and this work, contact Ada's policy team: hello@adalovelaceinstitute.org**

## Background

The Government's agenda for the AI Safety Summit is focused primarily on technical methods for avoiding hypothetical 'extreme risks' that could emerge from the misuse or loss of control of advanced 'frontier' AI systems. While this focus has broadened somewhat to include other types of risk in the weeks leading to the Summit, many within industry, academia and civil society have rejected the Summit's focus as overly narrow and insufficiently attentive to the wide range of AI harms people are currently experiencing – without adequate protection.[1]

In other domains of cross-economy importance – such as medicine, transport and food – we take a more expansive approach to governance. Regulation in these sectors is designed to ensure that systems and technologies function as intended, that the harms they present are proportionate, and that they enjoy public trust. In this way, regulation is an enabler of innovation – rather than an inhibitor – as it ensures products and services are safe for people to use.[2]

As we use AI systems more and more in our daily lives, they begin to form a part of our critical products and services. This means failures in technology design and deployment can cascade down into the contexts they are used in, and lead to severe consequences.

Recent advances in foundation models – defined as a single AI model capable of a wide range of tasks – exacerbate this challenge as AI is integrated into the digital economy. These models are capable of a range of general tasks (such as text synthesis, image manipulation and audio generation). Notable examples are OpenAI's GPT-3 and GPT-4, the foundation models that underpin the conversational chat agent ChatGPT. They provide a foundation on which downstream products and services can be created.

If AI systems become more integrated into different parts of our lives, our AI governance institutions will need to reflect the role AI plays in our societies and economies. This policy briefing provides an introduction to how regulation operates in other sectors, laying out early answers to some key questions that Ada will explore in the coming months:

- What roles are different types of AI likely to play in our future economy and society?
- What are the goals of regulation in other sectors that play a similar role?
- What regulatory mechanisms are used to achieve these goals?
- What are the key features of these regulatory systems that enable their success?

[1] Seth Lazar and Alondra Nelson, 'AI safety on whose terms?' (2023) 381(6654) Science https://www.science.org/doi/10.1126/science.adi8982 accessed 25 October 2023.

[2] Ada Lovelace Institute, *Regulate to innovate* (2021) https://www.adalovelaceinstitute.org/wp-content/uploads/2021/12/Regulate-to-innovate-Ada-report.pdf accessed 25 October 2023.

Regulation in other sectors can inspire AI regulation, either by providing examples of best practices or examples of regulatory failures that should be avoided in the future. We hope that this briefing can inform conversations at Bletchley Park and elsewhere, during the Summit and into the future, and provide constructive lessons for the announcements brought forward by Government.

## What role are different types of AI systems likely to play in our future economy and society?

The AI Safety Summit is premised on the notion that the impact of AI on our society and economy will be transformational.[3]

In some ways, it already is: AI is being deployed in important areas of scientific discovery such as genomics,[4] and across important societal challenges such as climate change adaptation and mitigation.[5] In the UK, AI tools have been adopted by businesses in most sectors of the economy, with varying levels of uptake and success.[6]

But in other respects, there are reasons to be concerned. Due to a lack of transparency and reporting, it remains unclear whether the carbon costs of AI – which are considerable – outweigh any benefits that are accrued. For now, societal and economic gains from AI remain unevenly distributed and relatively small in scale. There is, however, optimism – from the public, from experts and from practitioners across the public and private sectors – that AI can be developed and used for public benefit.[7]

Foundation models, which require significant quantities of compute, energy and data to train, carry many of these concerns. Foundation models can be built 'on top of', to develop different applications for many purposes. They are being used to add novel features to applications that already have millions of users, ranging from search engines (like Bing) and productivity software (like Office365), to language learning tools (such as Duolingo Max) and video games (such as AI Dungeon). In many cases they can be accessed through application programming interfaces (APIs),

[3] Department for Science, Innovation & Technology, 'AI Safety Summit: introduction', (GOV.UK, 11 October 2023) https://www.gov.uk/government/publications/ai-safety-summit-introduction/ai-safety-summit-introduction-html accessed 25 October 2023.

