WRITTEN EVIDENCE FOR THE OFFICE OF ARTIFICIAL INTELLIGENCE

AI REGULATION: A PRO-INNOVATION ARPPROACH

JUNE 2023

MINDEROO CENTRE FOR TECHNOLOGY & DEMOCRACY

Written evidence for the Office for Artificial Intelligence
AI regulation: a pro-innovation approach

by Dr Ann Kristin Glenster, Senior Policy Advisor on Technology Governance and Law, Minderoo Centre for Technology and Democracy, University of Cambridge

Evidence submitted by the Minderoo Centre for Technology and Democracy, University of Cambridge

20 June 2023

The Minderoo Centre for Technology and Democracy is an academic research centre at the University of Cambridge, with world-leading expertise in the regulation and governance of emerging technologies. We submit the following evidence.

www.mctd.ac.uk

Summary of the Submission:

The Office for Artificial Intelligence has requested evidence on the proposed pro-innovation framework set out in the UK Government's AI White Paper.[1] The framework is based on a 'contextual, sector-based' approach whereby the regulation of commercial innovation of AI technologies will be distributed to existing regulators.[2] These will be tasked with overseeing the proportionate implementation of values-focused principles.

This submission argues that the framework should articulate the proposed principles and make these mandatory through statutory enactment. Further the submission proposes the adoption of clear transparency and explainability standards, mandatory risk assessments, and a clear guidance as to how human-centric values should be the framework's governing norm. The UK should adopt a safety-by-design to AI innovation and ensure that regulation is informed not just by industry, but also by civil society and academia in order to foster an environment for responsible AI innovation for the public good.

[1] Department for Science, Innovation & Technology (DSIT), *A pro-innovation approach to AI regulation* (CP 815 March 2023) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper.
[2] Michael Birtwistle and Matt Davis, Regulating AI in the UK: three tests for the Government's plans: Will the proposed regulatory framework for artificial intelligence enable benefits and protect people from harm? Ada Lovelace Institute, 13 June 2023 https://www.adalovelaceinstitute.org/blog/regulating-ai-uk-three-tests/.

Main points:

1. Legislation should be adopted making the proposed principles clear and mandatory
2. Organisations developing and using AI should have to comport with transparency and explainability standards
3. Organisations developing and using AI should be compelled to publish risk assessments
4. The proposed principles must be human-centric
5. Rather than a 'duty of having regard', the UK should adopt a mandatory safety-by-design based approach to the governance of AI innovation
6. Central functions need clear mandates and resources
7. Monitoring must include evidence from civil society and academia
8. The proposed principles should be made specific and mandatory in law, adopting a safety-by-design approach

Question 1: 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? (2) what other transparency measures would be appropriate, if any? (3) Do you agree that current routes to contestability or redress for AI-related harms are adequate? (4) How could routes to contestability or redress for AI-related harms be improved, if at all? (5) Do you agree that, when implemented effectively, the revised cross-sectoral principles will cover the risks posed by AI technologies? (6) What, if anything, is missing from the revised principles?

1. Legislation should be adopted making the proposed principles clear and mandatory

The UK Government's proposed regulatory framework sets out five values-focused cross-sectoral principles: (1) safety, security, and robustness, (2) appropriate transparency and explainability, (3) fairness, (4) accountability and governance, and (5) contestability and redress.[3] The intention of the framework is that individual regulators shall oversee the adherence to these principles in AI innovations in their respective sectors, while centralised functions will monitor, evaluate, and coordinate how regulators discharge this novel task.

It is problematic that the principles are not clearly articulated or placed on a statutory footing. From the Government's AI White Paper, it is unclear how much prominence regulators are expected to give the principles, with the risk that the

[3] DSIT *supra* note 1, p. 26.

principles are only considered once other priorities have been met or are challenged by judicial review in cases where a regulator has a statutory remit with limited discretion. It is highly unclear how the efficacy of the principles, or the lack thereof, will be assessed to determine whether to adopt these in statutes.

While laudable, the proposed principles are interdependent and therefore the framework should be amended to ensure that regulators enforce all, not just some, of the principles when monitoring and evaluating the development of AI technologies. As currently proposed, the framework will allow regulators to choose which principles applies to their remit. This flexibility undermines the overall effectiveness of the framework.

The principles are also vague which means that it is highly unclear how they are to be interpreted or weighed. For example, it is not clear if the principle of fairness would include a commitment to ensure that AI technologies do not include bias or produce discriminatory outcomes; or that a workforce of coders and developers should reflect the demographic diversity.

It is also not clear how the principles will be given effect in legislation (once and if legislation is adopted) as the Government's AI White Paper states that legislation will not go beyond making it mandatory for organisations to have a 'duty to have regard' to these. We are concerned that the lack of a plan for the mandatory adoption and enforcement of the principles will undermine any notable positive effect these principles may have in the emerging regulatory landscape governing AI.

