

**RSS RESPONSE TO DEPARTMENT FOR SCIENCE, INNOVATION AND TECHNOLOGY  
WHITE PAPER ‘A PRO-INNOVATION APPROACH TO AI REGULATION’**

21 June 2023

## **1 Introduction**

1.1.1 This is the Royal Statistical Society’s response to the government’s white paper on AI: [A pro-innovation approach to AI regulation](#). The RSS is a professional society for statisticians and data scientists, with over 10,000 members.

1.1.2 Artificial Intelligence is an area in which government policy seems to be moving so quickly that some elements of the white paper have been outpaced by other developments. Government policy in this area – as expressed both in the white paper and in subsequent developments – is broadly well-intentioned and heading in the right direction. Because of how fast-moving policy in this space is, our response will focus primarily on the things that we think are missing from the government’s approach that need to be addressed. These are:

- Investment in evaluation methodology and systems for AI and the creation of a new centre for AI evaluation methodology.
- Investment in a unit of open source software developers.
- Strengthen government leadership on AI – involving organisations with deep expertise to improve intelligence and capacity.
- Provide strategic direction – this is an area where policy seems to have moved since the white paper, with the prime minister speaking of making the UK the home of global AI safety regulation.

1.1.3 Our response is split into two sections. The first section sets out why we feel that a focus on improving these four aspects of the government’s response is important. In the second part, we address the specific questions posed in the white paper where we have a view.



## 2 RSS priorities for the UK's AI strategy

- 2.1.1 In the view of the RSS the most important issue that is not addressed by the AI white paper is around evidence standards and evaluation methods. This is important because an ever-increasing number of businesses will be looking to use AI systems and it is vital that there is an agreed methodology for evaluating their impact.
- 2.1.2 There are similarities between the urgency of the current need for investment in evaluation methodologies in the context of AI as there was in the early 1990s to invest in methodology in the context of health. Then, there were a great many trials that were being funded but there was no clear centre of expertise on methodology for evaluating the evidence in terms of cost/benefit. As a result the [NHS R&D Health Technology Assessment Programme](#) was set up in 1993 to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies was produced in the most efficient way for those who use, manage and provide care. This was a massive success, establishing the UK as a world leader in evaluating health care. The Covid pandemic again showed the necessity for robust evaluation methodology, with the RECOVERY platform trial using a platform trial design with adaptive randomisation both to rule out ineffective treatments such as hydroxychloroquine, and efficiently demonstrating the effectiveness of simple interventions such as high dose dexamethasone. This saved possibly a million lives and again put the UK as the world-leader in Covid treatment evaluation.
- 2.1.3 The present situation in AI is similarly urgent and the RSS urges the government to make a commensurate investment in a centre for AI evaluation methodology. Users of AI systems should be able to judge the trustworthiness of (a) the claims made about the system by its developers (b) its conclusions in a specific situation. Without proper evaluation, we are back to a situation similar to that of medicine before the thalidomide disaster in the 1960s, when claims about the efficacy and safety drug treatments were based on authority rather than evidence from adequately controlled clinical trials.
- 2.1.4 Good evidence will support the implementation of proposed principles for established features akin to efficacy and safety, as well as more distributional notions of fairness which are harder to operationalise. Explanation, accountability and, ultimately, redress will also rely on a



comprehensive and consensus approach to evidence, which needs to anticipate more complex societal applications and interactions between them.

- 2.1.5 “Events” often drive the introduction of regulation, for instance the impact of thalidomide on regulatory guidance noted above. The proposed proactive approach being taken is something that the RSS supports. A risk-based approach is required, whereby AI is assessed in relation to how it is used (for instance, to recommend, inform, decide, create or entertain) and the consequences and impact of its use. It is also important to distinguish misinformation from disinformation and sloppiness from malevolence. A risk-based approach lends itself to identifying key systems that require greater scrutiny whilst not undermining innovation and agility.
- 2.1.6 Evaluation is a core part of the work of statisticians and this is an area where the profession can help. This is an area where the RSS is keen to constructively engage with the government to help ensure effective regulation of AI. In addition to the Society’s large membership, we have specific networks of statisticians and data scientists focusing on both data ethics and governance as well as AI in the context of data science – we would be keen to leverage this network to help support the government’s work on AI in this way.
- 2.1.7 Our view is that it is especially important for the government to engage with the professions that interface with AI because it is not clear that, without engaging with external experts with deep expertise, the government’s AI bodies are in a position to comprehensively identify the emerging opportunities and risks. An example of where this seems to have happened is around large language models. In 2021, in our engagement with the AI Roadmap, the RSS argued that an important element of the AI strategy that was missing was investment in open source software development. This is precisely the resource that has been required to develop large language models – which seem to have only been identified as a pressing matter more recently with the existing structures. This is understandable to an extent – it is difficult to have the level of deep expertise required in government departments – but it does point to the importance of engaging with the professions to identify wider opportunities.
- 2.1.8 We remain of the view that investment in open source software development is vital and would propose a Government-funded unit of open source developers who would:



