

# Regulate to innovate

A route to regulation that reflects  
the ambition of the UK AI Strategy

November 2021



# Contents

- 3 Executive summary
- 9 How to read this report
- 11 Introduction
- 23 **Chapter 1**  
Goals of AI regulation
- 36 **Chapter 2**  
Challenges for regulating AI systems
- 46 **Chapter 3**  
Tools, mechanisms and approaches for regulating AI systems
- 63 **Chapter 4**  
Considerations for policymakers
- 70 **Chapter 5**  
Open questions for Government
- 74 **Chapter 6**  
Recommendations for the Government's White Paper  
on AI regulation
- 81 Annex: The anatomy of regulatory rules and systems,  
and how these apply to AI
- 90 Bibliography
- 101 Acknowledgements
- 102 About the Ada Lovelace Institute

---

# Executive summary

In its 2021 National AI Strategy, the UK Government laid out its ambition to make the UK an ‘AI superpower’, bringing economic and societal benefits through innovation. Realising this goal has the potential to transform the UK’s society and economy over the coming decades, and promises significant economic and societal benefits. But the rapid development and proliferation of AI systems also poses significant risks.

As with other disruptive and emerging technologies,<sup>1</sup> creating a successful, safe and innovative AI-enabled economy will be dependent on the UK Government’s ability to establish the right approach to governing and regulating AI systems. And as the UK AI Council’s Roadmap, published in January 2021, states, ‘the UK will only feel the full benefits of AI if all parts of society have full confidence in the science and the technologies, and in the governance and regulation that enable them.’<sup>2</sup>

The UK is well placed to develop the right regulatory conditions for AI to flourish, and to balance the economic and societal opportunities with associated risks,<sup>3</sup> but urgently needs to set out its approach to this vital, complex task.

---

1 Mazzucato, M. (2015). 'From Market Fixing to Market-Creating: A New Framework for Economic Policy', *SSRN Electronic Journal*. Available at: <https://doi.org/10.2139/ssrn.2744593>.

2 AI Council. (2021). *AI Roadmap*. UK Government. January 2021. Available at: [www.gov.uk/government/publications/ai-roadmap](https://www.gov.uk/government/publications/ai-roadmap) [accessed 11 October 2021].

3 Office for AI. (2021). *National AI Strategy*. UK Government. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1020402/National\\_AI\\_Strategy\\_-\\_PDF\\_version.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf)

However, articulating the right governance and regulatory environment for AI will not be easy.

By virtue of their ability to develop and operate independently of human control, and to make decisions with moral and legal consequences, AI systems present a uniform set of general regulatory and legal challenges concerning agency, causation, accountability and control. At the same time, the specific regulatory questions posed by AI systems vary considerably across the different domains and industries in which they might be deployed.

Regulators must therefore be able to find ways of accounting consistently for the general properties of AI while also attending to the peculiarities of individual use cases and business models.

While other states and economic blocs are already in the process of engaging with tough but unavoidable regulatory challenges through new draft legislation, the UK has still to commit to its regulatory approach to AI.

In September 2021, the Office for AI pledged to set out the Government's position on AI regulation in a White Paper, to be published in early 2022.

Over the course of 2021, the Ada Lovelace Institute convened a cross-disciplinary panel of experts to explore approaches to AI regulation, and inform the development of the Government's position.

Based on this, and Ada's own research, this report sets out how the UK might develop its approach to AI regulation in line with its ambition for innovation.

In this report we:

1. explore some of the aims and objectives of AI regulation that might have been considered alongside economic growth
2. outline some of the challenges associated with regulating AI
3. review the regulatory toolkit, and options for rules and system design, which address technologies, markets and use-specific issues
4. identify and evaluate some of the different tools and approaches that might be used to overcome the challenges of AI regulation
5. assess the institutional and legal conditions required for the effective regulation of AI
6. raise outstanding questions that the UK Government will have to answer in setting out and realising its approach to AI regulation.

The report also identifies a series of conclusions for policymakers, as well as specific recommendations for the Office for AI's White Paper on the regulation and governance of AI. To present a viable roadmap for the UK's regulatory ecosystem, the White Paper will need to make clear commitments in three important areas:

1. The development of new, clear regulations for AI.
2. Improved regulatory capacity and coordination.
3. Improved transparency standards and accountability mechanisms.

## The development of new, clear regulations for AI

We make the case for the UK Government to:

1. develop a clear description of AI systems that reflects its overall approach to AI regulation, and criteria for regulatory intervention
2. create a central function to oversee the development and implementation of AI-specific, domain-neutral statutory rules for AI systems that are rooted in legal and ethical principles
3. require individual regulators to develop sector-specific codes of practice for the regulation of AI.

## Improved regulatory capacity and coordination

We argue that there is a need for:

4. expanded funding for regulators to help them deal with analytical and enforcement challenges posed by AI systems
5. expanded funding and support for regulatory experimentation and the development of anticipatory and participatory capacity within individual regulators
6. the development of formal structures for capacity sharing, coordination and intelligence sharing between regulators dealing with AI systems
7. consideration of what additional powers regulators may need to enable them to make use of a greater variety of regulatory mechanisms.

## **Improving transparency standards and accountability mechanisms**

The impacts of AI systems may not always be visible to, or controllable by, policymakers and regulators alone. As such, regulation and regulatory intelligence gathering will have to be complemented by, and coordinated with extra-regulatory mechanisms such as standards, investigative journalism and activism. We argue that the UK Government should consider:

- 8.** using the UK's influence over international standards to improve the transparency and auditability of AI systems
- 9.** how best to maintain and strengthen laws and mechanisms to protect and enable journalists, academics, civil-society organisations, whistleblowers and citizen auditors to hold developers and deployers of AI systems to account.

Overall, this report finds that, far from being an impediment to innovation, effective, future-proof regulation will provide companies and developers with the space to experiment and take risks without being hampered by concerns about legal, reputational or ethical exposure.

Regulation is also necessary to give the public the confidence to embrace AI technologies, and to ensure continued access to foreign markets.

The report also highlights how regulation is an indispensable tool, alongside robust industry codes of practice and judicious public-funding and procurement decisions, to help navigate the narrow path between the risks and harms these technologies present.

We propose that the clear, unambiguous rules that regulation can provide are necessary if the UK is to embrace AI on terms that will be beneficial in the long term.

To support this approach, we should resist the characterisation that regulation is the enemy of innovation: modern, relevant, effective regulation will be the brakes that allow us to drive the UK's AI vehicle successfully and safely into new and beneficial territories.

Finally, this research outlines the major questions and challenges that will need to be addressed in order to develop effective and proportionate AI regulation. In addition to supporting the UK Government's thinking on how to become an 'AI superpower' in a manner that manages risk and results in broadly felt public benefit, we hope this report will contribute to live debates on AI regulation in Europe and the rest of the world.

# How to read this report

This report is principally aimed at influencing the emerging policy discourse around the regulation of AI in the UK, and around the world.

- In the **introduction** we argue that regulation represents the missing link in the UK's overall AI strategy, and that addressing this gap will be critical to the UK's plans to become an AI superpower.
- **Chapter 1** sets out the aims and objectives UK AI regulation should pursue, in addition to economic growth.
- **Chapter 2** reviews the generic regulatory toolkit, and sets out the different ways that regulatory rules and systems can be conceived and configured to deal with different kinds of problems, technologies and markets.
- **Chapters 3 and 4** review some of the specific challenges associated with regulating AI systems, and set out some of the tools and approaches that have the potential to help overcome or ameliorate these difficulties.
- **Chapter 5** articulates some general lessons for policymakers considering how to regulate AI in a UK context.
- **Chapter 6** sets out some specific recommendations for the Office for AI's forthcoming White Paper on the regulation and governance of AI.

## If you're a UK policymaker thinking about how to regulate AI systems

We encourage you to read the recommendations at the end of this document ([p. 74](#)), which set out some of the key pieces of guidance we hope the Office for AI will incorporate in their forthcoming White Paper.

## If you're from a regulatory body

Explore the mechanisms and approaches to regulating AI, set out in [chapter 3](#), which may provide some ideas for how your organisation can hold these systems more accountable.

## If you're a policymaker from outside of the UK

Many of the considerations articulated in this report are, despite the UK framing, applicable to other national contexts. The considerations for regulating AI that are set out in [chapters 1, 2 and 3](#) are universally applicable.

## If you're a developer of AI systems, or an AI academic

The [introduction](#) and the lessons for policymakers section ([p. 63](#)) set out why the UK needs to take a new approach to the regulation of AI.

**A note on terminology:** Throughout this report, we use 'regulation' to refer to the codified 'hard' rules and directives established by governments to control and govern a particular domain or technology. By contrast, we use the term 'governance' to refer to non-regulatory means by which a domain or technology might be controlled or influenced, such as norms, conventions, codes of practice and other 'soft' interventions.

The terms *ex ante* (before the event) and *ex post* (after the event) are used throughout this document. Here, '*ex ante*' regulation typically refers to regulatory mechanisms intended to prevent or ameliorate future harms, whereas '*ex post*' refers to mechanisms intended to remedy harms after the fact, or to provide redress.

---

This report is intended to support policymakers – and particularly the Office for AI – to develop a strategy on governing and regulating AI

# Introduction

In its 2021 National AI Strategy, the UK Government outlines three core pillars for setting the country on a path towards becoming a global AI and science superpower. These are:<sup>4</sup>

1. investing in the long-term needs of the AI ecosystem
2. supporting the transition to an AI-enabled economy
3. ensuring the UK gets the national and international governance of AI technologies right to encourage innovation, investment and protect the public and ‘fundamental values’.<sup>5</sup>

As part of its third pillar, the strategy states the Office for AI will set out a ‘national position on governing and regulating AI’ in a White Paper in early 2022. This report seeks to help the Office for AI develop this forthcoming strategy, setting out some of the key challenges associated with the regulation of AI, different options for approaching the task and a series of concrete recommendations for the UK Government.

The publication of the new AI strategy represents an important articulation of the UK’s ambitions to cultivate and utilise the power of AI. It provides welcome detail on the Government’s proposed approach to AI investment, and their plans to increase the use of AI systems throughout different parts of the economy. Whether the widespread adoption of AI systems *will* increase economic growth remains to be seen, but it is a belief that underpins this Government’s strategy, and this paper does not seek to explore that assumption.<sup>6</sup>

The strategy also highlights some areas that will require further policy thinking and development in the near future. The chapter ‘Governing

---

4 Office for AI. (2021). *National AI strategy*. UK Government. Available at: [www.gov.uk/government/publications/national-ai-strategy](http://www.gov.uk/government/publications/national-ai-strategy).

5 The strategy uses, but does not define a range of terms related to values, including ‘fundamental values’, ‘our ethical values’, ‘our democratic values’, ‘UK values’, ‘fundamental UK values’ and ‘open society values’. It also refers to ‘values such as fairness, openness, liberty, security, democracy, rule of law and respect for human rights’.

6 One challenge is whether increasing AI adoption may only serve to consolidate the power of a handful of US-based tech companies who use their resources to acquire AI-based start ups. A 2019 UK Government review of digital competition found that ‘over the last 10 years the 5 largest firms have made over 400 acquisitions globally. See Furman, J. (2019). *Unlocking digital competition, Report of the Digital Competition Expert Panel*. HM Treasury. Available at: [www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel](http://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel).

AI effectively’, notes some of the challenges associated with governing and regulating AI systems that are top of mind for this Government and surveys some of the different regulatory approaches that could be taken, but remains agnostic on which might work best for the UK.

Instead, it asks whether the UK’s current approach to AI regulation is adequate, and commits to set out ‘the Government’s position on the risks and harms posed by AI technologies and our proposal to address them’ in a White Paper in early 2022. In making a commitment to set out the UK’s ‘national position on governing and regulating AI’, the Government has set itself an ambitious timetable for articulating how it intends to address one of the most important gaps in current UK AI policy.

This report explores how the UK’s National AI Strategy might address the regulation and governance of AI systems. It is informed by the Ada Lovelace Institute’s own research and analysis into mechanisms for regulating AI, as well as two expert workshops that the Institute convened in April and May 2021. These convenings brought together academics, public and civil servants, regulators and representatives from civil society organisations to discuss:

1. How the UK’s regulatory and governance mechanisms may have to evolve and adapt into order to serve the needs and ambitions of the UK’s approach to AI.
2. How Government policy can support the UK’s regulatory and governance mechanisms to undergo these changes.

The Government is already in the process of drawing up and consulting on plans for the future of UK data regulation and governance, much of which relates to the use of data for AI systems.<sup>7</sup> While relevant to AI, data-protection law does not holistically address the kinds of risks and impacts AI systems may present – and is not enough on its own to provide AI developers, users and the public with the clarity and protection they need to integrate these technologies into society with confidence.

Where work to establish a supporting ecosystem for AI is already underway, the Government has so far focused primarily on developing

---

<sup>7</sup> Department for Digital, Culture Media & Sport. (2021). *Data: A new direction*. UK Government. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1016395/Data\\_Reform\\_Consultation\\_Document\\_\\_Accessible\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1016395/Data_Reform_Consultation_Document__Accessible_.pdf).

To move forward, the UK Government will have to answer fundamental questions on the regulation of AI systems in the forthcoming White Paper

and setting out AI-governance measures, such as the creation of bodies like the Centre for Data Ethics and Innovation (CDEI), with less attention and activity on specific approaches to the regulation of AI systems.<sup>8</sup>

To move forward, **the UK Government will have to answer fundamental questions on the regulation of AI systems in the forthcoming White Paper**, including:

- What should the goal of AI regulation be, and what kinds of regulatory tools and mechanisms can help achieve those objectives?
- Do AI systems require bespoke regulation, or can the regulation of these systems be wrapped into existing sector-specific regulations, or a broader regulatory package for digital technologies?
- Should regulating AI require the creation of a single AI regulator, or empower existing regulatory bodies with the capacity and resources to regulate these systems?
- What kinds of governance practices work for AI systems, and how can regulation incentivise and empower these kinds of practices?
- How can regulators best address some of the underlying root causes of the harms associated with AI systems?<sup>9</sup>

For the UK's AI industry it will be vital that the Government provides actionable answers to these questions. Creating a world-leading AI economy will require consistent and understandable rules, clear objectives and meaningful enforcement mechanisms.

Other world leaders in AI development are already establishing regulations around AI. In April 2021, the European Commission released a draft proposal for the regulation of AI (part of a suite of regulatory proposals for digital markets and services), which proposes a risk-based model for establishing certain requirements on the sale and deployment

8 The first UK AI strategy (called the UK AI Sector Deal), published in 2017 and updated in 2019, makes relatively little mention of the role of regulation and governance. In discussing how to build trust in the adoption of AI and address its challenges, the strategy is limited to calls for the creation of the Centre for Data Ethics and Innovation to 'ensure safe, ethical and ground-breaking innovation in AI and data-driven technologies.' Though the CDEI, since its inception, has produced various helpful pieces of evidence and guidance on ethical best practice around AI (such as a review into bias in algorithmic decision-making and an adoption guide for privacy-enhancing technologies), thinking on how regulation, specifically, might support the responsible development and use of AI remains less advanced.

9 Balayan, A., and Gürses, S., (2021). *Beyond Debiasing: Regulating AI and its inequalities*. European Digital Rights. Available at: <https://edri.org/our-work/if-ai-is-the-problem-is-debiasing-the-solution>.

of AI technologies.<sup>10</sup> While this draft is still subject to extensive review, it has the potential to set a new global standard for AI regulation that other countries are likely to follow.

In August 2021, the Cyberspace Administration of China passed a set of draft regulations for algorithmic systems,<sup>11</sup> which includes requirements and standards for the design, use and kinds of data that algorithmic systems can use.<sup>12</sup> The USA is taking a slower and more fragmented route to the regulation of AI, but is also heading towards establishing its own approach.<sup>13</sup>

Throughout 2021, the US Congress has introduced several pieces of federal AI governance and data-protection legislation, such as the Information Transparency and Personal Data Control Act, which would establish similar requirements to the EU GDPR.<sup>14</sup> In October 2021, the White House Office of Science and Technology Policy announced its intention to develop a ‘bill of rights’ to ‘clarify the rights and freedoms [that AI systems] should respect.’<sup>15</sup> Moreover, it is looking increasingly likely that geostrategic considerations will push the EU and the USA into closer regulatory proximity over the coming years, with EU President von der Leyen having recently pushed for the EU and the USA to start collaborating together on the promotion and governance of AI systems.<sup>16</sup>

- 
- 10 European Commission. (2021). *A Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. Available at: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52021PC0206> [accessed 4 October 2021].
- 11 Cyberspace Administration of China (国家互联网信息办公室). (2021). *Notice of the State Internet Information Office on the Regulations on the Management of Recommendations for Internet Information Service Algorithms (Draft for Solicitation of Comments)*. 27 August. Available at: [wwwcacgovcntranslategoog/20210827/c\\_1631652502874117.htm?\\_x\\_tr\\_sch=http&\\_x\\_tr\\_sl=zh-CN&\\_x\\_tr\\_tl=en&\\_x\\_tr\\_hl=en&\\_x\\_tr\\_pto=ajax,nv,elem](http://wwwcacgovcntranslategoog/20210827/c_1631652502874117htm?_x_tr_sch=http&_x_tr_sl=zh-CN&_x_tr_tl=en&_x_tr_hl=en&_x_tr_pto=ajax,nv,elem) [Accessed 16 September 2021].
- 12 For an interesting analysis, see Schaefer, K. (2021). 27 August. Available at: <https://twitter.com/kendraschaefer/status/1431134515242496002> [accessed 22 October 2021].
- 13 Since 2019, numerous government offices – including the White House’s Office of Science and Technology Policy, the National Institute of Standards and Technology, and the Department of Defence Innovation Board – have set out positions and principles for a national framework on AI.
- 14 US Congress. (2021). *H.R.1816 – Information Transparency & Personal Data Control Act*. Available at: [www.congress.gov/bill/117th-congress/house-bill/1816/text?q=%7B%22search%22%5B%22hr+1816%22%5D%7D&r=1&s=10](http://www.congress.gov/bill/117th-congress/house-bill/1816/text?q=%7B%22search%22%5B%22hr+1816%22%5D%7D&r=1&s=10).
- 15 Lander, E., and Nelson, A. (2021). ‘Americans need a bill of rights for an AI-powered world,’ *Wired*, 10 October. Available at: [www.wired.com/story/opinion-bill-of-rights-artificial-intelligence](http://www.wired.com/story/opinion-bill-of-rights-artificial-intelligence) [accessed 11 October 2021].
- 16 In a November 2020 speech at the Council on Foreign Relations. See Branson, A. (2020). ‘European Commission woos US over AI agreement.’ *Global Government Forum*. Available at: [www.globalgovernmentforum.com/european-commission-woos-us-over-ai-agreement](http://www.globalgovernmentforum.com/european-commission-woos-us-over-ai-agreement).

As a complex, novel technology, the benefits of AI are yet to be evenly distributed to all members of society, yet there is a growing body of evidence around the ways they can cause harm

**As the positions of the world's most powerful states and economic blocs on the regulation of AI become clearer, more developed and potentially more aligned, it will be increasingly incumbent on the UK to set out its own plans, or risk getting left behind.** Unless the UK carves out its own approach towards the regulation of AI, it risks playing catch-up with other nations, or having to default to approaches developed elsewhere that may not align with the Government's particular strategic objectives. Moreover, if domestically produced AI systems do not align with regulatory standards adopted by other major trade blocs, this could have significant implications for companies operating in the UK's domestic AI sector, who could find themselves excluded from non-UK markets.

As well as trade considerations, **a clear regulatory strategy for AI will be essential to the UK Government's stated ambitions to use AI to power economic growth, raise living standards and address pressing societal challenges like climate change.** As the UK has learned from a variety of different industries, from its enduringly strong life-sciences sector,<sup>17</sup> to recent successes in fintech,<sup>18</sup> a clear and robust regulatory framework is essential for the development and diffusion of new technologies and processes. A regulatory framework would ensure developers and deployers of AI systems know how to operate in accordance with the law and protect against the kinds of well-documented harms associated with these technologies,<sup>19</sup> which can undermine public confidence in their development and use.

### **The need for clear and comprehensive AI regulation is pressing.**

As a complex, novel technology, the benefits of AI are yet to be evenly distributed to all members of society, yet there is a growing body of evidence around the ways they can cause harm.<sup>20</sup> Across the world, AI systems are being increasingly used in high-stakes settings such as determining which job applicants are successful,<sup>21</sup> what public

17 Kent, C. (2019). 'UK Healthcare Industry Analysis 2019: Why Britain Is a World Leader'. *Pharmaceutical Technology*. Available at: <https://www.pharmaceutical-technology.com/sponsored/uk-healthcare-industry-analysis-2019/> [accessed 20 September 2021].

18 McLean, A., and Wood, I. (2015). 'Do Regulators Hold the Key to FinTech Success?', *Financier Worldwide* Available at: [www.financierworldwide.com/do-regulators-hold-the-key-to-fintech-success](http://www.financierworldwide.com/do-regulators-hold-the-key-to-fintech-success) [accessed 20 September 2021].

19 McGregor, S. (2020). *When AI Systems Fail: Introducing the AI Incident Database*. Partnership on AI. Available at: <https://partnershiponai.org/aiincidentdatabase>.

20 Pownall, C. (2021). *AI, algorithmic and automation incidents and controversies*. Available at: <https://charliepownall.com/ai-algorithmic-incident-controversy-database>.

21 Dattner, B., Chamorro-Premuzic, T., Buchband, R., and Schettler, L. (2019). 'The Legal and Ethical Implications of Using AI in Hiring', *Harvard Business Review*, 25 April 2019. Available at: <https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring> [accessed 20 September 2021].

---

AI systems present a uniform set of general regulatory and legal challenges

benefits residents are eligible to claim,<sup>22</sup> what kind of loan a prospective financial-services client can receive,<sup>23</sup> or what risk to society a person may potentially pose.<sup>24</sup> In many of these instances, AI systems have not yet been proven capable of addressing these kind of tasks fairly or accurately; in others, they have not been properly integrated into the complex social environments in which they have been deployed.

