HOUSE OF LORDS

Communications and Digital Committee

1st Report of Session 2023–24

Large language models and generative AI

Ordered to be printed 29 January 2024 and published 2 February 2024

Published by the Authority of the House of Lords

HL Paper 54

Communications and Digital Committee

The Communications and Digital Committee is appointed by the House of Lords in each session "to consider the media, digital and the creative industries and highlight areas of concern to Parliament and the public".

Membership

The Members of the Communications and Digital Committee are:

Baroness Featherstone
Lord Foster of Bath
Baroness Fraser of Craigmaddie
Lord Griffiths of Burry Port
Lord Hall of Birkenhead
Baroness Harding of Winscombe
Baroness Healy of Primrose Hill
Lord Kamall
The Lord Bishop of Leeds
Lord Lipsey
Baroness Stowell of Beeston (Chair)
Baroness Wheatcroft
Lord Young of Norwood Green

Declaration of interests

See Appendix 1.

A full list of Members' interests can be found in the Register of Lords' Interests:
http://www.parliament.uk/mps-lords-and-offices/standards-and-interests/register-of-lords-interests

Publications

All publications of the Committee are available at:
https://committees.parliament.uk/committee/170/communications-and-digital-committee/publications/

Parliament Live

Live coverage of debates and public sessions of the Committee's meetings are available at:
http://www.parliamentlive.tv

Further information

Further information about the House of Lords and its Committees, including guidance to witnesses, details of current inquiries and forthcoming meetings is available at:
http://www.parliament.uk/business/lords

Committee staff

The staff who worked on this inquiry were Daniel Schlappa (Clerk), David Stoker (Policy Analyst) and Rita Cohen (Committee Operations Officer).

Contact details

All correspondence should be addressed to the Communications and Digital Committee, Committee Office, House of Lords, London SW1A 0PW. Telephone 020 7219 2922.
Email: holcommunications@parliament.uk

X (formerly known as Twitter)

You can follow the Committee at: @LordsCommsCom.

CONTENTS

Evidence is published online at https://committees.parliament.uk/work/7827/large-language-models/publications/ and available for inspection at the Parliamentary Archives (020 7219 3074).

Q in footnotes refers to a question in oral evidence.

EXECUTIVE SUMMARY

The world faces an inflection point on AI. Large language models (LLMs) will introduce epoch-defining changes comparable to the invention of the internet. A multi-billion pound race is underway to dominate this market. The victors will wield unprecedented power to shape commercial practices and access to information across the world. Our inquiry examined trends over the next three years and identified priority actions to ensure this new technology benefits people, our economy and society.

We are optimistic about this new technology, which could bring huge economic rewards and drive ground-breaking scientific advances.

Capturing the benefits will require addressing risks. Many are formidable, including credible threats to public safety, societal values, open market competition and UK economic competitiveness.

Far-sighted, nuanced and speedy action is therefore needed to catalyse innovation responsibly and mitigate risks proportionately. We found room for improvement in the Government's priorities, policy coherence, and pace of delivery here.

We support the Government's overall approach and welcome its successes in positioning the UK among the world's AI leaders. This extensive effort should be congratulated. But the Government has recently pivoted too far towards a narrow focus on high-stakes AI safety. On its own this will not deliver the broader capabilities and commercial heft needed to shape international norms. The UK cannot hope to keep pace with international competitors without a greater focus on supporting commercial opportunities and academic excellence. A rebalance is therefore needed, involving a more positive vision for the opportunities and a more deliberate focus on near-term risks.

Concentrated market power and regulatory capture by vested interests also require urgent attention. The risk is real and growing. It is imperative for the Government and regulators to guard against these outcomes by prioritising open competition and transparency.

We have even deeper concerns about the Government's commitment to fair play around copyright. Some tech firms are using copyrighted material without permission, reaping vast financial rewards. The legalities of this are complex but the principles remain clear. The point of copyright is to reward creators for their efforts, prevent others from using works without permission, and incentivise innovation. The current legal framework is failing to ensure these outcomes occur and the Government has a duty to act. It cannot sit on its hands for the next decade and hope the courts will provide an answer.

There is a short window to steer the UK towards a positive outcome. We recommend the following:

- Prepare quickly: The UK must prepare for a period of protracted international competition and technological turbulence as it seeks to take advantage of the opportunities provided by LLMs.

- Guard against regulatory capture: There is a major race emerging between open and closed model developers. Each is seeking a beneficial regulatory framework. The Government must make

market competition an explicit AI policy objective. It must also introduce enhanced governance and transparency measures in the Department for Science, Innovation and Technology (DSIT) and the AI Safety Institute to guard against regulatory capture.

- Treat open and closed arguments with care: Open models offer greater access and competition, but raise concerns about the uncontrollable proliferation of dangerous capabilities. Closed models offer more control but also more risk of concentrated power. A nuanced approach is needed. The Government must review the security implications at pace while ensuring that any new rules support rather than stifle market competition.

- Rebalance strategy towards opportunity: The Government's focus has skewed too far towards a narrow view of AI safety. It must rebalance, or else it will fail to take advantage of the opportunities from LLMs, fall behind international competitors and become strategically dependent on overseas tech firms for a critical technology.

- Boost opportunities: We call for a suite of measures to boost computing power and infrastructure, skills, and support for academic spinouts. The Government should also explore the options for and feasibility of developing a sovereign LLM capability, built to the highest security and ethical standards.

- Support copyright: The Government should prioritise fairness and responsible innovation. It must resolve disputes definitively (including through updated legislation if needed); empower rightsholders to check if their data has been used without permission; and invest in large, high-quality training datasets to encourage tech firms to use licenced material.

- Address immediate risks: The most immediate security risks from LLMs arise from making existing malicious activities easier and cheaper. These pose credible threats to public safety and financial security. Faster mitigations are needed in cyber security, counter terror, child sexual abuse material and disinformation. Better assessments and guardrails are needed to tackle societal harms around discrimination, bias and data protection too.

- Review catastrophic risks: Catastrophic risks (above 1000 UK deaths and tens of billions in financial damages) are not likely within three years but cannot be ruled out, especially as next-generation capabilities come online. There are however no agreed warning indicators for catastrophic risk. There is no cause for panic, but this intelligence blind spot requires immediate attention. Mandatory safety tests for high-risk high-impact models are also needed: relying on voluntary commitments from a few firms would be naïve and leaves the Government unable to respond to the sudden emergence of dangerous capabilities. Wider concerns about existential risk (posing a global threat to human life) are exaggerated and must not distract policymakers from more immediate priorities.

- Empower regulators: The Government is relying on sector regulators to deliver the White Paper objectives but is being too slow to give them the tools. Speedier resourcing of Government-led central support teams is needed, alongside investigatory and sanctioning powers for some regulators, cross-sector guidelines, and a legal review of liability.

- Regulate proportionately: The UK should forge its own path on AI regulation, learning from but not copying the US, EU and China. In doing so the UK can maintain strategic flexibility and set an example to the world—though it needs to get the groundwork in first. The immediate priority is to develop accredited standards and common auditing methods at pace to ensure responsible innovation, support business adoption, and enable meaningful regulatory oversight.

Large language models and generative AI

CHAPTER 1: THE GOLDILOCKS PROBLEM

Our inquiry

1. The world is facing an inflection point in its approach to artificial intelligence (AI). Rapid advances in large language models (LLMs) have generated extensive discussion about the future of technology and society. Some believe the developments are over-hyped. Others worry we are building machines that will one day far outstrip our comprehension and, ultimately, control.

2. We launched this inquiry to examine likely trajectories for LLMs over the next three years and the actions required to ensure the UK can respond to opportunities and risks in time. We focused on LLMs as a comparatively contained case study of the issues associated with generative AI. We focused on what is different about this technology and sought to build on rather than recap the extensive literature on AI.[1]

3. We took evidence from 41 expert witnesses, reviewed over 900 pages of written evidence, held roundtables with small and medium sized businesses hosted by the software firm Intuit, and visited Google and UCL Business.[2] We were assisted by our specialist adviser Professor Michael Wooldridge, Professor of Computer Science at the University of Oxford. We are grateful to all who participated in our inquiry.

The challenge

4. Large language models are likely to introduce some epoch-defining changes. Capability leaps which eclipse today's state-of-the-art models are possible within the next three years. It is highly likely that openly available models with increasingly advanced capability will proliferate. In the right hands, LLMs may drive major boosts in productivity and deliver ground-breaking scientific insights. In the wrong hands they make malicious activities easier and may lay the groundwork for qualitatively new risks.[3]

5. The businesses that dominate the LLM market will have unprecedented powers to shape access to information and commercial practices across the world. At present US tech firms lead the field, though that may not hold true forever. The UK, alongside allies and partners, must carefully consider the implications of ceding commercial advantage to states which do not share our

1 See for example Artificial Intelligence Committee, *AI in the UK: ready, willing and able?* (Report of Session 2017–19, HL Paper 100), Science, Innovation and Technology Committee, *The governance of artificial intelligence: interim report* (Ninth Report, Session 2022–23, HC 1769), DSIT, 'Frontier AI' (25 October 2023): https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper [accessed 8 January 2024] and Department for Digital, Culture, Media and Sport, *National AI Strategy*, CP 525 (September 2021): https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf [accessed 25 January 2024].

2 See Appendix 4.

3 Q 3 (Dr Jean Innes and Ian Hogarth), written evidence from the Alan Turing Institute (LLM0081) and DSIT (LLM0079)

values.[4] We believe there are strong domestic and foreign policy arguments favouring an approach that supports (rather than stifles) responsible innovation to benefit consumers and preserve our societal values.[5]

6. The revolution in frontier AI will take place outside Government. But the work involved in building and releasing models will take place in specific geographies—not least because the developers will need access to energy, compute and consumers. National governments and regulators will therefore play a central role in shaping what kind of companies are allowed to flourish. The most successful will wield extensive power. Professor Neil Lawrence, DeepMind Professor of Machine Learning at the University of Cambridge, believed governments have a rare moment of "steerage" and the ramifications of decisions taken now will have impacts far into the future.[6]

7. Getting this steerage right will be difficult. It is common for technological developments to outpace policy responses (as well as raise ethical questions). But the latest advances in foundation models suggest this divide is becoming acute and will continue to widen.[7] This presents difficulties for governments seeking to harness this technology for good. Too much early intervention and they risk introducing laws akin to the 'Red Flag Act' of 1865, which required someone to walk in front of the new motorcars waving a red flag.[8] This did not age well. But too much caution around sensible rules is also harmful: seatbelts were invented in 1885 but drivers were not required to wear them until 1983.[9]

8. Solving this 'Goldilocks' problem of getting the balance right between innovation and risk, with limited foresight of market developments, will be one of the defining challenges for the current generation of policymakers. Our report proposes a series of recommendations to help the Government, regulators and industry navigate the challenges ahead.