- Contribute to open source software projects deemed to be of strategic importance to the UK.
- Offer internships to help build and transfer skills throughout the UK private, public and third sectors.
- Support the use of open source in the UK through knowledge sharing, training and building the community.

2.1.9 The RSS is concerned that splitting responsibilities for regulating the use of AI between existing regulators does not meet the scale of the challenge – and will also not help businesses in the UK to act effectively in this space. Central leadership is required to give a clear, coherent and easily communicable framework that can be applied to all sectors. The speed of change within AI means that maintaining a consistent and coherent system across multiple existing regulators is challenging and inevitably some regulators will lag behind others: consistency, coherence and horizon scanning are vital to deal with this.

### 3 Responses to questions

3.1.1 This section provides the RSS's response to the specific questions posed in the white paper where we have an interest.

#### 3.2 *The revised cross-sectoral AI principles*

1. Do you agree that requiring organisations to make it clear when they are using AI would adequately ensure transparency?

3.2.1 This may work in some circumstances, but it assumes visibility of all processes. However in business-to-business commerce people whose data is being used do not always have sight of that fact and this will also be true when AI is involved as part of the process. So the requirement is only sufficient for direct consumer services. Risk assessments should be conducted when AI is being used, which identify people who may be affected and provide a means for alerting people to this if necessary. There are three distinct components to consider. First, the information (data) ingested, the legal gateway to do so, and its provenance. Second, the code that uses the information (data) to generate various outputs, including text, images, sounds. Third, how the outputs will be used in various ways at the individual and group levels. Transparency is required for all components to attain adequate transparency. Recent areas of



concern with generative AI (LLMs) include hallucinogenic AI whereby data sources are quoted that do not exist.

2. What other transparency measures would be appropriate, if any?

3.2.2 Evidence in support of validation of systems should be published according to the standards of intelligent transparency – proactively taking an open, clear and accessible approach. This might mean advancing specific explanation criteria, eg, by setting out how an AI system was each of autonomous and adaptive – this would keep it practical. It is important that evidence is accessible: some consumer information is already overwhelmingly detailed, so user testing is essential to ensure accessibility.

4. How could routes to contestability or redress for AI-related harms be improved, if at all?

3.2.3 Routes to contestability or redress could be improved by sharing information about the harms caused – so putting restrictions on what can be withheld from the public domain where redress has been obtained. In drug development, product labels describe potential harms and the expected incidence. The labels also contain information under which circumstances a product should not be used.

5. Do you agree that, when implemented effectively, the revised cross-sectoral principles will cover the risks posed by AI technologies?

3.2.4 The transformative impact of systems working together and co-optimising is likely to lead to more complex challenges. Further, the definition of fairness used in the white paper is so unspecific that it may not be sufficient to safeguard against the disadvantaging of vulnerable groups. (See also response to Q1.) The whole system impacts on the behaviour of humans, and further development of new systems and professional capability of developers will need consideration soon.

6. What, if anything, is missing from the revised principles?

3.2.5 The ethical idea of human dignity in respect of systems is overlooked. More broadly, the principles are heavily focused on impact, yet there needs to be some foresight so that AI systems are not deployed recklessly. The principles are heavily focused on consumers and



individuals when some of the challenges are about and the transformation of whole sectors and services. There needs to be further attention to the group-level ethics relevant to AI systems.

### 3.3 A statutory duty to regard

7. Do you agree that introducing a statutory duty on regulators to have due regard to the principles would clarify and strengthen regulators' mandates to implement our principles, while retaining a flexible approach to implementation?

3.3.1 We expect regulators to deliver their responsibilities in a trustworthy manner: ie, to be honest, competent and reliable. The need for honesty in regulation is axiomatic, competence can be supported by the central capability. Being reliable is about an ethical commitment and that can be seen in more challenging cases, with more complex systems and economic incentives. What is needed is a suitable commitment to have regard in these more difficult cases, and that is not clearly provided by the framework: it needs leadership. It is important that regulatory oversight includes a series of checks and balances that are enforced through audit of high risk projects. A trustworthy regulator would need to demonstrate competence, reliability and integrity (this is the view of trustworthiness developed by Onora O'Neil in, eg, her [Reith lecture](#)). That would point to a diverse and representative body, that enables innovative and agile AI whilst avoiding mishaps.