**But building such a regulatory framework for AI will not be easy.**

In virtue of their ability to develop and operate independently of human control, and to make decisions with moral and legal consequences, AI systems present a uniform set of regulatory and legal challenges concerning agency, causation, accountability and control.<sup>25</sup>

At the same time, the specific regulatory questions posed by AI systems vary considerably across the different domains and industries in which they might be deployed. Regulators must find ways of accounting consistently for the general properties of AI, while also attending to the peculiarities of individual use-cases and business models.

In these contexts, AI systems raise unprecedented legal and regulatory questions, such as their ability to automate morally significant decision-making processes in ways that can be difficult to predict, and their capacity to develop and operate independently of human control.

AI systems are also frequently complex and opaque, and often fail to fall neatly within the contours of existing regulatory systems – they either straddle regulatory remits, or else fall through the gaps in between them. And they are developed for a variety of purposes in different domains, where their impacts, benefits and risks may vary considerably.

These features can make it extremely difficult for existing regulatory bodies to understand if, how and in what manner to intervene.

---

22 Martinho-Truswell, E. (2018). 'How AI Could Help the Public Sector', *Harvard Business Review*, 26 January 2018. Available at: <https://hbr.org/2018/01/how-ai-could-help-the-public-sector> [accessed 20 September 2021].

23 Faggella, D., (2020). 'Artificial Intelligence Applications for Lending and Loan Management', *Emerj*. Available at: <https://emerj.com/ai-sector-overviews/artificial-intelligence-applications-lending-loan-management/> [accessed 20 September 2021].

24 Tashea, J. (2017). 'Courts Are Using AI to Sentence Criminals. That Must Stop Now', *Wired*. Available at: [www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/](http://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/) [accessed 20 September 2021].

25 Turner, J. (2018). *Robot Rules: Regulating Artificial Intelligence*. Palgrave Macmillan.

As a result of this ubiquity and complexity, there is no pre-existing regulatory framework – from finance, medicine, product safety, consumer regulation or elsewhere – that can be reworked to readily apply to an overall, cross-cutting approach to UK AI regulation, nor any that look capable of playing such a role without substantial modifications. Instead, a coherent, effective, durable regulatory framework for AI will have to be developed from first principles, borrowing and adapting regulatory techniques, tools and ideas where they are relevant and developing new ones where necessary.

Difficulties posed by the intrinsic features of AI systems are compounded by the current nature of the business practices of many companies that develop AI systems. The developers of AI systems often fail to sit neatly within any one geographic jurisdiction, and face few existing regulatory requirements to disclose details of how and where their systems operate. Moreover, the business models of many of the largest and most successful firms that develop AI systems tend towards market dominance, data agglomeration and user disempowerment.

All this makes the Office for AI's task of using their forthcoming White Paper to set out the UK's position on governing and regulating AI a substantial challenge. Even if the Office for AI limits itself to the articulation of a high-level direction of travel for AI regulation, doing so will involve adjudicating between competing values and visions of the UK's relationship to AI, as well as between differing approaches to addressing the multiple regulatory challenges posed by the technology.

Over the course of 2021, the Ada Lovelace Institute has undertaken multiple research projects and convened expert conversations on many of issues relevant to how the UK should approach the regulation of AI. These included:

- two expert workshops exploring the potential underlying goals of a regulatory system for AI in the UK, the different ways it might be designed, and the tools and mechanisms it would require
- workshops considering the EU's emerging approach to AI regulation
- research on algorithmic accountability in the public sector and on transparency methods of algorithmic decision-making systems.

This report sets out to answer a series of questions on how the UK might go about developing its approach to the regulation of AI

Drawing on the insights generated, and on our own research and deliberation, **this report sets out to answer the following questions on how the UK might go about developing its approach to the regulation of AI:**

1. What might the UK want to achieve with a regulatory framework for AI?
2. What kinds of regulatory approaches and tools could support such outcomes?
3. What are the institutional and legal conditions needed to enable them?

As well as influencing broader policy debates around AI regulation, it is our hope that these considerations are useful in informing the development of the Office for AI's White Paper, the publication of which presents a critical opportunity to help ensure that regulation delivers on its promise to help the UK live up to its ambitions of becoming an 'AI superpower' – and ensuring that such a status delivers economic and societal benefits.

## Expert workshops on the regulation of AI

In April and May 2021, the Ada Lovelace Institute (Ada) convened two expert workshops, bringing together academics, AI researchers, public and civil servants and civil-society organisations to explore how the UK Government should approach the regulation of AI. The insights gained from these workshops have, alongside Ada's own research and deliberation, informed the discussions presented in this report.<sup>26</sup>

These discussions were initially framed around the approach of the UK's National AI Strategy to AI regulation. In practice, they became broader dialogues about the UK's relationship to AI, what the goals of Government policy regarding AI systems should be and the UK's approach to their regulation.

- **Workshop one:** Explored the underlying goals and aims of UK AI policy, particularly with regards to regulation and governance. A key aim here was to establish what long-term objectives, alongside economic growth, the UK should aspire to achieve through AI policy.

<sup>26</sup> Any references in this report to the views and insights of 'expert participants' are references to the discussions in the two workshops.

- **Workshop two:** Concentrated on identifying the specific mechanisms and policy changes that would be needed for the realisation of a successful, joined-up approach to AI regulation. Participants were encouraged to consider the challenges associated with the different objectives of AI policy, as well as broader challenges associated with regulating AI. They then discussed what regulatory approaches, tools and techniques might be required to address them. Participants were also invited to consider whether the UK's regulatory infrastructure itself may need to be adapted or supplemented.

The workshops were conducted under Chatham House rules. With the exception of presentations given by expert participants, none of the insights produced by these workshops are attributed specifically to individual people or organisations.

Expert participants are listed out in full in the acknowledgements section on p.101.

Representatives from the Office for AI also attended the workshops as observers.

## UK AI strategies and regulation

*The UK Government's thinking on the regulation of AI has developed significantly over the past five years. This box sets out some of the major milestones in the Government's position on the regulation and governance of AI over this time, with the aim of putting the 2021 UK AI Strategy into the context of recent history.*

### 2017-19 UK AI strategy

The original UK AI strategy (called the UK AI Sector Deal), published in 2017 and updated in 2019, makes relatively little mention of the role of regulation.<sup>27</sup> In discussing how to build trust in the adoption of AI and address its challenges, the strategy is limited to calls for the creation of the Centre for Data Ethics and Innovation (ODEI) to 'ensure safe, ethical and ground-breaking innovation in AI and data-driven technologies'. The report also calls for the creation of the Office for AI to help the UK Government implement this strategy. The UK Government has since created guidance on the ethical adoption of data-driven technologies and the mitigation of potential harms, including guidelines, developed jointly with the Alan Turing Institute, for ethical AI use in the public sector,<sup>28</sup> a review into bias in algorithmic decision-making<sup>29</sup> and an adoption guide for privacy-enhancing technologies.<sup>30</sup>

### 2021 UK AI roadmap

In January 2021, the AI Council, an independent-expert committee that provides advice to the Office for AI on the AI ecosystem and its AI strategy implementation, published a roadmap with 16 recommendations for how the UK can develop a revised national AI strategy.<sup>31</sup>

The roadmap states that:

- A revised AI strategy presents an important opportunity for the UK Government to develop a strategy for the regulation and governance of AI technologies produced and sold in the UK, with the goal improving safety and public confidence in their use.
- The UK must become 'world-leading in the provision of responsible regulation and governance'.
- Given the rapidly changing nature of AI's development, the UK's systems of governance must be 'ready to respond and adapt more frequently than has typically been true of systems of governance in the past'.

27 European Commission. *United Kingdom AI Strategy Report*. Available at: [https://knowledge4policy.ec.europa.eu/ai-watch/united-kingdom-ai-strategy-report\\_en](https://knowledge4policy.ec.europa.eu/ai-watch/united-kingdom-ai-strategy-report_en).

28 Leslie, D. (2019). *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute. Available at: <https://doi.org/10.5281/zenodo.3240529>.

29 Centre for Data Ethics and Innovation. (2020). *Review into bias in algorithmic decision-making*. Available at: [www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making](http://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making).

30 Centre for Data Ethics and Innovation. (2021). *Privacy Enhancing Technologies Adoption Guide*. Available at: <https://cdeiuk.github.io/pets-adoption-guide>.

31 AI Council. (2021). *AI Roadmap*. UK Government. Available at: [www.gov.uk/government/publications/ai-roadmap](http://www.gov.uk/government/publications/ai-roadmap).

The Council recommends ‘commissioning an independent entity to provide recommendations on the next steps in the evolution of governance mechanisms, including impact and risk assessments, best-practice principles, ethical processes and institutional mechanisms that will increase and sustain public trust’.

### 2021 Scottish AI strategy

Some parts of the UK have further articulated their approach to the regulation of AI. In March 2021, the Scottish Government released an AI strategy that includes five principles that ‘will guide the AI journey from concept to regulation and adoption to create a chain of trust throughout the entire process.<sup>32</sup> These principles draw on the Organisation for Economic Cooperation and Development’s (OECD’s) five complementary values-based principles for the responsible stewardship of trustworthy AI. These are:<sup>33</sup>

1. AI should benefit people and the planet by driving inclusive growth, sustainable development and wellbeing.
2. AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.
3. There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.
4. AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed.
5. Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

The Scottish strategy also calls for the Government to ‘develop a plan to influence global AI standards and regulations through international partnerships’.

### 2021 Digital Regulation Plan

In July 2021, the Department for Digital, Culture, Media, and Sport (DCMS) released a policy paper outlining their thinking on the regulation of digital technologies, including AI.<sup>34</sup> The paper provides high-level considerations, including the establishment of three principles that should guide future plans for the regulation of digital technologies. These are:

1. **Actively promote innovation:** Regulation should ‘be designed to minimise unnecessary burdens on businesses’, be ‘outcomes-focused’, backed by clear evidence of harm, and consider the effects on innovation (a concept the paper does not define). The Government’s approach to regulation should also consider non-regulatory interventions like technical standards first.

<sup>32</sup> Digital Scotland. (2021) *Scotland’s AI Strategy: Trustworthy, Ethical and Inclusive*. Available at: [www.scotlandaistrategy.com](http://www.scotlandaistrategy.com).

<sup>33</sup> Organisation for Economic Co-operation and Development. (2019). *OECD Principles on Artificial Intelligence*. Available at: [www.oecd.org-going-digital/ai/principles](http://www.oecd.org-going-digital/ai/principles).

<sup>34</sup> Department for Digital, Culture, Media & Sport. (2021). *Plan for Digital Regulation*. UK Government. Available at: [www.gov.uk/government/publications/digital-regulation-driving-growth-and-unlocking-innovation](http://www.gov.uk/government/publications/digital-regulation-driving-growth-and-unlocking-innovation).

2. **Achieve forward-looking and coherent outcomes:** This section states regulation should be coordinated across regulators to reduce undue burdens or duplicating existing regulation. Regulation should take a 'collaborative approach' by working with businesses to test out new interventions and business models. Approaches to regulation should 'address underlying drivers of harm rather than symptoms, in order to protect against future changes'.
3. **Exploit opportunities and address challenges in the international arena:** Regulation should be interoperable with international regulations, and policymakers should 'build in international considerations from the start', including via the creation of international standards.

The Digital Regulation Plan includes several mechanisms for putting these principles into practice, including plans to create more regulatory coordination and cooperation, engagement in international forums, and plans to embed these principles across government. However, this policy paper stops short of providing specific recommendations, approaches or frameworks for the regulation of AI systems, and provides only a broad set of considerations that are top of mind for this Government. It does not address specific regulatory tools, mechanisms or approaches the UK should consider towards AI, nor does it provide specific guidance for the overall approach the UK should take towards regulating these technologies.

### 2021 UK AI Strategy

Released in September 2021, the most recent UK AI Strategy sets out three pillars to lead the UK towards becoming an AI science superpower, including:

- investing in the long-term needs of the AI ecosystem
- supporting the transition to an AI-enabled economy
- ensuring the UK gets the national and international governance of AI technologies right to encourage innovation, investment and protect the public and fundamental values.

Sections one and two of the strategy include plans to launch a National AI Research and Innovation (R&I) programme to align funding priorities across UK research councils, plans to publish a Defence AI Strategy articulating military uses of AI, and other investments to expand investment in the UK's AI sector. The third pillar on governance includes plans to pilot an AI Standards Hub to coordinate UK engagement in AI standardisation globally, fund the Alan Turing Institute to update guidance on AI ethics and safety in the public sector, and increase the capacity of regulators to address the risks posed by AI systems. In discussing AI regulation, it makes references to embedding values such as fairness, openness, liberty, security, democracy, the rule of law and respect for human rights.

# Chapter 1: Goals of AI regulation

Recent policy debates around AI have emphasised cultivating and utilising the technology's potential to contribute to economic growth. This focus is visible in the newly published AI strategy's approach to regulation, which stresses the importance of ensuring that the regulatory system fosters public trust and a stable environment for businesses without unduly inhibiting AI innovation.

Although it is prominent in the current Government's AI policy discussions, economic growth is just one of several underlying objectives for which the UK's regulatory approach to AI could be configured. As experts in our workshops pointed out, policymakers may also, for instance, want to stimulate the development of particular forms of AI, single out particular industries for disruption by the technology, or avoid particular consequences of the technology's development and adoption.

Different underlying objectives will not necessarily be mutually exclusive, but prioritisation matters – choices about which to explicitly include and which to emphasise will have a significant effect on downstream policy choices. This is especially the case with regulation, where new regulatory institutions, approaches and tools will need to be chosen and coordinated with broader strategic goals in mind.

The first of the two expert workshops identified and debated desirable objectives for the regulation of AI in addition to economic growth – and explored what adopting these would mean, in concrete terms, for the UK's regulatory system.<sup>35</sup>

A clear point of consensus among the workshop participants, and an important recommendation of this report, was that the Government's approach to AI must not be focused exclusively on fostering economic growth, and must consider the unique properties of how AI systems are developed, procured and integrated.

<sup>35</sup> As set out above, the expert workshops considered the question of how the UK should approach the regulation of AI through the lens of the UK National AI Strategy, though the discussion quickly expanded to cover the UK's regulatory approach to AI more generally.

The UK should build in as an objective the protection and advancement of human rights and societally important values, such as agency, democracy, the rule of law, equality and privacy

Rather than concentrating exclusively on increasing the rate and extent of AI development and use, expert participants stressed that the Government's approach to AI must also be attentive to the technology's unique features, the particular ways it might manifest itself, and the specific effects it might have on the country's economy, society and power structures.

The need to take account of the unique features of AI is a reason for developing a bespoke, codified regulatory approach to the technology – rather than accommodating it within a broader, technology-neutral, industrial strategy. Perhaps more importantly, though, workshop participants were keen to highlight that many of AI's most significant opportunities can only be utilised, and many of its risks can only be mitigated, with the help of an overarching Government strategy that sets out intentions for the use, regulation and governance of these systems. By attending to AI's specific properties, it will be easier for Government to steer the beneficial development and use of AI to address societal challenges, and for the potential risks posed by the technology to be effectively managed.

In light of the specific challenges and opportunity AI poses, expert participants identified four additional objectives that might be usefully built into any AI strategy (outlined below). A common theme cutting across the discussion was that the **UK should build in as an objective the protection and advancement of human rights and societally important values, such as agency, democracy, the rule of law, equality and privacy.**

---

### Objective 1: Ensure AI is used and developed in accordance with specific values and norms

A common refrain among participants was that the UK AI policy should articulate a set of high-level norms or ethical principles to govern the country's desired relationship with AI systems. As several experts pointed out, other countries' national AI strategies, including that of Scotland, have articulated a set of values.<sup>36</sup> The purpose of these

---

<sup>36</sup> Digital Scotland. (2021). *Scotland's AI Strategy: Trustworthy, Ethical and Inclusive*. Available at: [https://static1.squarespace.com/static/5dc00e9e32cd095744be7634/t/606430e006dc4a462a5fa1d4/1617178862157/Scotlands\\_AI\\_Strategy\\_Web\\_updated\\_single\\_page\\_aps.pdf](https://static1.squarespace.com/static/5dc00e9e32cd095744be7634/t/606430e006dc4a462a5fa1d4/1617178862157/Scotlands_AI_Strategy_Web_updated_single_page_aps.pdf) [accessed 22 October 2021].

The articulation of clear, universal and specific values in a prominent AI-policy document (such as an AI strategy) can help establish a common language and set of principles

principles would be to inform specific policy decisions in relation to AI, including the development of regulatory policy and sector-specific guidance and best practice.

The articulation of clear, universal and specific values in a prominent AI-policy document (such as an AI strategy) can help establish a common language and set of principles that could be referenced in future policy and public debates regarding AI. In this instance, the principles would set out how the Government should cultivate and direct the development of the technology, as well as how its use should be governed. They may also extend to the programming and decision-making architecture of AI systems themselves, setting out the values and priorities the UK public would want the developers and deployers of AI systems to uphold when putting them in operation.<sup>37</sup>

In its latest AI strategy, the UK Government makes brief references to several values, including fairness, openness, liberty, security, democracy, the rule of law and respect for human rights.<sup>38</sup> While the values and norms articulated by a national AI strategy would not themselves be able to adjudicate between competing interests and views on specific questions, they do create a framework for weighing and justifying particular courses of action. Medical ethics is a good example of the value of a common language and framework, as it provides medical practitioners with a toolkit to think about different value-laden decisions they might encounter in their practice.<sup>39</sup> In the AI strategy, the values are not well defined enough to underpin this function, nor are they translated into clearly actionable steps to support their being upheld.

There are already a number of AI ethics principles developed by national and international organisations that the UK could draw from to further define and articulate its values for AI regulation.<sup>40</sup> One example mentioned by expert participants is the Organisation for Economic Cooperation and Development's (OECD's) five complementary

<sup>37</sup> Public opinion on these values and priorities would be determined empirically through, for instance, deliberative public engagement.

<sup>38</sup> Office for AI. (2021). *National AI strategy*. UK Government. P. 50. Available at: [www.gov.uk/government/publications/national-ai-strategy](https://www.gov.uk/government/publications/national-ai-strategy).

<sup>39</sup> British Medical Association. (n.d.). *Ethics*. Available at: [www.bma.org.uk/advice-and-support/ethics](https://www.bma.org.uk/advice-and-support/ethics) [accessed 20 September 2021].

<sup>40</sup> Jobin, A., Ienca, M. & Vayena, E. (2019). 'The global landscape of AI ethics guidelines'. *Nature Machine Intelligence*. 1.9 pp. 389–99. Available at: <https://doi.org/10.1038/s42256-019-0088-2>.

Expert participants suggested that deliberative public-engagement exercises, such as citizens' assemblies and juries, could be used to set high-level values

values-based principles for the responsible stewardship of AI,<sup>41</sup> which the Scottish AI strategy draws on heavily.<sup>42</sup>

Another idea raised by the expert participants was that UK AI policy (and industrial strategy more broadly), should aim to establish and support democratic, inclusive mechanisms for resolving value-laden policy and regulatory decisions. Here, expert participants suggested that deliberative public-engagement exercises, such as citizens' assemblies and juries, could be used to set high-level values, or to inform particularly controversial, value-laden policy questions. In addition, participatory mechanisms should be embedded in the development and oversight of governance approaches to AI and data – a topic explored in a recent Ada Lovelace Institute report on participatory data stewardship.<sup>43</sup>

Expert participants noted that sustained public trust in AI will be vital, and the existence of such processes could be a useful means of ensuring that policy decisions regarding AI are aligned with public values.

However, it is important to note that while 'building public trust' in AI is a common and valuable objective surfaced in AI-policy debates, this framing also places the burden of responsibility onto the public to 'be more trusting', and does not necessarily address the root issue: the trustworthiness of AI systems.

Public participation in UK AI policy must therefore be recognised as effective not only at framing or refining existing policies in ways that will be considered more acceptable to the public, but to define the fundamental values that underpin those policies. Without this, there is a significant risk that AI will not align with public hopes, needs and concerns, and this will undermine trust and confidence.

41 Organisation for Economic Co-operation and Development. (2019). *OECD Principles on Artificial Intelligence*. Available at: [www.oecd.org/go-going-digital/ai/principles/](http://www.oecd.org/go-going-digital/ai/principles/) [accessed 22 October 2021].

42 Digital Scotland. (2021). *Scotland's AI Strategy: Trustworthy, Ethical and Inclusive*. Available at: [https://static1.squarespace.com/static/5dc00e9e32cd095744be7634/t/606430e006dc4a462a5fa1d4/1617178862157/Scotlands\\_AI\\_Strategy\\_Web\\_updated\\_single\\_page\\_aps.pdf](https://static1.squarespace.com/static/5dc00e9e32cd095744be7634/t/606430e006dc4a462a5fa1d4/1617178862157/Scotlands_AI_Strategy_Web_updated_single_page_aps.pdf) [accessed 22 October 2021].

43 Ada Lovelace Institute (2021). *Participatory data stewardship*. Available at: [www.adalovelaceinstitute.org/report/participatory-data-stewardship](http://www.adalovelaceinstitute.org/report/participatory-data-stewardship) [accessed 20 September 2021].

## Objective 2: Avoid or ameliorate specific risks and harms

Another commonly voiced view from workshop participants was that UK AI policy should be configured explicitly with a view to reduce, mitigate or completely avoid particular harms and categories of harms associated with AI and its business models. In outlining the particular kinds of harm that AI policy – and particularly regulation – should aim to address, reference was made to the following:

- harms to individuals and marginalised groups
- distributional harms
- harms to free, open societies.

