# CARNEGIE UK RESPONSE TO HOUSE OF LORDS COMMUNICATIONS AND DIGITAL COMMITTEE INQUIRY INTO LARGE LANGUAGE MODELS

**September 2023**

## Summary


1. We welcome the Lords Committee's inquiry into this topic and apologise for this late but brief submission which we hope will still be of value to the Committee's considerations. At Carnegie UK, we have built up a reputation over the last five years as an influential voice on digital regulation: our work on the proposal for a statutory duty of care for online harm reduction influenced the UK Government's approach to designing an online safety regime and underpins much of the structure of the Online Safety Bill, which will shortly receive Royal Assent.
2. A regulatory approach founded on robust risk assessment, mitigation and reporting provides transparency to users and consumers while ensuring corporate accountability and compliance through oversight and enforcement by an independent, evidence-based regulator. This approach can be deployed whether principles for the regulation of AI generally are sought – so, for example, foundational models, - or where those models are deployed in specific contexts. The adaptability of this approach is demonstrated by the fact it is by no means a novel approach: large swathes of the industrial sector are regulated in this way. Technology firms and the services they provide are not unique or sufficiently different from other firms whose products can present the risk of harm to those who use them, or to society as a whole. Specific sectoral regulation can, where necessary, sit against this general background – but care must be taken to ensure coherence of sectoral regimes between themselves and with the underpinning general approach. Existing laws apply to AI, and LLMs in particular, both in their development and deployment, save where there are relevant exceptions in those laws.
3. Embedding "safety by design" into the development, testing and deployment of such products reduces this risk. An even older concept – the "polluter pays principle" - ensures that the societal costs of those products is transferred back to the producers and is as relevant to AI companies (with vast capitalisation) who should be made to bear the full costs of their tools to that the costs are returned to the production decision (eg testing for safety, forbearance in releasing code, honesty about data, ensuring that learning data sets are not biased etc). And the "precautionary principle" - deployed by the UK Government in response to previous crises when the fast-moving pace of novel scientific or technological developments risked outpacing regulators' ability to keep up – is also tried and tested. Crucially – given the current government's focus on realising the benefits of Artificial Intelligence without regulatory interference – it enables innovation to flourish within a system that still provides protection while the evidence base of risk or harm is amassed.
4. The Select Committee does not need tutorials from us on these approaches, nor are we the best source of expertise on current capabilities and future trends in LLMs. Instead,

we set out briefly below some thoughts, specifically in relation to question 3, on the way forward for regulation in this area with a view to shaping the Committee's investigations during its inquiry.

## Domestic regulation

**How adequately does the AI White Paper (alongside other Government policy) deal with large language models? Is a tailored regulatory approach needed?**

5. In relation to foundational models, including LLMs, the AI White Paper suggests that specific regulatory actions "would be premature" at this point (Part 3, section 3.3.3) due to the risk of stifling innovation. We note however that the White Paper also highlights the importance of ongoing monitoring functions, risk assessment and an adaptive regulatory approach that accounts for the evolving landscape of AI technology. It is also silent on existing general laws that apply – this feels like a worrying oversight, given that AI-driven products and services are already in widespread use; for example, see para 9 below re the application of health and safety law to workplace deployment of AI (here, those based on LLMs).

6. This approach is not entirely aligned with the approach of the Inter-Departmental Liaison Group on Risk Assessment (ILGRA) to the precautionary principle as a counterweight to the unrestrained development of technology, where evidence of harm may be evident, but not conclusive of causation. This piece of work became the canon for scientific regulators and the Health and Safety Executive, a considered and impactful response to the science and public confidence crises of the 2000s – BSE, GM Foods, nanotech, etc – which drove UK government scientists to develop a version of the precautionary principle that enabled rapid progress in risky science with real-time supervisory regulation to maintain public trust. It was a clever device that acknowledged emergent evidence of harms but allowed progress where scientific certainty was lacking within the time frame for decision-making. The government work on AI is strangely a-historical. If the Committee is so minded, it would be interesting to ask either the former Chief Scientific Adviser or his successor whether there was an assessment made of the risks of diverging from this approach in both the CSA's paper on a "pro-innovation approach" to technology regulation and the AI White Paper itself.