2. Organisations developing and using AI should have to comport with transparency and explainability standards

In order to deliver on the principle of appropriate transparency and explainability, organisations should be compelled to not only inform the public of the fact that they are using AI, but also explain how the AI systems are used, including how they make decisions affecting individuals. To that effect, organisations should adopt a clear approach to explaining how the AI systems work, building on the emerging XAI standards in law.[4] Transparency and explainability are fundamental pillars of trust, and organisations and regulators should be obliged to ensure that information is

[4] See for example Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S. & Yang, G-Z. (2019). XAIExplainable artificial intelligence. Science Robotics, 4(37), eaay7120. doi: 10.1126/scirobotics.aay7120.

given to the public in a timely fashion and in an easily digestible format, adopted to the capabilities of the specific groups concerned.

One of the central functions of government should be to coordinate action between the regulators to develop a standardised framework for explainability so the public and individuals are able to navigate similar information from several organisations with ease. The central functions should include activities to publish the work of regulators on a central platform. This requirement should therefore go beyond simply encouraging regulators to publish their activities independently of one another.

3. Organisations developing and using AI should be compelled to publish risk assessments

Organisations should be compelled to publish risk assessments; and members of the public and interested parties, such as civil rights organisations, should have a right to challenge those risk assessments with the appropriate regulator. Regulators should have a duty to assess the risk assessments and issue specific regulatory guidance to named organisations when the organisations' risk assessments are found to be inadequate. Regulators should also have the powers to demand organisations explain how they have adopted appropriate mitigating safeguards based on the risk assessments approved by the regulators, and information about these mitigation measures should be made available to the public.

4. The proposed principles must be human-centric

It is concerning that none of the principles refer to human-centric AI or the role of humans in overseeing and benefitting from AI as an overarching goal. We have a major concern regarding the limitations and decontextualization of the framework, as: "The proposed regulatory framework does not seek to address all the wider societal and global challenges that may relate to the development or use of AI. This includes issues relating to access to data, compute capability, and sustainability, as well as the balancing of the rights of content producers and AI developers".[5] It is not clear why these issues are not addressed in the framework or how future frameworks addressing these issues will sit with the current proposal.

Scholars have critiqued how the framework promotes a model of deregulation where public trust is instrumentalised to ensure acceptance for the Government's

[5] DSIT *supra* note 1, p. 20.

objective of industry growth and innovation.[6] The instrumentalization of trust is normatively hollow as it suggest that the values that underpin the principles are mere 'slogans' with no mooring in, for example, the protection of fundamental rights. By coming up short in giving the principles legal effect, it is not clear why the public should have trust in their efficacy. The exclusion of wider concerns regarding the introduction of AI in ways that have severe impact on fundamental rights – for example, surveillance of the workplace – should be included in any framework on the governance of novel AI technologies.[7]

It is further concerning that the framework does not address the need for AI innovation in the public interest. The framework promotes the conditions for AI innovation for industry, which will enhance their economic growth, but it does not address how AI is needed to deliver functions in the public interest, and how to ensure that the industry is not at liberty to hold public functions, especially hose providing critical infrastructure, 'hostage' by imposing unreasonable price demands or terms and conditions for the use of new commercial AI technologies.

AI innovation is expensive and demands access to datasets, computational powers, coders, and energy at scale. The proposed framework does not explain how it will benefit innovation by other (British) companies than established, predominantly foreign-based- and owned technology behemoths, who are likely to do little more than try to impose the principles onto their existing data-extracting business models.[8] The monopolistic features of the technology industry illustrate how the UK needs to bolster competition law to ensure that competitive AI services and products will reach the market, yet the framework does not seem to offer any tangible solutions to these challenges.

The framework does not protect the public assets that will be used in private-sector AI innovation, such as data, whether that data comes from individuals or public agencies, or imposes obligations on organisations to share appropriate information and data in the public interest. Scholars from the UCL Institute for Innovation and Public Purpose and Stanford Cyber Police Center write, "…AI investment will

[6] Andre Charlesworth, Kit Fotheringham, Colin Gavaghan, Albert Sanches-Graells, and Clare Torrible, *Response to the UK's March 2023 White Paper "A pro-innovation approach to AI regulation",* Centre for Global Law and Innovation, University of Bristol Law School, 19 June 2023.
[7] Henry Parkes, Watching Me, Watching You: Worker Surveillance in the UK After the Pandemic, Institute for Public Policy Research (IPPR), March 2023, https://www.ippr.org/research/publications/worker-surveillance-after-the-pandemic.
[8] By contrast see the Obama Administration's innovation diversity approach, cited by Mazzucato, M., Schaake, M., Krier, S. and Entsminger, J. (2022). Governing artificial intelligence in the public interest. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2022-12). https://www.ucl.ac.uk/bartlett/public-purpose/wp2022-12, p. 11.