### Harms to individuals and marginalised groups

In discussing the potential harms to individuals and marginalised groups associated with AI, participants highlighted the fact that AI systems:

- **Can exhibit bias**, with the result that individuals may experience AI systems treating them unfairly or drawing unfair inferences about them. Bias can take many forms, and be expressed in several different parts of the AI product development lifecycle – including ‘algorithmic’ bias in which an AI system’s outputs unfairly bias human judgement.<sup>44</sup>
- **Are often more effective or more accurate for some groups than for others.**<sup>45</sup> This can lead to various kinds of harm, ranging from individuals having false inferences made about their identity or characteristics,<sup>46</sup> to individuals being denied or locked out of services due to the failure of AI systems to work for them.<sup>47</sup>

---

44 Selwyn, N. (2021). *Deb Raji on what ‘algorithmic bias’ is (...and what it is not)*. Data Smart Schools. Available at: <https://data-smart-schools.net/2021/04/02/deb-raji-on-what-algorithmic-bias-is-and-what-it-is-not>.

45 Buolamwini, J., Gebru, T. (2018). ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.’ *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, PMLR 81:77 91. Available at: <https://proceedings.mlr.press/v81/buolamwini18a.html>.

46 Hill, K. (2020). ‘Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match.’ *New York Times*. Available at: [www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html](http://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html).

47 Ledford, H. (2019). ‘Millions of black people affected by racial bias in health-care algorithms.’ *Nature*. 574. 7780 pp. 608–9. Available at: [www.nature.com/articles/d41586-019-03228-6](http://www.nature.com/articles/d41586-019-03228-6).

Workshop participants listed several specific kinds of distributional harms that AI systems can raise

- **Tend to be optimised for particular outcomes.**<sup>48</sup> There is a tendency on the part of those developing AI systems to forget, or otherwise insufficiently consider, how the outcomes for which systems have been optimised might affect underrepresented groups within society.
- **Can cause, and often rely on, the violation of individual privacy rights.**<sup>49</sup> A lack of privacy can impede an individual's ability to interact with other people and organisations on equal terms and can cause individuals to change their behaviour.<sup>50</sup>

## Distributional harms

Many of the harms associated with AI systems relate to the capacity of AI and its associated business models to drive and exacerbate economic inequality. Workshop participants listed several specific kinds of distributional harms that AI systems can raise:

- **The business models of leading AI companies tend towards monopolisation and concentration of market share.** Because machine-learning algorithms base their outcomes on data, well-established AI companies that can collect proprietary datasets tend to have an advantage over newer companies, which can be self-perpetuating. In addition, the large amounts of data required to train some machine-learning algorithms present a high barrier of entry into the market, which can incentivise mergers, acquisitions and partnerships.<sup>51</sup> As several recent critiques have pointed out, addressing the harms of AI must look at the wider social, political and economic power underlying the development of these systems.<sup>52</sup>

48 Leslie, D. (2019). *Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector*. The Alan Turing Institute.. Available at: <https://doi.org/10.5281/ZENODO.3240529>.

49 Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

50 Solove, D. J. (2011). *Nothing to Hide: The False Tradeoff Between Privacy and Security*. New Haven London: Yale University Press.

51 Furman, J., Coyle, D., Fletcher, A., McAuley, D., and Marsden, P. (2019). *Unlocking digital competition, Report of the Digital Competition Expert Panel*. HM Treasury. Available at: [www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel](http://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel)

52 Balayan, A., Gürses, S. (2021). *Beyond Debiasing: Regulating AI and its inequalities*. European Digital Rights. Available at: <https://edri.org/our-work/if-ai-is-the-problem-is-debiasing-the-solution>.

- **Labour's declining share of GDP.** Related to the tendency of AI-business models towards monopolisation, some economists have suggested that one reason for labour's declining share of GDP in developed countries is that 'superstar' tech firms, which employ relatively few workers but produce significant dividends for investors, have come to represent an increasing share of overall economic activity.<sup>53</sup>
- **Skills-biased technological change and automation.** Expert participants also cited the potential for automation and skills-biased technological change driven by AI to lead to greater inequality. While it is contested whether the rise of AI will necessarily lead to greater economic inequality in the long term, economists have argued that the short-term disruption caused by the transition from one 'techno-economic paradigm' to a new one will lead to significant inequality unless policy responses are developed to counter these tendencies.<sup>54</sup>
- **AI systems' capacity to undermine the bargaining power between workers and employers, and to exacerbate inequalities between participants in markets.** Finally, participants cited the ability of AI systems to undermine worker power and collective-bargaining capacity.<sup>55</sup> The use of AI systems to monitor and feedback on worker performance, and the application of AI to recruitment and pay-setting processes are two means by which AI could tip the balance of power further towards employers rather than workers.<sup>56</sup>

53 Autor, D., Dorn, D., Katz, L., Patterson, C., and Van Reenen, J. (2020). 'The Fall of the Labor Share and the Rise of Superstar Firms', *The Quarterly Journal of Economics*, 135.2, 645–709. Available at: <https://doi.org/10.1093/qje/qjaa004>.

54 Perez, C. (2015). 'Capitalism, Technology and a Green Global Golden Age: The Role of History in Helping to Shape the Future', *The Political Quarterly*, 86 pp. 191–217. Available at: <https://doi.org/10.1111/1467-923X.12240>.

55 Institute for the Future of Work. (2021). *The Amazonian Era: The gigification of work*. Available at: <https://www.ifow.org/publications/the-amazonian-era-the-gigification-of-work>

56 Partnership on AI. (2021). *Redesigning AI for Shared Prosperity: an Agenda*. Pp 23–24. Available at: <https://partnershiponai.org/wp-content/uploads/2021/08/PAI-Redesigning-AI-for-Shared-Prosperity.pdf>

---

Expert participants also pointed to the capacity of AI systems to undermine many of the necessary conditions for free, open and democratic societies

## Harms to free, open societies

Our expert participants also pointed to the capacity of AI systems to undermine many of the necessary conditions for free, open and democratic societies. Here, participants cited:

- **The use of AI-driven systems to distort competitive political processes.** AI systems that tailor content to individuals based on their data profile or behaviour (mostly through social media or search platforms) can be used to influence voter behaviour and the direction of democratic debates. This is recognised as problematic because access to these systems is likely to be unevenly distributed across the population and political groups, and because the opacity of content creation and sharing can undermine the democratic ideal of a commonly shared and accessible political discourse – as well as ideals about public debate being subject to public reason.<sup>57</sup>
- **The use of AI-driven systems to undermine the health and competitiveness of markets.** In the market sphere, AI-enabled functions such as real-time, A/B testing,<sup>58</sup> hypernudge,<sup>59</sup> and personalised pricing and search<sup>60</sup> undermine the ability of consumers to choose freely between competing products in a market, and can significantly skew the balance of power between consumers and large companies.
- **Surveillance, privacy and the right to freedom of expression and assembly.** The ability of AI-driven systems to monitor and surveil citizens has the potential to create a powerful negative effect on citizens exercising their rights to free expression and discourse – negatively affecting the tenor of democracies.

---

57 Quong, J. (2018). 'Public Reason' in Zalta, E. N. and Hammer, E. (eds) *The Stanford Encyclopedia of Philosophy*. Stanford: Center for the Study of Language and Information. Available at: [www.scirp.org/reference/referencespapers.aspx?referenceid=2710060](http://www.scirp.org/reference/referencespapers.aspx?referenceid=2710060) [accessed 20 September 2021].

58 Where two or more options are presented to users to determine which is more preferable.

59 Where an individual's data and responses to stimuli is used to inform how choices are framed to them, with a view towards predisposing them towards particular choices. See: Yeung, K. (2017). "Hypernudge": Big Data as a Mode of Regulation by Design', *Information, Communication & Society*, 20.1 pp.118–136. Available at: <https://doi.org/10.1080/1369118X.2016.1186713>.

60 Where the prices or search results seen by a consumer are determined by their data profile. See: Competition and Markets Authority. (2018). *Pricing algorithms: Economic working paper on the use of algorithms to facilitate collusion and personalised pricing*, p. 63. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/746353/Algorithms\\_econ\\_report.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746353/Algorithms_econ_report.pdf).

A country's approach to AI regulation could be informed by its stated priorities and objectives for the use of AI in society

- **The use of AI systems to police and control citizen behaviour.** It was noted that many AI systems could be used for more coercive methods of controlling or influencing citizens. Participants cited 'social-credit' schemes, such as the one being implemented in China, as an example of the kind of AI system that seeks to manipulate or enforce certain forms of social behaviour without adequate democratic oversight or control.<sup>61</sup>

---

### Objective 3: Use AI to contribute to the solution of grand societal challenges

Another common view of workshop participants was that a country's approach to AI regulation could be informed by its stated priorities and objectives for the use of AI in society. One of the common aims of many existing national AI strategies is to articulate how a country can leverage its AI ecosystem to develop solutions to, and means of addressing substantial, society-wide challenges facing individual nations – and indeed humanity – in coming decades.<sup>62</sup>

Candidates for these challenges range from decarbonisation and dealing with the effects of climate change, navigating potential economic displacement brought about by AI systems (and the broader context of the 'fourth industrial revolution'), to finding ways to manage the difficulties, and make best use, of an ageing population – which is itself one of the UK's 2017 Industrial Strategy grand challenges. Workshop participants also referred to the potential for AI to be deployed to address the long-term effects of the COVID-19 pandemic and its potential to ameliorate future public-health crises.

Workshop participants emphasised that the purpose of articulating grand-societal challenges that AI can address was to provide an effective way to think about the coordination of different industrial strategy

---

61 The authors of this paper note that many of the claims about the efficacy and goals of the Chinese social-credit system have been exaggerated in the Western media. See Matsakis, L. (2019). 'How the West Got China's Social Credit System Wrong.' *Wired*. Available at: [www.wired.com/story/china-social-credit-score-system](http://www.wired.com/story/china-social-credit-score-system).

62 Dutton, T. (2018). 'An Overview of National AI Strategies', *Politics + AI*. Available at: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>

---

By articulating grand challenges that AI can address, the UK Government can help establish funding and research priorities for applications of AI that show high reward and proven efficacy

levers, from R&D and regulatory policy, to tax policy and public-sector procurement. This approach would sidestep the risk of an AI national strategy that commands more AI for the sake of AI, or a strategy that places too much hope on the potential benefit of AI to bring positive societal change across all economic and societal sectors.

By articulating grand challenges that AI can address, the UK Government can help establish funding and research priorities for applications of AI that show high reward and proven efficacy. As an example, the French national AI strategy articulates several grand challenges as areas of focus for AI, including addressing the COVID-19 pandemic and fighting climate change.<sup>63</sup>

A reservation to consider with the societal-challenge approach is that it absolves Government of articulating a sense of direction when it comes to the UK's relationship to AI. Setting out that we want AI to be used to address particular problems, and how AI is to be supported and guided to develop in a manner conducive to their solution, does not provide any indication of the level of risk we are willing to tolerate, the kinds of applications of AI we may or may not want to encourage or permit (all else remaining equal) or how our industrial and regulatory policy should address difficult, values-based trade-offs.

---

#### Objective 4: Develop AI regulation as a sectoral strength

A fourth suggestion put forward by some workshop participants was that the UK should seek to develop AI regulation as a sectoral strength. There was limited agreement on what this goal might entail in practice, and whether it would be feasible.

Despite the UK's strengths in academic AI research, most participants agreed that, because of existing market dynamics in the tech industry – in which a combination of mostly US and Chinese firms dominate the market, it will be very difficult to the UK market to create the next industry powerhouse.

---

63 European Commission. (2021). *Knowledge for Policy: France AI Strategy Report*. Available at: [https://knowledge4policy.ec.europa.eu/ai-watch/france-ai-strategy-report\\_en](https://knowledge4policy.ec.europa.eu/ai-watch/france-ai-strategy-report_en)

---

The UK could potentially become world leading in flexible, innovative and ethical approaches to the regulation of AI

However, an idea that emerged in the first workshop was that the UK could potentially become world leading in flexible, innovative and ethical approaches to the regulation of AI. The UK Government has expressed explicit ambitions to lead the world in tech and data ethics since at least 2018.<sup>64</sup> Workshop participants noted that the UK already has an established reputation for regulatory innovation, and that the country is potentially well placed to develop an approach to the regulation of AI that is compatible with EU standards, but more sophisticated and nuanced.

This idea received additional scrutiny in the second workshop, which saw a more sustained and critical discussion, detailed below, of what cultivating a niche in the regulation of AI might look like in practice, and of the benefits it might bring.

## Why is leadership in AI regulation desirable?

Some participants challenged whether leadership in the regulation of AI would actually be desirable, and if so how.

It was noted that, in some cases, a country that drives the regulatory agenda for a particular technology or science will be in a good position to attract greater levels of expertise and investment. For instance, the UK is a world leader in biomedical research and technology, in large part because it has a robust regulatory system that ensures a high quality of accuracy, safety and public trust.<sup>65</sup> It was cautioned, however, that the UK's status with the regulation of biomedical technology is the product of the combination of demanding standards, a pragmatic approach to the interpretation of those standards and a rigorously enforced institutional regime.

Some expert panellists suggested that, despite the fact that many regulatory rules have been set at an EU level, the UK has become a leader in the regulation of the life sciences because it combined those high ethical and legal standards with sufficient flexibility to enable genuine innovation – rather than because it relaxed regulatory standards.

---

64 Kelion, L. (2018). 'UK PM seeks 'safe and ethical' artificial intelligence.' *BBC News*. 25 January. Available at: [www.bbc.co.uk/news/technology-42810678](http://www.bbc.co.uk/news/technology-42810678).

65 Calvert, M. J., Marston, E., Samuels, M., Cruz Rivera, S., Torlinska, B., Oliver, K., Denniston, A. K., and Hoare, S. (2019). 'Advancing UK regulatory science and innovation in healthcare', *Journal of the Royal Society of Medicine*. 114.1. pp. 5-11. Available at: <https://doi.org/10.1177/0141076820961776>

## The UK can't compete on regulatory substance, but could compete on some aspects of regulatory procedure and approach

There was a degree of scepticism among expert panellists about whether the model that has enabled the UK to achieve leadership in the regulation of the biomedical-sciences industry would be replicable or would yield the same results in the context of AI regulation. In contrast to the biomedical sciences – where there are strict and clearly defined routes into practice – it is difficult for a regulator to understand and control actors developing and deploying AI systems. The scale and the immediacy of the impacts of AI technologies also tends to be far greater than in biomedical sciences, as is the number of domains in which AI systems could potentially be deployed.

In addition to this, it was noted that the EU also has ambitions to become a global leader in the ethical regulation of AI, as demonstrated by the European Commission's proposed AI regulations.<sup>66</sup> It is therefore unclear what the UK might leverage to position itself as a distinct leader, alongside a larger, geographically adjacent and more influential economic bloc with a good track record of exporting its regulatory standards, which also has ambitions to occupy this space. The EU's proposal of a comprehensive AI regulation also means that the UK does not have a first-mover advantage when it comes to the regulation of AI.

Many participants of our workshops thought it was unlikely that the UK would be able to compete with the EU (or other large economic blocs) on regulatory substance, or the specific rules and regulations governing AI. Some workshop participants observed that the comparatively small size of the UK market would mean that approval from a UK regulatory body is of less commercial value to an AI company than regulatory approval from the EU.

In terms of regulatory substance, some participants considered whether the UK could make itself attractive as a place to develop AI products by lowering regulatory standards, but other participants noted this would be undesirable and would go against the grain of the UK's

---

<sup>66</sup> European Commission. (2021). *A Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. Available at: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52021PC0206> [accessed 4 October 2021].

---

It may be possible for the UK to lead on the processes and procedures for regulating AI

strengths in the flexible enforcement of exacting regulatory standards. Moreover, participants suggested that a ‘race to the bottom’ approach would be counter-productive, given the size of the UK market and the higher regulatory standards that are already developing elsewhere. Adopting this approach could mean that UK-based AI developers would not be able to sell their services and products in regions with higher regulatory standards.

Despite the limited prospects for the UK leading the world in the development of regulatory standards for AI, some workshop participants argued that it may be possible for the UK to lead on the processes and procedures for regulating AI. The UK does have a good reputation for following regulatory processes and for regulatory process innovation (as exemplified by regulatory sandboxes, a model that has been replicated by many other jurisdictions, including the EU).<sup>67</sup>

While sandboxes no longer represent a unique selling point for the UK, the UK may be able to make itself more attractive to AI firms by establishing a series of regulatory practices and norms aimed at ensuring that companies have better guidance and support in complying with regulations than they might receive elsewhere. These sorts of processes are particularly appealing to start-ups and small- to medium-sized enterprises (SMEs), who may struggle to navigate and comply with regulatory processes more than their larger counterparts.

A final caveat that several expert participants made was that, although more supportive regulatory processes might be enough to attract start-ups and early-stage AI ventures to the UK, keeping such companies in the UK as they grow will also require the presence of the right financial, legal and a research-and-development supportive ecosystem. While this report does not seek to answer the question of what this wider ecosystem should look like, it is clear that a regulatory framework is a necessary condition for the realisation of the Government’s stated ambition of developing a world-leading AI sector, closely coordinated with policies to nurture and maintain these other enabling conditions.

---

<sup>67</sup> Privacy & Information Security Law Blog. (2021). *Regulatory Sandboxes are Gaining Traction with European Data Protection Authorities*. Hunton Andrews Kurth. Available at: <https://www.huntonprivacyblog.com/2021/02/25/regulatory-sandboxes-are-gaining-traction-with-european-data-protection-authorities/>

---

# Chapter 2: Challenges for regulating AI systems

Given AI's relative novelty, complexity and applicability across both domains and industries, the effective and consistent regulation of AI systems presents multiple challenges. This chapter details some of the most significant of these, as highlighted by our expert workshop participants, and sets out additional analysis and explanation of these issues. The following chapter, 'Tools, mechanisms and approaches for regulating AI', details some ways these challenges might be dealt with or overcome. Additional details on some of the different considerations when designing and configuring regulatory systems, which may be a useful companion to these two chapters, can be found in the annex, (p. 81).

The table below maps the regulatory challenges identified with the relevant tools, mechanisms and approaches for overcoming them.

## Regulatory challenges and relevant tools, mechanisms and approaches

---

Challenges for regulating AI systems	Potentially useful approach, tool or mechanism
AI regulation demands bespoke, cross-cutting rules	Regulatory capacity building Regulatory coordination
The incentive structures and power dynamics of AI-business models can run counter to regulatory goals and broader societal values	
It can be difficult to regulate AI systems in a manner that is proportionate	Risk-based regulation Professionalisation
Many AI systems are complex and opaque	Regulatory capacity building Algorithmic impact assessment Transparency requirements Inspection powers External-oversight bodies International standards Domestic standards (e.g. via procurement)
AI harms can be difficult to separate from the technology itself	Moratoria and bans

---

Regulating [AI] successfully is likely to require the development of new, domain-neutral laws and regulatory principles

## AI regulation demands bespoke, cross-cutting rules

Perhaps one of the biggest challenges presented by AI is that regulating it successfully is likely to require the development of new, domain-neutral laws and regulatory principles. There are several, interconnected reasons for this:

1. AI presents novel challenges for existing legal and regulatory principles
2. AI presents systemic challenges that require a coordinated response
3. horizontal regulation will help avoid boundary disputes and aid industry-specific policy development
4. effective, cross-cutting legal and regulatory principles won't emerge organically
5. the challenges of developing bespoke, horizontal rules for AI.

### **1. AI presents novel challenges for existing legal and regulatory principles**

One argument for developing new laws and regulatory principles for AI is that those in existence are not fit for purpose.

AI has two features that present difficulties for contemporary legal principles. The first is its tendency to fully or partially automate moral decision-making processes in ways that can be opaque, difficult to explain and difficult to predict. The second is the capacity of AI systems to develop and operate independently of human control. For these reasons, AI systems can challenge legal notions of agency and causation as the relationship between the behaviour of the technology and the actions of the user or developer can be unclear, and some AI systems may change independently of human control and intervention.

While these principles have been unproblematically applied to legal questions concerning other emerging technologies, it is not clear that they will apply readily to those presented by AI. As barrister Jacob Turner explains, in contrast to AI systems, ‘a bicycle will not re-design itself to become faster. A baseball bat will not independently decide to hit a ball or smash a window.’<sup>68</sup>

---

68 Turner, J. (2018). *Robot Rules: Regulating Artificial Intelligence*. Palgrave Macmillan. P. 79.

## 2. AI presents systemic challenges that require a coordinated response

In addition to demanding new approaches to legal principles of agency and causation the effective regulation and governance of AI systems will require high levels of coordination.

As a powerful technology that can operate at scale and be applied in a wide range of different contexts, AI systems can manifest impacts at the level of the whole economy and the whole of society, rather than being confined to particular domains or sectors. Among policymakers and industry professionals, AI is regularly compared to electricity, with claims that it can transform a wide range of different sectors.<sup>69</sup> Whether or not this is hyperbole, the ambition to integrate AI systems across a wide variety of core services and applications raises risks of significant negative outcomes. If governments aspire to use regulation and other policy mechanisms to control the systematic impacts of AI, they will have to coordinate legal and regulatory responses to particular uses of AI. Developing a general set of principles to which all regulators must adhere when dealing with AI is a practical way of doing this.

## 3. Horizontal regulation will help avoid boundary disputes and aid industry-specific policy development

There are also practical arguments for developing cross-cutting legal and regulatory principles for AI. The gradual shift from narrow to general AI will mean that attempts to regulate the technology exclusively through the rules applied to individual domains and sectors will become increasingly impractical and difficult. A fully vertical or compartmentalised approach to the regulation of AI would be likely to lead to boundary disputes, with persistent questions about whether particular applications or kinds of AI fall under the remit of one regulator or another – or both, or neither.

---

<sup>69</sup> Lynch, S. (2017). *Andrew Ng: Why AI Is the New Electricity*. Stanford Graduate School of Business. Available at: [www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity](http://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity)

#### 4. Effective, cross-cutting legal and regulatory principles won't emerge organically

Clear, cross-cutting legal and regulatory principles for AI will have to be set out in legislation, rather than developed through, and set out in common law. Perhaps the most important reason for this is that setting out principles in statute makes it possible to protect against the potential harms of AI in advance (*ex ante*), rather than once things have gone wrong (*ex post*) – something a common law approach would be incapable of doing. Given the potential gravity and scope of the sorts of harms AI is capable of producing, it would be very risky to wait until harms occur to develop legal and regulatory protections against them.