[4] Ada Lovelace Institute, DNA.I. - Early findings and emerging questions on the use of AI in genomics (2023), https://www.adalovelaceinstitute.org/report/dna-ai-genomics/ accessed 25 October 2023.

[5] Emily Clough, *Net zero or net hero? The role of AI in the climate crisis* (Ada Lovelace Institute 2023) https://www.adalovelaceinstitute.org/resource/climate-change-ai/#using-ai-to-address-climate-change-8 accessed 25 October 2023.

[6] Andrew Evans & Anja Heimann, 'AI Activity in UK businesses' (Capital Economics, Department for Digital, Culture, Media, and Sport, 2022) https://assets.publishing.service.gov.uk/media/61d87355e90e07037668e1bd/AI_Activity_in_UK_Businesses_Report__Capital_Economics_and_DCMS__January_2022__Web_accessible_.pdf accessed 25 October 2023.

[7] *Ada Lovelace Institute, Foundation models in the public sector* (2023) https://www.adalovelaceinstitute.org/evidence-review/foundation-models-public-sector/ accessed 25 October 2023; Ada Lovelace Institute & Alan Turing Institute, How do people feel about AI? (2023) https://www.adalovelaceinstitute.org/report/public-attitudes-ai/ accessed 25 October 2023.

which enable businesses to integrate them into their own services; in other cases models are open sourced online.[8]

As major technology companies offer foundation models as a service, this means that they may be increasingly integrated into products, services and organisational workflows. Alongside business use, there is already considerable evidence of piloting within the public sector, where potential uses include document analysis, decision support, policy drafting and public knowledge access.[9] An example of this is the recently reported testing of a gov.uk chatbot that would guide users in navigating and accessing public services like receiving benefits and paying tax.[10]

Existing foundation models such as GPT-4 are already powerful, although *how* powerful is contested, and developer claims of their performance have been criticised as misleading.[11] The rapid increase in availability and uptake of foundation-model-based systems means we are yet to understand the full extent of their impact on society and the economy.

It is unclear whether today's enthusiasm for AI marks the beginning of an epochal technological transition, or simply the peak of a hype cycle that has yet to reach a trough of disillusionment. There have been AI winters in the past[12] and it is not certain that current methods of AI development, premised on ever-increasing demands for data and compute, will continue to yield performance improvements at the current rate. And, importantly, many AI systems are still exhibiting impressive capabilities in controlled settings while behaving erratically or simply not working at all in applied contexts.[13]

It is also unlikely that the current trajectory of AI development and deployment is sustainable in either economic or environmental terms. Some estimates suggest that combined compute costs for AI development would likely outstrip the entire GDP of the United States by 2037, if they were to continue without improvements to efficiency.[14] The environmental consequences of advanced AI development and deployment are similarly severe, with substantial costs in terms of greenhouse gas emissions, water and land use, and rare minerals.[15] Yet accounts of AI's potential uses – and of 'AI safety' – often overlook these 'hidden' costs.

[8] Ada Lovelace Institute, *Explainer: What is a foundation model?* (2023) https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/ accessed 25 October 2023.

[9] Ada Lovelace Institute (n 7).

[10] Sunak to launch AI chatbot to help Britons with taxes and pensions (telegraph.co.uk)

[11] See, for example: OpenAI (2023) GPT-4 technical report. arXiv:2303.08774, although it is worth noting that these claims have been queried, e.g. https://www.aisnakeoil.com/p/gpt-4-and-professional-benchmarks

[12] Jim Howe, 'Artificial Intelligence at Edinburgh University: A Perspective' (www.inf.ed.ac.uk, November 1994). https://www.inf.ed.ac.uk/about/AIhistory.html accessed 25 October 2023.

[13] Inioluwa Deborah Raji and others, 'The Fallacy of AI Functionality', *2022 ACM Conference on Fairness, Accountability, and Transparency* (2022) http://arxiv.org/abs/2206.09511 accessed 30 October 2023

[14] Lennart Heim, 'This Can't Go On(?) – AI Training Compute Costs' (heim.xyz, 1 June 2023), https://blog.heim.xyz/this-cant-go-on-compute-training-costs accessed 25 October 2023. AI Now Institute 'Computational Power and AI' (2023) https://ainowinstitute.org/publication/policy/compute-and-ai accessed 25 October 2023.