4 Written evidence from Andreessen Horowitz (LLM0114)
5 Written evidence from Google and Google DeepMind (LLM0095), Meta (LLM0093), Microsoft (LLM0087), the Market Research Society (LLM0088), Oxford Internet Institute (LLM0074) and Andreessen Horowitz (LLM0114)
6 Q 3
7 Q 2 (Dr Jean Innes) and written evidence from the Open Data Institute (LLM0083)
8 The Open University, 'The Red Flag Act': https://law-school.open.ac.uk/blog/red-flag-act [accessed 20 December 2023]
9 Department for Transport and Stephen Hammond MP, 'Thirty years of seatbelt safety' (January 2013): https://www.gov.uk/government/news/thirty-years-of-seatbelt-safety [accessed 20 December 2023]

CHAPTER 2: FUTURE TRENDS

9. This chapter sets out capabilities and future trends in large language models (LLMs). The purpose is to summarise how they work, distinguish hype from reality, and provide the groundwork for our subsequent assessments of opportunity, risk and regulation. We do not attempt to provide exhaustive technical detail.

What is a large language model?

Box 1: Key terms

Artificial intelligence (AI): there is no universally accepted definition, though AI is commonly used to describe machines or systems performing tasks that would ordinarily require human brainpower. Smartphones, computers, and many online services use AI tools.

Deep learning: a method used in developing AI systems which involves processing data in ways inspired by how the human brain works.

Foundation model: a type of AI which typically uses deep learning and is trained on large datasets. It is characterised in part by its ability to adapt to a wide range of tasks. Many use a deep learning model, known as a transformer, developed by Google in 2017.

Generative AI: Closely related to foundation models, generative AI is a type of AI capable of creating a range of outputs including text, images or media.

Large language model: a subset of foundation models focused on language (written text). Examples of LLMs include OpenAI's GPT, Google's PaLM 2 and Meta's LLaMA.

Multi-modal model: a subset of foundation models which can handle more than one modality (for example images, video, code).

Frontier AI: a term used to describe the most powerful and cutting-edge general-purpose AI tools that match or exceed today's most advanced capabilities.

Compute: we use this term to refer to the hardware, software and infrastructure resources required for advanced AI processes.

Hallucination: a term describing LLMs producing inaccurate responses, many of which can sound plausible.

Model cards: a short document used in AI to provide information about how a model works, how it was developed and how it should be used.

Source: Written evidence from the Alan Turing Institute (LLM0081), Alan Turing Institute, 'Frequently asked questions': https://www.turing.ac.uk/about-us/frequently-asked-questions [accessed 17 January 2024], House of Lords Library, 'Artificial intelligence: Development, risks and regulation' (18 July 2023): https://lordslibrary.parliament.uk/artificial-intelligence-development-risks-and-regulation/ [accessed 17 December 2023] and Amazon Web Services, 'What is compute?': https://aws.amazon.com/what-is/compute/ [accessed 20 December 2023]

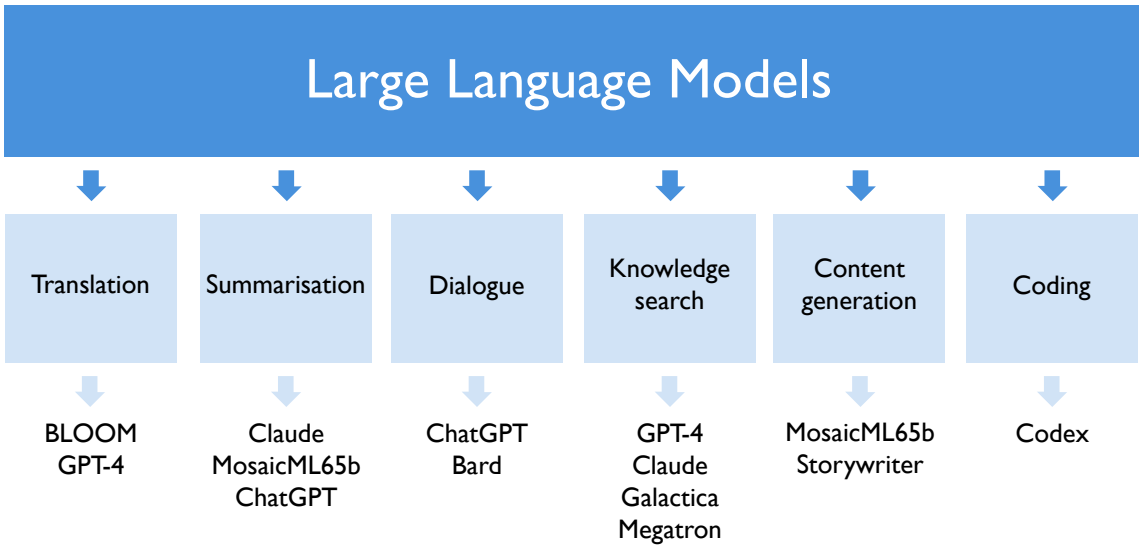
10. Large language models are a type of general purpose AI. They are designed to learn relationships between pieces of data and predict sequences. This makes them excellent at generating natural language text, amongst many other things.[10] LLMs are, at present, structurally designed around probability and plausibility, rather than around creating factually accurate assessments which correspond to the real world. This is partly responsible for the phenomenon

10 Written evidence from Dr P Angelov et al (LLM0032), Alan Turing Institute (LLM0081) and Google and Google DeepMind (LLM0095)

of 'hallucinations' whereby the model generates plausible but inaccurate or invented answers.[11]

11. LLMs can nevertheless perform a surprisingly wide range of economically useful tasks. They can already power chatbots, translation services and information retrieval systems; speed up office tasks by auto-generating documents, code and marketing materials; and catalyse research by synthesising vast amounts of data, and reviewing papers to identify patterns and insights.[12] OpenAI told us that LLMs will deliver "immense, tangible benefits to society".[13] Fundamentally new products remain nascent, though there is speculation that a highly capable autonomous personal assistant could emerge that can operate across a range of different services.[14]

Figure 1: Sample of LLM capabilities and example products



Source: Alan Turing Institute, 'Large Language Models and Intelligence Analysis' (2023): https://cetas.turing.ac.uk/publications/large-language-models-and-intelligence-analysis [accessed 14 December 2023]

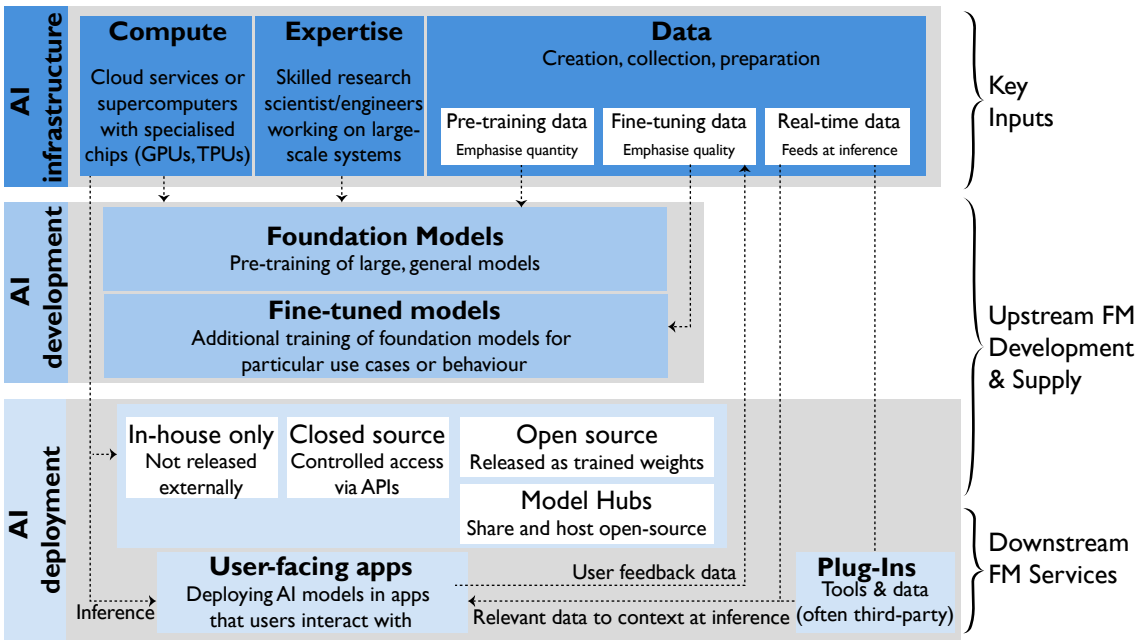
12. Developing an LLM is complex and costly. First, the underlying software must be designed and extensive data collected, often using automated bots to obtain text from websites (known as web crawling).[15] The model is pre-trained using parameters (known as model weights) which are adjusted to teach the model how to arrive at answers.[16]

13. Further fine-tuning may be undertaken to improve model performance and its ability to handle more specialised tasks.[17] The process for arriving at an answer is typically described as a 'black box' because it is not always possible

11 Q 97 (Jonas Andrulis)
12 Q 15 (Dr Zoë Webster), written evidence from the Market Research Society (LLM0088), MIT Technology Review, 'Large language models may speed drug discovery' (22 August 2023): https://www.technologyreview.com/2023/08/22/1076802/large-language-models-may-speed-drug-discovery/ [accessed 28 November 2023]
13 Written evidence from OpenAI (LLM0113)
14 Competition and Markets Authority, *AI Foundation Models Review* (2023): https://assets.publishing.service.gov.uk/media/65045590dec5be000dc35f77/Short_Report_PDFA.pdf [accessed 14 December 2023]
15 Web crawlers search and index content online for search engines.
16 Written evidence from Dr P Angelov et al (LLM0032) and Microsoft (LLM0087)
17 Q 75 (Rob Sherman) and written evidence from Dr P Angelov et al (LLM0032)

to trace exactly how a model uses a particular input to generate particular outputs, though efforts are underway to improve insight into their workings.[18]