The Law Society's evidence submission to the House of Commons Science and Technology Select Committee summarises some of reasons to favour a statutory approach to regulating and governing AI:

'One of the disadvantages of leaving it to the Courts to develop solutions through case law is that the common law only develops by applying legal principles after the event when something untoward has already happened. This can be very expensive and stressful for all those affected. Moreover, whether and how the law develops depends on which cases are pursued, whether they are pursued all the way to trial and appeal, and what arguments the parties' lawyers choose to pursue. The statutory approach ensures that there is a framework in place that everyone can understand.'<sup>70</sup>

<sup>70</sup> The Law Society. (2016). *Written evidence submitted by the Law Society (ROB0037)*. Available at: <http://data.parliament.uk/writtenEvidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/robotics-and-artificial-intelligence/written/32616.html> [accessed 20 September 2021].

---

Waiting for harms to become clear and manifest before embarking on regulatory interventions can come with significant risks

## 5. The challenges of developing bespoke, horizontal rules for AI

The need to develop new, domain-neutral, AI-specific law raises several difficult questions for policymakers. Who should be responsible for developing these legal and regulatory principles? What values and priorities should these principles reflect? How can we ensure that those developing the principles have a good enough understanding of the ways AI can and might develop and impact on society?

**It can be difficult to regulate AI systems in a manner that is proportionate**

Given the range of applications and uses of AI, a critical challenge in developing an effective regulatory approach is ensuring that rules and standards are strong enough to capture potential harms, while not being unjustifiably onerous for more innocuous or lower-risk uses of the technology.

The difficulties of developing proportionate regulatory responses to AI are compounded because, as with many emerging technologies, it can be difficult for a regulatory body to understand the potential harms of a particular AI system before that system has become widely deployed or used. However, waiting for harms to become clear and manifest before embarking on regulatory interventions can come with significant risks. One risk is that harms may transpire to be grave, and difficult to reverse or compensate for. Another is that, by the time the harms of an AI system have become clear, these systems may be so integrated into economic life that *ex post* regulation becomes very difficult.<sup>71</sup>

**The incentive structures and power dynamics created by AI-business models can run counter to regulatory goals and broader societal values**

Several expert participants also noted that an approach to regulation must acknowledge the current reality around the market and business

---

<sup>71</sup> Liebert, W., and Schmidt, J. C. (2010). 'Collingridge's Dilemma and Technoscience: An Attempt to Provide a Clarification from the Perspective of the Philosophy of Science', *Poiesis & Praxis*, 7:1–2 pp. (2010), 55–71 Available at: <https://doi.org/10.1007/s10202-010-0078-2>.

dynamics for AI systems. As many powerful AI systems rely on access to large datasets, the business models of AI developers can be heavily skewed towards accumulating proprietary data, which can incentivise both extractive data practices and restriction of access to that data.

Many large companies now provide AI 'as a service', raising the barrier to entry for new organisations seeking to develop their own independent AI capabilities.<sup>72</sup> In the absence of strong countervailing forces, this can create incentive structures for businesses, individuals and the public sector that are misaligned with the ultimate goals of regulators and the values of the public. Expert participants in workshops and follow-up discussions identified two of these possible perverse incentive structures: data dependency and the data subsidy.

## Data dependency

The principle of universal public services under democratic control is undermined by the public sector's incentives to rely on large, private companies for data analytics, or for access to data on service users. These services promise efficiency benefits, but threaten to disempower the public-service provider, with the following results:

- Public-service providers may feel incentivised to collect more data on their service users than they can use to inform AI services.
- By relying on data analytics provided by private companies, public services give up control of important decisions to AI systems over which they have little oversight or power.
- Public-service providers may feel increasingly unable to deliver services effectively without the help of private tech companies.

## The data subsidy

The principle of consumer markets that provide choice, value and fair treatment is undermined by the public's incentives to provide their data for free or cheaper services (the 'data subsidy'). This can result in phenomena like personalised pricing and search, which undermine

---

72 Cobbe, J., and Singh, J. (2021). 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges'. *SSRN Electronic Journal*. Available at: <https://ssrn.com/abstract=3824736> or <http://dx.doi.org/10.2139/ssrn.3824736>.

---

It can be very difficult for a regulator to understand exactly how an AI system operates, whether there is the potential for it to cause harm, and whether it has done so

consumer bargaining power and *de facto* choice, and can lead to the exploitation of vulnerable groups.

## Many AI systems are complex and opaque

Another significant difficulty concerning the regulation of AI concerns the complexity and opacity of many AI systems. In practice, it can be very difficult for a regulator to understand exactly how an AI system operates, whether there is the potential for it to cause harm, and whether it has done so. The difficulty in understanding AI systems poses serious challenges, and in looking for solutions, it is helpful to distinguish between some of the sources of these challenges, which may include:

1. regulators' technical capacity and resources
2. the opacity of AI developers
3. the opacity of AI systems themselves.

### 1. Regulators' technical capacity and resources

Firstly, many expert participants, including some from regulatory agencies, noted that existing regulatory bodies struggle to regulate AI systems due to a lack of capacity and technical expertise.

There are over 90 regulatory agencies in the UK that enforce legislation in sectors like transportation, public utilities, financial services, telecommunications, health and social services and many others. As of 2016, the total annual expenditure on these regulatory agencies was around £4 billion – but not all regulators receive the same amount, with some like the Competition and Markets Authority (CMA) or the Office of Communications (Ofcom) receiving far more than smaller regulators like the Equalities and Human Rights Commission (EHRC).<sup>73</sup>

Some regulators like the CMA and the Information Commissioner's Office (ICO) already have some in-house employees specialising in data science and AI techniques, to reflect the nature of the work they do and kinds of organisations they regulate. But as AI systems become more

---

<sup>73</sup> National Audit Office. (2017). *A short guide to regulation*. UK Government. Available at: [www.nao.org.uk/wp-content/uploads/2017/09/A-Short-Guide-to-Regulation.pdf](http://www.nao.org.uk/wp-content/uploads/2017/09/A-Short-Guide-to-Regulation.pdf).

widely used in various sectors of the UK economy, it becomes more urgent for regulators of all sizes to have access to the technical expertise required to evaluate and assess these systems, along with the powers necessary to investigate AI systems.

This poses questions about how regulators might best build their capacity to understand and engage with AI systems, or secure access to this expertise consistently.<sup>74</sup>

## 2. The opacity of AI developers

Secondly, many of the difficulties regulators have in understanding AI systems result from the fact that much of the information required to do so is proprietary, and that AI developers and tech companies are often unwilling to share information that they see as integral to their business model. Indeed, many prominent developers of AI systems have cited intellectual property and trade secrets as reasons to actively disrupt or prevent attempts to audit or assess their systems.<sup>75</sup>

While some UK regulators do have powers to inspect AI systems, where those systems are developed by regulated entities, the inspection of systems becomes much more difficult when those systems are provided by third parties. This issue poses questions about the powers regulators might need to require information from AI developers or users, along with standards of openness and transparency on the part of such groups.

## 3. The opacity of AI systems themselves

Finally, in some cases, there are also deeper issues concerning the ability of anyone, even the developers of an AI system, to understand the basis on which it may make decisions. The biggest of these is the fact that non-symbolic AI systems, which are the kind of AI responsible

<sup>74</sup> Ada Lovelace Institute (Forthcoming). *Technical approaches for regulatory inspection of algorithmic systems in social media platforms*. Available at: <https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection>.

<sup>75</sup> Facebook, for example, has recently shut down independent attempts to monitor and assess their platform's behaviour. See: Kayser-Bril, N. (2021). *AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook*. Algorithm Watch. Available at: <https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/>, and Bobrowsky, M. (2021). Facebook Disables Access for NYU Research Into Political-Ad Targeting. *Wall Street Journal*. Available at: [www.wsj.com/articles/facebook-cuts-off-access-for-nyu-research-into-political-ad-targeting-11628052204](http://www.wsj.com/articles/facebook-cuts-off-access-for-nyu-research-into-political-ad-targeting-11628052204).

for some of the most recent impressive advances in the field, tend to operate as ‘black boxes’, whose decision-making sequences are difficult to parse. In some cases, it may be the case that certain types of AI systems may not be appropriate for deployment in settings where it is essential to be able to provide a contestable explanation.

These difficulties in understanding AI systems’ decision-making processes become especially problematic in cases where a regulator might be interested in protecting against ‘procedural’ harms, or ‘procedural injustices’. In these cases, a harm is recognised not because of the nature of the outcome, but because of the unfair or flawed means by which that outcome was produced.

While there are strong arguments to take these sorts of harms seriously, they can be very difficult to detect without understanding the means by which decisions have been made and the factors that have been taken into account. For instance, looking at who an automated credit-scoring system considers to be most and least creditworthy may not reveal any obvious unfairness – or at the very least will not provide sufficient evidence of procedural harm, as any discrepancies between different groups could theoretically have a legitimate explanation. It is only when considering how these decisions have been made, and whether the system has taken into account factors that should be irrelevant, that procedural unfairness can be identified or ruled out.

## AI harms can be difficult to separate from the technology itself

The complexity of the ways that AI systems can and could be deployed means that there are likely to be some instances when regulators are unsure of their ability to effectively isolate potential harms from potential benefits.

These doubts may be caused by a lack of information or understanding of a particular application of AI. There will inevitably be some instances in which it is very difficult to understand exactly the level of risk posed by a particular form of the technology, and if and how the risks posed by it might be mitigated or controlled, without undermining the benefits of the technology.

In other cases, these doubts may be informed by the nature of the application itself, or by considerations of the likely dynamics affecting its development. There may be instances where, due to the nature of the form or application of AI, it seems difficult to separate the harms it poses from its potential benefits. Regulators might also doubt whether particular high-risk forms or uses of AI can realistically be contained to a small set of heavily controlled uses. One reason for this is that the infrastructure and investment required to make limited deployments of a high-risk application possible create long-term pressure to use the technology more widely: the industry developing and providing the technology is incentivised to advocate for a greater variety of uses. Government and public bodies may also come under pressure to expand the use of the technology to justify the cost of having acquired it.

---

No single mechanism or approach will be sufficient to regulate AI effectively

## Chapter 3: Tools, mechanisms and approaches for regulating AI systems

To address some of the challenges outlined in the previous section, our expert workshop participants identified a number of tools, mechanisms and approaches to regulation that could potentially be deployed as part of the Government's efforts to effectively regulate AI systems at different stages of the AI lifecycle.

Some mechanisms can provide an *ex ante* pre-assessment of an AI system's risk or impacts, while others provide ongoing monitoring obligations and *ex post* assessments of a system's behaviour. It is important to understand that **no single mechanism or approach will be sufficient to regulate AI effectively** – but that regulators will need a variety of tools in their toolboxes to draw on as needed.

Many of the mechanisms described below follow the National Audit Office's *Principles of effective regulation*,<sup>76</sup> which we believe may offer a useful guide for the Government's forthcoming White Paper.

---

### Regulatory infrastructure – capacity building and coordination

#### Capacity building and coordination

The 2021 UK AI Strategy acknowledges that regulatory capacity and coordination will be a major area of focus for the next few years. Our expert participants also proposed sustained and significant expansion of the regulatory system's overall capacity and levels of coordination, to support successful management of AI systems.

---

<sup>76</sup> National Audit Office. (2021). *Principles of effective regulation*. UK Government. Available at: [www.nao.org.uk/wp-content/uploads/2021/05/Principles-of-effective-regulation-SOff-interactive-accessible.pdf](http://www.nao.org.uk/wp-content/uploads/2021/05/Principles-of-effective-regulation-SOff-interactive-accessible.pdf).

---

There are different models for allocating regulatory resources, and for improving the system's overall capacity, flexibility and cohesiveness

If the UK's regulators are to adjust to the scale and complexity of the challenges presented by AI, and control the practices of large, multinational tech companies effectively, they will need greater levels of expertise, greater resourcing and better systems of coordination.

Expert participants were keen to stress that calls for the expansion of regulatory capacity should not be limited to the cultivation of technical expertise in AI, but should also extend to better institutional understanding of legal principles, human-rights norms and ethics. Improving regulators' ability to understand, interrogate, predict and navigate the ethical and legal challenges posed by AI systems is just as important as improving their ability to understand and scrutinise the workings of the systems themselves.<sup>77</sup>

Expert participants also emphasised some of the limitations of AI-ethics exercises and guidelines that are not backed up by hard regulation and the law<sup>78</sup> – and cited this as an important reason to embed ethical thinking within regulators specifically.

There are different models for allocating regulatory resources, and for improving the system's overall capacity, flexibility and cohesiveness, any model will need:

- a means to allocate additional resources efficiently, avoiding duplication of effort across regulators, and guarding against the possibility of gaps and weak spots in the regulatory ecosystem
- a way for regulators to coordinate their responses to the applications of AI across their respective domains, and to ensure that their actions are in accordance with any cross-cutting regulatory principles or laws regarding AI
- a way for regulators to share intelligence effectively and conduct horizon-scanning exercises jointly.

One model would be to have centralised regulatory capacity that individual regulators could draw upon. This could consist of AI experts

---

77 Yeung, K., Howes, A., and Pogrebna, G. (2020). 'AI Governance by Human Rights-Centred Design, Deliberation, and Oversight: An End to Ethics Washing', in Dubber, M. D., Pasquale, F., and Das, S. (eds) *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press. pp. 75–106. Available at: <https://doi.org/10.1093/oxfordhb/9780190067397.013.5>.

78 Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., and Cave, S. (2019). *Ethical and societal implications of algorithms, data, artificial intelligence: A roadmap for research*. London: Nuffield Foundation. Available at: [www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf](http://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf).

---

A much-discussed means of improving regulatory capacity is the formation of a new, dedicated AI regulator

and auditors, as well as funding available to support capacity building in individual regulators. A key advantage of a system of centralised regulatory capacity is that regulators could draw on expertise and resources as and when needed, but the system would have to be designed to ensure that individual regulators had sufficient expertise to understand when they needed to call in additional resources.

An alternative way of delivering centralised regulatory capacity is a model where experts on AI and related disciplines are distributed within individual regulators and circulate around, reporting back cross-cutting intelligence and knowledge. This would build expert capacity and understanding of the effects AI is having on different sectors and parts of the regulatory system, to identify common trends and to strategise and coordinate potential responses.

Another method would be to have AI experts permanently embedded within individual regulators, enabling them to develop deep expertise of the particular regulatory challenges posed by AI in that domain. In this model experts would have to communicate and liaise across regulatory bodies to prevent siloed thinking.

Finally, a much-discussed means of improving regulatory capacity is the formation of a new, dedicated AI regulator. This regulatory body could potentially serve multiple functions, from setting general regulatory principles or domain-specific rules for AI regulation, to providing capacity and advice for individual regulators, to overseeing and coordinating horizon-scanning exercises and coordinating regulatory responses to AI across the regulatory ecosystem.

Most expert participants did not feel that there would be much benefit from establishing an independent AI regulator for the purposes of setting and enforcing granular regulatory rules. There are some common and consistent questions that all kinds of AI systems raise around issues of accountability, fairness, explainability of automated decisions, the relationship between machine and human agency, privacy and bias.

However, most expert participants agreed that regulatory processes and rules need to be specific to the domain in which AI is being deployed. Some participants acknowledged that there may be some need for an entity to develop and maintain a common set of principles and standards for the regulation of AI, and to ensure that individual regulators apply those principles in a manner that is consistent – by maintaining

---

'Anticipatory regulation' [is] a set of techniques and principles intended to help regulators be more proactive, coordinated and democratic in their approach to emerging technologies

an overview of the coherence of all the regulatory rules governing AI, and by providing guidance for individual regulators on how to interpret the cross-industry regulatory principles.

None of the above models should be seen as mutually exclusive, nor substitutes for more money and resources being given to all regulators to deal with AI. Creating pooled-regulatory capacity that individual regulators can draw on need, and should not, come at the expense of improving levels of expertise and analytic capacity within individual regulatory bodies.

With regards to regulatory coordination, several participants noted that existing models aimed at helping regulators work together on issues presented by AI systems should be continued and expanded. For example, the Digital Regulation Cooperation Forum functions with the CMA, ICO, Ofcom and the Financial Conduct Authority (FCA) to 'ensure a greater level of cooperation given the unique challenges posed by regulation of online platforms'<sup>79</sup>

## Anticipatory capacity

If the regulatory system is to have a chance of addressing the potential harms posed by AI systems and business models effectively, it will need to better understand and anticipate those harms. The ability to anticipate AI harms is also fundamental to overcoming the difficulty of designing effective *ex ante* rules to protect against harms that have not yet necessarily occurred on a large scale.

One promising approach to help regulators better understand and address the challenges posed by AI is 'anticipatory regulation', a set of techniques and principles intended to help regulators be more proactive, coordinated and democratic in their approach to emerging technologies.<sup>80</sup> These techniques include horizon-scanning and futures exercises, such as scenario mapping (especially as collaborations between regulators and other entities), along with iterative, collaborative

---

79 Digital Regulation Cooperation Forum. (2021). UK Government. Available at: [www.gov.uk/government/collections/the-digital-regulation-cooperation-forum](http://www.gov.uk/government/collections/the-digital-regulation-cooperation-forum).

80 Armstrong, H., Gorst, C., Rae, J. (2019). *Renewing Regulation: 'anticipatory regulation' in an age of disruption*. Nesta. Available at: [www.nesta.org.uk/report/renewing-regulation-anticipatory-regulation-in-an-age-of-disruption](http://www.nesta.org.uk/report/renewing-regulation-anticipatory-regulation-in-an-age-of-disruption).

approaches, such as regulatory sandboxes. They may also include participatory-futures exercises like citizen juries that involve members of the public, particularly those from traditionally marginalised communities, to help anticipate potential scenarios.

There is already support for regulators to experiment with anticipatory techniques, such as that provided by the Regulators' Pioneer Fund, and initiatives to embed horizon scanning and futures thinking into the regulatory system, such as the establishment of the Regulatory Horizons Council.<sup>81</sup> However, for these techniques to become the norm among regulators, Government support for anticipatory methods will have to be more generous, provided by default and long term.

Workshop participants noted that harms posed by emerging technologies can be overlooked because policymakers lack understanding of how new technologies or services might affect particular groups. Given this, some participants suggested that efforts to bring in a variety of perspectives to regulatory policymaking processes, via public-engagement exercises or through drives to improve the diversity of policymakers themselves, would have a positive effect on the regulators' capacity to anticipate and understand harms and unintended consequences of AI.<sup>82</sup>

## Developing a healthy ecosystem of regulation and governance

Several participants in our workshops noted the need for the UK to adopt a regulatory approach to AI that enables an 'ecosystem' of governance and accountability that rewards and incentivises self-governance, and makes possible third-party, independent assessments and reviews of AI systems.

Given the capacity for AI technologies to be deployed in a range of settings and contexts, no single regulator may be capable of assessing an AI system for all kinds of harms and impacts. The Competition and Markets Authority, for example, seeks to address issues of competition

---

<sup>81</sup> UK Government. (2021). *Regulatory Horizons Council (RHC)*. Available at: [www.gov.uk/government/groups/regulatory-horizons-council-rhc](http://www.gov.uk/government/groups/regulatory-horizons-council-rhc).

<sup>82</sup> For some ideas on the kinds of participatory mechanisms policymakers could use, please read Ada Lovelace Institute. (2021). *Participatory data stewardship*. Available at: [www.adalovelaceinstitute.org/report/participatory-data-stewardship](http://www.adalovelaceinstitute.org/report/participatory-data-stewardship).

Successful regulatory frameworks enable an ecosystem of governance and accountability by empowering regulators, civil society organisations, academics and members of the public to hold systems to account

and enable a healthy digital market. The Information and Commissioners Office seeks to address issues of data protection and privacy, while the Equalities and Human Rights Commission seeks to address fundamental human rights issues across the UK. AI systems can raise a variety of different risks which may fall under different regulatory bodies.

One major recommendation from workshop participants, and one evidenced in our research into assessment and auditing methods,<sup>83</sup> is that successful regulatory frameworks enable an ecosystem of governance and accountability by empowering regulators, civil society organisations, academics and members of the public to hold systems to account. The establishment of whistleblower laws, for example, can empower tech workers who identify inherent risks to come forward to a regulator.<sup>84</sup>

A regulatory framework might also enable greater access to assess a system's impacts and behaviour by civil-society organisations and academic labs, who are currently responsible for the majority of audits and assessments that have identified alarming AI-system behaviour. A regulatory framework that empowers other actors in the ecosystem can help remove the burden from individual regulators to perform these assessments entirely on their own.

---

## Regulatory approaches – risk-based approaches to regulating AI

In 2021, the European Commission released a draft risk-based framework to regulate AI systems that identifies what risk a system poses and assigns specific requirements for developers to meet based on that risk level.<sup>85</sup> Like the EU, the UK could consider adopting a risk-based approach to the regulation of AI systems, based on their impacts on society. Importantly, the levels of risk in the Commission's proposed framework are not based on the underlying technological

---

83 Ada Lovelace Institute and DataKind UK. (2020). *Examining the Black Box: Tools for Assessing Algorithmic Systems*. Available at: [www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/](http://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/) [accessed 11 October 2021].

84 Johnson, K. (2020). 'From whistleblower laws to unions: How Google's AI ethics meltdown could shape policy'. *VentureBeat*. Available at: <https://venturebeat.com/2020/12/16/from-whistleblower-laws-to-unions-how-googles-ai-ethics-meltdown-could-shape-policy>.

85 European Commission. (2021). *Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>.

If the UK chooses a risk-based approach to regulating AI, it should include a framework for defining and assessing risk that includes a participatory process

method used (for example, deep learning vs. reinforcement learning), but on the potential impact on ‘fundamental rights’.<sup>86</sup>

The EU model creates four tiers of risks posed by the use of AI in a particular context – unacceptable risk (uses that are banned), high, moderate and minimal risk. Each tier comes with specific requirements for developers of those systems to meet. High-risk systems, for example, must undergo a self-conformity assessment and be listed on a European-wide public register.

While the EU AI regulation states the protection of fundamental rights is a core objective, another clear aim of this regulation is to develop harmonised rules of AI regulation for all member states to adopt.

