The British Academy

British Academy & Office for AI Roundtable: AI Governance and Regulation - Summary

Introduction

The British Academy has had a long-standing interest in the challenges and opportunities posed by AI technologies, most recently explored in our collaborative project with UCL Public Policy on Artificial intelligence (AI) and the future of work. AI brings opportunities and uncertainties for UK society, including those related to the future of work, wellbeing, productivity, and social inequality, and across a broad range of sectors. Changes due to AI will sit alongside changes related to climate change, as well as the immediate impacts of COVID-19 and the UK exiting the EU. The role of AI should also be considered in this wider context of overlapping, interdependent factors to ensure an equitable transformation. Government has an important role to play in cultivating a national environment that shapes AI for the benefit of all. Effective governance and regulation of AI in the coming decade will be a vital component of this role. In the near term, the UK Office for AI is currently developing the UK's national position on governing and regulating AI, with a White Paper planned for 2022. This roundtable summary highlights insights that are both relevant to the White Paper and of wider interest for the AI policy and governance landscape in the coming decade.

On Wednesday 23rd February 2022, the British Academy hosted a virtual roundtable on AI governance and regulation, convening the policymakers from the Office for AI with expertise from SHAPE disciplines (Social Sciences, Humanities and the Arts for People and the Economy) that exists in the British Academy's Fellowship and wider networks, as well as expertise on AI in the other National Academies and relevant professional bodies. This note is a summary of the main discussion points and key insights that emerged from the roundtable discussion.[1]

Original Questions

[1] These points will also help to inform two new programmes of work (on governance and on digital society) within the British Academy's policy directorate.

Below is the list of prompt questions sent in advance to attendees at the roundtable.

1. How is the current approach to governing and regulating AI in the UK working from your perspective?
2. Are there any areas where greater clarity is needed, or is anything missing?
3. What are the unique capabilities and risks posed by AI?
4. What opportunities and challenges do SHAPE disciplines identify for AI governance in the UK?
5. In what ways do we need to consider how AI governance and regulation overlaps with other areas of governance and regulation (such as health, education, and the media)?
6. What does a good future regulatory landscape look like, and what additional tools or powers might regulators require to achieve this?
7. What would your top priority be for the UK's White Paper on AI governance and Regulation?
8. What role do you see non-regulatory tools, such as global technical standards, playing in the future governance of AI?
9. Which areas of AI could benefit from technical standards (e.g. product safety, security etc), and why?

Synthesis summary of roundtable discussion

The summary below is a synthesis of the central points that emerged throughout the roundtable discussion. These broadly divide into four sections: firstly, identifying useful **principles** with which to structure or think about AI governance and regulation; secondly, the importance of **structuring regulation** with both **horizontal and vertical components**; thirdly, the **tools for regulation** available in a diverse regulatory toolbox; and finally, some ongoing **challenges for regulatory enforceability.**

The challenges explored below are ones that technical expertise cannot address on its own. These are not questions of the narrow technical construction of AI tools, but of the wider social environment in which they might be or are used. Insights from SHAPE disciplines are therefore vital for understanding how, where, and why different people, businesses and organisations interact with AI technologies.[2]

1 Principles for Regulation

Articulating the endpoint vision

Government has an important role to play in articulating the vision of a desirable future society that we want to create through AI governance and regulation. Identifying and articulating such a vision, and how it will benefit citizens, households, businesses, and communities across the UK, is a vital starting step that allows government to then steer

[2] The Ofqual algorithm fiasco was given as an example of this. The algorithm was highly effective, met its specifications and made good decisions based on the data it used. However, a failure to take account of the context in which it would be used and the reaction to its use led to a massive pushback publicly and within the education sector. There was a lack of public confidence in the tool, despite the large role that algorithms play in the ordinary everyday human process of grading exams, where marks are adjusted, standardised, and partitioned according to often fairly elaborate algorithms.

innovation with a clear sense of direction. Regulation can thereby be formulated as productive and proactive, facilitating innovation rather than restricting or generating barriers for it.