Figure 2: Building, releasing and using a large language model

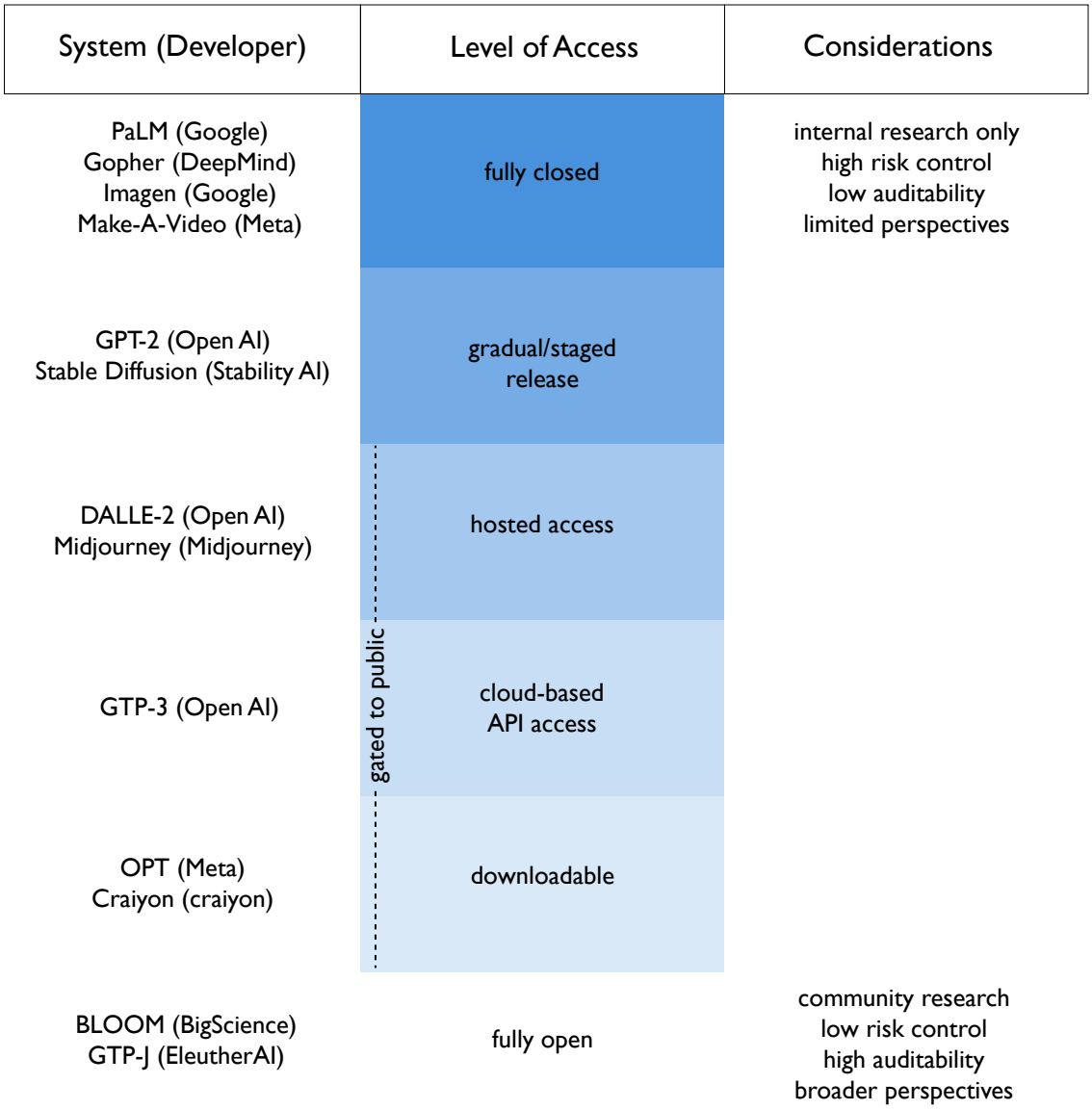


Source: Competition and Markets Authority, AI Foundation Models Review (2023): https://assets.publishing.service.gov.uk/media/65045590dec5be000dc35f77/Short_Report_PDFA.pdf [accessed 14 December 2023]

14. Models may be released in a variety of open or closed formats. Those on the open end of the spectrum tend to make more of the underlying system code, architecture and training data available.[19] The parameters may also be published, allowing others to fine-tune the model easily.[20] Those on the closed end of the spectrum tend to publish less information about how it has been developed and the data used.[21] The use of the term 'open source' model remains contested. We therefore use the term 'open access'.[22]

18 Written evidence from Sense about Science (LLM0046)
19 Written evidence from OpenUK (LLM0115)
20 Written evidence from Hugging Face (LLM0019)
21 Written evidence from Google and Google DeepMind (LLM0095), Microsoft (LLM0087) and OpenUK (LLM0115)
22 Our use of the term 'open access' is in line with definitions provided by the Oxford Internet Institute (LLM0074)

Figure 3: The scale of open and closed model release

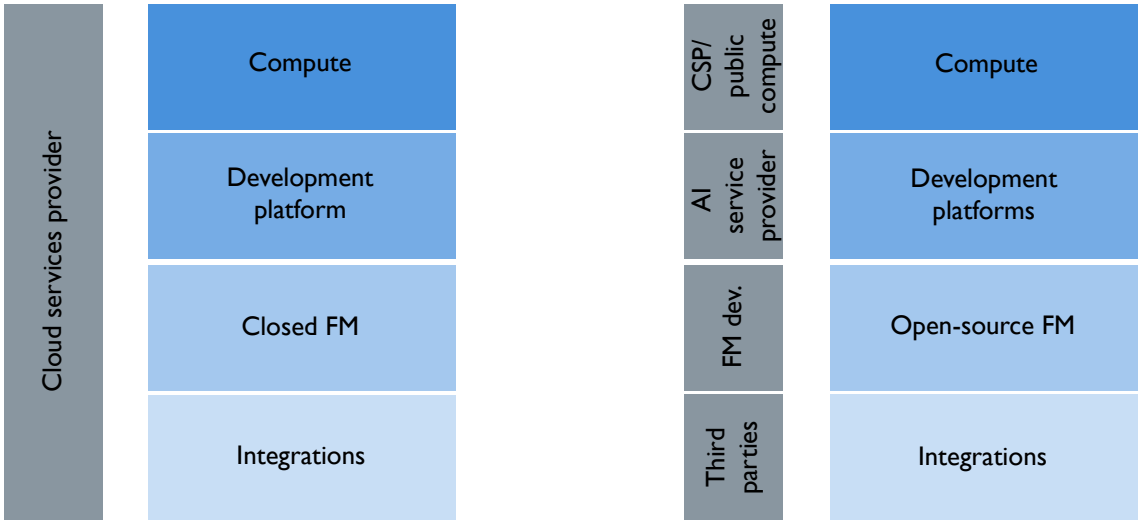


Source: Irene Solaiman, The Gradient of Generative AI Release (February 2023): https://arxiv.org/pdf/2302.04844.pdf [accessed 14 December 2023]

15. The building blocks and distribution channels for LLMs are likely to vary considerably. Some large tech firms might own the entire process from development to distribution. Others are likely to have different businesses working on each part of the model development and deployment.[23]

23 Competition and Markets Authority, 'AI Foundation Models: initial review' (2023): https://www.gov.uk/cma-cases/ai-foundation-models-initial-review [accessed 20 December 2023]

Figure 4: Level of vertical integration in model development and deployment



Source: Competition and Markets Authority, AI Foundation Models Review (2023): https://assets.publishing.service.gov.uk/media/65045590dec5be000dc35f77/Short_Report_PDFA.pdf [accessed 14 December 2023]

Trends

16. Models will get bigger and more capable. The amount of computing power used in training has expanded over the past decade by a factor of 55 million. Training data use has been growing at over 50 per cent per year.[24] Ian Hogarth, Chair of the (then) Frontier AI Taskforce, anticipated up to six orders of magnitude increase in the amount of compute used for next-generation models in the next decade, yielding "breath-taking capabilities".[25]

17. Costs will grow significantly. EPOCH, a research initiative, estimates the costs for developing state-of-the-art models could reach between $600 million and $3 billion over the next three years.[26]

18. Fine-tuned models will become increasingly capable and specialised. The Royal Academy of Engineering believed models trained on high quality curated datasets are likely to have "superior accuracy, consistency, usability and accountability" than general-purpose LLMs.[27]

19. Smaller models will offer attractive alternatives. These could deliver capable systems with much lower compute costs and data requirements. Some might even be run locally on a smartphone.[28]

20. Open access models will proliferate over the next three years. There is a clear trend towards ever greater numbers of open access models with increasingly

24 DSIT, *Capabilities and risks from frontier AI* (October 2023), p 11: https://assets.publishing.service.gov.uk/media/65395abae6c968000daa9b25/frontier-ai-capabilities-risks-report.pdf [accessed 17 December 2023]. Computing power is typically measured in floating-point operations per second (FLOPs).
25 Q 3
26 Written evidence from EPOCH (LLM002). Note that further infrastructure costs could be substantial.
27 Written evidence from the Royal Academy of Engineering (LLM0063)
28 Written evidence from the Royal Statistical Society (LLM0055), Royal Academy of Engineering (LLM0063) and TechTarget, 'Small language models emerge for domain-specific use cases' (August 2023): https://www.techtarget.com/searchbusinessanalytics/news/366546440/Small-language-models-emerge-for-domain-specific-use-cases [accessed 20 December 2023]

sophisticated capabilities, driven in part by the growing ease and falling costs of development and customisation.[29] They are unlikely to outclass cutting edge closed source models within the next three years if judged on a suite of benchmarks, but will offer attractive options for those who do not require cutting edge capabilities.[30] Consumer trust is likely to be a factor affecting uptake.

21. Integration with other systems will grow. Models are likely to gain more widespread access to the internet in real time, which may improve the accuracy and relevance of their outputs.[31] Better ways of linking LLMs both with other tools that augment their capacities (for example calculators), and with other real-world systems (for example email, web search, or internal business processes) are also expected.[32] The availability of existing infrastructure suggests this will occur faster than in previous waves of innovation.[33]

22. The timeline and engineering pathway to widespread integration of LLMs in high-stakes areas remains uncertain. LLMs continue to hallucinate, exhibit bias, regurgitate private data, struggle with multi-step tasks, and pose difficulties for interpreting black-box processes.[34] In light of these issues it is unclear how quickly LLMs should be integrated into high-stakes applications (for example in critical national infrastructure). Improvements to bias detection, memory, complex task execution, error correction and interpretability are major areas of research and some improvements within three years are highly likely.[35]

---

29 See for example written evidence from Market Research Society (LLM0088), Edward J. Hu et al, 'LoRA: 'Llow-Rank Adaptation of Large Language Models' (June 2021): https://arxiv.org/abs/2106.09685 [accessed 20 December 2023] and IEEE Spectrum, 'When AI's Large Language Models Shrink' (March 2023): https://spectrum.ieee.org/large-language-models-size [accessed 20 December 2023].

30 Written evidence from the Royal Academy of Engineering (LLM0063), Stability AI (LLM0078), TechTarget, 'Small language models emerge for domain-specific use cases' (August 2023): https://www.techtarget.com/searchbusinessanalytics/news/366546440/Small-language-models-emerge-for-domain-specific-use-cases [accessed 20 December 2023] and IEEE Spectrum, 'When AI's Large Language Models Shrink' (March 2023): https://spectrum.ieee.org/large-language-models-size) [accessed 20 December 2023]

31 See for example OpenAI, 'ChatGPT Plugins' (March 2023): https://openai.com/blog/chatgpt-plugins [accessed 28 November 2023] and TechCrunch, 'You.com launches new apis to connect LLMs to the web' (November 2023): https://techcrunch.com/2023/11/14/you-com-launches-new-apis-to-connect-llms-to-the-web/ [accessed 28 November 2023].

32 Q 98 (Jonas Andrulis), written evidence from the Royal Statistical Society (LLM0055), Dr P Angelov et al (LLM0032), Alan Turing Institute (LLM0081), Google and Google DeepMind (LLM0095) and DSIT, *Capabilities and risks from frontier AI* (October 2023): https://assets.publishing.service.gov.uk/media/65395abae6c968000daa9b25/frontier-ai-capabilities-risks-report.pdf [accessed 17 December 2023]

33 Written evidence from the Bright Initiative (LLM0033)

34 Written evidence from Oxford Internet Institute (LLM0074), Royal Statistical Society (LLM0055), Royal Academy of Engineering (LLM0063), Microsoft (LLM0087), Google and Google DeepMind (LLM0095), NCC Group (LLM0014)

35 Written evidence from the Alan Turing Institute (LLM0081), Google and Google DeepMind (LLM0095), Professor Ali Hessami et al (LLM0075). See also research interest in related areas, for example Jean Kaddour et al, 'Challenges and Applications of Large Language Models' (July 2023): https://arxiv.org/abs/2304.05332 [accessed 20 December 2023], Noah Shinn et al, 'Reflexion: Language Agents with Verbal Reinforcement Learning' (March 2023): https://arxiv.org/abs/2303.11366 [accessed 8 January 2024] and William Saunders et al, 'Self-critiquing models for assisting human evaluators' (June 2022): https://arxiv.org/abs/2206.05802 [accessed 8 January 2024].