7. In our 2019 paper, we set out the merits of an ILGRA-derived precautionary approach which 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. It is noteworthy that reference to the precautionary principle was also made by The Lord Bishop of Oxford, Lord Freyberg and Lord Bassam in the Advanced AI debate (on 24 July 2023).

8. The government is actively engaged in addressing the existing challenges posed by technology, including the utilisation of algorithms and AI-driven systems, through a range of legislative measures currently progressing through Parliament (though some of these contexts may relate to the deployment of LLMs rather than their development). Notable among these are the Online Safety Bill, which has been confirmed to cover AI, the Data Protection and Digital Information Bill, the Digital Markets, Competition and Consumers Bill, as well as the Media Bill. Also, the government's Product Safety Review expressed particular concerns over products incorporating AI-type features (p. 42), some of which

could result in a product (and potentially its safety characteristics) changing over time. The government has also committed to bring forward a targeted package of measures in respect of online advertising, another large area where algorithmic-driven consumer harm is rising exponentially. These are all key foundations on which AI regulation can be built, but it is critical to review the intersection of all these pieces of legislation through an AI-lens and consider their collective impact as well as the requirements they place on regulators. Implementing an ILGRA-based precautionary strategy within a robust and well-structured legal framework would help effectively manage potential risks arising from the wider implementation and utilisation of AI, both presently and in the future, and would position the UK exceptionally well on the global stage.

9. There are also some fundamentals of regulation that currently safeguard individuals against harms stemming from AI-generated sources. But the AI White Paper does not seem to mention the cross-cutting protections already available for individuals and employees through laws like the Health and Safety at Work Act 1974. As we argued in our Response to AI Strategy White Paper,

> *"there is already an established regulatory baseline in the Health and Safety at Work Act 1974 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" (para. 21).*

10. 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).2 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.

Other areas of law may be relevant too – consider, for example, the relevance of consumer protection rules in relation to consumer-facing chatbots.

11. We also refer to our evidence to the recent Science and Technology Committee inquiry into AI governance, in which we stressed that the need for regulation arises when the costs of using AI in production decisions impact society at large, rather than just the company itself, affecting workers, customers, and third parties. Various existing regulatory

mechanisms should be explored before considering new models. A successful strategy involves risk-based, proportionate regulation or self-regulation centred on desired outcomes. This adaptable model, exemplified by the Health and Safety at Work Act 1974, remains fit for purpose across so many other sectors and can be effectively applied to AI, focusing on service design outcomes rather than specific technologies. This approach provides a flexible foundation across AI use and can be tailored to sector-specific needs, including LLMs. It can be adopted in relation to the development of those models as well as in relation to their deployment – and in our view both developers and those using those models in other applications should have responsibility for their choices in how to develop or how to use these models respectively. (For example, in the Bridges case in 2020, which involved the deployment of live automated facial recognition by the South Wales Police Force, the Courts said that the police should have checked that the models they were using for AI weren't biased. While this was in the context of the Equalities Act, it is in an example of a deployer of the technology having to take responsibility over its choice of tool.)

12. This approach allows for alignment between sector rules and general AI regulations. Our proposal for a statutory duty of care to reduce online harms, partially adopted by the Government in the Online Safety Bill, draws inspiration from the Health and Safety legislation's longstanding approach. We believe this approach can be applied to AI in various industries, possibly with additional refinements for specific risks and harms in certain sectors.

13. The UK government's approach in the AI White Paper, however, seems to lean more towards waiting for clear evidence of harm before implementing regulations. This delay in action reinforces the perception that AI is a novel technology developed and implemented independently from already tightly regulated products and services, especially in terms of safety considerations. This approach could potentially allow risks to materialise before effective safeguards are put in place and hinder the ability to respond quickly and effectively to emerging risks.

14. The emphasis on a pro-innovation stance overlooks the globally accepted risk management approach highlighted in the UNGPs and OECD guidance on Responsible Business Conduct. This shared international principle should have been a key basis for the Government's strategy. As we have previously argued, there is already an established regulatory baseline in the Health and Safety at Work Act 1974 that all technology-deploying companies must follow. Reinforcing this could have directed more industry attention to immediate risk management and harm reduction, allowing the government to then develop broader pro-innovation strategies to sit on top of this.