[15] Emily Clough (n 5).

Foundation models are only one pathway through which AI may become more integrated and infrastructural. A challenge for Government – particularly where the stakes are high – is that many of the risks of AI are not related to the 'capabilities' of a base model but rather the functionality of a specific AI system in an applied context.[16] There is already considerable evidence of AI use and piloting within the UK public sector, with recent reports indicating government applications across benefit decisions, Home Office risk assessments[17] and retrospective facial recognition in policing.[18]

It is therefore important to engage critically with claims from industry and governments about the future centrality of AI technologies to national and global economies, and assess them against the best available evidence. Nonetheless, such claims should be taken seriously. It is plausible – if far from guaranteed – that AI models will soon occupy a fundamental role in the operation of critical products and services, both public and private, from healthcare to the provision of benefits. This, in turn, would mean AI effectively serving as a critical product and service for societies and economies at large.

This raises the important question of the current domination of the AI market by a small number of companies. Only a handful of players are truly competitive at the leading edge of AI development, as well as in key markets that supply important AI inputs such as compute and data.[19] [20] A potential consequence of this is that a small number of companies may end up steering the trajectory and security of what could in time become central digital infrastructure. We share concerns from regulators and civil society[21] about the oversight and market power implications this would have both in the digital economy and more broadly, and have previously made recommendations on how this could be rebalanced.[22]

It will also have important implications for the 'safety' conversation that unfolds at Bletchley Park. Governments often take a close interest in infrastructure as part of their responsibilities towards citizens. Some types of infrastructure are run directly by the public sector, and many are highly regulated to manage potential harms and ensure public benefit. If the Summit's premise of AI's transformational potential is to be taken at face value, we can learn from the governance of similarly consequential systems.

[16] Seth Lazar and Alondra Nelson (n 1).

[17] Stacey K, 'UK Risks Scandal over "Bias" in AI Tools in Use across Public Sector' *The Guardian* (23 October 2023) https://www.theguardian.com/technology/2023/oct/23/uk-risks-scandal-over-bias-in-ai-tools-in-use-across-public-sector

[18] 'Letter to Police on AI Enabled Facial Recognition Searches' (GOV.UK) https://www.gov.uk/government/news/letter-to-police-on-ai-enabled-facial-recognition-searches

[19] Competition and Markets Authority, 'AI Foundation Models: Initial Review' (GOV.UK, 4 May 2023) https://www.gov.uk/cma-cases/ai-foundation-models-initial-review accessed 25 October 2023.

[20] AI Now Institute (n 14).

[21] Competition and Markets Authority (n 19); 'Artificial Intelligence for Public Value Creation: Introducing Three Policy Pillars for the UK AI Summit' (IPPR, 25 October 2023) https://www.ippr.org/research/publications/ai-for-public-value-creation accessed 30 October 2023; AI Now Institute, '2023 Landscape: Confronting Tech Power' (2023) https://ainowinstitute.org/2023-landscape

[22] Ada Lovelace Institute, *Rethinking data and rebalancing digital power* (2022) https://www.adalovelaceinstitute.org/report/rethinking-data/

## What are the goals of regulation in sectors that play a similarly critical role?

The Government's agenda for the Summit is focused primarily on technical methods for avoiding hypothetical 'extreme risks' that could emerge from the misuse or loss of control of advanced 'frontier' AI systems. This focus has been queried by many within industry, academia and civil society as overly narrow, and insufficiently attentive to the most pressing AI harms.[23]

Addressing the challenges of AI is not the first time that regulators have grappled with governing highly complex technologies that play a central societal and economic role. As we stated in *Regulate to innovate*,[24] there is no perfect analogue for AI, but looking at how 'safety' is employed in other domains – such as medicines, transport and food – can help to inform strategies for AI regulation.