Regulation as providing a license to innovate

There was a consensus that good regulation cultivates rather than hampers innovation by providing innovators with a license to do things that would otherwise be unacceptable or untenable (for example, by ensuring an aircraft is safe to fly). In other words, regulation does not have to be thought of as in tension with innovation and instead can provide a safe framework that people can use to take things to market and expand markets through setting common standards and aligning behaviour.

Similarly, participants noted that while de-biasing data or designing unbiased algorithms can be seen as a cost for developers and businesses to burden, it is also an investment. Businesses can benefit from finding and processing data that is unbiased, because often poor data does not reflect the true dynamics of users and citizens, so there is a business incentive to having more ethical and unbiased algorithms.

The discussion raised a suggestion related to this - that policymakers carefully consider the impact of regulations on the experiences of the various end users. For instance, how will AI regulation be seen by the owner of a pub, and how might they react to it, or how would a YouTube content creator relate to this piece of legislation? Will it even reach them?

2 Structuring Regulation: Designing regulation for AI with both horizontal and vertical components

The roundtable discussion stressed both the **horizontal** dimension of AI – in other words, AI as a distinctive technology that is deployed across a range of sectors (in transport, music and culture, marketing, healthcare, etc.) – as well as its **vertical** dimension, where understanding the context specific applications and impacts of AI at different scales *within* a sector is important. AI technologies function both in a general digital infrastructure that enables AI resource provision and access (horizontal element), and as sector-specific applications. [3] Ideally, regulation will address issues across the whole 'socio-technical' lifecycle of AI technologies, from those that arise at the stage of conception and development through to those that exist within the context of deployment (with recognition that what happens at one stage can impact upon other stages).

While regulating by sector can allow for the flexibility to adapt regulation to specific contexts, it also carries risks if regulation is too product and outcome orientated in its approach (product orientated regulation refers to regulating individual AI products in the same way as equivalent non-AI products). There were concerns that this approach could lead to a regulatory philosophy that does not account for unknown or unanticipated shocks and may not adequately grapple with the uncertainties of emerging technologies. Moreover,

[3] In the first instance, there is multi-layered digital infrastructure for enabling AI processing and application development that comprises key cloud services and platforms, AI platforms such as TensorFlow, complementary technologies, data resources, machine learning algorithms, pre-trained machine learning (ML) models and so on. On the other side, there are sector-specific applications and services for AI application development, which is more specialised and specific to individual sectors.

some sector specific regulators may face cultural challenges, as well as a shortage of suitable regulations. For example, for medical devices and aviation, regulation is conducted by testing a system exhaustively and then freezing it, and such approaches may be culturally poorly aligned with a regulatory system that learns and improves.

On the other hand, process-oriented regulation (which refers to regulating AI products as distinct from equivalent non-AI products) can help to encourage market competition around emerging technologies across sectors, particularly if undertaken from an early stage (introducing it too late can create significant barriers to entry for smaller, newer companies seeking to enter markets). Meanwhile, safety work in a process-based system could begin by identifying those distinct hazards and potential harms associated with AI technologies in different sector contexts.

However, it was noted that there is a risk that many AI services (e.g., the training of machine learning algorithms, the cloud resources for AI model training and operation, and the data resources required for the training of machine learning models) may become highly centralised and offered through a 'cloud oligopoly' dominated by digital incumbents such as Google, Amazon Web Services, and others. Regulation from the Competition and Markets Authority will be a crucial form of intervention required to prevent the lock-in advantages for incumbents (and disadvantages for smaller or newer businesses) that could arise from such an environment.

Finally, the participants widely agreed upon the crucial point that a purely sector led approach does not account for dynamics generated across the horizontal landscape, and thereby risks compounding multiple forms of disadvantage that might exist beyond a particular sector for certain groups (who may be subject to discrimination, exclusion, and so on).

