

Carnegie UK Response to AI Strategy White Paper

June 2023

1. We welcome the opportunity to submit to this consultation. In recent years, Carnegie UK has been at the forefront of the debate in the UK on digital regulation. We created the approach that underpins the Online Safety Bill, which is shortly to complete its Parliamentary passage: a statutory duty of care which focuses on systems and processes and is enforced by a regulator. Since 2018 we have been working with parliamentarians, civil society and the government to develop and refine proposals for online harm reduction. William Perrin (Carnegie UK Trustee) and Prof Lorna Woods (University of Essex) have given evidence on the Bill to numerous Parliamentary committees and were awarded OBEs for their work in 2020.
2. Our work can be found [here](#); and a dedicated resource page for our analysis and publications on the Online Safety Bill specifically is [here](#).
3. The Government's AI Strategy White Paper has already been overtaken by events since publication – both national and international developments and scrutiny relating to the technology and its deployment and the Prime Minister's (markedly more cautious) statements¹ on its proposed approach to regulation.
4. This consultation response does not therefore go through the White Paper detail line by line but instead makes some high-level observations about the existing models for regulating fast-moving technologies and the Government's regulatory options, based on our expertise in this area. In summary, we would urge the Government to:
 - a. Adopt a long-term approach to AI regulation that is built on the precautionary principle informed by the experience of prior science based moral panics as defined by the Inter-Departmental Liaison Group on Risk Assessment (ILGRA)
 - b. Rapidly assess the legislative levers they already have in play and map the gaps where AI-generated harms (especially via the impact of algorithms) may emerge
 - c. Ensure a more robust approach to regulatory funding and coordination than envisaged in the AI White Paper – both for current and future harm identification and mitigation
 - d. Redress the apparent lack of civil society engagement in this agenda
 - e. Ensure that – while taking the global lead on AI – its proposals continue to evolve as other jurisdictions bring in their own legislative approaches.

The precautionary principle

5. The AI White Paper is built on a "pro-innovation" approach to regulation. It states in paragraph 9 that "Our framework is designed to build the evidence base so that we can learn from experience and continuously adapt to develop the best possible regulatory regime. Industry has praised our pragmatic and proportionate approach."
6. We would strongly advise the Government not to proceed along these lines. As we set out in a detailed reference paper back in 2019, where we brought together all our work to date on the proposal for an approach to online harms regulation founded on a statutory duty of care enforced by an independent regulator: "The continual evolution of online services, where software is updated almost continuously makes traditional evidence gathering such as long-term randomised control trials problematic. New services adopted rapidly that potentially cause harm illustrate long standing tensions between science and public policy."²

1 <https://www.theguardian.com/technology/2023/may/18/uk-will-lead-on-guard-rails-to-limit-dangers-of-ai-says-rishi-sunak>

2 "Online Harm Reduction: a statutory duty of care and a regulator" – found here: https://d1ssu070pg2vqi.cloudfront.net/pex/carnegie_uk_trust/2019/04/08091652/Online-harm-reduction-a-statutory-duty-of-care-and-regulator.pdf