These foundational sectors act as critical platforms that other essential parts of our economies and societies depend on – in other words, infrastructure. Their governance is therefore designed to ensure that these systems and technologies are trustworthy: that they function as intended, that the risks they pose are managed holistically, and that the goods and benefits they produce are widely available at a reasonable cost.

Examples of this include:

- **Medicines and medical devices:** The UK's Medicines and Healthcare products Regulatory Agency's (MHRA) goal is to 'ensure the safety and effectiveness of medical products within the UK […] we aim to safeguard patient well-being and maintain public trust in the healthcare sector'. It has a wide range of responsibilities, from upholding 'applicable standards of safety, quality and efficacy' to securing supply chains and enabling beneficial research and development.[25]

- **Food:** The UK's food system is overseen by the Food Standards Agency, whose mission is 'Food you can trust', which it delivers by 'safeguard[ing] public health and protect[ing] the interests of consumers'.[26]

- **Financial services:** The Financial Conduct Authority (FCA) protects customers, promotes 'healthy competition' between providers, and works to maintain the stability of the system as a whole.[27]

[23] Seth Lazar and Alondra Nelson (n 1).

[24] Ada Lovelace Institute (n 2).

[25] 'MHRA Annual Report and Accounts' (Medicines & Healthcare products Regulatory Agency, GOV.UK 2023) https://assets.publishing.service.gov.uk/media/65378cc7e839fd000d867417/230809_MHRA_Annual_Report_2023_ACCESSIBLE_CC_Normal_print_231023.pdf accessed 25 October 2023.

[26] 'Food Standards Agency' (GOV.UK) https://www.food.gov.uk/ accessed 25 October 2023.

[27] 'About the FCA' (FCA.ORG.UK 2023), https://www.fca.org.uk/about/what-we-do/the-fca accessed 25 October 2023.

- **Transport:** in the UK various regulators are tasked with ensuring the safety of different modes of transport, including aviation, rail and roads.

  — **Aviation:** The Civil Aviation Authority (CAA) makes sure that 'the aviation and aerospace industry meet the highest safety standards', 'consumers have choice, value for money, are protected and treated fairly when they fly', 'the environmental impact of aviation is effectively managed', and 'the aviation industry manages security risks effectively'.[28] The CAA states that cooperation between nations is 'vital, as is trust and compliance with international standards of safety, helping each nation to respect the safety levels and oversight of others'.[29]

  — **Road and rail:** The Office of Rail and Road (ORR) describes its core purpose as 'protect[ing] the interests of rail and road users, improving safety, value and performance of railways and roads, today and in the future'.[30] The ORR does so by regulating health and safety standards across the whole rail industry, and holding National Highways to account on its commitments to improving the performance of England's strategic road network.

- **Energy:** There are two major regulators in the UK energy sector.

  — **Gas and electricity:** Ofgem, the UK's energy regulator, works 'to protect energy consumers, especially vulnerable people, by ensuring they are treated fairly and benefit from a cleaner, greener environment'.[31] Ofgem's primary responsibility is to protect the interests of consumers, especially those who are vulnerable, but its objectives also include 'to deliver a net zero economy' and 'enabling competition and innovation'.[32]

  — **Nuclear energy:** The Office for Nuclear Regulation's (ONR) mission is to protect society by securing safe nuclear options. The ONR has legal authority to regulate nuclear safety, nuclear security and conventional health and safety at nuclear sites. A key theme in the ONR's strategy is 'inspiring public confidence in its regulation of the nuclear industry'.[33]

[28] 'Annual Report & Accounts 2022/2023' (Civil Aviation Authority 2023) https://www.caa.co.uk/media/clviohxz/annualreport2022-23.pdf accessed 24 October 2023

[29] Ibid.

[30] 'Business plan 2023-24' (Office of Rail and Road 2023). https://www.orr.gov.uk/sites/default/files/2023-04/ORR-business-plan-summary-2023-2024.pdf accessed 24 October 2023.

[31] 'Welcome to Ofgem' (Ofgem 2023) https://www.ofgem.gov.uk/ accessed 24 October 2023.