**Do the UK's regulators have sufficient expertise and resources to respond to large language models? If not, what should be done to address this?**

15. The AI White Paper is silent on the need for the Government – for it is only the Government who can do this – to join up the *existing* areas of legislative policy and regulatory development that they have set in train. This includes: the Online Safety Bill; the Data Protection and Digital Information Bill; the Digital Markets and Consumers Bill; the Media Bill. Moreover, it continues to rely on the Digital Regulation Cooperation Forum (DRCF) – a worthwhile and competent but ultimately informal grouping of regulatory leads – as the sole mechanism to ensure cooperation and collaboration, with no statutory requirement on Ofcom to cooperate with its fellow domestic regulators. Indeed, each of the regulators who are members of the DRCF have different powers and obligations in this regard. It is unclear how the DRCF members are expect to cooperate (where appropriate) with sector regulators outside the DRCF. The Committee might find it instructive to ask whether the Government has made an assessment, with regard to AI, of the gaps between these legislative vehicles and the regulators assigned to enforce them?

16. It is sensible that the Government have not created a new AI-specific, cross-sector regulator, and instead that the regulatory bodies with oversight of individual sectors should retain the lead in oversight of how these sectors are using AI. But we doubt that current expert regulators possess the necessary resources or powers to fulfil the government's requirement for cross-sectoral collaboration.

17. It is striking that Ofcom, in order to proportionately regulate social media services' severe negative impacts, has been authorised to recruit 300 staff (making up for nearly one-third of its current workforce), with Exchequer support until a fee structure is established, allowing Ofcom to recover costs from major companies based on a "polluter pays" model. The Vallance review draws attention to the struggle of existing regulators to recruit AI-proficient personnel. To facilitate their expanded AI-related roles, increased funding is needed for larger regulators, allowing competitive recruitment and fees from AI suppliers for risk assessment and mitigation work, following a similar "polluter pays" framework.

18. The Government's call for regulators to apply cross-cutting principles and conduct comprehensive risk analysis and enforcement within their domains raises the question of how they will be incentivised or compelled to share the evidence and learnings with other regulators dealing with similar issues/harms in different sectors. The Government has mentioned the DRCF as a potential solution. While we have seen in recent years good evidence of the participating regulators undertaking high-quality evidence-gathering and horizon-scanning in areas of cross-cutting interest (including algorithmic processing), this cooperative effort lacks the statutorily-underpinned power of information exchange and collaboration. But existing regulators do not have either the resources or powers to deliver the cross-sectoral cooperation and coordination required by the Government.

19. The Government's White Paper involves a narrow set of information-sharing provisions between regulators that will lead to evidence of harm being miss or left unaddressed. The Government should instead introduce statutorily-underpinned information-sharing and cooperation powers, as there is already precedent for in the Communications Act 2003. The Vallance review notes existing regulators find it hard to recruit AI-skilled staff. For existing regulators to take on new work on AI, they should introduce a one-off uplift in funding for larger regulators, enabling recruitment at competitive rates. AI suppliers could

then be charged for work on AI risk assessment and mitigation on a polluter pays basis.

**What are the greatest opportunities and risks over the next three years?**

20. Given the arguments we have set out above, we return now to an earlier question in the call for evidence in the context of the preparations for the AI summit in November this year. The summit will not deliver on the Prime Minister's objectives to deliver credible global leadership if it does not actively invite representatives of civil society to participate. Civil society engagement has been absent in both the Vallance review and the announcement on the upcoming AI summit. A call for expressions of interest to participation in the £100 million Foundation Model Taskforce was solely invited to those with technical skills and was circulated in haste, using a google form, via the social media networks of the new Government's new AI adviser. When asked by the Committee at its inaugural hearing (on 12 September 2023) why there was no ethicist or civil society representative on the Foundation Model Taskforce's Advisory Board, its Chair could only name a medical professional. The budget allocated to the taskforce has so far only recruited technical experts to work within government. No wider stakeholder engagement – reaching out into the vast expertise of civil society and social sciences – has been undertaken or is intended (insofar as we are aware). A Washington Post article recently characterised the UK tech policymaking sphere as "chummy and insular", only seeking out and amplifying the entrepreneurs and tech bros. The Government's lack of engagement with civil society on AI reflect this characterisation.

21. We would remind the Committee that the great British successes in innovation-enabling regulation in life sciences came from independent multidisciplinary bodies and reviews (like Warnock and Nuffield Council) not top-down government policy or narrow technical research projects.