7. We are at this point with artificial intelligence. The cross-sectoral, "wait and see" approach proposed in the White Paper, suggesting that individual AI risks will emerge in different areas of industry or public service and can be dealt with by sector-specific regulators, with some kind of knitting together of approaches for post-hoc consistency, is fundamentally flawed. The bottom-up structure reinforces a siloed sectoral approach to regulation. The scale and pace of the developments in AI means that the elements which cause harm – both individual and societal – will be entrenched long before the regulators have time to gather and review the evidence in order to act.
8. The BSE, GM Foods and the nanotechnology scares of the late 1990s were conceptually very similar to the current AI moral panic – a failure of scientists to communicate risks to the public and the amplification of those risks real and imagined by the media. These science/risk scares led to real economic damage to the UK. After the event, UK scientists and administrators worked on theoretical models for communication and understanding of risk. A part of this was the Inter-Departmental Liaison Group on Risk Assessment (ILGRA) which published a superb exposition of the precautionary principle, designed to enable economic and scientific progress where evidence was lacking about serious risks but to do so in such a manner to maintain public confidence. The ILGRA work (Annex A) stood in stark contrast to more conventional understanding of the precautionary principle which can seem to be deploying it to stop things rather than enable them. This ILGRA work was until recently at the heart of the HSE approach to risk management and was inexplicably overlooked by the Vallance paper.
9. As we set out in our 2019 paper: "For decades scientists and politicians have wrestled with commercial actions for which there is emergent evidence of harms: genetically modified foods, human fertilisation and embryology, mammalian cloning, nanotechnologies, mobile phone electromagnetic radiation, pesticides, bovine spongiform encephalopathy. In 2002, risk management specialists reached a balanced definition of the precautionary principle that allows economic development to proceed at risk in areas where there is emergent evidence of harms but scientific certainty is lacking within the time frame for decision making.³ ... The precautionary principle provides the basis for policymaking in this field, where evidence of harm may be evident, but not conclusive of causation."
10. With regard to social media regulation, we advised that "companies should embrace the precautionary principle" because it prevented the need for banning particular types of content and instead took a systemic approach to regulation, founded on risk assessment. (While the Online Safety Bill retains some of the aspects of this approach in its foundational structure, we would argue that – had the Government followed the precautionary principle throughout the long evolution of the legislation, it would have saved itself many of the intractable political controversies around the nature of "harm" online and led to a much simpler Bill.) We went on to ask how regulation and economic activity proceed in the face of indicative harm but where scientific certainty cannot be achieved in the time frame available for decision making?:

This is not the first time the government has been called to act robustly on possible threats to public health before scientific certainty has been reached. After the many public health and science controversies of the 1990s, the UK government's Interdepartmental Liaison Group on Risk Assessment (ILGRA) published a fully worked-up version of the precautionary principle for UK decision makers. 'The precautionary principle should be applied when, on the basis of the best scientific advice available in the time-frame for decision-making; there is good reason to believe that harmful effects may occur to human, animal or plant health, or to the environment; and the level of scientific uncertainty about the consequences or likelihoods is such that risk cannot be assessed with sufficient confidence to inform decision-making.' The ILGRA document advises regulators on how to act when early evidence of harm to the public is apparent, but before unequivocal scientific advice has had time to emerge, with a particular focus on novel harms. ILGRA's work focuses on allowing economic activity that might be harmful to proceed 'at risk', rather than a more simplistic, but often short-term politically attractive approach of prohibition.

11. There is much to learn in these historic science/risk dilemmas. We would urge the Government to properly consider the merits of an ILGRA approach to the precautionary approach as it develops its global proposal ahead of the AI summit⁴ in the autumn and not repeat the same mistakes it made in approaching online safety regulation. We are heartened to see that one of those quoted in the Government's press release for the summit – Demis Hassabis of Deep Mind – had already argued that "as with any transformative

3 <https://publications.parliament.uk/pa/ld200304/ldselect/ldsctech/110/110we29.htm>

4 <https://www.gov.uk/government/news/uk-to-host-first-global-summit-on-artificial-intelligence>

technology we should apply the precautionary principle, and build & deploy it with exceptional care".⁵

12. The relevance of the precautionary principle to technology is set out in our 2019 exposition on this topic. We quote it in full here for ease of reference:

One of the recurrent arguments put forward for not regulating social media and other online companies is that they are unique or special: a complex, fast-moving area where traditional regulatory approaches will be blunt instruments that stifle innovation and require platform operators to take on the role of police and/or censors. Another is that the technology is so new, sufficient evidence has not yet been gathered to provide a reliable foundation for legislation; where there is a body of evidence of harm, in most cases the best it can do is prove a correlation between social media use and the identified harm, but not causation. We believe that the traditional approach of not regulating innovative technologies needs to be balanced with acting where there is good evidence of harm.