[32] 'Ofgem Forward Work Programme' (Ofgem 2022) https://www.ofgem.gov.uk/publications/202223-ofgem-forward-work-programme accessed 24 October 2023

[33] 'Strategy 2020-25' (The Office for Nuclear Regulation 2020) https://www.onr.org.uk/documents/2020/onr-strategy-2020-2025.pdf accessed 24 October 2023.

- **Communications:** Ofcom is the UK's regulator for communication services, which includes broadband, home phone and mobile services, TV and radio, airwaves and the universal postal service. It ensures, among other responsibilities, that people can access communications services, that viewers and listeners are protected from harmful or offensive material and that people are protected from unfair treatment and do not have their privacy invaded.[34] Ofcom works to provide 'media we can trust and value [...] and accurate and impartial news that we can trust'.[35]

A recurring theme across each of these regulatory regimes is the goal of ensuring systems are not only 'safe', but that they enjoy public trust. This necessarily entails proving that those systems are *trustworthy*, which means different things in different sectors, but ultimately means that they can be relied on by the people that interact with and depend on them. Evidencing trustworthiness can require the use of a wide range of mechanisms to ensure greater transparency, robust routes to accountability and redress for when things go wrong, and that the risk of harms occurring is proportionate to benefits.

## How are risks and harms managed in other sectors?

There are many typologies of AI risks and harms[36] (this is an active area of research for the Ada Lovelace Institute), but one way of thinking is to group them into four broad categories:

- **supply chain harms** from the processes and inputs used to develop AI, such as poor labour practices, environmental impacts and the inappropriate use of personal data or protected intellectual property

- **accidental harms** from AI systems failing or acting in unanticipated ways, such as self-driving car crashes, or discrimination when sifting job applications

- **misuse harms** from AI systems being used in malicious ways, such as bad actors generating misinformation using 'generative' AI applications such as ChatGPT and Midjourney

- **structural or systemic harms** from AI systems altering social, political and economic systems, such as the creation of unequal power dynamics (for example through market concentration or inequitable access to AI systems), or the aggregate effect of misinformation on democratic institutions.

[34] 'What is Ofcom?' (Ofcom 2023) https://www.ofcom.org.uk/about-ofcom/what-is-ofcom accessed 24 October 2023.

[35] 'The Office of Communications Annual Report and Accounts 2022-2023' (Ofcom 2023). https://www.ofcom.org.uk/__data/assets/pdf_file/0022/264136/22-23-annual-report.pdf accessed 24 October 2023.

[36] See, for example: Laura Weidinger and others, 'Taxonomy of Risks Posed by Language Models', *2022 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2022) https://doi.org/10.1145/3531146.3533088 accessed 30 January 2023; Renee Shelby and others, 'Identifying Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction' (arXiv, 8 February 2023) http://arxiv.org/abs/2210.05791 accessed 27 March 2023.

In some cases, these harms are already well-evidenced – such as the tendency of certain AI systems to reproduce harmful biases – but in others they can refer to harms from technologies that do not yet exist, and are therefore difficult or impossible to prove. These include, for example, the potential for mass unemployment resulting from AI-enabled job automation or augmentation, or claims that powerful AI systems may pose extreme or 'existential' risks to future human society.

All these types of harms could reasonably be considered in scope for the goal of achieving 'AI safety', and the extent of their prevention and mitigation would all have a significant impact on the trustworthiness of the systems involved. In other sectors, a variety of interlocking mechanisms are used to address harms not dissimilar to those that AI produces:

| Type of harm | Description | Example from non-AI sector | Mitigation |
|---|---|---|---|
| **Supply chain** | When the processes or inputs used to develop a product or service produce negative consequences. | Risks to health and safety of personnel operating at nuclear sites. | Strict health and safety requirements to protect staff, including regulations that restrict the exposure of workers to radiation. |
| **Accidental** | When a product or service fails, or acts in unintended ways. | Pharmaceuticals that are ineffective for their intended purpose, or otherwise harmful to health. | Requirements on developers of drugs or medical devices to provide (sufficiently positive) evidence on the safety risks, efficacy, and accessibility of those products before they are approved to be sold in a market or continue to the next development phase (referred to as pre-market approval or pre-approval). |
| **Misuse** | When a product or service is used in unintended or malicious ways. | Use of financial services for money laundering. | 'Know Your Customer' regulations, which aim to prevent illicit activities through the monitoring of transactions and the imposition of reporting requirements, with heavy fines or sanctions levied against institutions that fail to comply with regulatory standards. |
| **Systemic or structural** | When products and services are operating as intended, as is the system supplying them, but negative societal or economic consequences are still produced. | Electricity and gas supply creates greenhouse gas emissions and causes negative environmental impacts. | As part of its statutory duty to contribute to reaching the 2050 Net Zero goal for greenhouse gas emissions, Ofgem has introduced renewable electricity schemes, renewable heat schemes and energy efficiency schemes to reduce greenhouse emissions in gas and electricity supply. |

The table above is not exhaustive – and should not be seen as an endorsement of particular measures, or as a judgement of their effectiveness. It is intended to illustrate that ample mechanisms exist in other sectors for addressing many of the harms associated with AI. While there are novel problems associated with regulating AI (some of which we have highlighted in our previous research),[37] it is not the case that Government or regulators are starting from a position of limited or no knowledge. Claims that regulation needs to wait because of the novelty and complexity of AI ignore the ways that we already govern to manage uncertainty and risk of existing technologies.

## What are the key features of regulatory systems in sectors that provide important societal and economic infrastructure?

Enforcing the sorts of mechanisms referenced above requires empowered, well-resourced regulatory institutions with access to the appropriate information. We are currently investigating the factors determining the effectiveness of regulatory systems in other sectors, and our early research has identified common themes.

### Powers

All regulatory domains described in this briefing are supported by statute. None of them contain self-regulatory approaches – in which companies voluntarily subscribe to agreed standards – considered sufficient for the management of harms and benefits for such societally critical technologies and industries. While it is welcome that the UK Government has published a set of safety practices for foundation model developers, these should not be seen as a substitute for hard regulation.

Legislation provides for the institutional legitimacy and scope of regulators in other sectors, as well as granting them powers to enforce rules and shape the behaviour of relevant actors. These vary greatly in the UK: some regulators have broad powers to request information from key actors (such as large companies and industry bodies) and to impose pre-release requirements, while others are highly proscribed in what action they can take. *Our Regulating AI in the UK* report highlights significant gaps in the UK's current proposals for governing AI through existing regulators. Many of these gaps relate to the capacity of regulatory actors to 'reach' different contexts in which AI will be used, and to shape the practices of actors further up the value chain, such as developers and hosts of AI systems, particularly for foundation models.

[37] Ada Lovelace Institute (n 2); Ada Lovelace Institute, *Regulating AI in the UK* (2023) https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/

Among other functions, legislation in the UK could level this playing field by supporting all regulators with a common set of powers to address harms from AI. This could include *ex ante* powers to place conditions and penalties on AI developers, as well as greater powers to request information on people, policies, practices, data and models from companies developing, deploying or using AI systems, and to compel those organisations to make that information available more widely when appropriate. It could also include requiring companies to undertake *ex-ante* independent audits and evaluations of an AI system's performance and potential impacts before a system is deployed, and in some cases require post-deployment monitoring obligations to evaluate a system's risks after it is released. It additionally includes the establishment of legal liability throughout an AI supply chain to create clear modes of redress for when things go wrong,[38] and to disincentivise bad behaviour.

## Resources

Given that some forms of AI are increasingly being used in critical products and services that could lead to them being treated as infrastructure, the scale of support required for governing this general-purpose technology is likely to be in the same order of magnitude as that of regulators responsible for key parts of our economic, social and technological infrastructure. Looking at the revenue, expenditure and staffing of such regulators gives an indication of volume of resource that will be needed to govern AI effectively.