23. There is a realistic possibility of integration with systems capable of kinetic movement. There is some evidence of progress already, though sci-fi scenarios of a robot apocalypse remain implausible.[36]

24. There is a realistic possibility of unexpected game-changing capability leaps in solving real-world problems. These remain difficult to forecast as there is not a predictable relationship between improvements to inputs and problem-solving capabilities.[37]

25. Some automation of model development may occur. This would involve using AI to build AI. Such progress might speed up some aspects of model development significantly, though at the cost of fewer humans involved in the process.[38]

26. High quality data will be increasingly sought after. EPOCH expects developers to exhaust publicly available high-quality data sources such as books, news, scientific articles and open source repositories within three years, and turn to lower quality sources or more innovative techniques.[39] Professor Zoubin Ghahramani, Vice President of Research at Google DeepMind, said there was ongoing research into using machine-generated synthetic data, but thought this could also lead to a degraded information environment,[40] or model malfunction.[41]

27. The level of market competition remains uncertain. A multi-billion pound race to dominate the market is underway. Many leading AI labs emerged outside big tech firms, though there has been subsequent evidence of trends towards consolidation.[42] It is plausible that a small number of the largest cutting-edge models will be used to power an extensive number of smaller models, mirroring the existing concentration of power in other areas of the digital economy.[43]

28. **Large language models (LLMs) will have impacts comparable to the invention of the internet. *The UK must prepare for a period of heightened technological turbulence as it seeks to take advantage of the opportunities.***

---

36 Jean Kaddour et al, 'Challenges and Applications of Large Language Models' (July 2023): https://arxiv.org/abs/2304.05332 [accessed 20 December 2023]

37 Government Office for Science, *Future risks of frontier AI* (October 2023): https://assets.publishing.service.gov.uk/media/653bc393d10f3500139a6ac5/future-risks-of-frontier-ai-annex-a.pdf [accessed 25 January 2024]. See also AI Alignment Forum, 'What a compute-centric framework says about AI takeoff speeds' (January 2023): https://www.alignmentforum.org/posts/Gc9FGtdXhK9sCSEYu/what-a-compute-centric-framework-says-about-ai-takeoff [accessed 20 December 2023] and Lukas Finnveden, 'PaLM-2 & GPT-4 in "Extrapolating GPT-N performance"' (May 2023): https://www.alignmentforum.org/posts/75o8oja43LXGAqbAR/palm-2-and-gpt-4-in-extrapolating-gpt-n-performance [accessed 8 January 2024].

38 Daniil A Boiko et al, 'Emergent autonomous scientific research capabilities of large language models' (2023): https://arxiv.org/ftp/arxiv/papers/2304/2304.05332.pdf [accessed 21 December 2023], Drexler, 'Reframing superintelligence' (2019): https://www.fhi.ox.ac.uk/reframing/ [accessed 21 December 2023] and Tom Davidson, 'Continuous doesn't mean slow' (April 2023): https://www.planned-obsolescence.org/continuous-doesnt-mean-slow/ [accessed 25 January 2024]

39 Written evidence from EPOCH (LLM002)

40 Q 99

41 Ilia Shumailov et al, 'The curse of recursion' (May 2023): https://arxiv.org/abs/2305.17493 [accessed 21 December 2023]

42 Open Markets Institute, 'AI in the public interest' (15 November 2023): https://www.openmarketsinstitute.org/publications/report-ai-in-the-public-interest-confronting-the-monopoly-threat [accessed 21 December 2023]

43 Competition and Markets Authority, *AI Foundation Models Review*

CHAPTER 3: OPEN OR CLOSED?

29. Competition dynamics will play a defining role in shaping who leads the market and what kind of regulatory oversight works best. At its heart, this involves a contest between those who operate 'closed' ecosystems, and those who make more of the underlying technology openly accessible. We examined whether the Government should adopt an explicit position favouring one or the other, and how it should navigate concerns about regulatory capture.

Open and closed models

30. The arguments were nuanced and shaped in part by stakeholders' particular interests. Closed models are associated with the most advanced capabilities developed in a small number of research laboratories such as OpenAI (backed by Microsoft), Anthropic (which has relationships with Amazon and Google), and Google DeepMind.[44] A range of smaller fine-tuned products may be built on top of a base model. But closed models offer fewer downstream opportunities for other businesses to examine and experiment with the underlying technology.[45]

31. Open access models tend to be cheaper and more accessible.[46] Dr Draief, Managing Director of Mozzilla.ai, argued that open models provided a "virtuous circle" by enabling more people to experiment with the technology.[47] Irene Solaiman, Global Policy Director of the open access platform Hugging Face, said openness provided better transparency and opportunities for community-led improvements.[48] Open models have however lagged behind the most advanced closed models on full-spectrum benchmarks[49] and have fewer options to recall and fix harmful products.[50]

32. Microsoft and Google said they were in general very supportive of open access technologies but believed the security risks arising from openly available powerful LLMs were so significant that more guardrails are needed.[51] OpenUK said there were many different types of 'open' technologies, in the same way that cars and lorries are different types of vehicle, and suggested nuanced regulatory proposals were essential.[52] Getty Images cautioned against "gaps in the fabric of regulations" that might exempt open models from obligations.[53]

33. Our evidence suggested a nuanced and iterative approach will be essential. Our review of risks in Chapter 5 suggests that the release and deployment of models without guardrails may pose credible risks. Equally, a market

44 Written evidence from OpenAI (LLM0113), Microsoft (LLM0087), Google and Google DeepMind (LLM0095) and Reuters, 'Amazon steps up AI race with Anthropic investment' (29 September 2023): https://www.reuters.com/markets/deals/amazon-steps-up-ai-race-with-up-4-billion-deal-invest-anthropic-2023–09-25/ [accessed 9 January 2024]
45 Written evidence from OpenUK (LLM0115) and the Bright Initiative (LLM0033)
46 Written evidence from OpenUK (LLM0115) and Hugging Face (LLM0019)
47 Q 66
48 Q 67
49 Some smaller open models compare favourably to the largest models when judged on a narrower range of capability assessments. But they tend to lag behind when judged against a wider 'full spectrum' range of benchmarks. See for example Stack Exchange, 'How do open source LLMs compare to GPT-4?' (July 2023): https://ai.stackexchange.com/questions/41214/how-do-open-source-llms-compare-to-gpt-4 [accessed 8 January 2024].
50 Q 10 (Ian Hogarth)
51 Google and Google DeepMind (LLM0095) and Microsoft (LLM0087)
52 Written evidence from OpenUK (LLM0115)
53 Written evidence from Getty Images (LLM0054)

dominated by closed models presents other risks around overreliance, single points of failure and concentrated market power.[54]

34. A recent report by the Competition and Markets Authority concluded that positive market outcomes would require "a range of models pushing at the frontier … on both an open-source and closed-source basis".[55]

35. We heard concerns however that the exploitation of first-mover advantage among large developers could lead to entrenched market power.[56] Ben Brooks of Stability AI said that limited action to ensure digital competition in the past had resulted in "one search engine, two social media platforms and three cloud computing providers". He believed there was a "serious risk of repeating these mistakes" and called on the Government to make "open innovation and competition in AI an explicit policy objective".[57]

36. Professor Neil Lawrence, DeepMind Professor of Machine Learning at the University of Cambridge, drew parallels with the early days of disruption around the internet:

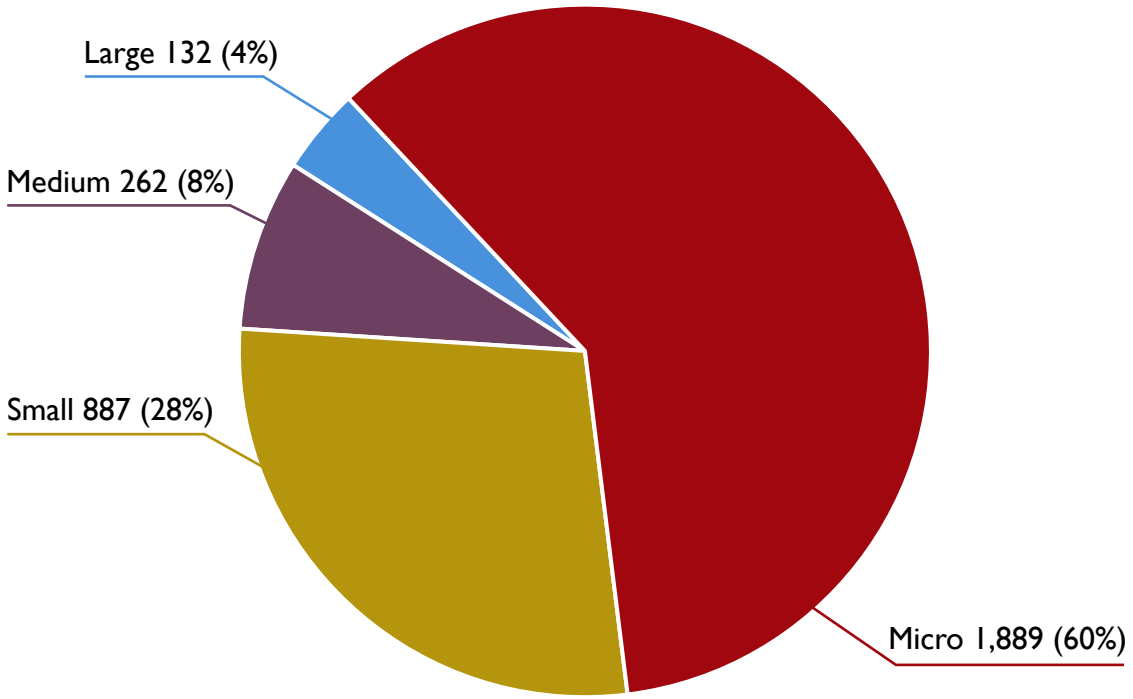
> "one of the most important aspects was the system of open-source software that enabled companies such as Google to compete with Microsoft. You can see that there is a lot of interest among the big tech companies in maintaining closed ecosystems, because they do not want to be disrupted."[58]

37. We heard that the structure of the UK's economy may lend itself to a strategy which helps smaller businesses experiment with open access technologies,[59] coupled with risk mitigations,[60] and incentives for a smaller number of firms (such as Google DeepMind) to operate at the cutting edge of research.[61]

38. The UK has around 3,170 AI companies, of which 60 per cent are dedicated AI firms and 40 per cent use AI in their products and services. These figures include US firms with bases in the UK. At present only a small proportion are likely to focus on building LLMs, though a larger number may in time focus on using them to improve products and services.