Existing regulatory and legislative approaches

13. The Government has a number of legislative vehicles at various stages of Parliamentary passage which address many of the current harms arising from technology, including use of algorithms and AI-driven technology. These include: the Online Safety Bill, which the Government has confirmed covers AI⁶; the Data Protection and Digital Information Bill; the Digital Markets, Competition and Consumers Bill; and the Media Bill. It also has commitments to publish its long-delayed Online Advertising Programme and its Product Safety Review – two large areas where consumer harm is rising exponentially and where algorithms drive targeted advertising. (With the recent Machinery of Government changes, it is not yet clear how the Department for Science, Innovation and Technology (DSIT) and Department for Culture, Media and Sport (DCMS) are going to address their shared interests here going forward.)
14. These are all vital pieces of the foundation on which AI regulation can be built but the attention being given to them at the centre of Government does not appear to be equal to the Prime Minister's enthusiasm for tech innovation and putting down a marker for the UK in the global AI debate. Good government is about getting the fundamentals right. Has DSIT reviewed the intersection of all these pieces of legislation through an AI-lens and considered their collective impact, the potential gaps and the requirements they make on regulators? This is particularly relevant to the aspects of the various Bills that cover algorithmic design and transparency, data collection, use and protection as well as requirements on regulated companies to ensure adequate risk assessment and mitigation strategies and the suite of new powers given to the respective regulators for information gathering and enforcement. An ILGRA-derived precautionary approach sitting on top of a solid, coherent legislative framework would put the UK in a uniquely strong international position and give the Government legitimacy as it attempts to consolidate its global leadership in this space to address emerging and future risks from the broader application and use of AI.
15. There are also some fundamentals of regulation which provide existing protections for individuals from AI-generated harms. We acknowledge that the approach set out in the AI White Paper does suggest that the Government's favoured (initial) approach was to ensure that existing specialist regulators would take the lead in their respective sectors working within existing regulatory frameworks. But we see no mention of the cross-cutting protections that already exist for individuals and employees via legislation such as the Health and Safety at Work Act 1974 – indeed the Government themselves have confirmed, via a PQ⁷, that this is the case. AI is the latest of a long line of technologies that when deployed by companies could harm workers in the workplace and other people. HSAW1974 already has a logical framework to deal with things that might be harmful but that might be too complex for a company deploying them to assess fully the risk of deployment. HSAW74 Section 6 requires that the supplier of things for use at work tests them for safety (See below).
16. This approach is micro economically efficient as it returns the substantial cost of testing to the originator of the potential risk. One can assume that testing of things supplied for use of work forms part of insurance cover acting as a driver for adoption. The government has confirmed that HSAW Section 6 applies to AI.

⁵ <https://twitter.com/demishassabis/status/1663553039960256512?s=20>

⁶ DSIT questions 14/6/23: <https://hansard.parliament.uk/commons/2023-06-14/debates/82DBB544-6A19-4A47-BF09-DA2D32F016CE/ArtificialIntelligenceRegulation>; OSB Committee day 10: <https://hansard.parliament.uk/lords/2023-05-25/debates/619A6A9E-6D1F-4007-8171-C00A2BB0B01D/OnlineSafetyBill>

⁷ <https://questions-statements.parliament.uk/written-questions/detail/2018-05-23/HL8200>

A sensible area of focus for the government would be to work towards an international AI safety testing institute that is sufficiently independent from AI companies. We note that the HSE produced a paper on safety management in expert systems over 20 years ago. ('Industrial use of safety-related expert systems' Robertson and Fox 2000). The government should ensure that the Health and Safety Executive is adequately funded to support enforcement of this Section of HSAW74 in respect of AI.