[38] Ian Brown, *Allocating accountability in AI supply chains: a UK-centred regulatory perspective* (Ada Lovelace Institute 2023) https://www.adalovelaceinstitute.org/resource/ai-supply-chains/

| Regulator | Approximate annual revenue/ expenditure (2022) | Approximate number of full-time equivalent employees (2022) |
|---|---|---|
| Civil Aviation Authority | £140m[39] | 1,270[40] |
| Food Standards Agency | £100m[41] | 2,945[42] |
| Medicines and Healthcare products Regulatory Agency | £120m[43] | 1,177[44] |
| Office for Nuclear Regulation | £90m[45] | 671[46] |
| Office of Road and Rail | £36m[47] | 336[48] |
| Ofgem | £142m[49] | 1,187[50] |
| Ofcom | £154m[51] | 1,102[52] |

The central functions announced in the UK's AI Regulation white paper are not yet costed, but in terms of resource committed to date by the UK Government, the above figures can be compared with £100m currently committed over 18 months for the Frontier AI Taskforce — a one-off lump sum rather than ongoing annual expenditure.

[39] Civil Aviation Authority (n 28).

[40] Ibid.

[41] 'Accounts' (Food Standards Agency 2023) https://www.food.gov.uk/about-us/accounts accessed 25 October 2023.

[42] 'Staff Report' (Food Standards Agency 2023) https://www.food.gov.uk/about-us/staff-report accessed 25 October 2023.

[43] Medicines & Healthcare products Regulatory Agency (n 25).

[44] Ibid.

[45] 'Annual Reports and Accounts 2021/2022' (Office for Nuclear Regulation 2022) https://www.onr.org.uk/documents/2022/onr-annual-report-and-accounts-2021-22.pdf accessed 25 October 2023.

[46] Ibid.

[47] 'Annual report and accounts 2022 to 2023: Performance report – Performance overview' (Office of Road and Rail 2023) https://www.orr.gov.uk/orr-annual-reports-and-accounts/2022-2023/performance-overview accessed 25 October 2023.

[48] 'ORR organogram and datasets' (Office of Road and Rail 2023) https://www.orr.gov.uk/about/corporate-data/orr-organogram-and-data-sets#:~:text=Our%20budgeted%20staff%20job%20complement,percentage%20of%20number%20of%20posts accessed 25 October 2023.

[49] 'Ofgem Annual Report and Accounts 2021-22' (Ofgem 2022) https://www.ofgem.gov.uk/publications/ofgem-annual-report-and-accounts-2021-22 accessed 25 October 2023.

[50] Ibid.

[51] 'Ofcom Annual Report and Accounts 2021/22' (Ofcom 2022) https://www.ofcom.org.uk/__data/assets/pdf_file/0022/240727/annual-report-2021-22.pdf accessed 25 October 2023.

[52] Ibid.

## Information

A necessary feature of effective governance regimes in other sectors is monitoring and horizon-scanning capabilities. Governments and regulators need to understand what is happening in critical sectors that they are responsible for, and what harms are likely to emerge over time.

At present, the Government is largely reliant on external expertise from industry for these insights when it comes to AI. 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.

The Government's decision to invest in better understanding AI risks is therefore a welcome one – although it should not be a cause to delay urgent and necessary action on harms that are already well understood, including by delivering on and improving their white paper proposals.

In other sectors there are strong examples for providing world-class evidence and advisory capacity to Government on technical issues. One example is the Committee on Climate Change (CCC), formed under the Climate Change Act (2008) to advise the United Kingdom and devolved governments and parliaments on tackling and preparing for climate change.[53]

The CCC has certain features that could make it an attractive model for a body conducting horizon-scanning on AI opportunities and risks. It is independent from Government and accountable to Parliament, insulating it from political churn and enabling it to take a longer-term perspective.[54]

It has a broad yet clear remit ranging across both climate change adaptation and mitigation, and the resources to collate expertise from a variety of sectors and academic disciplines.[55]

In this regard an independent national-level entity like the CCC may prove a better model than the Intergovernmental Panel on Climate Change, to which a similar body has been proposed for providing oversight and governance of emergent AI challenges. The IPCC it is not politically independent: its summary reports are subject to line-by-line approval by governments, and so while highly respected, it is generally understood to be the most cautious and conservative about climate change. The history of the IPCC is also instructive for this moment in AI – despite decades of evidence of human-caused climate change, national governments chose to problematise the issue as a research question rather than regulate.[56] Governments must be careful to avoid the same mistake for AI.