54 Tech Policy, 'Monopoly Power Is the Elephant in the Room in the AI Debate' (October 2023): https://www.techpolicy.press/monopoly-power-is-the-elephant-in-the-room-in-the-ai-debate/ [accessed 8 January 2024] and written evidence from Andreessen Horowitz (LLM0114)
55 Competition and Markets Authority, *AI Foundation Models Review*
56 Q 8 (Ben Brooks)
57 *Ibid.*
58 Q 3
59 Q 66 (Dr Moez Draief)
60 Written evidence from Martin Hosken (LLM0009)
61 Q 66 (Dr Draief) and Q 111 (Jonas Andrulis)

Figure 5: The structure of UK's AI ecosystem



Source: DSIT, Artificial Intelligence Sector Study (March 2023): https://assets.publishing.service.gov.uk/media/641d71e732a8e0000cfa9389/artifical_intelligence_sector_study.pdf [accessed 28 November 2023]

39. Viscount Camrose, Minister for AI and Intellectual Property, acknowledged the "innovation arguments both ways". He said that if open access models could be frozen in their current state "we would be very much in favour of it", but worried about the security risks of next-generation tools empowering malicious actors.[62]

40. **Fair market competition is key to ensuring UK businesses are not squeezed out of the race to shape the fast-growing LLM industry. The UK has particular strengths in mid-tier businesses and will benefit most from a combination of open and closed source technologies.**

41. ***The Government should make market competition an explicit policy objective. This does not mean backing open models at the expense of closed, or vice versa. But it does mean ensuring regulatory interventions do not stifle low-risk open access model providers.***

42. ***The Government should work with the Competition and Markets Authority to keep the state of competition in foundation models under close review.***

62 Q 141

Regulatory capture

43. Throughout our inquiry we encountered mounting concern about regulatory capture.[63] This might occur through lobbying or because officials lack technical know-how and come to rely on a narrow pool of private sector expertise to inform policy and standards. Similar problems may emerge from groupthink.[64] These might lead to regulatory frameworks which favour a select group of commercial rather than public interests, for example by creating barriers to new competitors entering the market.[65]

44. Current trends suggest growing private sector influence. Witnesses emphasised the limited extent of public sector expertise and the necessity of closer industry links, including staff exchanges.[66] Big tech firms are reportedly funding the salaries of US Congress staff working on AI policy.[67] Forums representing the positions of open and closed market leaders are proliferating, including the Frontier Model Forum (led by Google, Microsoft, OpenAI and Anthropic); and the "open science" AI Alliance (backed by Meta and IBM).[68]

45. There has been further concern that the AI safety debate is being dominated by views narrowly focused on catastrophic risk, often coming from those who developed such models in the first place.[69] Critics say this distracts from more immediate issues like copyright infringement, bias and reliability.[70]

46. Andreessen Horowitz, a venture capital firm, cautioned that large AI businesses must "not [be] allowed to establish a government-protected cartel that is insulated from market competition due to speculative claims of AI risk".[71] Professor Neil Lawrence also warned of "a very serious danger of regulatory capture".[72] The Open Markets Institute similarly raised concerns that incumbents may "convert their economic heft into regulatory influence" and distract policymakers "with far-off, improbable risks".[73]

---

63 Q 3 (Professor Neil Lawrence), written evidence from Nquiring Minds (LLM0073), OpenUK (LLM0115), Andreessen Horowitz (LLM0114). See also Bloomberg, 'Google DeepMind chief calls Meta's AI criticisms preposterous (1 November 2023): https://www.bloomberg.com/news/articles/2023–11-01/google-deepmind-chief-calls-meta-s-ai-criticisms-preposterous [accessed 20 December 2023].

64 This may involve a dominant intellectual viewpoint emerging which is not exposed to systematic challenge. See for example Public Administration Committee, *Lessons still to be learned from the Chilcot inquiry: Government Response* (Tenth Report, Session 2016–17, HC 656).

65 'Setting rules for AI must avoid regulatory capture by Big Tech', *Financial Times* (27 October 2023): https://www.ft.com/content/6a1f796b-1602-4b07-88cd-4aa408cf069a [accessed 20 December 2023]

66 Q 5 (Ian Hogarth) and Q 119 (Professor Dame Angela McLean)

67 Politico, 'Key Congress staffers in AI debate are funded by tech giants like Google and Microsoft' (12 March 2023): https://www.politico.com/news/2023/12/03/congress-ai-fellows-tech-companies-00129701 [accessed 20 December 2023]

68 Frontier Model Forum, 'Frontier Model Forum: Advancing Safe AI Development': https://www.frontiermodelforum.org/ [accessed 20 December 2023] and The AI Alliance, 'Members': https://thealliance.ai/members [accessed 20 December 2023]

69 MIT Technology Review, 'It's time to talk about the real AI risks' (12 July 2023): https://www.technologyreview.com/2023/06/12/1074449/real-ai-risks/ [accessed 20 December 2023]

70 Politico, 'How Silicon Valley doomers are shaping Rishi Sunak's AI plans' (14 September 2023): https://www.politico.eu/article/rishi-sunak-artificial-intelligence-pivot-safety-summit-united-kingdom-silicon-valley-effective-altruism/ [accessed 20 December 2023]

71 Written evidence from Andreessen Horowitz (LLM0114)

72 Q 3

73 Max von Thun, 'Monopoly powers is the elephant in the room in the AI debate' (23 October 2023): https://www.techpolicy.press/monopoly-power-is-the-elephant-in-the-room-in-the-ai-debate/ [accessed 20 December 2023]

47. We heard a concerted effort is needed to guard against such outcomes.[74] Some parts of Government use a variety of techniques including red teaming to ensure decisions are subject to systematic challenge and review.[75] We asked the Minister what steps the Department was taking. He did not suggest there would be enhanced governance measures, though he did emphasise "remov[ing] barriers to innovation for smaller companies".[76]

48. **The risk of regulatory capture is real and growing. External AI expertise is becoming increasingly important to regulators and Government, and industry links should be encouraged. But this must be accompanied by stronger governance safeguards.**

49. ***We recommend enhanced governance measures in DSIT and regulators to mitigate the risks of inadvertent regulatory capture and groupthink. This should apply to internal policy work, industry engagements and decisions to commission external advice. Options include metrics to evaluate the impact of new policies and standards on competition; embedding red teaming, systematic challenge and external critique in policy processes; more training for officials to improve technical know-how; and ensuring proposals for technical standards or benchmarks are published for consultation.***

Conflicts of interest

50. External AI expertise will become increasingly important to the Government and regulators. We heard that deeper engagement with academia and industry will help policymakers navigate the complexities of AI, and should be encouraged.[77]

51. But doing so will also bring challenges: many experts appointed from the private sector to lead major Government initiatives will inevitably have significant financial conflicts of interest requiring appropriate mitigations. As outlined earlier, concerns are growing about the potential for corporate influence over policy choices in such a critical sector.[78] Transparency around the process for managing conflicts of interest will therefore become increasingly important to uphold public confidence in the integrity of the Government's work on AI, and to protect the individuals who enter public roles from the private sector.

52. The position of the Chair of the Frontier AI Taskforce (and now AI Safety Institute) is illustrative. We noted the Chair and his investment platform had previously made extensive financial investments in businesses directly

74 Written evidence from Connected by Data (LLM0066), OpenUK (LLM0115). See also Public Administration Committee, *Lessons still to be learned from the Chilcot inquiry: Government Response* (Tenth Report, Session 2016–17, HC 656).

75 Red teaming involves a structured process for challenging ideas from an adversarial perspective. See for example Cabinet Office, 'Skills: Wargaming and Red Teaming—How the MoD is challenging defence thinking' (6 November 2023): https://moderncivilservice.blog.gov.uk/2023/11/06/skills-wargaming-and-red-teaming-how-the-mod-is-challenging-defence-thinking/ [accessed 20 December 2023].

76 Q 141

77 Q 5 (Ian Hogarth) and Q 119 (Professor Dame Angela McLean)

78 'Setting rules for AI must avoid regulatory capture by Big Tech', *Financial Times* (27 October 2023): https://www.ft.com/content/6a1f796b-1602-4b07-88cd-4aa408cf069a [accessed 20 December 2023], Max von Thun, 'Monopoly powers is the elephant in the room in the AI debate' (23 October 2023): https://www.techpolicy.press/monopoly-power-is-the-elephant-in-the-room-in-the-ai-debate/ [accessed 20 December 2023], written evidence from Nquiring Minds (LLM0073), OpenUK (LLM0115) and Andreessen Horowitz (LLM0114)

associated with the policy area he would be leading.[79] The Department confirmed there were various "mitigations to manage potential conflicts of interest … with effect from the start of his role".[80]

53. We discussed the matter with the Chair in public and with the Permanent Secretary in private.[81] We commended the Chair's commitment to public service and acknowledged the financial loss and heightened scrutiny that this entails. We were reassured by the seriousness with which Government treats these issues.

54. There is no suggestion of a link between his investments, his appointment in June 2023 and subsequent changes to Government policy set out in Chapter 4.

55. Nonetheless, it was clear to us that more transparency is needed for high-profile positions in AI. There was not a deadline for confirming publicly that the mitigations have been completed, for example.[82] Nor was there sufficient public information on the types of mitigations being implemented.[83] We acknowledge the need to balance privacy and transparency. But providing more transparency upfront would do much to address questions about the integrity of the Government's work on AI and ensure those entering public life are empowered to address questions about financial conflicts directly and with confidence.[84]

56. **The perception of conflicts of interest risks undermining confidence in the integrity of Government work on AI. Addressing this will become increasingly important as the Government brings more private sector expertise into policymaking. Some conflicts of interest are inevitable and we commend private sector leaders engaging in public service, which often involves incurring financial loss. But**

79 See for example Ian Hogarth, 'About': https://www.ianhogarth.com/about [accessed 20 December 2023] and Crunchbase, 'Conjecture': https://www.crunchbase.com/organization/conjecture [accessed 20 December 2023].

80 DSIT confirmed in a press statement that Mr Hogarth had agreed to a series of mitigations including "divestments of personal holdings in companies building foundation models or foundation model safety tools. Mitigations are being put in place to address each of the potential conflicts with effect from the start of his role". See DSIT, 'Tech entrepreneur Ian Hogarth to lead UK's AI Foundation Model Taskforce' (18 June 2023): https://www.gov.uk/government/news/tech-entrepreneur-ian-hogarth-to-lead-uks-ai-foundation-model-taskforce [accessed 19 January 2024].

81 Q 7. See also public correspondence on this issue letter from Baroness Stowell of Beeston, Chair of the Communications and Digital Committee to Sarah Munby, Permanent Secretary at DSIT (22 September 2023): https://committees.parliament.uk/publications/41564/documents/204778/default/, letter from Sarah Munby, Permanent Secretary to Baroness Stowell of Beeston, Chair of the Communications and Digital Committee (19 October 2023): https://committees.parliament.uk/publications/41895/documents/207713/default/ and letter from Baroness Stowell of Beeston, Chair of the Communications and Digital Committee to Sarah Munby, Permanent Secretary (30 November 2023): https://committees.parliament.uk/publications/42388/documents/210602/default/.