The proposed regulation seeks to ensure a consistent approach across all member states, and so pre-empt and overrule the development of national regulation of AI systems. To achieve this, it relies heavily on EU-standards bodies to establish specific requirements for certain systems to meet based on their risk category. As several academics have noted, these standards bodies are often inaccessible to civil-society organisations, and may be poorly suited for the purposes of regulating AI.<sup>87</sup>

A risk-based approach to regulating AI will ensure not all uses of AI are treated the same, which may help avoid unnecessary regulatory scrutiny and wasting of resources on uses of AI that are low risk.

However, risk-based systems of regulation come with their own challenges. One major challenge relates to the identification of risks.<sup>88</sup> How should a regulatory system determine what qualifies as a high-risk, medium-risk or low-risk application of a technology? Who gets to make this judgement, and according to what framework of risk? Risks are social constructs, and what may present a risk to one individual in society may benefit another. To mitigate this, if the UK chooses a risk-based approach to regulating AI, it should include a framework for defining and assessing risk that includes

86 Lum, K., and Chowdhury, R. (2021). ‘What is an “algorithm”? It depends whom you ask’. *MIT Technology Review* Available at: [www.technologyreview.com/2021/02/26/1020007/what-is-an-algorithm](https://www.technologyreview.com/2021/02/26/1020007/what-is-an-algorithm).

87 Veale, M., and Zuiderveen Borgesius, F. (2021). ‘Demystifying the Draft EU Artificial Intelligence Act’. *Computer Law Review International*. 22 (4). Available at: <https://osf.io/preprints/socarxiv/38p5f/>; Cath-Speth, C. (2021). Available at: [https://twitter.com/c\\_\\_\\_cs/status/1412457639611600900](https://twitter.com/c___cs/status/1412457639611600900).

88 Baldwin, R., and Black, J. (2016). Driving Priorities in Risk-Based Regulation: What’s the Problem? *Journal of Law and Society*. 43.4 pp. 565–95. Available at: <https://onlinelibrary.wiley.com/doi/10.1111/jols.12003>.

---

Mechanisms like impact assessments or participatory methods of citizen engagement offer a promising pathway for the UK to develop an effective tier-based system of regulation

a participatory process involving civil-society organisations and those who are likely to be affected by those systems.

Some AI systems are dynamic technologies that can be used in different contexts, so assessing the risk of a system – like an open-source facial recognition API – may miss the unique risks it poses when deployed in different contexts or for different purposes. For example, identifying the presence of a face for a phone camera will create different risks than if the system is used in the creation of a surveillance apparatus for a law-enforcement body. This suggests that there may need to be different mechanisms for assessing the risk of an AI system and its impacts at different stages of its ‘lifecycle’.

Some of the mechanisms described below have the potential to help both developers and regulators assess the risk of a system in early research and development stages, while others may be useful for assessing the risk of a system after it has been procured or deployed. Mechanisms like impact assessments or participatory methods of citizen engagement offer a promising pathway for the UK to develop an effective tier-based system of regulation that captures risk at different stages of an AI system’s lifecycle. However, more work is needed to determine the effectiveness of these mechanisms.

---

## Regulatory tools and techniques

This section provides some examples of mechanisms and tools for the regulation of AI that our expert participants discussed, and draws heavily on a recent report documenting the ‘first wave’ of public-sector algorithm accountability mechanisms.<sup>89</sup>

This section is not a holistic description of all the mechanisms that regulators might use – sandboxes, for example, are notably absent – but rather seeks to describe some existing, emerging mechanisms for AI systems that are less well-known, and provides some guidance

---

<sup>89</sup> Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). *Algorithmic Accountability for the Public Sector*. Available at: [www.opengovpartnership.org/documents/algorithmic-accountability-public-sector](http://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector).

for the UK Government when considering the forthcoming White Paper and in its forthcoming AI Assurance Roadmap.<sup>90</sup>

## Algorithmic impact assessments (AIAs)

To assess the potential impacts of an AI system on people and society, regulators will need new powers to audit, assess and inspect such systems. As the Ada Lovelace Institute's report *Examining the Black Box* notes, the auditing and assessment of AI systems can occur prior to a system's deployment and after its deployment.<sup>91</sup>

Impact assessments have a lengthy history of use in other sectors to assess human rights, equalities, data protection, financial and environmental impacts of a policy or technology *ex ante*. Their purpose is to provide a mechanism for holding developers and procurers of a technology more accountable for its impacts, by enabling greater external scrutiny of its risks and benefits.

Some countries and developers have begun to use algorithmic impact assessments (AIAs) as a mechanism to explore the impacts of an AI system prior to its use. AIAs offer a way for developers or procurers of a technology to engage members of affected communities about what impacts they might foresee an AI system causing, and to document potential impacts. They can also provide developers of a technology with a standardised mechanism for reflecting on intended uses and design choices in the early stages, enabling better organisational practices that can maximise the benefits of a system and minimise its harms. For example, the Canadian Directive on Automated Decision-Making is a public-sector initiative that requires federal public agencies to conduct an AIA prior to the production of an AI system.<sup>92</sup>

While there is no one-size-fits-all approach to conducting AIAs, recent research has identified ten constitutive elements to any AIA process that ensure meaningful accountability.<sup>93</sup> These include the establishment of a clear independent assessor, the public posting of the results of the AIA, and the establishment of clear methods of redress.

90 In its 2021 National AI Strategy, the UK Government states the Centre for Data Ethics and Innovation will publish a roadmap for 'AI Assurance' which sets out a number of different governance mechanisms and roles for different actors to play in holding AI systems more accountable.

91 Ada Lovelace Institute and DataKind UK. (2020). *Examining the Black Box: Tools for Assessing Algorithmic Systems*. Available at: [www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems](http://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems).

92 As of the date of this report, only two AIAs have been completed by Canadian federal agencies under this directive. Treasury Board of Canada Secretariat, Government of Canada. (2019). *Directive on Automated Decision Making*. Available at: [www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592](http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592).

93 Moss, E., Watkins, E.A., Singh, R., Elish, M.C., and Metcalf, J. (2021). *Assembling Accountability Through Algorithmic Impact Assessment*. Data & Society Research Institute. Available at: <http://datasociety.net/library/assembling-accountability>.

Conducting a meaningful regulatory inspection of an algorithmic system would require regulators to have powers to accumulate specific types of evidence

## Auditing and regulatory inspection

While impact assessments offer a promising method for an *ex ante* assessment of an AI system's impacts on people and society, auditing and regulatory inspection powers offer a related method to assess an AI system's behaviour and impacts *ex post* and over time.

Regulatory inspections are used by regulators in other sectors to investigate potentially harmful behaviours. Financial regulatory inspections, for example, enable regulators to investigate the physical premises, documents, computers and systems of banks and other financial institutions. Regulatory inspections of AI systems could involve the use of similar powers to assess a system's performance and accuracy, along with its broader impacts on society.<sup>94</sup>

Conducting a meaningful regulatory inspection of an algorithmic system would require regulators to have powers to accumulate specific types of evidence, including information on:

- **Policies** – company policies and documentation that identify the goals of the AI system, what it seeks to achieve, and where its potential weaknesses lie.
- **Processes** – assessment of a company's process for creating the system, including what methods they chose and what evaluation metrics they have applied.
- **Outcomes** – the ability to assess the outcomes of these systems on a range of different users of the system.<sup>95</sup>

Regulatory inspections may make use of technical audits of an AI system's performance or behaviour over a period of time. Technical-auditing methods can help to answer several kinds of questions relating to an AI system's behaviour, such as whether a particular system is producing biased outputs or what kind of content is being amplified to a particular user demographic by a social media platform.

94 Ada Lovelace Institute and DataKind UK. (2020). *Examining the Black Box: Tools for Assessing Algorithmic Systems*. Available at: [www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/](http://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/).

95 Ada Lovelace Institute and Reset. (2020). *Inspecting algorithms in social media platforms*. Available at: <https://www.adalovelaceinstitute.org/report/inspecting-algorithms-in-social-media-platforms/>.

---

A major challenge with regulating AI systems is the lack of transparency about where these systems are being used in both the public and private sectors

In order to conduct technical audits of an AI system, regulators will need statutory powers granting them the ability to access, monitor and audit specific technical infrastructures, code and data underlying a platform or algorithmic system. It should be noted that most technical auditing of AI systems is currently undertaken by academic labs and civil-society organisations, such as the Gender Shades audit that identified racial and gender biases in several facial-recognition systems.<sup>96</sup>

### Transparency requirements

Several expert participants noted a major challenge with regulating AI systems is the lack of transparency about where these systems are being used in both the public and private sectors. Without disclosure of the existence of these systems, it is impossible for regulators, civil-society organisations, or members of the public to understand what AI-based decisions are being made about them or how their data is being used.

This lack of transparency creates an inherent roadblock for regulators to assess the risk of certain systems effectively, and anticipate future risk down the line. A lack of transparency may also undermine public trust in institutions that use these systems, diminishing trust in government institutions and consumer confidence in UK businesses that use AI systems. The public outcry over the 2020 Ofqual A-level algorithm was in response to the deployment of an algorithmic system that had insufficient public oversight.<sup>97</sup>

### External-oversight bodies

Another mechanism a UK regulatory framework might consider implementing is a wider adoption of external-oversight bodies that review the procurement or use of AI systems in particular contexts. The West Midlands Police Department currently uses an external-ethics

---

96 Buolamwini, J. and Gebru, T. (2018). Gender shades: intersectional accuracy disparities in commercial gender classification.

In: *Conference on Fairness, Accountability, and Transparency*, 81, p1–15. New York: PLMR. Available at: <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.

97 Office for Statistics Regulation. (2021). *Ensuring statistical models command public confidence*. Available at: <https://osr.statisticsauthority.gov.uk/publication/ensuring-statistical-models-command-public-confidence/>.

---

The UK already exerts considerable influence over international standards in AI, but can and should aspire to do so more systematically

committee – consisting of police officials, ethicists, technologists and members of the local community – to review department requests to procure AI-based technologies, such as live facial-recognition systems and algorithms designed to predict an individual's likelihood to commit a crime.<sup>98</sup> While the committee's decisions are non-binding, they are published on the West Midlands Police website.

External-oversight bodies can also serve the purpose of ensuring a more participatory form of public oversight of AI systems. By enabling members of an affected community to have a say in the procurement and use of these systems, external-oversight bodies can ensure the procurement, adoption and integration of AI-systems is carried out in accordance with democratic principles. Some attempts to create external-oversight bodies have been in bad faith, and these types of bodies must be given meaningful oversight and fair representation if they are to succeed.<sup>99</sup>

## Standards

In addition to laws and regulatory rules, standards for AI systems, products and services have the potential to form an important component of the overall governance of the technology.

One notable potential use of standards is around improving the transparency and explainability of AI systems. Regulators could develop standards, or standards for tools, to ensure data provenance (knowing where data came from), reproducibility (being able to recreate a given result) and data versioning (saving snapshot copies of the AI in specific states with a view to recording which input led to which output).

At an international level, the UK AI Strategy states that the UK must get more engaged in international standard-setting initiatives,<sup>100</sup> a conclusion that many expert participants also agreed with. The UK already exerts considerable influence over international standards on AI, but can and should aspire to do so more systematically.

---

98 West Midlands Police and Crime Commissioner (2021). *Ethics Committee*. Available at: [www.westmidlands-pcc.gov.uk/ethics-committee/](http://www.westmidlands-pcc.gov.uk/ethics-committee/).

99 Richardson, R. ed. (2019). *Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force*. AI Now Institute. Available at: <https://ainowinstitute.org/ads-shadowreport-2019.html>.

100 Office for AI. (2021). National AI Strategy. UK Government. Available at: <https://www.gov.uk/government/publications/national-ai-strategy>.

At a domestic level, the UK could enforce specific standards of practice around the development, use and procurement of AI systems by public authorities. The UK Government has developed several non-binding guidelines around the development and use of data-driven technologies, including the UK's Data Ethics Framework that guides responsible data use by public-sector organisations.<sup>101</sup> Guidelines and principles like these can help developers of AI systems identify what kinds of approaches and practices they should use that can help mitigate harms and maximise benefits. While these guidelines are currently voluntary, and are largely focused on the public sector, the UK could consider codifying them into mandatory requirements for both public- and private-sector organisations.

A related mechanism is the development of standardised public procurement requirements that mandate developers of AI systems undertake certain practices. The line between public and private development of AI systems is often blurry, and in many instances public-sector organisations procure AI systems from private agencies who maintain and support the system. Local authorities in the UK often procure AI systems from private developers, including for many high-stakes settings like decisions around border control and the allocation of state benefits.<sup>102</sup>

Procurement agreements are a crucial pressure point where public agencies can place certain requirements around data governance, privacy and assessing impacts on a developer. The City of Amsterdam has already created standardised language for this purpose in 2020. Called the 'Standard Clauses for Municipalities for Fair Use of Algorithmic Systems', this language places certain conditions on the procurement of data-driven systems, including that underlying data quality of a system is assessed and checked.<sup>103</sup> The UK might therefore consider regulations that codify and enforce public-procurement criteria.

<sup>101</sup> Central Digital and Data Office. (2018). *Data Ethics Framework*. UK Government. Available at: [www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework](http://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework).

<sup>102</sup> BBC News. (2020). 'Home Office Drops "racist" Algorithm from Visa Decisions'. 4 August. Available at: [www.bbc.com/news/technology-53650758](http://www.bbc.com/news/technology-53650758); BBC News. (2021). 'Council Algorithms Mass Profile Millions, Campaigners Say'. 20 July. Available at: [www.bbc.com/news/uk-57869647](http://www.bbc.com/news/uk-57869647).

<sup>103</sup> Municipality Amsterdam. (2020). *Standard Clauses for Municipalities for Fair Use of Algorithmic Systems*. Gemeente Amsterdam. Available at: [www.amsterdam.nl/innovatie/](http://www.amsterdam.nl/innovatie/).

---

The UK's regulatory system will need to get the balance between standards and rules right

Despite the importance of standards in any regulatory regime for AI, they have several important limitations when it comes to addressing the challenges posed by AI systems. First, standards tend to be developed through consensus, and are often developed at an international level. As such, they can take a very long time to develop and modify. A flexible regulatory system capable of dealing with issues that arise quickly or unexpectedly, should therefore avoid overreliance on standards, and will need other means of addressing important issues in the short term.

Moreover, standards are not especially well-suited to dealing with considerations of important and commonly held values such as agency, democracy, the rule of law, equality and privacy. Instead they are typically used to moderate the safety, quality and security of products. While setting standards on AI transparency and reporting could be instrumental in enabling regulators to understand the ethical impacts of AI systems, the qualitative nature of broader, values-based considerations could make standards poorly suited to addressing such questions directly.

It will therefore be important to avoid overreliance on standards, instead seeing them as a necessary but insufficient component of a convincing regulatory response to the challenges posed by AI.

The UK's regulatory system will need to get the balance between standards and rules right, and will need to be capable of dealing with issues pertaining to ethical and societal questions posed by AI as well as questions of safety, quality, security and consumer protection. Equally, it will be important for the regulatory system to have mechanisms to respond to both short- and long-term problems presented by AI systems.

Though standards do have the potential to improve transparency and explainability, some participants in our expert workshops noted that the opaque nature of some AI systems places hard limits on the pursuit of transparency and explainability, regardless of the mechanism used to pursue these goals. Given this, it was suggested that the regulatory system should place more emphasis on methods that sidestep the problem of explainability, looking at the outcomes of AI systems, rather than the processes by which those outcomes are achieved.<sup>104</sup>

---

<sup>104</sup> This is a relatively common approach taken by regulators currently, who understandably do not want to, or feel under-qualified to get into the business of auditing code. A difficulty with this approach is that the opacity of AI systems can make it difficult to predict and assess the outcomes of their use in advance. As a result, 'outcomes-based' approaches to regulating AI need to be grounded in clear accountability for AI decisions, rather than attempts to configure AI systems to produce more desirable outcomes.

---

Establishing professional standards could offer a way for regulators to enforce and incentivise particular governance practices

A final caveat concerning standards is that standard setting is also currently heavily guided and influenced by industry groups, with the result that standards tend to be developed with a particular set of concerns and in mind.

Standards could potentially be a more useful complement to other regulatory and governance activity were their development to be influenced by a broader array of actors, including civil-society groups, representatives of communities particularly affected by AI, academics and regulators themselves. Should the UK become more actively involved in standard setting for AI systems, this would present a good opportunity to bring a greater diversity of voices and groups to the table.

## Professionalisation

Another suggested mechanism by which the UK regulatory system could seek to address the risks and harms posed by AI systems was the pioneering of an ethical-certification and training framework for those people designing and developing AI systems. Establishing professional standards could offer a way for regulators to enforce and incentivise particular governance practices, giving them more enforcement ‘teeth’.

There are several important differences between AI as a sector and domain of practice, and some of the sectors where training and professional accreditation have proven the most successful, such as medicine and the law. These professionalised fields have a very specific domain of practice, the boundaries of which are clear and therefore easy to police. There are also strong and well established social, economic and legal sanctions for acting contrary to a professional code of practice.

Some expert panellists argued there is potentially a greater degree of tension between the business models for AI development and potential contents of an ethical certification for AI developers. Some expert participants noted that the objections to certain AI systems lie not in how they are produced but in their fundamental business model, which may rely on practices like the mass collection of personal data or the development of mass-surveillance systems that some may see as objectionable. This raises questions about the scope and limits of professionalised codes of practice and how far they might be able to help.

---

One challenge with establishing bans and moratoria for certain technological uses is the necessity of developing a process for assessing the risks and benefits of these technologies

Another common concept when discussing the professionalisation of the AI industry is that of fiduciary duties, which oblige professionals to act solely in the best interest of a client who has placed trust and dependence in them. However, some expert participants pointed out that though this model works well in industries like law and finance, it is less readily applicable to data-driven innovation and AI, where it is not the client of the professional who is vulnerable, but the end consumer or subject of the product being developed. The professional culture of ethics exemplified by the fiduciary duty exists within the context of particular, trusting relationship between professional and client which isn't mirrored in most AI business models.

## Moratoria and bans

In response to worries about instances in which it may be impossible for regulators to assure themselves that they can successfully manage the harms posed by high-risk applications of AI, it may be desirable for the UK to refrain entirely from the development or deployment of particular kinds of AI technology, either indefinitely or until such a time as risks and potential mitigations are better understood.

Facial recognition was cited by our expert workshop participants as an example of a technology that, in some forms, could pose sufficiently grave risks to an open and free society as to warrant being banned outright – or at the very least, being subjected to a moratorium. Other countries, including Morocco, have put in place temporary moratoria on the use of these kinds of systems until existing legal frameworks can be established.<sup>105</sup> Similar bans exist on city uses of facial recognition in the US cities of Portland and San Francisco, though these have come with some criticism around their scope and effectiveness.<sup>106</sup>

One challenge with establishing bans and moratoria for certain technological uses is the necessity of developing a process for assessing the risks and benefits of these technologies, and endowing a regulator

---

<sup>105</sup> National Control Commission for the Protection of Personal Data. (2020). *Press release accompanying the publication of deliberation No. D-97-2020 du 26/03/2020* (in French). Available at: <https://www.cndp.ma/fr/presse-et-media/communique-de-presse/661-communique-de-presse-du-30-03-2020.html> [accessed 22 October 2021].

<sup>106</sup> Simonite, T., and Barber, G. (2019). 'It's Hard to Ban Facial Recognition Tech in the iPhone Era'. *Wired*. Available at: [www.wired.com/story/hard-ban-facial-recognition-tech-iphone](http://www.wired.com/story/hard-ban-facial-recognition-tech-iphone).

with the power to enact these restrictions. Currently, the UK has not endowed any regulator with explicit powers to make these bans of AI systems, nor with the capacity to develop a framework for assessing in which contexts certain uses of a technology would be worthy of a ban or moratoria. If the UK is to consider this mechanism, one initial step would be to develop a framework for the kinds of systems that may meet an unreasonable bar of risk.

Another worry expressed by some expert participants was whether bans and moratoria could end up destroying the UK's own research and commercial capacity in a particular emerging technological field. Would a ban on facial-recognition systems, for example, be overly broad and risk creating a chilling effect on potential positive uses of the underlying technology?

Other expert participants were far less concerned with this possibility, and argued that bans and moratoria should focus on specific uses and outcomes of a technology rather than its underlying technique. A temporary moratoria could be restricted to specific high-risk applications that require additional assessment of their effectiveness and impact, such as the use of live facial recognition in law enforcement settings. In the UK, current bans and moratoria on live facial recognition have been dealt with by court challenges like the recent decision on the New South Wales Police use of the technology.<sup>107</sup>

---

<sup>107</sup> Courts and Tribunals Judiciary. (2020). *R (on the application of Edward Bridges) v. The Chief Constable of South Wales Police and the Secretary for the State for the Home Department*. Case No: C1/2019/2670. Available at: <https://www.judiciary.uk/wp-content/uploads/2020/08/R-Bridges-v-CC-South-Wales-ors-Judgment.pdf> [accessed 22 October 2021].

---

Done properly, AI regulation is a prerequisite, rather than an impediment to the development of a flourishing UK AI ecosystem

# Chapter 4: Considerations for policymakers

This section sets out some general considerations for policymakers, synthesised from our expert workshops and the Ada Lovelace Institute's own research and deliberations. These are not intended to be concrete policy recommendations (see chapter 5, p. 70), but are general lessons about the parameters within which the Government's approach to AI regulation and governance will need to be developed, and the issues that need to be addressed with the current regulatory system.

In summary, policymakers should consider the following:

1. Government ambitions for AI will depend on the stability and certainty provided by robust, AI-specific regulation and law.
2. High regulatory standards and innovative, flexible regulatory processes will be critical to supporting AI innovation and use.
3. A critical challenge with regulating AI systems is that risks can arise at various stages of an AI system's development and deployment.
4. The UK's approach to regulation could involve a combination of a unified approach to the governance of AI, with new, cross-cutting rules set out in statute, and sectoral approaches to regulation.
5. Substantial regulatory capacity building will be unavoidable.
6. Promising regulatory approaches and tools will need to be refined and embedded into regulatory systems and structures.
7. New tools need to be 'designed into' the regulatory system.