17. On this point, we refer to the evidence we submitted to the recent Science and Technology Committee inquiry into AI governance:

A need for regulation implies that there are costs arising from the use of AI in a production decision which do not fall on the company but on wider society – such as workers, customers and third parties. This causes allocative inefficiency or individual or social harm. There are many regulatory mechanisms for regulation of external costs – these should be explored before reaching for new uncertain models. A common and highly successful approach to returning external costs to the production decision is risk-based, proportionate regulation or self-regulation focused on the outcomes of a company process. Such a model should operate whether or not AI is involved. A strong and effective example of this model is the Health and Safety at Work Act 1974. A statutory duty of care enforced by a regulator has proven effective and futureproofed. The Act firmly applies to AI in the workplace as we describe below. This approach is flexible and future-proofed, focusing as it does on the outcomes that arise from service design - the company systems and processes, rather than the specifics of the technology that underpins them. One advantage with this approach is that it could apply as a base-level principle across the use of AI generally, but could also be deployed in sector-specific contexts, so as to allow the particular characteristics and risks of those contexts to be taken into account. ... Having a common approach potentially allows the interconnection between sector specific rules and general AI rules to occur.

Our proposal for a statutory duty of care for online harm reduction which has been adopted, in part, by the Government in the Online Safety Bill, drew on the approach that, for nearly 50 years, has underpinned Health and Safety legislation in the UK. We are of the view that this approach is applicable to AI and its application in many different industrial sectors, albeit that some of those sectors may require additional considerations or refinements to be added to the regulatory framework to take account of their specific risks and potential harms.

The Government, it would seem, agrees with our view. In 2018, we worked with Lord Stevenson of Balmacara on PQ UIN HL8200, tabled on 23 May 2018, about whether the Health and Safety at Work etc. Act 1974 applied to 'artificial intelligence' and algorithms used in the workplace (which might cause safety concerns). We set out the Government's answer in full here:

Section 6 of the Health and Safety at Work etc. Act 1974 places duties on any person who designs, manufacturers, imports or supplies any article for use at work to ensure that it will be safe and without risks to health, which applies to artificial intelligence and machine learning software. Section 6(1)(b) requires such testing and examination as may be necessary to ensure that any article for use at work is safe and without risks but does not specify specific testing regimes. It is for the designer, manufacturer, importer or supplier to develop tests that are sufficient to demonstrate that their product is safe.

The Health and Safety Executive's (HSE) Foresight Centre monitors developments in artificial intelligence to identify potential health and safety implications for the workplace over the next decade. The Centre reports that there are likely to be increasing numbers of automated systems in the workplace, including robots and artificial intelligence. HSE will continue to monitor the technology as it develops and will respond appropriately on the basis of risk.

This fact that HSAW74 applies to AI has languished in obscurity. It might suit some to ignore it while others prefer speculating about new AI laws rather than applying existing law, which remains fit for purpose across so many other sectors. Similarly, the existing data protection regime applies to AI. (The same is true of data protection as it is for user safety; the Information Commissioner has set out clearly how "the underlying data protection questions for even the most complex AI project are much the same as with any new project. Is data being used fairly, lawfully and transparently? Do people understand how their data is being used? How is data being kept secure?")

18. It is regrettable that in the Government's desire to reassure industry that its "pro-innovation" approach would defer any legislative or regulatory moves, the White Paper has further entrenched the idea that AI is some kind of shiny and new technology that is developed and deployed separately to products and services that

are already heavily regulated, particularly with regard to safety. Reminding industry of the need to focus on outcomes wherever AI is deployed – and that regulatory compliance is already required in many such areas – would have been an important signal. As we set out in our Science and Technology Committee submission: “Using HSAW74 as the baseline regulation would have simplified the regulatory approach and prevented companies using arguments that AI-driven technology is somehow novel or special in order to avoid scrutiny or oversight”. It would also have guarded against – as the White Paper potentially bakes in – the encouragement of “multiple sector-specific regulatory frameworks or different compliance requirements”.