3 Tools for Regulation: A diverse regulatory toolbox

There was a broad consensus that regulation could be supplemented and strengthened by fostering cultures of responsible and ethical innovation through principles that set technical or ethical standards (both in and across sectors). For instance, some non-sector specific universal principles could be formulated and applied to address asymmetries of information and power between citizens and decision makers that might be generated by AI.

Incentives can also be powerful tools alongside regulation. For example, corporate governance incentives could be used alongside regulation across supply chains – as 'tools from the regulatory toolbox' that help to ensure a level of consistency across sector regulators. When there is no alignment of incentives between people deploying machine learning systems and the users affected by them, there is a risk that automated decisions have a concerning impact on people's lives. Governance frameworks could address this by incorporating evaluative processes that can shape the types of incentives offered across and within different sector contexts.

Professional codes of conduct and professional standards are another kind of governance mechanism that could be cultivated across sectors. At present, some professions have a more embedded professional identity and sense of professional standards in relation to AI (such as engineering). Even if professional standards and codes of conduct may necessarily differ from sector to sector, fostering professional identity that brings with it ethos and codes that would apply to practitioners, organisations, and sectors (and to qualifications that enable

people to be practitioners in a sector) could strengthen the AI governance landscape through leveraging the power of professional norms. In some areas of application (for example, health, law, information and culture), it could be made professional requirement to articulate the underlying normative principles by which AI systems are to be guided, an approach akin to the 'value-informed design' movement within engineering and computer science.

4 Challenges to regulatory enforceability: for further consideration

Participants with legal expertise agreed that AI technologies pose a particular challenge for the courts. One of the guiding principles of judicial review is that a decision was made by a person (or institution, or person in an institution). If an AI system generates a decision, this could create an obstacle for accountability in a context where citizens would usually be entitled to a hearing regarding decisions about them. Policymakers might therefore give careful consideration to the compatibility of AI regulatory strategies with principles of judicial review, as approaches that ignore the issue of the opacity of AI risk being challenged in court. Moreover, the question was raised around whether the enforceability of responses to breaches of regulatory regimes can be consistent across sectors whilst also providing a sense of surety for each sector.

The roundtable participants agreed about the importance of maintaining a role for humans in AI decision making, with some participants suggesting that this should be made mandatory, so that legal and ethical responsibility for AI decisions can be determined within legal contexts. GDPR was noted as a good practice example of legislation that does this, though the group acknowledged that GDPR has its own ambiguity and interpretation problems. AI governance structures might also include ways to encourage developers to identify and articulate the role of human oversight – and where ultimate responsibility lies for AI algorithms – as part of the design and implementation processes. Moving forward, it will be key to have mechanisms for identifying who is responsible when technological systems go wrong, be it autonomous vehicles or diagnostic tools that identify cancer.[4]

Finally, the discussion pointed out that regulatory approaches will have implications for international competition. For instance, if UK governance and regulatory regimes are more restrictive than those of certain other nations, this could put the UK at a risk of economic disadvantage, as over-prescriptive regulations might impose legislative burdens on potentially beneficial advances in AI. The AI environment is an internationally competitive one. A particular characteristic of AI research and application is the speed of change. This is true of data science more generally. While it is true that legislation necessarily lags innovation, too great a lag causes damage as well as missed opportunities.[5]

[4] See the following article for more information: Human control of AI and autonomy: the art of the possible - Assuring Autonomy International Programme, University of York

[5] There are recent international developments in this vein. On March 1st, China introduced various rules such as making it illegal to use algorithms that register fake users, generate fake likes, manipulate topic lists, or impose unreasonable restrictions on other Internet information service providers. Some of these regulations have things in common with the GDPR, but there are some original developments. Incidentally, the UNESCO Recommendations on the ethics of AI explicitly bans the use of AI systems for social scoring and mass surveillance.