1. Government ambitions for AI will depend on the stability and certainty provided by robust, AI-specific regulation and law

One of the clearest conclusions to be drawn from the considerations in the previous two sections is that, done properly, AI regulation is a prerequisite, rather than an impediment to the development of a flourishing UK AI ecosystem.

---

The Government's approach to the governance and regulation of AI needs to be attentive to the specific features and potential impact of the technology

Government ambitions to establish the UK as a 'science superpower' and use emerging technologies such as AI to drive broadly felt, geographically balanced economic growth will rely on the ability of the UK's regulatory system to provide stability, certainty and continued market access for innovators and businesses, and accountability and protection from harms for consumers and the public.

In particular, without the confidence, guidance and support provided by a robust regulatory system for AI, companies and organisations developing AI or looking to exploit its potential will have to grapple with the legal and ethical ramifications of systems on their own. As AI systems become more complex and capable – and as a greater variety of entities look to develop or make use of them – the existence of clear regulatory rules and a well-resourced regulatory ecosystem will become increasingly important in de-risking the development and use of AI, helping to ensure that it is not just large incumbents that are able to work with the technology.

Critically, the Government's approach to the governance and regulation of AI needs to be attentive to the specific features and potential impacts of the technology. Rather than concentrating exclusively on increasing the rate and extent of AI development and diffusion, the UK's approach to AI regulation must also be attentive to the particular ways the technology might manifest itself, and the specific effects it stands to have on the country's economy, society and power structures.

In particular, a strategy for AI regulation needs to be designed with the protection and advancement of important and commonly held values, such as agency, human rights, democracy, the rule of law, equality and privacy, in mind. The UK's AI Strategy already makes reference to some of these values, but a strategy for regulation must provide greater clarity on how these should apply to the governance of AI systems.

## 2. High regulatory standards and innovative, flexible regulatory processes will be critical to supporting AI innovation and use

In practice, creating the stability, certainty and continued market access needed to cultivate AI as a UK strength will require the Government to commit to developing and maintaining high, flexible regulatory standards for AI.

As observed by our workshop panellists, there is limited scope for the UK to develop more permissive regulatory standards than its close allies and neighbours, such as the USA and the European Union. Notably, as well as undermining public confidence in a novel and powerful technology, aspiring to regulatory standards that are lower than those of the European Union would deprive UK-based AI developers of the ability to export their products and services not only to the EU, but to other countries likely to adopt or closely align with the bloc's regulatory model.

There are, nonetheless, significant opportunities for the UK to do AI regulation differently to, and more effectively than, other countries. While the UK will need to align with its allies on regulatory *standards*, the UK is in a good position to develop more flexible, resilient and effective regulatory *processes*. The UK has an excellent reputation and track record in regulatory innovation, and the use of flexible, pragmatic approaches to monitoring and enforcement. This expertise, which has in part contributed to British successes in fields such as bioscience and fintech, should be leveraged to produce a regulatory ecosystem that supports and empowers businesses and innovators to develop and exploit the potential of AI.

### 3. A critical challenge with regulating AI systems is that risks can arise at various stages of an AI system's development and deployment

Unlike most other technologies, AI systems can raise different kinds of risks at different stages of a system's development and deployment. The same AI system applied in one setting (such as a facial scan for authenticating entry to a private warehouse) can raise significantly different risks when applied in another (such as authenticating entry to public transport). Similarly, some AI systems are dynamic, and their impacts can change drastically when fed new kinds of data or when deployed in a different context. An *ex ante* test of a system's behaviour in 'lab' settings may therefore not provide an accurate assessment of that system's actual impacts when deployed 'in the wild'.

Many of the proposed models for regulating AI focus either on *ex ante* assessments that classify an AI system's risk, or *ex post* findings of harm in a court of law. One option the UK might consider is an approach to AI regulation that includes regulatory attention at all stages of an AI system's development and deployment. This may, for example, involve

using *ex ante* algorithmic impact assessments (AIAs) of a system's risks and benefits pre-deployment, along with post-deployment audits of that system's behaviour.

If the UK chooses to follow this model, it will have to provide regulators with the necessary powers and capacity to undertake these kinds of holistic regulatory assessments. The UK may also consider delegating some of these responsibilities to independent third parties, such as algorithmic-auditing firms.

#### 4. The UK's approach to regulation could involve a combination of a unified approach to the governance of AI, with new, cross-cutting rules set out in statute, and sectoral approaches to regulation

A common challenge raised by our expert participants was whether the UK should adopt a unified approach to regulating AI systems involving a central function that oversees all AI systems, or if regulation should be left to individual regulators who approach these issues on a sectoral or case-by-case basis.

One approach the UK Government could pursue is a combination of the two. While individual regulators can and should develop domain- and sector-specific regulatory rules for AI, there is also a need for a more general, overarching set of rules, which outline if and under what circumstances the use of AI is permissible. The existence of such general rules is a prerequisite for a coherent, coordinated regulatory and legal response to the challenges posed by AI.

If they are to provide the stability, predictability and confidence needed for UK to get the most out of AI, these new, AI-specific regulatory rules will probably have to be developed and set out in statute.

The unique capacity of AI systems to develop and change independently of human control and intervention means that existing legal and regulatory rules will be likely to prove inadequate. While the UK's common-law system may develop to accommodate some of these features, this will only happen slowly (if it happens at all) and there is no guarantee the resulting rules will be clear or amount to a coherent response to the technology.

---

The successful management of AI will require a sustained and significant expansion of the regulatory system's overall capacity and levels of coordination

## 5. Substantial regulatory capacity building will be unavoidable

The successful management of AI will require a sustained and significant expansion of the regulatory system's overall capacity and levels of coordination.

There are several viable options for how to organise and allocate additional regulatory capacity, and to improve the ability of regulators to develop sector-specific regulatory rules that amount to a coherent whole. Regardless of the specific institutional arrangements, any capacity building and coordination efforts must ensure that:

1. additional resources can be allocated without too much duplication of effort, and that gaps and blind spots in the regulatory system are avoided
2. regulators are able to understand how their responses to AI within their specific domains contribute to the broader regulatory environment, and are provided with clear guidance on how their policies can be configured to complement those of other regulators
3. regulators are able to easily share intelligence and jointly conduct horizon-scanning exercises.

## 6. Promising regulatory approaches and tools will need to be refined and embedded into regulatory systems and structures

There are a number of tools and mechanisms that already exist, or that are currently being developed, that could enable regulators to effectively rise to the challenges presented by AI – many of which were pioneered by UK entities.

These include tools of so called 'anticipatory regulation', such as regulatory sandboxes, regulatory labs and coordinated horizon scanning and foresight techniques, as well as deliberative mechanisms for better understanding informed public opinion and values regarding emerging technologies, such as deliberative polling, citizens' juries and assemblies.

Some of these tools are still emerging and should be tested further to determine their value, such as the use of transparency registers to disclose where AI systems are in operation, or algorithmic impact

---

There needs to be a substantial, long-term commitment to supporting regulatory innovation and experimentation

assessments to provide an *ex ante* assessment of an AI system's benefits and harms. While many of the above tools have the potential to prove invaluable in helping regulators and lawmakers rise to the challenges presented by AI, many are still nascent or have only been used in limited circumstances. Moreover, many of the tools needed to help regulators address the challenges posed by AI do not yet exist.

To ensure that regulators have the tools they require, there needs to be a substantial, long-term commitment to supporting regulatory innovation and experimentation, and to supporting the diffusion of the most mature, proven techniques throughout the regulatory ecosystem. This ongoing experimentation will be crucial to ensure that the regulatory system does not become overly dependent on particular kinds of regulatory interventions, but instead has a toolkit that allows it to respond quickly to emerging harms and dangers, as well as being able to develop more nuanced and durable rules and standards in the longer term.

## 7. New tools need to be 'designed into' the regulatory system

As well as helping cultivate and refine new regulatory tools and techniques, further work is required to understand how regulatory structures and processes might be configured to best enable them.

This is particularly true of anticipatory and participatory mechanisms. The value of techniques like sandboxing, horizon scanning and citizen juries is unlikely to be fully realised unless the insight gained from these activities is systematically reflected in the development and enforcement of broader regulatory rules.

A good example of how these tools are likely to be most useful if 'designed into' regulatory systems and processes is provided by risk-based regulation. Given the variety of applications of AI systems, the UK may choose to follow the approach of the European draft AI regulation and adopt some form of risk-based regulation to prevent gross over or under regulation of AI systems. However, if such an approach is to avoid creating gaps in the regulatory system, in which harmful practices escape appropriate levels of regulatory scrutiny, the system's ability to make and review judgements about which risk categories different AI systems should fall into will need to be improved.

One element of this will be using anticipatory mechanisms to help predict harms and unintended consequences that could arise from different uses of AI. Participatory mechanisms that involve regulators working closely with local-community organisations, members of the public and civil society may also help regulators identify and assess risks to particular groups.

Perhaps the bigger challenge, though, will be to design processes by which the risk tiers into which different kinds of systems fall are regularly reviewed and updated, so that AI systems whose risk profiles may change over time do not end up being over or under regulated.

---

In the coming months, we encourage the Office for AI to engage closely with members of the public, academia, civil society and regulators to further develop these ideas

# Chapter 5: Open questions for Government

This section sets out a series of open questions that we believe the White Paper on AI regulation and governance should respond to, before making a series of more specific recommendations about things that we believe it should commit to.

We acknowledge that these open questions touch on complex issues that cannot be easily answered. In the coming months, we encourage the Office for AI to engage closely with members of the public, academia, civil society and regulators to further develop these ideas.

## Open questions

AI systems present a set of common, novel regulatory challenges, which may manifest differently in different domains, and which demand holistic solutions. A coherent regulatory response to AI systems therefore requires a combination of general, cross-cutting regulatory rules and sector-specific regulations, tailored to particular uses of AI.

Finding the right balance between these two will depend on how the UK chooses to answer several open questions relating to the regulation of AI. A more detailed discussion around some of these questions, along with other considerations when designing and configuring regulatory systems, can be found in the annex, [\(p. 81\)](#).

## What to regulate?

First, the UK Government must determine what kinds of AI systems it seeks to regulate, and what definition it will use to classify AI systems appropriately. Some possible options include:

1. **Regulating all AI systems equally.** Anything classified as an 'AI system' must follow common rules. This may require the UK choosing a more precise definition of 'AI system' to ensure particular

kinds of systems (such as those used to augment or complement human decision-making) are included. This may prove resource intensive both for regulators and for new entrants seeking to build AI, but this approach could ensure no potentially harmful system avoids oversight.

2. **Regulating higher-risk systems.** This would involve creating risk tiers and regulating ‘higher-risk’ systems more intensely than lower risk, and could involve the UK adopting a broader and more encompassing definition of AI systems. A challenge with risk-based approaches to regulation comes in identifying and assessing the level of risk, particularly when risks for some members of society may be benefits for others. The UK could consider assigning risk tiers in a number of ways, including:
  - a. **Enumerating certain domains** (such as credit scoring, or public services) that are inherently higher risk.<sup>108</sup> This approach could be easily bypassed by a developer seeking to classify their system in a different domain, and it may not capture ‘off-label’ uses of a system that could have harmful effects.
  - b. **Enumerating certain uses** (such as facial-recognition systems that identify people in public places) as higher risk. This approach could also be easily bypassed by a developer who reclassifies the use of their system, and would require constant updating of new high-risk uses and a process for determining that risk.
  - c. **Enumerating certain criteria** for assigning higher risk. These could include *ex ante* assessments of the foreseeable risk of a system’s intended and reasonably likely uses, along with *ex post* assessments of a system’s actual harms over time.

---

<sup>108</sup> This is the approach the EU’s Draft AI Regulation takes. See Annex III of the European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021) 206 final).

## Who to regulate?

The UK Government must similarly choose who is the focus of AI regulation. This could include any of the following actors, with different obligations and requirements applying to each one:

1. **Developers:** Those who create a system. Regulatory rules that enforce *ex ante* requirements about a system's design, intended use or oversight could be enforced against this group.
2. **Adapters:** A sub-developer who creates an AI system based on building blocks provided by other developers. For example, a developer who uses the Google Cloud ML service, which provides machine-learning models for developers to use, could be classified as an adapter. Similarly, a developer who utilises 'foundation' models like OpenAI's GPT-3 to train their model could be classified as an adapter.<sup>109</sup>
3. **Deployers:** The person who is responsible for putting a system into practice. While a deployer may have procured this system from a developer, they may not have access to the source code or data of that system.<sup>110</sup>

## How and when to regulate?

Part of the challenge with regulating AI systems is that risks and harms may arise in different stages of a product's lifecycle. Addressing this challenge requires a combination of both *ex ante* and *ex post* regulatory interventions. Some options the UK Government could consider include:

1. ***Ex ante* criteria that all AI systems must meet.** These could be both technical requirements around the quality of datasets an AI system is trained on, along with governance requirements including

---

<sup>109</sup> For a discussion about the opportunities and risks of 'foundation models,' see Bommasani, R., et al. (2021). *On the opportunities and risks of foundation models*. Stanford FSI. Available at: <https://fsi.stanford.edu/publication/opportunities-and-risks-foundation-models>.

<sup>110</sup> The EU's Draft AI regulation attempts to distinguish between developers and 'users,' a term that can be confused with those who are subject to an AI system's decisions. See Smuha, N. et al. (2021). *How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act*. Available at SSRN: <https://ssrn.com/abstract=3899991> or <http://dx.doi.org/10.2139/ssrn.3899991>.

documentation standards (such as the use of model cards) and bias assessments. A regulatory system could ensure developers of an AI system meet these requirements through either:

- a. **Self-certification:** A developer self-certifies they are meeting these requirements. This raises a risk of certification becoming a checkbox exercise that is easily gameable.<sup>111</sup>
  - b. **Third-party certification:** The UK Government could require developers to obtain a certification from a third-party, either a regulator or Government-approved independent certifier. This could enable more independent certification, but may become a barrier for smaller firms.
2. ***Ex ante* sectoral codes of practice.** Certain sectors may choose to implement additional criteria on an AI system before it enters the market. This may be essential for certain sectors like healthcare that require additional checks for patient safety and operability of a system. This could include checks about how well a system has been integrated into a particular environment, or checks on how a system is behaving in a sandbox environment.
  3. ***Ex post* auditing and inspection requirements.** Regulators could evaluate the actual impacts and risks of a system post-deployment by inspecting and auditing its behaviour. This may require expanding on existing multi-regulator coordination efforts like the Digital Regulation Cooperation Forum to identify gaps and share information, and to create longitudinal studies on the risk and behaviour of an AI system over time.
  4. **Novel forms of redress.** This could include the creation of an ombudsman or form of consumer champion for intakeing and raising complaints about an AI system on behalf of people and society, and ensuring the appropriate regulator has dealt with them.

---

<sup>111</sup> The EU's proposed regulations follow this same approach.

# Chapter 6: Recommendations for the Government's White Paper on AI regulation

With the above open questions in mind, we recommend the Government focuses on taking action in the following three areas in their forthcoming White Paper on AI regulation:

1. The development of new, clear regulations for AI.
  2. Improved regulatory capacity and coordination.
  3. Improving transparency standards and accountability mechanisms.
- 

## 1. The development of new, clear regulations for AI

### **Recommendation 1:**

**The Government should establish a clear definition of AI systems that matches their overall approach towards regulation.**

How broad and encompassing this definition may be will depend on what kind of regulatory approach the Government chooses (for example, risk-based vs all-encompassing), what criteria the Government chooses to trigger intervention (such as systems they classify as 'high risk' vs 'low risk') and which actors the Government chooses to target regulation at (such as the developers of AI or the deployers).

- In their White Paper, the Government should explore the possibility of combining sectoral and risk-based approaches, and should commit to engaging with civil society on these questions.
- The Government should commit to ensuring the definition and approach to AI they choose will be subject to parliamentary scrutiny.

## Recommendation 2:

**Government should consider creating a central function to oversee the development and implementation of AI-specific, domain-neutral statutory rules for AI systems.** These rules should be subject to regular parliamentary scrutiny.

These domain-neutral statutory rules could:

- **set out consistent ways for regulators to approach common challenges posed by AI systems** (such as accountability for automated decision-making, the encoding of contestable, value-laden judgements into AI systems, AI bias, the appropriate place for human oversight and challenge of AI systems, the problems associated with understanding, trusting and making important choices on the basis of opaque AI decision-making processes). The proposed approaches should be rooted in legal concepts and ethical values such as fairness, liberty, agency, human rights, democracy and the rule of law.

The specific understanding of these concepts and values should be informed not just by the existing discourse on AI ethics, but also by engagement with the public. The Government should commit to co-developing these rules with members of the public, civil society and academia. These rules should:

- **include and set out a requirement for, and mechanism by which the central function must regularly revisit the definition of AI, the criteria for regulatory intervention and the domain-neutral rules themselves.** The central function should be required to provide an annual report to Parliament on the status and operation of these rules.
- **provide a means of requiring individual regulators to attend to, and address the systemic, long-term impacts of AI systems.** While the regulatory system as a whole is a potentially critical lever in addressing them, many of the most significant impacts of AI systems – such as how they affect democracies and alter the balance of power between different groups in society – are not covered by the narrow, domain-bounded remits of individual regulators. The provision of domain-neutral rules for AI regulation would be one way to require and mandate individual regulators to make regulatory decisions with a view to addressing these larger, more systemic issues – and could be a way of guiding regulators to do so in a coordinated manner.

- **provide a means for regulators to address all stages of an AI system's lifecycle**, from research to product development to procurement and post-deployment. This would require regulators to use *ex ante* regulatory mechanisms (such as impact assessments) to assess the potential impacts of an AI system on people and society, along with *ex post* mechanisms (such as regulatory inspections and audits) to determine the actual impact of an AI system's behaviour on people and society. Regulators could also be required to use anticipatory methods to assess the potential future risks posed by AI systems in different contexts.
- **be intended to supplement, rather than replace, existing laws governing AI systems.** These rules should complement existing health and safety, consumer protection, human rights and data-protection regulations and law.

In addition to developing and updating the domain-neutral rules, the central function could be responsible for:

- **leading cross-regulatory coordination** on the regulation of AI systems, along with cross-regulatory horizon-scanning and foresight exercises to provide intelligence on potential harms and challenges posed by AI systems that may require regulatory responses
- **monitoring common challenges with regulating AI** and, where there is evidence of problems that require new legislation, **making recommendations to Parliament to address gaps in the law.**

### **Recommendation 3:**

**Government should consider requiring regulators to develop sector-specific codes of practice for the regulation of AI.**

These sector-specific codes of practices would:

- **lay out a regulator's approach to setting and enforcing regulatory rules covering AI systems in particular contexts or domains**, as well as the general regulatory requirements placed on developers, adapters and deployers of those systems

---

Recommendations 2 and 3 could individually each bring benefits to the regulatory system's capacity to deal with the challenges posed by AI, we believe that they would be most beneficial if implemented together

- **be developed and maintained by individual regulators**, who are best placed to understand the particular ways in which AI systems are deployed in regulatory domains, the risks involved in those deployments, their current and future impacts, and the practicality of different regulatory interventions
- **be subject to regular review** to ensure that they keep pace with developments in AI technologies and business models.

### Potential synergy between recommendations 2 and 3

While recommendations 2 and 3 could individually each bring benefits to the regulatory system's capacity to deal with the challenges posed by AI, we believe that they would be most beneficial if implemented together, enabling a system in which cross-cutting regulatory rules inform and work in tandem with sector-specific codes of practice.

On p. 80 we illustrate one potential way that the **central function**, **domain-neutral statutory rules** and **sector-specific codes of practice** could be combined to improve the coordination and responsiveness of the regulatory system with regards to AI systems.

On this model:

- The **central function** would create **domain-neutral statutory rules**.
- Individual regulators would be required to take the **domain-neutral statutory rules** into account when developing and updating the **sector-specific codes of practice**. These sector-specific codes of practice would apply the domain-neutral statutory rules to specific kinds of AI systems, or the use of those systems in specific contexts. These codes of practice should include enforcement mechanisms that address all stages of an AI system's lifecycle, including *ex ante* assessments like impact assessments and *ex post* audits of a system's behaviour.
- Careful adherence to the **domain-neutral statutory rules** when developing the **sector-specific codes of practice** would help ensure that the multiple different AI codes of practice, developed across different regulators, all approached AI regulation with the same high-level goals in mind.

- The **central function** would have a duty to advise and work with individual regulators on how best to interpret the **domain-neutral statutory rules** when developing **sector-specific codes of practice**.
- 

## 2. Improved regulatory capacity and coordination

AI systems are often complex, opaque and straddle regulatory remits. For the regulatory system to be able to deal with these challenges, significant improvements will need to be made to regulatory capacity (both at the level of individual regulators and the whole regulatory system) and to improve coordination and knowledge sharing between regulators.

### **Recommendation 4:**

**Government should consider expanded funding for regulators to deal with analytical and enforcement challenges posed by AI systems.** This funding will support building regulator capacity and coordination.

### **Recommendation 5:**

**Government should consider expanded funding and support for regulatory experimentation, and the development of anticipatory and participatory capacity within individual regulators.** This will involve bringing in new forms of public engagement and futures expertise.

### **Recommendation 6:**

**Government should consider developing formal structures for capacity sharing, coordination and intelligence sharing between regulators dealing with AI systems.**

These structures could include a combination of several different models, including centralised resources of AI knowledge, experts rotating between regulators and the expansion of existing cross-regulator forums like the Digital Regulation Cooperation Forum.

**Recommendation 7:**

**Government should consider granting regulators the powers needed to enable them to make use of a greater variety of regulatory mechanisms.**

These include providing statutory powers for regulators to engage in regulatory inspections of different kinds of AI systems. The Government should commission a review of the powers different regulators will need to conduct *ex ante* and *ex post* assessments of an AI system before, during, and after its deployment.