8. Is there an alternative statutory intervention that would be more effective?

3.3.2 Rather than an alternative, what is needed is leadership, to support ethical commitment to reliable implementation of the principles, in challenging situations. While the professions will be part of this, looking to the role that engineering professions already take, there is a scrutiny role for parliament. To the extent that the principles are codifiable and require political independence and expertise to implement them effectively, parliamentary scrutiny is well suited to the accountability of regulator for the duty to have regard to them. Select committees in parliament would need considerable development of expertise to do this effectively, as understanding of uses of models of which are not both autonomous and adaptive has been very limited as yet. The RSS would be pleased to work with relevant bodies in supporting and developing the understanding of parliamentary staff.





### **3.4 New central functions to support the framework**

#### 9. Do you agree that the functions outlined in section 3.3.1 [of the white paper] would benefit our AI regulation framework if delivered centrally?

3.4.1 The Alan Turing Institute have produced a good paper on this topic ([Common Regulatory Capacity for AI](#)), and the functions drawing on that are very important. All of these would be beneficial, but we note that CDEI was intended to do some of this work and has not been able to deliver that role. We submit that this is very much more difficult than has been anticipated and remains so.

#### 10. What, if anything, is missing from the central functions?

3.4.2 Leadership, as noted in the industry demand for coherence but not listed in activities. As this is considered part of the professional responsibilities of engineers, the Alliance for Data Science Professionals will consider taking up a role in leadership for the profession, as the government cannot. However, this would be where the AI Council and Foundation Models Taskforce could have a very useful new role, working with industry and government professions and we encourage the government to convene such.

#### 11. Do you know of any existing organisations who should deliver one or more of our proposed central functions?

3.4.3 Many self-contained activities, like the Evaluation Task Force, the Analysis Function, the Regulatory Horizons Council, the Digital Regulation Cooperation Forum, the What Works Network and the Government Office for Science should have a role within government. Some outside bodies like the Science Media Centre are stepping in to inform the public, and organisations like Which? will respond to what consumer information needs are. Academic bodies will also take a role as seems to be suitable, but coordinating across the landscape is central, and the Office for Statistics Regulation is developing its own role in relation to other bodies in respect of regulating AI applications in official statistics.

#### 12. Are there additional activities that would help businesses confidently innovate and use AI Technologies?



3.4.4 As we detail in §2, evidence standards and evaluation methods are crucial. Particularly methods for evaluating the impact of AI systems, and the stages which review should go through – these would then be able to support work in specific sandboxes.

12.1. If so, should these activities be delivered by government, regulators or a different organisation?

3.4.5 Standards are in development with bodies such as National Physical Laboratory, but the review stages are better set by sector specific regulators. However, methods work should be more ambitious, analogous to the investment in research centres in clinical trials methodology. These should be research council funded standing investments, in cross council collaborations.

13. Are there additional activities that would help individuals and consumers confidently use AI Technologies?

3.4.6 The universal education espoused by the AI Council in its roadmap, ‘what everyone needs to know’ is helpful starting point. It would be valuable to test the information that is available to assess its suitability for informing the decisions of consumers with a wide range of data literacy alongside the general provisions proposed.

13.1. If so, should these activities be delivered by government, regulators or a different organisation?

3.4.7 There are a small number of academic organisations which work on public understanding. The Science Media Centre ought to be able to adapt to respond to news stories about AI, from industry and government as well as academic sources. It is not obvious that further activity by government on any of this would be trusted by consumers needing it.

14 How can we avoid overlapping, duplicative or contradictory guidance on AI issued by different Regulators?

3.4.8 An overarching regulative framework – produced with a view to consistency, coherence and horizon scanning – would seem to be the natural way to avoid these challenges. Leadership would also help considerably on this point. Professional networks can support information sharing and norms promoting the public interest. Parliament also ought to have a role, and





parliamentarians will need better information about what is happening to be able to effectively fulfil this function.

### **3.5 Monitoring and evaluation of the framework**

#### 15. Do you agree with our overall approach to monitoring and evaluation?

3.5.1 An effective monitoring and evaluation system must be able to identify issues early on and work to prevent serious mishaps. It is not clear what the overall approach proposed is and less clear still that it will work to identify issues at an early stage. We think work is needed to define it and would be happy to support in this.