naturally tend toward capital-rich environments rather than flowing towards addressing unmet social needs."[9] They find that, "… the private sector is never incentivised to fund and develop longer-term innovation that may be globally beneficial but are commercially risky."[10] In its current conceptualisation, the Government's proposed framework is likely to widen the gap between the access to AI technologies by the monied commercial sector and the lack of access by civil society and the public sector. Ultimately, this gap will be so wide that it will become a democratic problem and risk undermining not just the public's trust in AI but also the very functioning of our public institutions.[11]

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

5. Rather than a 'duty of having regard', the UK should adopt a mandatory safety-by-design approach to the governance of AI innovation

Only "having regard" to a statutory duty to consider the principles would weaken the principles' effectiveness and robustness, with a knock-on effect of also likely undermining the public's trust in AI. Instead of instituting a 'duty to have regard to' the principles as a statutory duty, we recommend a greater focus on a safety.by-design approach whereby organisations developing and using AI should be compelled to demonstrate how they have identified and addressed specific risks pertaining to their use of AI systems. The regulators should have a duty to enforce the adoption of robust and forward-looking mitigation strategies and be able to fine organisations for lack of compliance.

The time it would take for the principles to be evaluated, then introduced as a 'duty of care' on the statute books would likely undermine any benefits of the proposed principles as the AI systems, with the potential to cause considerable harm, would already be in place. There are no mechanisms in the framework to ban or suppress harmful AI once it has been released into the real world. At the point that a 'duty to have regard' would be adopted in statute, it is likely that other, more stringent, sector-specific rules would have come into effect, further diminishing the value of

[9] *ibid,* p. 5.
[10] *ibid,* p. 10.
[11] *ibid,* p. 14

such a duty. There is therefore a great risk that a 'duty to have regard' would be perceived by organisations, regulators, and the public as little more than a tick-box exercise.

New central functions to support the framework: (9) Do you agree that the functions outlined in section 3.3.1 would benefit our AI regulation framework if delivered centrally? (10) What, if anything, is missing from the central functions? (11) Do you know of any existing organisations who should deliver on one or more of our proposed central functions? (12) Are there additional activities that would help businesses confidentially innovate and use AI technologies? If so, should these activities be delivered by government, regulators, or a different organisation? (13) Are there additional activities that would help individuals and consumers confidently use AI technologies? If so, should these activities be delivered by government, regulators or a different organisation? (14) How can we avoid overlapping, duplicative or contradictory guidance on AI issued by different regulators?

6. Central functions need clear mandates and resources

The central functions that are set out in the proposed framework are crucial to its delivery and should be bolstered. From the Government's AI White Paper, it is not clear how this regulatory capacity will be nurtured within the existing framework.[12] The central functions should include a responsibility to oversee researchers' access to data and to collect information on the use of AI systemically, including its impact on sustainability. The central functions should be empowered to compel regulators to develop frameworks for enforcement actions in their sectors. Scholars have noted that there currently is no robust mapping of the regulatory landscape governing AI and that the difference in remits and powers makes it difficult to see how the intention of the framework would be realised.[13]

Unless the principles are given legal weight, they would be competing for attention and resources with already set and oversubscribed regulatory priorities. Some commentators have also noted that the framework does not explain how regulatory

[12] For example, through the Office for Artificial Intelligence or under the remit of the Minister for Artificial Intelligence and Intellectual Property of the Department for Science, Technology and Innovation (DSIT).
[13] Andre Charlesworth, Kit Fotheringham, Colin Gavaghan, Albert Sanches-Graells, and Clare Torrible, *Response to the UK's March 2023 White Paper "A pro-innovation approach to AI regulation",* Centre for Global Law and Innovation, University of Bristol Law School, 19 June 2023

capacity will be ensured, especially given the high risk of capture by industry.[14] Without proper funding and allocated resources, the public cannot have confidence that the centralised functions will indeed deliver on the vision set out in the framework.

Monitoring and evaluating the framework: (15) Do you agree with our overall approach to monitoring and evaluation? (16) What is the best way to measure the impact of our framework? (17) Do you agree that our approach strikes the right balance between supporting AI innovation; addressing known, priorities risks, and future-proofing the AI regulation framework? (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?

7. Monitoring must include evidence from civil society and academia

The success of the principles is contingent on there being a common understanding of what they mean, and a willingness of organisations and AI developers to integrate the principles in their practices. The monitoring activities in the framework are essential to foster an environment in which the principles can take effect. It is, however, concerning that the principles are free-floating and not tied to overall Governments' objectives or values, for example addressing sustainability or inequality.