---

### 3. Improving transparency standards and accountability mechanisms

The impacts of AI systems may not always be visible to, or controllable by policymakers and regulators alone. This means that regulation and regulatory intelligence gathering will need to be complemented by and coordinated with extra-regulatory mechanisms, such as standards, investigative journalism and activism.

**Recommendation 8:**

**Government should consider how best to use the UK's influence over international standards to improve the transparency and auditability of AI systems.**

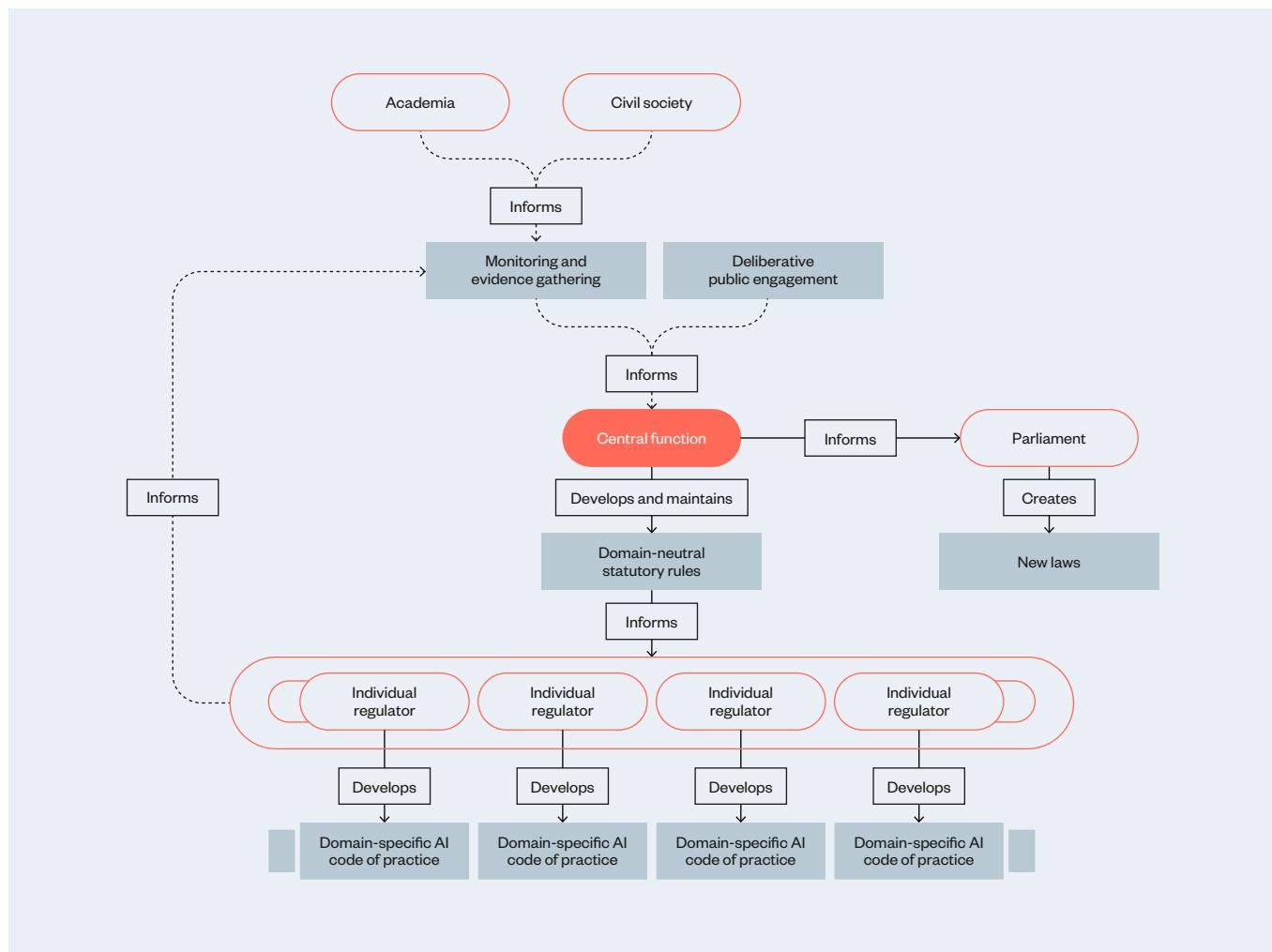
While these are not a silver bullet, they can help ensure the UK's approach to regulation and governance remains interoperable with approaches in other regions.

### Recommendation 9:

**Government should consider how best to maintain and strengthen laws and mechanisms to protect and enable journalists, academics, civil-society organisations, whistleblowers and citizen auditors to hold developers and deployers of AI systems to account.**

This could include passing novel legislation to require the disclosure of AI systems when in use, or requirements for AI developers to disclose data around systems' performance and behaviour.

### A potential model for horizontal and vertical regulation of AI



---

This section sets out some important variables in the design of regulatory systems, and how they might apply specifically to the regulation of AI

## Annex: The anatomy of regulatory rules and systems, and how these apply to AI

To explore how the UK's regulatory system might adapt to meet the needs of the Government's ambitions for AI, it is useful to consider ways in which regulatory systems (and sets of regulatory rules) can vary.

This section sets out some important variables in the design of regulatory systems, and how they might apply specifically to the regulation of AI. It is adapted from a presentation given at the second of the expert workshops, by Professor Julia Black, who has written extensively on this topic.<sup>112</sup>

The following section addresses the challenges some of these variables may pose for the regulation of AI.

### Why to regulate: The underlying aims of regulation

Regulatory systems can vary in terms of their underlying aims. Regulatory systems may have distinct, narrowly defined aims (such as maximising choice and value for consumers within a particular market, or ensuring a specific level of safety for a particular category of product), and may also have been driven by different broader objectives.

In the context of the regulation of AI, some of the broader values that could be taken into consideration by a regulatory system might include economic growth, the preservation of privacy, the avoidance of concentrations of market power and distributional equality.

---

<sup>112</sup> Black, J., and Murray, A. D. (2019). 'Regulating AI and machine learning: setting the regulatory agenda'. *European Journal of Law and Technology*, 10 (3). Available at: [http://eprints.lse.ac.uk/102953/4/722\\_3282\\_1\\_PB.pdf](http://eprints.lse.ac.uk/102953/4/722_3282_1_PB.pdf).

## When to regulate: The timing of regulatory interventions

A second important variable in the design of a regulatory system concerns the stage at which regulatory interventions take place. Here, there are three, mutually compatible, options:

**Before:** A regulator can choose to intervene prior to a product or service entering a market, or prior to it receiving regulatory approval. In the context of AI, examples of *ex ante* regulation might include pre-market entry requirements, such as audits and assessments of AI systems by regulators to ascertain the levels of accuracy and bias.<sup>113</sup> It might also include bans on specific uses of AI in particular, high-risk settings.

**During:** A regulator can also intervene during the course of the operation of a business model, product or service. Here, this will be stipulating requirements that need to be met by the product during the course of its operation. Typically, this type of intervention will require some form of inspection regime to ensure ongoing compliance with the regulator's requirements. In the context of AI, it might involve establishing mechanisms by which regulators can inspect algorithmic systems, or requirements for AI developers to disclose information on the performance of their systems – either publicly or to the regulator.

**After:** A regulator can intervene retrospectively to remedy harms, or address breaches of regulatory rules and norms. Retrospective regulation can take the form of public enforcement, undertaken by regulators with statutory enforcement powers, or private-sector enforcement pursued via contract, tort and public-law remedies. An AI-related example might be regulators having the power to issue fines to developers or users of AI systems for breaches of regulatory rules, or as redress for particular harms done to individuals or groups resulting from failure to comply with regulation.

---

<sup>113</sup> Ada Lovelace Institute and DataKind UK. (2020). *Examining the Black Box: Tools for Assessing Algorithmic Systems* Available at: [www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems](http://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems).

## What to regulate: Targets of regulatory interventions

A third important variable concerns the targets of regulatory interventions. Here, regulators and regulatory systems can be configured to concentrate on any of the following:

**Conduct and behaviour:** One of the most common forms of intervention involves regulating the conduct or behaviour of a particular actor or actors. On the one hand, regulation of conduct can be directed at suppliers of goods, products or services, and often involves stipulating:

1. rules for how firms should conduct business,
2. requirements to provide information or guidance to consumers, or
3. responsibilities that must be borne by particular individuals.

Regulation of conduct can also be directed towards consumers, however. Attempts to regulate consumer behaviour typically involve the provision of information or guidance to help consumers better navigate markets. This kind of regulation may also involve manipulation of the way that consumers' choices are presented and framed, known as 'choice architecture', with a view towards 'nudging' consumers to make particular choices.

**Systems and processes:** A second target of regulation are the systems and processes followed by companies and organisations. Regulators may look to dictate aspects of business processes and management systems or else introduce new processes that companies and organisations have to follow, such as health-and-safety checks and procedures. Regulators may also target technical and scientific processes, for example, the UK Human Fertilisation and Embryology Authority addresses the scientific processes that can be adopted for human fertilisation.

**Market structure:** A third target of regulation is the overall dynamics and structure of a market with the aim of addressing current or potential market failures. Regulation of market structure may be aimed at preventing monopolies or other forms of anti-competitive behaviours or structures, or at more specific goals, such as avoiding moral hazard or containing the impact of the collapse of particular companies or sectors. These can be achieved through competition law or through the imposition of sector-specific rules.

---

Technological infrastructure should be a key concern for regulators of AI

Technological infrastructure should be a key concern for regulators of AI, particularly given that the majority of AI systems and ‘cloud’ services are going to be built and dependent on physical infrastructure provided by big tech. Regulators will want to consider control of the infrastructure necessary for the functioning of AI (and digital technologies more generally), as well as the competition implications of this trend.

It is worth noting that the early 2020s is likely to be a time of significant change in approaches to competition law – particularly in relation to the tech industry. In the USA, the Biden administration has shown greater willingness than any of its recent predecessors to reform competition law, though the extent and direction of any changes remains unclear.<sup>114</sup> In the EU, the Digital Markets Act<sup>115</sup> is set to change the regulatory landscape dramatically. For a UK Government eager to stimulate and develop the UK tech sector, getting the UK regulatory system’s approach to competition law right will be imperative to success.

**Calculative methods:** A particularly important target of regulation in the context of AI is calculative and decision-making models. These can range from simple mathematical models that set the prices of consumer products, to more complex algorithms used to rate a person’s credit worthiness, or the artificial-neural networks used to power self-driving vehicles.

Regulation of calculative methods can be undertaken by directly stipulating the requirements for the model (for instance stating that a decision-making model should have a particular accuracy threshold), or else by regulating the nature of the calculative or decision-making models themselves. For instance, in finance, a regulator might stipulate the means by which a bank calculates its liabilities – the cash reserves it must set aside as contingency.

---

<sup>114</sup> Bietti, E. (2021). ‘Is the Goal of Antitrust Enforcement a Competitive Digital Economy or a Different Digital Ecosystem?’ *Ada Lovelace Institute*. Available at: [www.adalovelaceinstitute.org/blog/antitrust-enforcement-competitive-digital-economy-digital-ecosystem/](http://www.adalovelaceinstitute.org/blog/antitrust-enforcement-competitive-digital-economy-digital-ecosystem/) [accessed 20 September 2021].

<sup>115</sup> Tambiama, M. (2021). *Digital Markets Act – Briefing*, May 2021, p. 12. Available at: [www.europarl.europa.eu/RegData/etudes/BRIE/2021/690589/EPRS\\_BRI\(2021\)690589\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/BRIE/2021/690589/EPRS_BRI(2021)690589_EN.pdf).

## How widely to regulate: The scope of regulatory intervention

An important related variable that is particularly salient in the context of a general-purpose technology like AI is the scope of regulation. Here, it is useful to distinguish between:

1. **The scope of the aims of regulation:** One the one hand, a regulatory intervention might aim for the use of AI in a particular context to avoid localised harms, and for the use of AI in a particular domain to be consistent with the functioning of that domain. On the other, individual regulators might also be concerned with how the use of AI in their particular enforcement domain affects other domains, or how the sum of all regulatory rules concerning AI across different industries or domains affects the technology's overall impact on society and the economy.
2. **The institutional scope of regulation:** Closely related is the question of the extent to which regulators and other institutions see and develop regulatory rules as part of a coherent whole, or whether they operate separately.
3. **The geographical scope of regulation:** Is regulation set at a national or a supranational level?

As a general rule, regulation with a narrow scope is easier for individual regulators to design and enforce, as it provides regulatory policy development and evaluation with fewer variables and avoids difficult coordination problems. Despite these advantages, narrow approaches to regulation have significant setbacks, which are of particular relevance to a general-purpose technology like AI, and may make the difficulties of more holistic, integrated approaches worth considering:

- Regulatory systems that focus on addressing narrowly defined issues can often be blind to issues that are only visible in the aggregate.
- Regulatory systems characterised by regulators with narrow areas of interest are more prone to blind spots in between domains of regulation.
- The existence of regulators and regulatory regimes with narrow geographical or market scope can increase the risks of arbitrage (where multinational firms exploit the regulatory differences between markets to circumvent regulation).

## How to regulate: Modes of regulatory intervention, and tools and techniques

A final variable is the tools, approaches and techniques used by a regulator or regulatory system.

The different mechanisms by which regulators can achieve their objectives can be divided up into the following categories:

- norms
- numbers
- incentives and sanctions
- regulatory approach
- trust and legitimacy.

### Norms

Perhaps the most common means of regulating is by setting norms. Regulatory norms can take the form of specific rules, or more general principles. The latter can be focused either on the outcomes the regulated entity should produce, or the nature of the processes or procedures undertaken. In terms of scope, norms can be specific to particular firms or industries, or can be cross sectoral or even cross jurisdictional.

While norms do tend to require enforcement, there are many cases where norms are voluntarily adhered to, or where norms create a degree of self-regulation on the part of regulated entities. In the context of AI, regulatory policy (and AI policy more generally) may attempt to encourage norms of data stewardship,<sup>116</sup> greater use of principles of data minimisation and privacy-by-design, and transparency about when and how AI systems are used. In some cases, however, the nature of the incentive structures and business models for tech companies will place hard limits on the efficacy of reliance on norms. (For instance, corporations' incentives to maximise profits and to increase shareholder value in the short term may outweigh considerations about adherence to specific norms).

---

<sup>116</sup> Ada Lovelace Institute. (2020). *Exploring norms for data stewardship*. Available at: [www.adalovelaceinstitute.org/project/exploring-principles-for-data-stewardship](http://www.adalovelaceinstitute.org/project/exploring-principles-for-data-stewardship) [accessed 11 October 2021].

## Numbers

Another important means of regulatory intervention is by stipulating prices for products in a market, or by stipulating some of the numerical inputs to calculative models. For instance, if a company uses a scorecard methodology to make a particular, significant decision, a regulator might decide to stipulate the confidence threshold.

These kinds of mechanisms may be indirectly relevant to AI systems used to set prices within markets, and could be directly relevant for symbolic AI systems, where particular numerical inputs can have a significant and clear effect on outputs. However, recent literature on competition law and large technology companies highlights that a fixation on price misses other forms of competition concern.<sup>117</sup>

## Incentives and sanctions

Regulators can also provide incentives or impose penalties to change the behaviours of actors within a market. These might be pegged to particular market outcomes (such as average prices or levels of consumer satisfaction), to specific conduct (such as the violation of regulatory rules or principles) or to the occurrence of specific harms. Penalties can take the form of fines, requirements to compensate injured parties, the withdrawal of professional licenses or, in extreme cases, criminal sanctions. A prime example of the use of sanctions in tech regulation is provided by the EU's General Data Protection Regulation, which imposes significant fines on companies for non-compliance.<sup>118</sup>

## Regulatory approach

Finally, there are various questions of regulatory approach. Differences in regulatory approach might include whether a regulatory regime is:

- **Anticipatory**, whereby the regulator attempts to understand potential harms or market failures before they emerge, and to address them

---

<sup>117</sup> Khan, L. (2017). 'Amazon's Antitrust Paradox'. *Yale Law Journal*. Volume 126, No. 3. Available at: [www.yalelawjournal.org/note/amazons-antitrust-paradox](http://www.yalelawjournal.org/note/amazons-antitrust-paradox).

<sup>118</sup> General Data Protection Regulation. Available at: <https://gdpr.eu/fines>.

before they become too severe, or **reactive**, whereby regulators respond to issues once harms or other problems are clearly manifest. In the realm of technology regulation, anticipatory approaches are perhaps the best answer to the ‘Collingridge dilemma’: when new technologies and business models do present clear harms that require regulation, these often only become apparent to regulators well after they have become commonplace. By this time, the innovations in question have often become so integrated into economic life that *post hoc* regulation is extremely difficult.<sup>119</sup> However, anticipatory approaches tend to have to err on the side of caution, potentially leading to a greater degree of overregulation than reactive approaches – which can operate with a fuller understanding of the harms and benefits of new technologies and business models.

- **Compliance based**, where a regulator works with regulated entities to help them comply with rules and principles, or **deterrance based**, where regulatory sanctions provide the main mechanisms by which to incentivise adherence. This difference also tends to be more pronounced in the context of emerging technologies, where there is less certainty regarding what is and isn’t allowed under regulatory rules.
- **Standardised**, where all regulated products and services are treated the same, or **risk based**, whereby regulators monitor and restrict different products and services to differing degrees, depending on judgements of the severity or likelihood of potential harms from regulatory failure.<sup>120</sup> By creating different levels of regulatory requirements, the rules created by risk-based systems can be less onerous for innovators and businesses, but also depend on current (and potentially incorrect) judgements about the potential levels of risk and harms associated with different technologies or business models. Risk-based approaches come with the danger of creating gaps in the regulatory system, in which harmful practices or technologies can escape an appropriate level of regulatory scrutiny.

---

<sup>119</sup> Liebert, W., and Schmidt, J. C. (2010).

<sup>120</sup> In determining how to calibrate a regulatory response to a product or technology to the level of risk it presents, two of the most important factors are 1) If and to what extent the harms it could cause are reversible or compensatable; and 2) whether the harms done are contained in scope, or broader and more systemic.

## Trust and legitimacy

There are different things that different groups will require from a regulator or regulatory system in order for the system to be seen as trustworthy and legitimate. These include:

- **Expertise:** A regulator needs to have, and be able to demonstrate a sufficient level of understanding of the subject matter they are regulating. This is particularly important in industries or areas where asymmetries of information are common, such as AI. While relevant technical expertise is a necessity for regulators, in many contexts (and especially that of AI regulation) understanding the dynamics of sociotechnical systems and their effects on people and society will also be essential.
- **Normative values:** It is also important for a regulator to take into account societal values when developing and enforcing regulatory policy. For example, in relation to AI, it will be important for questions about privacy, distributional justice or procedural fairness to be reflected in a regulator's actions, alongside considerations of efficiency, safety and security.
- **Constitutional, democratic and participatory values:** A final important set of factors affecting the legitimacy and trustworthiness of a regulator concern whether a regulator's ways of working are transparent, accountable and open to democratic participation and input. Ensuring a regulator is open to meaningful participation can often be difficult, depending on its legal and practical ability to make decisions differently in response to participatory interventions, and on the accessibility of the decisions being made.

---

# Bibliography

Ada Lovelace Institute. (2020). *Examining the Black Box: Tools for Assessing Algorithmic Systems* Available at: [www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/](https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/) [accessed 11 October 2021]

Ada Lovelace Institute. (2021). *Participatory data stewardship*. Available at: <https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/> [accessed 20 September 2021]

Ada Lovelace Institute. (Forthcoming). *Technical methods for regulatory inspection of algorithmic systems in social media platforms*. Available at: <https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection>

Ada Lovelace Institute, AI Now, and Open Government Partnership. (2021). *Algorithmic Accountability for the Public Sector*. Available at: <https://www.opengovpartnership.org/wp-content/uploads/2021/08/algorithmic-accountability-public-sector.pdf>

Ada Lovelace Institute and DataKind UK. (2020). *Exploring principles for data stewardship*. September 2020. Available at: <https://www.adalovelaceinstitute.org/project/exploring-principles-for-data-stewardship/> [accessed 11 October 2021]

Ada Lovelace Institute and Reset. (2020). *Inspecting algorithms in social media platforms*. Available at: <https://www.adalovelaceinstitute.org/report/inspecting-algorithms-in-social-media-platforms/> [accessed 22 October 2021]

AI Council. (2021). *AI Roadmap*. UK Government. Available at: <https://www.gov.uk/government/publications/ai-roadmap> [accessed 11 October 2021]

Armstrong, H., Gorst, C., and Rae, J. (2019). *Renewing Regulation: 'Anticipatory Regulation' in an Age of Disruption*. Nesta. Available at: <https://www.nesta.org.uk/report/renewing-regulation-anticipatory-regulation-in-an-age-of-disruption/> [accessed 22 October 2021]

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). 'The Fall of the Labor Share and the Rise of Superstar Firms', *The Quarterly Journal of Economics*, 135.2 pp.645–709. Available at: <https://doi.org/10.1093/qje/qjaa004>

Balayn, A., and Gürses, S. (2021). *Beyond Debiasing: Regulating AI and its inequalities*. European Digital Rights. Available at: [https://edri.org/wp-content/uploads/2021/09/EDRI\\_Beyond-Debiasing-Report\\_Online.pdf](https://edri.org/wp-content/uploads/2021/09/EDRI_Beyond-Debiasing-Report_Online.pdf) [accessed 22 October 2021]

Baldwin, R., and Black, J. (2016). 'Driving Priorities in Risk-Based Regulation: What's the Problem?' *Journal of Law and Society*, 43.4 pp.565–95. Available at: <https://doi.org/10.1111/jols.12003>

BBC News. (2020). 'Home Office Drops "racist" Algorithm from Visa Decisions'. 4 August. Available at: <https://www.bbc.com/news/technology-53650758> [accessed 22 October 2021]

BBC News. (2021). 'Council Algorithms Mass Profile Millions, Campaigners Say'. 20 July. Available at: <https://www.bbc.com/news/uk-57869647> [accessed 22 October 2021]

Bietti, E. (2021). 'Is the Goal of Antitrust Enforcement a Competitive Digital Economy or a Different Digital Ecosystem?' *Ada Lovelace Institute*. Available at: <https://www.adalovelaceinstitute.org/blog/antitrust-enforcement-competitive-digital-economy-digital-ecosystem/> [accessed 20 September 2021]

Black, J., and Murray, A. D. (2019). 'Regulating AI and Machine Learning: Setting the Regulatory Agenda', *European Journal of Law and Technology*, 10.3. Available at: <http://eprints.lse.ac.uk/102953/> [accessed 22 October 2021]

Bommasani, R., et al. (2021). *On the opportunities and risks of foundation models*. Stanford FSI. Available at: <https://fsi.stanford.edu/publication/opportunities-and-risks-foundation-models>.