19. The same is true of international regulatory frameworks and principles. The White Paper mentions that the cross-sectoral principles “build on and reflect our commitment to, the Organisation for Economic Co-operation and Development (OECD) values-based AI principles, which promote the ethical use of AI”. We note that the summary of responses to the consultation references that many stakeholders “felt further use of international approaches by organisations such as the OECD or UNESCO would add more human focused benefits and aid companies working across jurisdictions”. In response, the Government said “The white paper clarifies the substance of the principles in [section 3.2.3](#). Human rights and environmental sustainability are not explicitly named in the revised principles as we expect regulators to adhere to existing law when implementing the principles. We have emphasised the social benefits alongside the economic opportunities we intend to unlock with our pro-innovation approach to AI regulation.”
20. Again, this focus on the “pro-innovation” approach misses the internationally-recognised approach to risk management that underlines both the UNGPs and the OECD guidance on Responsible Business Conduct. This is a global common denominator that – again – would have been an important part of a strong foundation to underpin the Government’s approach.⁸
21. We also do not understand why the former Chief Scientist’s review of regulatory approaches for emerging technology⁹ seemed to abandon years of practice on science risk management – leading to the Government summarising his report by saying that he had found that “rushed attempts to regulate AI too early would risk stifling innovation.” It was particularly striking that Sir Patrick called for a shift in the burden of proof on drones away from operators – which is both economically inefficient on a polluter pays basis and goes against the advice of the ILGRA work that arose from the great science risk crises. As we argue above, there is already an established regulatory baseline in the HSAW74 which all companies deploying the technology need to adhere to: reasserting this could have ensured a greater industry focus on managing risk and mitigating harm in the here and now, creating space for the Government to then design its wider “pro-innovation” approaches to sit on top of this.

The UK's options for regulation

22. We agree with the Government in one specific regard: “creating a new AI-specific, cross-sector regulator would introduce complexity and confusion, undermining and likely conflicting with the work of our existing expert regulators” (para 47). As we set out in our evidence to the Science and Technology Committee, “the regulatory bodies with oversight of individual industrial sectors should retain the lead in the oversight of how industries and companies within those sectors are using AI. This might involve, for example, scrutiny of the risk assessment and mitigation processes in place for the development and updating of new industrial techniques using AI software; risk assessments of (a) the potential for harm arising from the operation of AI, ML and other software that controls systems and processes and (b) monitoring and performance management of staff.”
23. However, we do not think – even under the Government’s very loose framework of coordination and cooperation set out in the White Paper and its stated commitment that “existing regulatory forums may need to be supplemented or adapted to successfully implement the cross-cutting principles” – that existing expert regulators have either the resources or the powers to deliver on the cross-sectoral cooperation and coordination that the Government requires. For example, if the Government wishes that they “apply the cross-cutting principles” or “conduct detailed risk analysis and enforcement activities within their areas of expertise”, how will they be empowered (or compelled) to share their learning and evidence with other regulators potentially dealing with the same issues and/or harms from the same technology in a different sector?
24. To partly answer this question, the Government refers a number of times to the Digital Regulation Cooperation

8 <https://mneguidelines.oecd.org/mneguidelines/>

9 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1142883/Pro-innovation_Regulation_of_Technologies_Review_-_Digital_Technologies_report.pdf

Forum. We welcome this initiative and have seen good evidence in recent years of the participating regulators undertaking high-quality evidence-gathering and horizon-scanning in areas of cross-cutting interest, including algorithmic processing and transparency. But this kind of collaboration is not the same as statutorily-underpinned information-sharing and cooperation powers – an issue on which the Government is strangely reluctant to yield in debates on amendments to the Online Safety Bill.¹⁰ Most recently, in the Committee debate on Lord Clement-Jones's amendment, Lord Parkinson said that "Ofcom already has a statutory footing to share information with UK regulators under the Communications Act 2003. Section 393 of that Act includes provisions for sharing information between Ofcom and other regulators in the UK, such as the Information Commissioner's Office, the Financial Conduct Authority and the Competition and Markets Authority. So, we believe the issues set out in the amendment are covered."¹¹ In the Government's White Paper approach – envisaging multiple sector-specific regulators assessing risk and managing mitigations of potentially the same technology or products but applied in different ways – it is readily apparent how such a narrow set of information-sharing provisions between one regulator and a handful of others will quickly cause problems and lead to evidence of harm being missed or left unaddressed.