82 DSIT, 'Tech entrepreneur Ian Hogarth to lead UK's AI Foundation Model Taskforce' (18 June 2023): https://www.gov.uk/government/news/tech-entrepreneur-ian-hogarth-to-lead-uks-ai-foundation-model-taskforce [accessed 19 January 2024]

83 We note there has been substantial public interest in the work of the Chair and his position on AI policy. See for example '"This is his climate change": The experts helping Rishi Sunak seal his legacy', *The Telegraph* (23 September 2023): https://www.telegraph.co.uk/business/2023/09/23/artificial-intelligence-safety-summit-sunak-ai-experts/ [accessed 17 January 2024] and Politico, 'How Silicon Valley doomers are shaping Rishi Sunak's AI plans' (14 September 2023): https://www.politico.eu/article/rishi-sunak-artificial-intelligence-pivot-safety-summit-united-kingdom-silicon-valley-effective-altruism/ [accessed 17 January 2024].

84 Letter from Baroness Stowell of Beeston, Chair of the Communications and Digital Committee to Sarah Munby, Permanent Secretary (30 November 2023): https://committees.parliament.uk/publications/42388/documents/210602/default/

their appointment to powerful Government positions must be done in ways that uphold public confidence.

57. ***We recommend the Government should implement greater transparency measures for high-profile roles in AI. This should include further high-level information about the types of mitigations being arranged, and a public statement within six months of appointment to confirm these mitigations have been completed.***

CHAPTER 4: A PRO-INNOVATION STRATEGY?

58. This chapter sets out the potential opportunities created by large language models (LLMs), followed by an assessment of how well the Government's strategy is positioning the UK to take advantage.

Benefiting organisations

59. LLM-powered services offer significant potential across a range of sectors. Examples include

- IT (code writing);[85]
- advertising (tailoring customer engagement);[86]
- product design (producing better ideas through LLM-supported brainstorming);[87]
- education (teaching aids calibrated to the learner's progress and abilities);[88]
- healthcare (analysing patient records and helping diagnostics);[89]
- legal (research and case work);[90] and
- finance (analysing financial and news data, and supporting clients), and much more.[91]

60. Goldman Sachs has estimated wider generative AI could add trillions of dollars to the global economy over the next decade.[92] The Advertising Association was "optimistic" about the UK's ability to take advantage of the opportunities.[93]

Benefitting society

61. Rachel Coldicutt OBE, Executive Director of Careful Industries, argued that LLMs "can and should contribute to a more equitable prosperous society for everyone", but emphasised this could only be achieved if more effort is made to ensure innovation is deliberately "calibrated to produce societal

85 Written evidence from the Oxford Internet Institute (LLM0074)
86 Written evidence from Advertising Association (LLM0056)
87 The Economist, 'Generative AI generates tricky choices for managers' (27 November): https://www.economist.com/business/2023/11/27/generative-ai-generates-tricky-choices-for-managers [accessed 21 December 2023]
88 Written evidence from Connected by Data (LLM0066)
89 Q 100
90 Solicitors Regulation Authority, 'SRA response to questions on large language models (October 2023), Legal Tech Hub, 'The use of large language models in legal tech' (18 February 2023): https://www.legaltechnologyhub.com/contents/the-use-of-large-language-models-in-legaltech/ [accessed 29 November 2023]
91 Letter from Bank of England and Prudential Regulation Authority to Baroness Stowell of Beeston, Chair of the Communications and Digital Committee (5 October 2023): https://committees.parliament.uk/publications/42157/documents/209538/default/. See also Bloomberg, 'Introducing BloombergGPT' (30 March 2023): https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/ [accessed 21 December 2023].
92 Goldman Sachs, 'Generative AI could raise global GDP by 7 per cent' (April 2023): https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html [accessed 8 January 2024]
93 Written evidence from the Advertising Association (LLM0056)

benefits".[94] This might involve a greater focus on ethical development, minimising environmental impacts, and developing socially valuable uses.[95] Rob Sherman, Vice President and Deputy Chief Privacy Officer for Policy at Meta, thought LLMs could be a major "force for inclusion", and gave the example of LLMs supporting computer vision systems for people with visual impairments.[96] Owen Larter, Director of Public Policy at Microsoft's Office for Responsible AI, said LLMs would be used to "address major societal challenges" and democratise access to technology.[97]

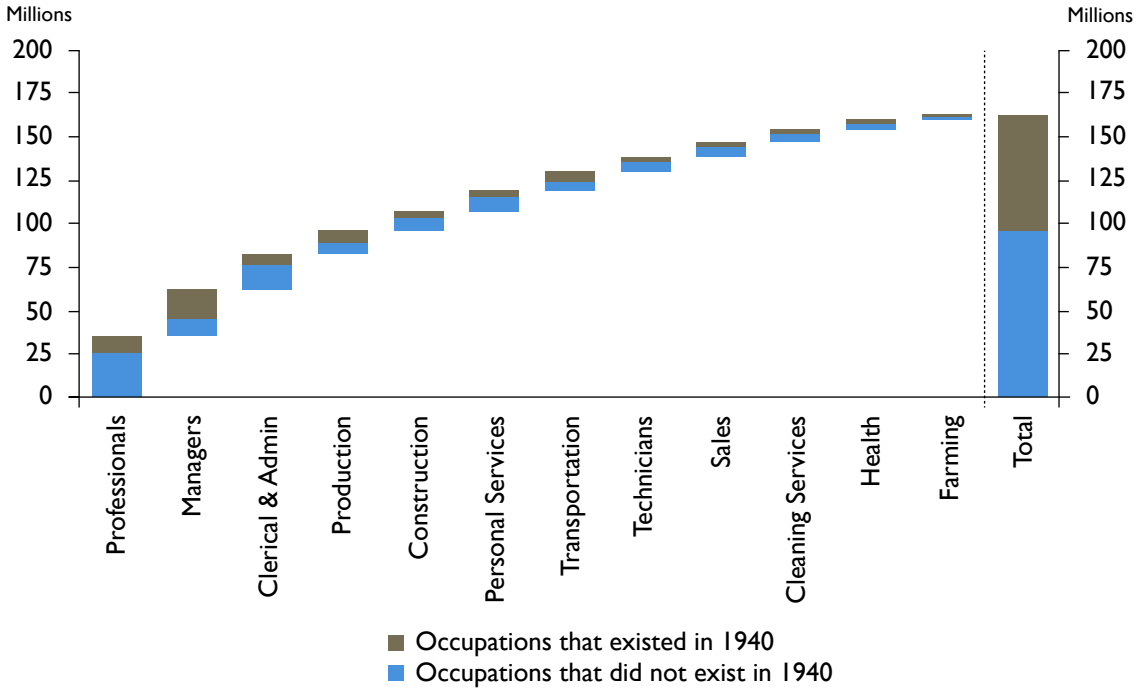
Benefitting workers

62. Labour market impacts remain uncertain. Some studies suggest jobs involving physical or interpersonal work are unlikely to experience much disruption. Others such as IT, administration and legal work could face substantial changes.[98] Some types of business model are also likely to come under pressure. Submissions from DMG Media, the Financial Times and the Guardian Media Group noted that print journalism may be significantly affected, particularly if advertising or subscription revenues drop as people turn to LLM tools for information rather than clicking through to news websites.[99] (See Chapter 8 for a discussion on copyright and implications for news media).

63. Other studies indicate previous waves of disruption have seen new jobs broadly offsetting losses.[100] Much of our evidence suggested initial disruption would give way to enhanced productivity (and see also Figure 6 below on the impact of technology on job creation). We did not find plausible evidence of imminent widespread AI-induced unemployment.[101]

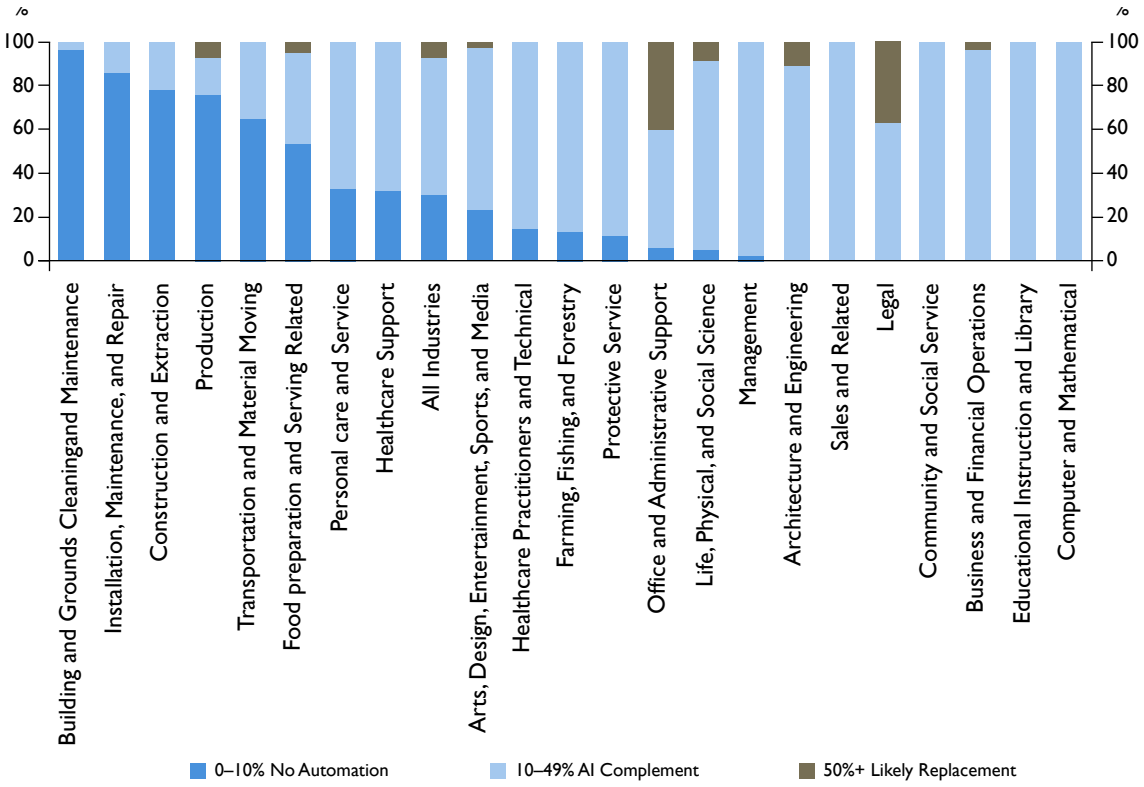
94 Written evidence from Rachel Coldicutt (LLM0041)
95 Written evidence from Martin Hosken (LLM009) and UCL Institute of Health Informatics (LLM0076)
96 Q 74
97 *Ibid.*
98 Goldman Sachs, *The potentially large effects of artificial intelligence on economic growth* (March 2023): https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html [accessed 30 November 2023]
99 Written evidence from DMG Media (LLM0068), Financial Times (LLM0034) and Guardian Media Group (LLM0108)
100 American Economic Association, 'Automation and new tasks: how technology displaces and reinstates labor' (2019): https://www.aeaweb.org/articles?id=10.1257/jep.33.2.3 [accessed 30 November 2023] and Goldman Sachs, *The potentially large effects of artificial intelligence on economic growth*
101 Written evidence from the Market Research Society (LLM0088), Creators Rights Alliance (LLM0039), Surrey Institute for People-Centred Artificial Intelligence (LLM0060) and Goldman Sachs, *The potentially large effects of artificial intelligence on economic growth*