Branson, A. (2020). 'European Commission Woos US over AI Agreement', *Global Government Forum*, Available at: [www.globalgovernmentforum.com/european-commission-woos-us-over-ai-agreement/](http://www.globalgovernmentforum.com/european-commission-woos-us-over-ai-agreement/) [accessed 22 October 2021]

British Medical Association. (n.d). *Ethics*. Available at: [www.bma.org.uk/advice-and-support/ethics](http://www.bma.org.uk/advice-and-support/ethics) [accessed 20 September 2021]

Buolamwini, J., and Gebru, T., (2018). 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (presented at the Conference on Fairness, Accountability and Transparency, PMLR, 2018), pp. 77–91. Available at: <https://proceedings.mlr.press/v81/buolamwini18a.html> [accessed 22 October 2021]

Calvert, M. J., Marston, E., Samuels, M., Cruz Rivera, S., Torlinska, B., Oliver, K., Denniston, A. K., and Hoare, S. (2021). 'Advancing UK Regulatory Science and Innovation in Healthcare', *Journal of the Royal Society of Medicine*, 114.1, pp. 5–11. Available at: <https://doi.org/10.1177/0141076820961776>

Cath-Speth, C. (2021). Available at: [https://twitter.com/c\\_\\_cs/status/1412457639611600900](https://twitter.com/c__cs/status/1412457639611600900) [accessed 22 October 2021]

Central Digital and Data Office. (2018). *Data Ethics Framework*. UK Government. Available at: <https://www.gov.uk/government/publications/data-ethics-framework>

Centre for Data Ethics and Innovation. (2021). *Privacy Enhancing Technologies - Adoption Guide*, Available at: <https://cdeiuk.github.io/pets-adoption-guide> [accessed 22 October 2021]

Centre for Data Ethics and Innovation. (2020). *Review into Bias in Algorithmic Decision-Making*. p. 151 Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/957259/Review\\_into\\_bias\\_in\\_algorithmic\\_decision-making.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957259/Review_into_bias_in_algorithmic_decision-making.pdf)

Cobbe, J., and Singh, J. (2021). 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges', *SSRN Electronic Journal*. Available at: <https://doi.org/10.2139/ssrn.3824736>

Courts and Tribunals Judiciary. (2020). *R (on the application of Edward Bridges) v. The Chief Constable of South Wales Police and the Secretary for the State for the Home Department*. Case No: C1/2019/2670. Available at: <https://www.judiciary.uk/wp-content/uploads/2020/08/R-Bridges-v-CC-South-Wales-ors-Judgment.pdf> [accessed 22 October 2021]

Competition and Markets Authority. (2018). *Pricing algorithms: Economic Working Paper on the Use of Algorithms to Facilitate Collusion and Personalised Pricing*. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/746353/Algorithms\\_econ\\_report.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746353/Algorithms_econ_report.pdf).

Cyberspace Administration of China (国家互联网信息办公室). (2021). *Notice of the State Internet Information Office on the 'Regulations on the Management of Recommendations for Internet Information Service Algorithms (Draft for Solicitation of Comments)' Public Solicitation of Comments*. 27 August. Available at: [https://www-cac-gov-cn.translate.goog/2021-08/27/c\\_1631652502874117.htm?\\_x\\_tr\\_sch=http&\\_x\\_tr\\_sl=zh-CN&\\_x\\_tr\\_tl=en&\\_x\\_tr\\_hl=en&\\_x\\_tr\\_pto=ajax,nv,elem](https://www-cac-gov-cn.translate.goog/2021-08/27/c_1631652502874117.htm?_x_tr_sch=http&_x_tr_sl=zh-CN&_x_tr_tl=en&_x_tr_hl=en&_x_tr_pto=ajax,nv,elem) [accessed 22 October 2021]

Dattner, B., Chamorro-Premuzic, T., Buchband, R., and Schettler, L. (2019). 'The Legal and Ethical Implications of Using AI in Hiring', *Harvard Business Review*, 25 April. Available at: <https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring> [accessed 20 September 2021]

Department for Digital, Culture, Media and Sport. (2021). *Data: A New Direction*. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1022315/Data\\_Reform\\_Consultation\\_Document\\_Accessible\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022315/Data_Reform_Consultation_Document_Accessible_.pdf)

Department for Digital, Culture, Media and Sport. (2021). *Digital Regulation: Driving Growth and Unlocking Innovation*. Available at: <https://www.gov.uk/government/publications/digital-regulation-driving-growth-and-unlocking-innovation/digital-regulation-driving-growth-and-unlocking-innovation> [accessed 22 October 2021]

Digital Regulation Cooperation Forum. (2021). UK Government. Available at: <https://www.gov.uk/government/collections/the-digital-regulation-cooperation-forum> [accessed 22 October 2021]

Digital Scotland. (2021). *Scotland's AI Strategy: Trustworthy, Ethical and Inclusive*. Available at: [https://static1.squarespace.com/static/5dc00e9e32cd095744be7634/t/606430e006dc4a462a5fa1d4/1617178862157/Scotlands\\_AI\\_Strategy\\_Web\\_updated\\_single\\_page\\_aps.pdf](https://static1.squarespace.com/static/5dc00e9e32cd095744be7634/t/606430e006dc4a462a5fa1d4/1617178862157/Scotlands_AI_Strategy_Web_updated_single_page_aps.pdf) [accessed 22 October 2021]

Dutton, T., (2018). 'An Overview of National AI Strategies', *Politics + AI*. Available at: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd> [accessed 11 October 2021]

European Commission. (2021). *Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. Available at: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52021PC0206> [accessed 4 October 2021]

European Commission. (2021). *Knowledge for Policy: France AI Strategy Report*. Available at: [https://knowledge4policy.ec.europa.eu/ai-watch/france-ai-strategy-report\\_en](https://knowledge4policy.ec.europa.eu/ai-watch/france-ai-strategy-report_en) [accessed 22 October 2021]

European Commission. (2021). *United Kingdom AI Strategy Report*. Available at: [https://knowledge4policy.ec.europa.eu/ai-watch/united-kingdom-ai-strategy-report\\_en](https://knowledge4policy.ec.europa.eu/ai-watch/united-kingdom-ai-strategy-report_en) [accessed 22 October 2021]

Faggella, D., 'Artificial Intelligence Applications for Lending and Loan Management', *Emerj*. Available at: <https://emerj.com/ai-sector-overviews/artificial-intelligence-applications-lending-loan-management/> [accessed 20 September 2021]

Furman, J., Coyle, D., Fletcher, A., McAuley, D., and Marsden, P. (2019). *Unlocking Digital Competition: Report of the Digital Competition Expert Panel*. HM Treasury. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/785547/unlocking\\_digital\\_competition\\_furman\\_review\\_web.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf) [accessed 22 October 2021]

Hill, K. (2020). 'Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match', *The New York Times*. Available at: [www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html](http://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html) [accessed 22 October 2021]

Institute for the Future of Work. (2021). *The Amazonian Era – The gigification of work*. Available at: <https://www.ifow.org/publications/the-amazonian-era-the-gigification-of-work>

Jobin, A., lenca, M., and Vayena, E. (2019). 'The Global Landscape of AI Ethics Guidelines', *Nature Machine Intelligence*, 1.9. pp. 389–99. Available at: <https://doi.org/10.1038/s42256-019-0088-2>

Johnson, K. (2020). 'From Whistleblower Laws to Unions: How Google's AI Ethics Meltdown Could Shape Policy', *VentureBeat*. Available at: <https://venturebeat.com/2020/12/16/from-whistleblower-laws-to-unions-how-googles-ai-ethics-meltdown-could-shape-policy/> [accessed 22 October 2021]

Kelion, L. (2018). 'UK PM Seeks "safe and Ethical" Artificial Intelligence', *BBC News*. 25 January. Available at: [www.bbc.com/news/technology-42810678](http://www.bbc.com/news/technology-42810678) [accesseed 22 October 2021]

Kent, C. (2019). 'UK Healthcare Industry Analysis 2019: Why Britain Is a World Leader'. *Pharmaceutical Technology*. Available at: [www.pharmaceutical-technology.com/sponsored/uk-healthcare-industry-analysis-2019/](http://www.pharmaceutical-technology.com/sponsored/uk-healthcare-industry-analysis-2019/) [accessed 20 September 2021]

Khan, L. (2017). 'Amazon's Antitrust Paradox', *Yale Law Journal*. Available at: <https://www.yalelawjournal.org/note/amazons-antitrust-paradox> [accessed 22 October 2021]

Lander, E., and Nelson, A. (2021). 'Americans Need a Bill of Rights for an AI-Powered World', *Wired*. 10 October. Available at: [www.wired.com/story/opinion-bill-of-rights-artificial-intelligence/](http://www.wired.com/story/opinion-bill-of-rights-artificial-intelligence/) [accessed 22 October 2021]

Ledford, H. (2019). 'Millions of Black People Affected by Racial Bias in Health-Care Algorithms', *Nature*, 574.7780. pp. 608–9. Available at: <https://doi.org/10.1038/d41586-019-03228-6>

Leslie, D. (2019). *Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector*. The Alan Turing Institute. Available at: <https://doi.org/10.5281/ZENODO.3240529>

Liebert, W, and Schmidt, J. C. (2010). 'Collingridge's Dilemma and Technoscience: An Attempt to Provide a Clarification from the Perspective of the Philosophy of Science', *Poiesis & Praxis*, 7.1–2 pp. 55–71. Available at: <https://doi.org/10.1007/s10202-010-0078-2>

Lum, L., and Chowdhury, R. (2021). 'What Is an "Algorithm"? It Depends Whom You Ask', *MIT Technology Review*. Available at: [www.technologyreview.com/2021/02/26/1020007/what-is-an-algorithm/](http://www.technologyreview.com/2021/02/26/1020007/what-is-an-algorithm/) [accessed 22 October 2021]

Lynch, S. (2017). *Andrew Ng: Why AI Is the New Electricity*, Stanford Graduate School of Business. Available at: <https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity> [accessed 22 October 2021]

Martinho-Truswell, E. (2018). 'How AI Could Help the Public Sector', *Harvard Business Review*, 26 January. Available at: <https://hbr.org/2018/01/how-ai-could-help-the-public-sector> [accessed 20 September 2021]

Matsakis, L. (2019). 'How the West Got China's Social Credit System Wrong.' *Wired*. Available at: [www.wired.com/story/china-social-credit-score-system](http://www.wired.com/story/china-social-credit-score-system).

Mazzucato, M. (2015). 'From Market Fixing to Market-Creating: A New Framework for Economic Policy', *SSRN Electronic Journal*. Available at: <https://doi.org/10.2139/ssrn.2744593>

McGregor, S. (2020). *When AI Systems Fail: Introducing the AI Incident Database*. Partnership on AI. Available at: <https://partnershiponai.org/aiincidentdatabase/> [accessed 22 October 2021]

McLean, A., and Wood, I. (2015). 'Do Regulators Hold the Key to FinTech Success?', *Financier Worldwide*. Available at: [www.financierworldwide.com/do-regulators-hold-the-key-to-fintech-success](http://www.financierworldwide.com/do-regulators-hold-the-key-to-fintech-success) [accessed 20 September 2021]

Moss, E., Watkins, E. A., Singh, R., Elish, M.C., and Metcalf, J. (2021). *Assembling Accountability - Algorithmic Impact Assessment for the Public Interest*. Data & Society Research Institute. Available at: <https://datasociety.net/wp-content/uploads/2021/06/Assembling-Accountability.pdf> [accessed 22 October 2021]

Municipality Amsterdam. (2020). *Standard Clauses for Municipalities for Fair Use of Algorithmic Systems*. Gemeente Amsterdam. Available at: <https://www.amsterdam.nl/innovatie/> [accessed 22 October 2021]

National Audit Office. (2017). *A Short Guide to Regulation*. UK Government. Available at: <https://www.nao.org.uk/wp-content/uploads/2017/09/A-Short-Guide-to-Regulation.pdf>

National Audit Office. (2021). *Principles of Effective Regulation*. UK Government. Available at: <https://www.nao.org.uk/wp-content/uploads/2021/05/Principles-of-effective-regulation-SOff-interactive-accessible.pdf>

National Control Commission for the Protection of Personal Data. (2020). *Press release accompanying the publication of deliberation No. D-97-2020 du 26/03/2020*’ (in French). Available at: <https://www.cndp.ma/fr/presse-et-media/communique-de-presse/661-communique-de-presse-du-30-03-2020.html> [accessed 22 October 2021]

Office for AI. (2021). *National AI Strategy*. UK Government. Available at: <https://www.gov.uk/government/publications/national-ai-strategy>

Office for Statistics Regulation. (2021). *Ensuring Statistical Models Command Public Confidence*. Available at: <https://osr.statisticsauthority.gov.uk/publication/ensuring-statistical-models-command-public-confidence/> [accessed 22 October 2021]

Organisation for Economic Co-operation and Development. (2019). *OECD Principles on Artificial Intelligence*. Available at: [www.oecd.org-going-digital/ai/principles/](http://www.oecd.org-going-digital/ai/principles/) [accessed 22 October 2021]

Partnership on AI. (2021). *Redesigning AI for Shared Prosperity: An Agenda*. Available at: <https://partnershiponai.org/wp-content/uploads/2021/08/PAI-Redesigning-AI-for-Shared-Prosperity.pdf> [accessed 22 October 2021]

Perez, C. (2015) ‘Capitalism, Technology and a Green Global Golden Age: The Role of History in Helping to Shape the Future’, *The Political Quarterly*, 86 pp. 191–217. Available at: <https://doi.org/10.1111/1467-923X.12240>

Pownall, C. (2021). *AI, Algorithmic and Automation Incidents and Controversies*. Available at: <https://charliepownall.com/ai-algorithmic-incident-controversy-database/> [accessed 22 October 2021]

Privacy & Information Security Law Blog. (2021). *Regulatory Sandboxes Are Gaining Traction with European Data Protection Authorities.*

Hunton Andrews Kurth. Available at: <https://www.huntonprivacyblog.com/2021/02/25/regulatory-sandboxes-are-gaining-traction-with-european-data-protection-authorities/> [accessed 4 October 2021]

Quong, J. (2018). ‘Public Reason’ in Zalta, E. N., and Hammer, E. (eds) *The Stanford Encyclopedia of Philosophy*. Stanford: Center for the Study of Language and Information. Available at: <https://plato.stanford.edu/archives/spr2018/entries/public-reason/> [accessed 20 September 2021]

Richardson, R. (2019). *Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force*. AI Now Institute. Available at: <https://ainowinstitute.org/ads-shadowreport-2019.pdf> [accessed 22 October 2021]

Schaefer, K. (2021). 27 August. Available at: <https://twitter.com/kendraschaefer/status/1431134515242496002> [accessed 22 October 2021]

Selwyn, N. (2021). ‘Deb Raji on What “Algorithmic Bias” Is (... and What It Is Not)’, *Data Smart Schools*. Available at: <https://data-smart-schools.net/2021/04/02/deb-raji-on-what-algorithmic-bias-is-and-what-it-is-not/> [accessed 22 October 2021]

Simonite, T, and Barber, G. (2019). ‘It’s Hard to Ban Facial Recognition Tech in the iPhone Era’, *Wired* Available at: <https://www.wired.com/story/hard-ban-facial-recognition-tech-iphone/> [accessed 22 October 2021]

Smuha, N. et al. (2021). *How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission’s Proposal for an Artificial Intelligence Act*. Available at SSRN: <https://ssrn.com/abstract=3899991> or <http://dx.doi.org/10.2139/ssrn.3899991>.

Solove, D. J., (2011). *Nothing to Hide: The False Tradeoff between Privacy and Security*. New Haven London: Yale University Press

Tambiama, M. (2021). *Digital Markets Act - Briefing*. May 2021. p. 12. Available at: [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/690589/EPRS\\_BRI\(2021\)690589\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/690589/EPRS_BRI(2021)690589_EN.pdf)

Tashea, J. (2017). 'Courts Are Using AI to Sentence Criminals. That Must Stop Now', *Wired*. Available at: [www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/](http://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/) [accessed 20 September 2021]

The Law Society. (2016). *Written Evidence submitted by the Law Society (ROB0037)*. Available at: <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/robotics-and-artificial-intelligence/written/32616.html> [accessed 20 September 2021]

Turner, J. (2018). *Robot Rules: Regulating Artificial Intelligence*. Palgrave Macmillan

UK Government. (2021). *Regulatory Horizons Council (RHC)*. Available at: <https://www.gov.uk/government/groups/regulatory-horizons-council-rhc> [accessed 22 October 2021]

US Congress. (2021). *H.R.1816 - Information Transparency & Personal Data Control Act*. Available at: <https://www.congress.gov/bill/117th-congress/house-bill/1816/text> [accessed 22 October 2021]

Veale, M, and Zuiderveen Borgesius, F. (2021). 'Demystifying the Draft EU Artificial Intelligence Act', *Computer Law Review International*. 22(4). Available at: <https://doi.org/10.31235/osf.io/38p5f>

West Midlands Police and Crime Commissioner. (2021). *Ethics Committee*. Available at: <https://www.westmidlands-pcc.gov.uk/ethics-committee/> [accessed 22 October 2021]

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., and Cave, S. (2019). *Ethical and Societal Implications of Data and Artificial Intelligence: A Roadmap for Research*. London: Nuffield Foundation. Available at: [www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf](http://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf)

Yeung, K. (2017). "Hypernudge": Big Data as a Mode of Regulation by Design', *Information, Communication & Society*, 20(1). pp.118–36. Available at: <https://doi.org/10.1080/1369118X.2016.1186713>

Yeung, K., Howes, A., and Pogrebna, G. (2020). 'AI Governance by Human Rights-Centered Design, Deliberation, and Oversight: An End to Ethics Washing', in Dubber, M. D., Pasquale, F., and Das, S. (eds) *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press. pp. 75–106  
Available at: <https://doi.org/10.1093/oxfordhb/9780190067397.013.5>

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

# Acknowledgements

We are grateful to the expert panellists who took part in our workshops in April and May 2021, the findings of which helped inform much of this report. Those involved in these workshops are listed below.

Workshop participant	Affiliation
Ghazi Ahamat	Centre for Data Ethics & Innovation
Mhairi Aitken	Alan Turing Institute
Haydn Belfield	Centre for the Study of Existential Risk
Elettra Bietti	Berkman Klein Center for Internet and Society
Reuben Binns	University of Oxford
Kate Brand	Competition and Markets Authority
Lina Dencik	Data Justice Lab, Cardiff University
George Dibb	Institute for Public Policy Research
Mark Durkee	Centre for Data Ethos & Innovation
Alex Georgiades	UK Civil Aviation Authority
Mohammed Gharbawi	Bank of England
Emre Kazim	University College London
Paddy Leerssen	University of Amsterdam
Samantha McGregor	Arts and Humanities Research Council
Seán ÓhÉigearaigh	Leverhulme Centre for the Future of Intelligence & Centre for the Study of Existential Risk
Lee Pope	Department for Digital, Culture, Media and Sport
Mona Sloane	New York University, Center for Responsible AI
Anna Thomas	Institute for the Future of Work
Helen Toner	Center for Security and Emerging Technology
Salomé Viljoen	Columbia Law School
Karen Yeung	Birmingham Law School & School of Computer Science

We are also grateful to those who, in addition to participating in the workshops, provided comments at different stages of this report and whose thinking, ideas and writing we have drawn on heavily, in particular: Professor Julia Black, London School of Economics; Jacob Turner, barrister at Falcon Chambers and Professor Lillian Edwards, University of Newcastle.

This report was lead authored by Harry Farmer, with substantive contributions from Andrew Strait and Imogen Parker.

---

# About the Ada Lovelace Institute

The Ada Lovelace Institute (Ada) was established by the Nuffield Foundation in early 2018, in collaboration with the Alan Turing Institute, the Royal Society, the British Academy, the Royal Statistical Society, the Wellcome Trust, Luminate, techUK and the Nuffield Council on Bioethics.

Ada is an independent research institute with a mission to make data and AI work for people and society.

We are working to create a shared vision of a world where AI and data are mobilised for good, to ensure that technology improves people's lives. We take a sociotechnical, evidence-based approach and use deliberative methods to convene and centre diverse voices. We do this to identify the ways that data and AI reorder power in society, and to highlight tensions between emerging technologies and societal benefit.

We are funded by the Nuffield Foundation, an independent charitable trust with a mission to advance social well-being. The Foundation funds research that informs social policy, primarily in education, welfare and justice. It also provides opportunities for young people to develop skills and confidence in STEM and research. In addition to the Ada Lovelace Institute, the Foundation is also the founder and co-funder of the Nuffield Council on Bioethics and the Nuffield Family Justice Observatory.

## Find out more:

Website: [Adalovelaceinstitute.org](http://Adalovelaceinstitute.org)

Twitter: [@AdaLovelaceInst](https://twitter.com/AdaLovelaceInst)

Email: [hello@adalovelaceinstitute.org](mailto:hello@adalovelaceinstitute.org)



Permission to share:

This document is published under a creative commons  
licence: CC-BY-4.0

Preferred citation: Farmer, H. Strait, A. and Parker, I.  
(2021). *Regulate to innovate*. Ada Lovelace Institute.  
Available at: [https://www.adalovelaceinstitute.org/  
report/regulate-innovate](https://www.adalovelaceinstitute.org/report/regulate-innovate)

ISBN 978-1-8382567-7-7