#### 16. What is the best way to measure the impact of our framework?

3.5.2 Funding a serious academic project to do so, through either the Arts and Humanities Research Council (AHRC) or the Economic and Social Research Council (ESRC).

#### 17. Do you agree that our approach strikes the right balance between supporting AI innovation; addressing known, prioritised risks; and future-proofing the AI regulation framework?

3.5.3 It is far too strongly focused on self-contained individual consumer technology. More complex and less visible AI systems augmenting other digital systems and interacting behind the scenes or operating as business-to-business (B2B) services pose much more of a challenge and have not received adequate focus.

#### 18. Do you agree that regulators are best placed to apply the principles and government is best placed to provide oversight and deliver central functions?

3.5.4 Government can deliver, coordinate and commission in support of central services. Where government is deficient is in providing oversight, and this has already been attempted with CDEI, so a new approach is needed. Sector regulators ought to be able to apply principles if there is leadership in support of coherence, given planned central functions. However this will only be effective if government brings in the right expertise and undertakes effective horizon scanning. Such expertise requires wider representation including expert groups containing people working at the cutting edge of the discipline.



### **3.6 Regulator capabilities**

20 Do you agree that a pooled team of AI experts would be the most effective way to address capability gaps and help regulators apply the principles?

3.6.1 Capability in a rapidly growing area is something that the sector regulators will need to continually develop and refresh. There will be enormous demand for technically capable individuals for their skills and experience putting them into practice. So although it is a pragmatic aim and likely to be able to support capacity building it will not necessarily be a team that is easy to grow and resource at a level to ensure that it is high-performing.

### **3.7 Tools for trustworthy AI**

21. Which non-regulatory tools for trustworthy AI would most help organisations to embed the AI regulation principles into existing business processes?

3.7.1 Case studies and training materials of suitable technical depth to develop skills and understanding of experienced industry staff. Although many people are being trained, most of the AI workforce will be existing staff automating systems, and they need support.

### **3.8 Final thoughts**

22. Do you have any other thoughts on our overall approach? Please include any missed opportunities, flaws, and gaps in our framework.

3.8.1 *Quis custodiet ipsos custodes* (who watches the watchmen)? Regulators are anticipated to use AI themselves, especially in monitoring and risk evaluation, but how is such RegTech to be overseen? Parliament ought to have a role, but lacks the technical support, individual capability and structures.

### **3.9 Legal responsibility for AI**

L2.i. Do you agree that the implementation of our principles through existing legal frameworks will fairly and effectively allocate legal responsibility for AI across the life cycle?



3.9.1 Here we would note that as well as effective legal frameworks it is also important to make sure that processes are suitable. For example, existing legal frameworks have been strained by the post office subpostmasters' accounting system. If the law is suitable, then processes need to be suitable to uphold it.

L.2.ii. How could it be improved, if at all?

3.9.2 Legal definitions of fairness and bias will be tested, and will rely heavily on the information available to challenge consumer experiences.

**3.10 Foundation models and the regulatory framework**

F1. What specific challenges will foundation models such as large language models (LLMs) or open source models pose for regulators trying to determine legal responsibility for AI outcomes?

3.10.1 The technology-neutral, outcomes-based approach advanced needs more commitment.

**3.11 AI sandboxes and testbeds**

S1. Which of the sandbox models described in section 3.3.4 would be most likely to support innovation?

3.11.1 Sandboxes should be aligned with the evaluation pipeline, for example from ideation to proof of product to preparing for regulatory approval, as in the FCA approach.

S2. What could government do to maximise the benefit of sandboxes to AI innovators?

3.11.2 There is a risk of fragmentation if AI sandboxes are implemented only by specific regulators and are not cross-sectoral. The government could consider centralising AI Sandboxes, coordinated out of for example the Cabinet Office, similar to the Evaluation Taskforce.

S3. What could government do to facilitate participation in an AI regulatory sandbox?

3.11.3 Setting restrictions on entry, so that access to a sandbox means endorsement of some aspect of the development, and making such criteria subject to public consultation.



S4. Which industry sectors or classes of product would most benefit from an AI sandbox?

3.11.4 There are a few areas where a sandbox would seem especially beneficial:

- Consumer-facing applications where firms or government agencies do not routinely collect data on protected characteristics, for example in markets like general insurance or consumer credit.
- Products that lead to decisions with important consequences.
- First of kind products – especially in new areas with a high risk assessment.
- Products where there will be little human oversight when implemented at scale.
- Areas where misinformation can spread quickly and cause harm.