[53] 'CCC assessment of recent announcements and developments on Net Zero' (Climate Change Committee) https://www.theccc.org.uk/ accessed 25 October 2023.

[54] Ibid.

[55] Ibid.

[56] Rich N, 'Losing Earth: The Decade We Almost Stopped Climate Change' *The New York Times* (1 August 2018) https://www.nytimes.com/interactive/2018/08/01/magazine/climate-change-losing-earth.html

It's important to recognise, too, that the IPCC forms only a part of the overall framework of climate change governance. The United Nations (UN) created the IPCC in 1988, after nearly 20 years of international scientific conversation around climate change. The IPCC and CCC both rely on interdisciplinary scientific input, but their position is founded on a clear consensus around the dangers that climate change poses. This consensus does not exist in the case of AI, so the creation of a new body should not be thought of as a panacea, and will need to incorporate a wide range of perspectives, including those of people affected by uses of these technologies.

These comparisons provide important lessons for any similar initiative on researching AI safety. It is imperative that any new body is politically independent and can take a long-term view without interference from the government of the day. It will need to take a broad approach that incorporates input from different disciplines – not only technical specialists but social science and humanities scholars – and public perspectives. And it should not be a blocker or substitute for robust regulation: rather, like the CCC and IPCC, it must be able to advise on and evaluate policies as they are implemented, helping to improve them over time.

## Conclusion and next steps

The AI Safety Summit is premised on the notion that AI is likely to have a transformative impact and ultimately assume a critical role in our economies and societies. If this is the case, then the conditions for success will be a governance system as comprehensive and robust as those that exist for other technologies and sectors that already play an equivalent infrastructural role. In the absence of such governance, AI harms will continue to be realised, gambling the potential loss of public trust in one of the most promising technologies we have, and the benefits of its use along with it.

Regulatory systems are not created overnight, and it would be unrealistic to expect the AI Safety Summit to address every angle of the AI governance question. But it would be equally unrealistic to suggest that a narrow conversation centred solely on technical mitigations for misuse and loss-of-control risks from a subset of AI systems is adequate to the challenges of the present moment. This week is an important opportunity to reassure people that, as AI assumes a greater role in their lives, it will be governed in the ways that we expect from other infrastructural technologies, and to set out a roadmap to that governance.

Over the coming months, the Ada Lovelace Institute will be diving deeper into these comparisons and evaluating what we can learn from other regulatory frameworks. Regulation in other sectors can inspire AI regulation, either by providing examples of best practices or examples of regulatory failures that should be avoided in the future.

We'll be exploring in more depth the questions raised in this briefing, such as:

- the types of objectives and public benefits that these regulatory regimes aim to achieve
- the mechanisms put in place at different stages of the value chain to achieve these objectives and benefits
- the distribution of liability and compliance burdens across value chains in these regulatory environments
- the impacts of regulation on innovation and market size.

If you would like more information on this policy briefing, or if you would like to discuss our research in this area, please contact our policy research team at hello@adalovelaceinstitute.org.

## Ada publications relating to AI safety

*Regulating AI in the UK:* Our work contextualising and summarising the UK's current plans for AI regulation, along with recommendations to help strengthen the proposed regulatory framework.

Policy briefing https://www.adalovelaceinstitute.org/policy-briefing/regulating-ai-in-the-uk/

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

*What do the public think about AI? Understanding public attitudes and how to involve the public in decision-making about AI* https://www.adalovelaceinstitute.org/evidence-review/what-do-the-public-think-about-ai/

'An EU AI Act that works for people and society' (policy briefing) https://www.adalovelaceinstitute.org/policy-briefing/eu-ai-act-trilogues/

**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**