Figure 6: The impact of technology on job creation



Goldman Sachs, The potentially large effects of artificial intelligence on economic growth

Figure 7: Labour exposure to automation by field



Source: Goldman Sachs, The potentially large effects of artificial intelligence on economic growth

64. As we highlighted in our reports on the creative industries and digital exclusion, it matters who is disrupted and how they are supported. Automating tasks commonly found in some roles risks reducing access routes for people to get a foot on the employment ladder, which in turn increases advantages for those with existing connections and finances to obtain experience.[102] Furthermore, the ongoing failure to address digital skills gaps perpetuates bottlenecks at the lower end of the supply chain and risks deepening societal divides between those able to take advantage of opportunities created by technological advances and those who are left behind.[103] The limited incentives for industry-led skills schemes suggests this challenge will require a concerted effort to address.[104]

65. **Large language models have significant potential to benefit the economy and society if they are developed and deployed responsibly. The UK must not lose out on these opportunities.**

66. **Some labour market disruption looks likely. Imminent and widespread cross-sector unemployment is not plausible, but there will inevitably be those who lose out. The pace of change also underscores the need for a credible strategy to address digital exclusion and help all sectors of society benefit from technological change.**

67. ***We reiterate the findings from our reports on the creative industries and digital exclusion: those most exposed to disruption from AI must be better supported to transition. The Department for Education and DSIT should work with industry to expand programmes to upskill and re-skill workers, and improve public awareness of the opportunities and implications of AI for employment.***

Government strategy and evolving priorities

68. The Government's approach to AI has evolved in recent years, shaped by the work of the AI Council (set up in 2019) and National AI Strategy (published in 2021).[105]

69. In March 2023 the Government published its "pro-innovation approach to AI regulation". This White Paper envisioned an "agile and iterative approach" structured around five principles:

- safety, security and robustness;
- appropriate transparency and explainability;
- fairness;

102 Communications and Digital Committee, *At risk: our creative future* (2nd Report, Session 2022–23, HL 125), para 53

103 Communications and Digital Committee, *Digital Exclusion* (3rd Report, Session 2022–23, HL Paper 219) and Government Response *to the Committee's report 'At risk: our creative future'*: https://committees.parliament.uk/publications/39303/documents/192860/default/. Letter from Baroness Stowell of Beeston, Chair of the Communications and Digital Committee to Lucy Frazer MP, Secretary of State (June 2023): https://committees.parliament.uk/publications/40617/documents/198054/default/ and written evidence from BT Group (LLM0090)

104 Q 114 and Communications and Digital Committee, *Digital Exclusion*

105 DSIT, 'AI Council' (7 July 2023): https://www.gov.uk/government/news/ai-council [accessed 18 January 2024] and Department for Digital, Culture, Media and Sport, *National AI Strategy*, CP 525 (September 2021): https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf [accessed 25 January 2024]

- accountability and governance; and
- contestability and redress.[106]

70. Existing regulators are expected to take account of these (non-statutory) principles when overseeing AI in their respective sectors. The Government committed to a range of actions including a set of "central functions" staffed by officials to provide co-ordination and support.[107] Some issues such as copyright, compute, and skills were not in the White Paper's scope.[108]

71. The framework was broadly welcomed by business communities for offering a flexible and pro-innovation framework,[109] though critiqued by others for expecting too much of regulators, and deferring decisions on regulation.[110]

72. The Government also set up a taskforce to address the most recent advances in AI, following the launch of ChatGPT. The timeline below suggests the strategic focus evolved from balancing innovation with risk towards a primary focus on AI safety throughout 2023:

- 29 March: the Government announces a "new expert taskforce to build the UK's capabilities in foundation models".[111]
- 24 April: the Government announces £100 million for the "Foundation Model Taskforce" which will be "responsible for accelerating the UK's capability in rapidly-emerging type[s] of artificial intelligence", "ensure sovereign capabilities" and encourage adoption of safe models.[112]
- 7 June: the Prime Minister announces the UK will host a global summit on AI safety, and will work with allies to make AI "safe and secure".[113]
- 18 June: DSIT announces the tech entrepreneur Ian Hogarth will lead the Foundation Model Taskforce.[114]
- 7 July: DSIT announces the AI Council has been disbanded.[115] It had been established in 2019. Its role included providing expert advice, and

106 DSIT, 'A pro-innovation approach to AI regulation' (August 2023): https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [accessed 8 January 2024]
107 DSIT, *A pro-innovation approach to AI regulation*. The Government anticipated introducing a statutory duty on regulators requiring them to have due regard to the principles in future.
108 DSIT, *A pro-innovation approach to AI regulation*
109 Written evidence from the Startup Coalition (LLM0089)
110 Written evidence from the National Union of Journalists (LLM0007), Glenlead Centre (LLM0051) and Surrey Institute for People-Centred Artificial Intelligence (LLM0060)
111 DSIT, 'UK unveils world leading approach to innovation in first artificial intelligence white paper to turbocharge growth' (29 March 2023): https://www.gov.uk/government/news/uk-unveils-world-leading-approach-to-innovation-in-first-artificial-intelligence-white-paper-to-turbocharge-growth [accessed 8 January 2024]
112 DSIT, 'Initial £100 million for expert taskforce to help UK build and adopt next generation of safe AI' (24 April 2023): https://www.gov.uk/government/news/initial-100-million-for-expert-taskforce-to-help-uk-build-and-adopt-next-generation-of-safe-ai [accessed 5 December 2023]
113 Prime Minister's Office, 'UK to host first global summit on Artificial Intelligence' (7 June 2023): https://www.gov.uk/government/news/uk-to-host-first-global-summit-on-artificial-intelligence [accessed 8 January 2024]
114 DSIT, 'Tech entrepreneur Ian Hogarth to lead UK's AI Foundation Model Taskforce' (18 June 2023): https://www.gov.uk/government/news/tech-entrepreneur-ian-hogarth-to-lead-uks-ai-foundation-model-taskforce [accessed 8 January 2024]
115 DSIT, 'AI Council' (7 July 2023): https://www.gov.uk/government/news/ai-council [accessed 8 January 2024]

supporting "the growth of AI in the UK [and promoting] its adoption and use in businesses and society".[116]

- 7 September: the Foundation Model Taskforce is renamed as the Frontier AI Taskforce, "explicitly acknowledging its role in evaluating risk at the frontier of AI".[117] Its progress update cites a new "expert advisory board spanning AI Research and National Security".[118]

- 9 September: the Centre for Data Ethics and Innovation (CDEI) advisory board is disbanded.[119] It had a remit for "identifying the measures we need to take to maximise the benefits of data and Artificial Intelligence for our society and economy".[120]

- 1–2 November: the Government holds the AI Safety Summit and confirms the Frontier AI Taskforce will become the new AI Safety Institute. The erstwhile "core parts of the Taskforce's mission" including boosting public sector AI use and strengthening UK capabilities will now "remain in DSIT as policy functions".[121]

73. The Government confirmed the AI Safety Institute would have a budget of circa £400 million to the end of the decade,[122] with £100 million allocated across 2023–24 and 2024–25. The majority of spending "will be on safety research and will be a mix of staffing costs, infrastructure and contractual arrangements". Its £35.5 million budget for 2023–24 allocates circa 86.6 per cent on capital and 13.4 per cent on resource departmental expenditure limit (which typically includes salaries and administration).[123]

74. Professor Dame Wendy Hall, Regius Professor of Computer Science, University of Southampton, thought AI safety was important but believed the Government had "pivoted" to "the tunnel of safety and security risks" in recent months.[124] The Open Data Institute noted the AI Safety Summit

116 HM Government, 'AI Council': https://www.gov.uk/government/groups/ai-council [accessed 8 January 2024]
117 HC Deb, 19 September 2023, vol 737WS
118 DSIT, 'Frontier AI Taskforce: first progress report' (7 September 2023): https://www.gov.uk/government/publications/frontier-ai-taskforce-first-progress-report/frontier-ai-taskforce-first-progress-report#we-have-established-an-expert-advisory-board-spanning-ai-research-and-national-security [accessed 8 January 2024]
119 According to a withdrawn transparency update. See CDEI, 'Transparency data, Advisory Board of the Centre for Data Ethics and Innovation' (12 September 2023): https://www.gov.uk/government/publications/advisory-board-of-the-centre-for-data-ethics-and-innovation/advisory-board-of-the-centre-for-data-ethics-and-innovation [accessed 9 January 2024]. This was subsequently confirmed by a blog on its website. See CDEI, 'Championing responsible innovation: reflections from the CDEI Advisory Board' (26 September 2023): https://cdei.blog.gov.uk/2023/09/26/championing-responsible-innovation-reflections-from-the-cdei-advisory-board/ [accessed 8 January 2024].
120 Department for Media, Culture and Sport, 'Centre for Data Ethics and Innovation: Government response to consultation' (November 2018): https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-government-response-to-consultation [accessed 8 January 2024]
121 DSIT, 'Introducing the AI Safety Institute' (November 2023): https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute [accessed 8 January 2024]
122 Q 134 (Viscount Camrose)
123 Letter from Viscount Camrose, Parliamentary Under Secretary of State Department for Science, Innovation & Technology to Baroness Stowell of Beeston, Chair of the Communications and Digital Committee (8 December 2023): https://committees.parliament.uk/publications/42737/documents/212659/default/
124 Q 30

agenda reflected a narrow view of AI risks shaped largely by big tech firms.[125] Dame Wendy thought that erstwhile priorities under the National AI Strategy to take a more holistic approach were "now slipping", notably around skills, industry adoption and support for disrupted sectors.[126]

75. This is problematic because our evidence suggested leadership in AI safety and commercial prowess are closely linked. Dr Moez Draief, Managing Director of Mozilla.ai, noted that the skills gained from working on commercial models were often those most needed in AI safety, and cautioned that "if the UK is not involved in building or testing models … it will not have the capability to take advantage".[127]

76. And it will be difficult for the Government to use AI specialists to boost public sector expertise if the brightest entrepreneurs and academics are tempted by more attractive offers overseas.[128] As the Royal Academy of Engineering warned:

> "Should the UK fail to develop rapidly as a hub for the development and implementation of LLMs, and other forms of AI, it is likely to lose influence in international conversations on standards and regulatory practices".[129]

77. We therefore welcomed the Government's achievements in convening the AI Safety Summit, but questioned the growing focus on making "AI systems safe", rather than the (arguably harder) task of catalysing responsible innovation and adoption.[130]

78. Professor Dame Angela McLean, Government Chief Scientific Adviser, said the Government's work remained balanced despite changes in public rhetoric.[131] We noted a number of workstreams supporting this position, including the CDEI's £400,000 Fairness Innovation Challenge, Research and Innovation (UKRI) funding for university research programmes, the BridgeAI programme to support adoption, and AI research fellowships.[132]