25. It is striking that Ofcom in order to regulate social media for its worst harms, in a proportionate manner has been authorised to recruit 300 staff against a current head count of 1,000. This is being underwritten by the Exchequer until a fees regime allows OFCOM to recover the cost from the largest companies on a micro economically efficient 'polluter pays' basis. The Vallance review notes that existing regulators find it hard to recruit AI-skilled staff. If the government was serious about existing regulators undertaking much new work on AI one would expect to see a one-off uplift in funding for the larger regulators, enabling recruitment at competitive rates (some are tied to public sector pay scales – OFCOM isn't) and then an ability to charge AI suppliers for work on AI risk assessment and mitigation on a polluter pays basis.

European and international considerations

26. Finally, we note that the Government states that "having exited the European Union we are free to establish a regulatory approach that enables us to establish the UK as an AI superpower. It is an approach that will actively support innovation while addressing risks and public concerns. The UK is home to thriving start-ups, which our framework will support to scale-up and compete internationally. Our pro-innovation approach will also act as a strong incentive when it comes to AI businesses based overseas establishing a presence in the UK. The white paper sets out our commitment to engaging internationally to support interoperability across different regulatory regimes. Not only will this ease the burden on business but it will also allow us to embed our values as global approaches to governing AI develop."
27. However, the European Union is already further ahead on this agenda than the UK. In the week in which the Prime Minister told London Tech Week that he wanted the UK to "not just be the intellectual but the geographical home of global AI safety regulation"¹², the European union signed off its negotiating mandate for an AI Act, with a series of specific and practical proposals to ban a number of potentially harmful and discriminatory uses.¹³
28. We also note the publication in the USA of the NIST risk management framework for AI. The framework is an elegant, simple document that applies common sense risk management to AI building on the long history of safe software development. Indeed, it seems to us that it is exactly the sort of document that the UK HSE could have produced had the government asked and funded it to do so.
29. We note that AI tools are built on data, some of which is likely to be personal data for the purposes of data protection. It is important that the emphasis on pro-innovation does not lead to data protection rights being undercut, especially against the context of data transfers under the GDPR. Products which cannot demonstrate compliance may be harder to market into a European market, on this basis.

10 <https://hansard.parliament.uk/Lords/2023-05-25/debates/003844E1-1E97-4A7B-AA27-DF1685DBA345/OnlineSafetyBill#contribution-DFBB7304-F11E-49D1-8F16-6A73D28F3C74>

11 <https://hansard.parliament.uk/Lords/2023-05-25/debates/003844E1-1E97-4A7B-AA27-DF1685DBA345/OnlineSafetyBill#>

12 <https://www.gov.uk/government/speeches/pm-london-tech-week-speech-12-june-2023#:~:text=AI%20can%20help%20us%20achieve.in%20the%20world%20for%20tech.>

13 <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>

Involving civil society

30. It has not gone unnoticed that civil society engagement has been absent in both the Vallance review and the announcement on the upcoming AI summit. Scant references to civil society involvement in the consultations that informed the White Paper are crowded out by self-congratulatory references as to how pleased industry is with the Government's "pro-innovation" proposals. We agree with the points made in this recent Washington Post ¹⁴ article on the UK's position within the global debate and its characterisation of the "chummy and insular" tendencies of the UK tech policymaking sphere – to only seek out and amplify the entrepreneurs and tech bros and to avoid the voices of civil society, academics and victims' groups. A few days later, a [call for expressions of interest](#) to participate in the £100m Foundation Model Taskforce solely invited those with technical skills to apply and appeared to be circulated in haste, using a google form, via the social media networks of the new Government's new AI adviser.
31. The AI Summit and all its related initiatives, however well-funded, will fail to deliver on the Prime Minister's objectives to deliver credible global leadership if they do not actively invite a wider cross-section of experts and representatives of civil society to participate.

Carnegie UK

June 2023

Contact: maeve.walsh@carnegieuk.org

14 https://www.washingtonpost.com/business/2023/06/16/let-the-british-government-write-ai-s-rules/92dd563a-0bfe-11ee-8132-a84600f3bbgb_story.html