The monitoring and evaluation framework has significant deficiencies in ensuring that it is future-proof. First, it does not adequately draw on and pay heed to academic research into AI technologies, their societal impact, and regulation. This is a failure to exercise joint-up-thinking. Second, the framework does not detail how regulators, government, or academics shall have access to adequate information from organisations using AI. Thus, the monitoring activities can be stifled by claims that providing salient information would violate legal rights to confidentiality or intellectual property. Clarification of the powers of interested parties to compel organisations to provide information and access to data is therefore needed.

Legal Responsibility for AI: (L1) What challenges might arise when regulators apply the principles across different AI applications and system? How could we address

[14] *ibid.* See also Ada Lovelace Institute, *Regulate to Innovate: A route to regulation that reflect the ambition of the UK AI Strategy, November 2021,* https://www.adalovelaceinstitute.org/report/regulate-innovate/#:~:text=These%20are%3A,the%20paper%20does%20not%20define).

these challenges through our proposed AI regulatory framework? (L2) 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? How could it be improved, if at all?

8. The proposed principles should be made specific and mandatory in law, adopting a safety-by-design approach

The values-focused principles will not produce the desired effect if they are not made enforceable in law. The principles are too vague and as highlighted in the Government's AI White Paper, likely to conflict. Currently, the flexibility of the framework undermines their effectiveness. Thus, more specific definitions are needed, and the principles should be tethered to specific legal provisions. Simply telling the public that AI is used as a way to discharge duties arising from the principle of transparency will not foster trust in these technologies unless the public also knows it can rely on the regulators and courts to uphold the other principles.

Implementing the principles in existing legal frameworks overseen by established regulators is a sensible approach. However, making the principles vague and voluntary undercuts their effectiveness and robustness. The lack of specificity and enforceability will undermine public trust in AI technologies. Regulators should have to devise clear and enforceable standards and frameworks for how organisation should be compelled to integrate how the principles in measurable ways in the design of their AI systems, taking a mandatory safety-by-design approach. Regulators should be given powers to fine organisations for failure to have regard to the principles. Organisations should only be permitted to avoid principles when they have a clear risk assessment which demonstrates how they have applied to criteria of proportionality, and the public should have a right to challenge that test with the regulators, and possibly even in the courts.

The Government's AI White Paper emphasises that the UK will pursue an agenda of 'global interoperability and international engagement.'[15] While we support the UK's ambition to take a lead in developing legal and regulatory frameworks to govern AI, we are concerned that the efforts in this fields are concentrated on international rather than national legal frameworks, which are unlikely to afford British citizens rights of protection they can enforce through domestic courts, or if they can do so, is likely to be too cumbersome or costly to give individuals a meaningful means of

[15] DSIT *supra* note 1, pp 68-70.

legal redress. Outsourcing legal accountability for responsible AI to an international framework will not suffice to engender trust in AI at home.

Conclusion

Trust can only be begotten from principles with concrete and actionable meaning. The vagueness of the 'values' that underpin the five proposed principles of the proposed pro-innovation framework may produce the opposite effect as the public will not be able to see how these will be used to ensure that innovation leads to responsible AI for the public good.

References

Ada Lovelace Institute, Regulate to Innovate: A route to regulation that reflect the ambition of the UK AI Strategy, November 2021, https://www.adalovelaceinstitute.org/report/regulate-innovate/#:~:text=These%20are%3A,the%20paper%20does%20not%20define)

Michael Birtwistle and Matt Davis, Regulating AI in the UK: three tests for the Government's plans: Will the proposed regulatory framework for artificial intelligence enable benefits and protect people from harm? Ada Lovelace Institute, 13 June 2023, https://www.adalovelaceinstitute.org/blog/regulating-ai-uk-three-tests/

Andre Charlesworth, Kit Fotheringham, Colin Gavaghan, Albert Sanches-Graells, and Clare Torrible, Response to the UK's March 2023 White Paper "A pro-innovation approach to AI regulation", Centre for Global Law and Innovation, University of Bristol Law School, 19 June 2023

Department for Science, Innovation & Technology, A pro-innovation approach to AI regulation (CP 815 March 2023) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S. & Yang, G-Z. (2019). XAIExplainable artificial intelligence. Science Robotics, 4(37), eaay7120. doi: 10.1126/scirobotics.aay7120

Mazzucato, M., Schaake, M., Krier, S. and Entsminger, J. (2022). Governing artificial intelligence in the public interest. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2022-12). https://www.ucl.ac.uk/bartlett/public-purpose/wp2022-12

Henry Parkes, Watching Me, Watching You: Worker Surveillance in the UK After the Pandemic, Institute for Public Policy Research (IPPR), March 2023, https://www.ippr.org/research/publications/worker-surveillance-after-the-pandemic