79. Mr Hogarth told us there was "a certain urgency to the national security challenge" and advocated addressing these first before "you can really start to think about the opportunities".[133] Viscount Camrose, Minister for AI and IP, acknowledged the "tone" of Government's work had veered between innovation and risk, and hoped to "talk with equal emphasis about safety and innovation" in future.[134] When questioned about the balance of external expert advisers, he stated that the disbanding of the AI Council and CDEI advisory board were due to a need for greater agility, and not because the

125 Written evidence from the Open Data Institute (LLM0083)
126 Q 30
127 Q 70
128 Q 33 (Dr Jeremy Silver), Q 119 (Professor Dame Angela McLean) and Surrey Institute for People-Centred Artificial Intelligence (LLM0060)
129 Written evidence from the Royal Academy of Engineering (LLM0063)
130 Q 24 (Professor Stuart Russell OBE), Q 30 (Professor Dame Wendy Hall and Dr Jeremy Silver) and written evidence from Kairoi Ltd (LLM0110)
131 Q 113
132 See for example DSIT, '£54 million boost to develop secure and trustworthy AI research' (14 June 2023): https://www.gov.uk/government/news/54-million-boost-to-develop-secure-and-trustworthy-ai-research [accessed 21 December 2023].
133 Q 5
134 Q 131

Government was losing interest in responsible innovation or believed these bodies provided insufficient support.[135]

80. **The Government is not striking the right balance between innovation and risk. We appreciate that recent advances have required rapid security evaluations and we commend the AI Safety Summit as a significant achievement. But Government attention is shifting too far towards a narrow view of high-stakes AI safety. On its own, this will not drive the kind of widespread responsible innovation needed to benefit our society and economy. The Government must also recognise that long-term global leadership on AI safety requires a thriving commercial and academic sector to attract, develop and retain technical experts.**

81. ***The Government should set out a more positive vision for LLMs and rebalance towards the ambitions set out in the National AI Strategy and AI White Paper. It otherwise risks falling behind international competitors and becoming strategically dependent on a small number of overseas tech firms. The Government must recalibrate its political rhetoric and attention, provide more prominent progress updates on the ten-year National AI Strategy, and prioritise funding decisions to support responsible innovation and socially beneficial deployment.***

82. **A diverse set of skills and people is key to striking the right balance on AI.** ***We advocate expanded systems of secondments from industry, academia and civil society to support the work of officials—with appropriate guardrails as set out in Chapter 3. We also urge the Government to appoint a balanced cadre of advisers to the AI Safety Institute with expertise beyond security, including ethicists and social scientists.***

Removing barriers to UK advantage

83. The Government's Science and Technology Framework lists ten areas required to make the most of technological progress. Five stand out for capitalising on LLM opportunities:

- infrastructure (notably compute);
- skills;
- financing (for spinout companies);
- innovative public sector use (for sovereign capabilities); and
- regulatory certainty.[136]

We cover the first four in this chapter and regulation in Chapter 7.

135 Q 136

136 DSIT, 'The UK Science and Technology Framework' (March 2023): https://www.gov.uk/government/publications/uk-science-and-technology-framework/the-uk-science-and-technology-framework [accessed 8 January 2024]

84. Compute: The UK needs to boost its compute capacity to enable researchers and businesses to keep pace with international competitors.[137] In March 2023 the Government announced £900 million for an 'exascale' supercomputer and AI Research Resource, followed by a further £500 million in November 2023.[138]

85. Professor Zoubin Ghahramani, Vice President of Research at Google DeepMind, said this provided the right "ingredients" for UK-led innovation,[139] though we noted the investments remain dwarfed by big tech. Microsoft alone is investing £2.5 billion over the next three years to expand next generation UK data centres.[140]

86. The UK's universities have long provided publicly beneficial AI research which drives UK international prominence, though high computing costs mean such work on LLMs is increasingly out of reach (see Figure 8 below).[141] Professor Dame Muffy Calder, Vice-Principal at the University of Glasgow and former Chief Scientific Adviser for Scotland, said a "national resource" was needed providing fair access for academic research on LLMs.[142]

137 DSIT, *Independent Review of The Future of Compute* (6 March 2023), Recommendations: https://www.gov.uk/government/publications/future-of-compute-review/the-future-of-compute-report-of-the-review-of-independent-panel-of-experts [accessed 29 November 2023]

138 DSIT, 'Government commits up to £3.5 billion to future of tech and science' (March 2023): https://www.gov.uk/government/news/government-commits-up-to-35-billion-to-future-of-tech-and-science [accessed 8 January 2024] and DSIT 'Science, Innovation and Technology backed in Chancellor's 2023 Autumn Statement' (23 November 2023): https://www.gov.uk/government/news/science-innovation-and-technology-backed-in-chancellors-2023-autumn-statement [accessed 25 January 2024]

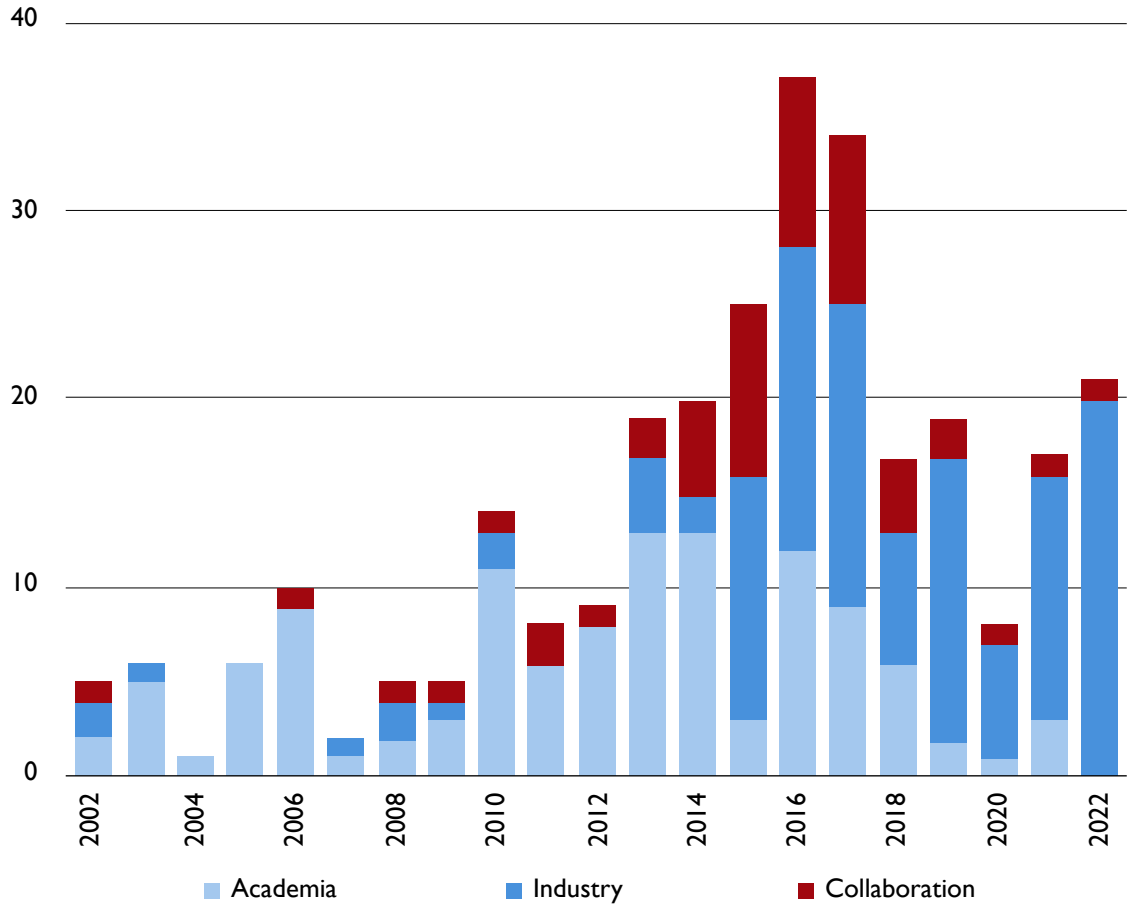
139 Q 104

140 Microsoft, 'Our investment in AI infrastructure, skills and security to boost the UK's AI potential' (November 2023): https://blogs.microsoft.com/on-the-issues/2023/11/30/uk-ai-skilling-security-datacenters-investment/ [accessed 8 January 2024]

141 McKinsey Global Institute, 'Artificial intelligence in the United Kingdom' (2019): https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Artificial%20intelligence%20in%20the%20United%20Kingdom%20Prospects%20and%20challenges/Artificial-intelligence-in-the-United-Kingdom-VF2.ashx [accessed 20 December 2023] and written evidence from Andreessen Horowitz (LLM0114)

142 Q 33

Figure 8: Affiliation of research teams building notable AI systems



Source: HM Government, Safety and security risks of generative artificial intelligence to 2025 (October 2023): https://assets.publishing.service.gov.uk/media/653932db80884d0013f71b15/generative-ai-safety-security-risks-2025-annex-b.pdf [accessed 9 January 2024]

87. Dr Jeremy Silver, CEO of Digital Catapult, an accelerator institute, thought the Government could not match big tech spending but could get the most out of investments by making its new compute capacity more accessible to SMEs.[143] Rachel Coldicutt OBE, Executive Director of Careful Industries, thought future investments should be designed thoughtfully to avoid current problems of overburdening local grid capacity, perhaps by powering facilities through excess renewable capacity.[144] Others similarly recommended guidelines and incentives to boost energy efficiency and environmentally responsible development.[145]

88. Skills: Professor Dame Angela McLean, Government Chief Scientific Adviser, said that skills gaps remained another significant barrier to AI leadership. She called for continued investments in skills throughout the career lifecycle; secondments between industry, government and regulators; and new cadres of technically adept public servants attracted through better pay and conditions.[146] Professor Dame Wendy Hall noted there were some

143 *Ibid.*
144 Written evidence from Careful Industries (LLM0041)
145 Written evidence from the Market Research Society (LLM0088) and Caution Your Blast (LLM0077)
146 QQ 114–119

skills programmes but was similarly concerned the UK was falling behind rivals in league tables.[147]

89. Spin-out companies: Dr Nathan Benaich, Founder of the AI venture capital firm Air Street Capital, outlined the UK's longstanding challenges around supporting spin-outs and incentivising businesses to remain in the UK.[148] Dr Silver said better pathways were needed to help academic spin-outs achieve sustainable commercialisation.[149] Value for money could be achieved by directing support at ventures addressing public service needs, for example in education and healthcare.[150] Dr Silver also suggested focusing on retaining business ownership in the UK, even if scaling up occurs in the US.[151]

90. During our visit to UCL Business we heard that changes to funding allocations for Centres for Doctoral Training meant there had been a significant drop in the number of funded AI PhD places.[152] Professor David Barber, Director of the UCL Centre for Artificial Intelligence, said the situation was "alarming", noting that successful centres with a track record of producing commercial spinouts were at significant risk.[153]

91. Overseas funding is likely to be the main alternative for many universities, and some reports indicate China is likely to be a key actor.[154] The Intelligence and Security Committee recently warned about the growing threat from China's influence in strategic sectors and raised concerns around intellectual property transfer as a condition of funding.[155]

92. **Recent Government investments in advanced computing facilities are welcome, but more is needed and the Government will struggle to afford the scale required to keep pace with cutting edge international competitors. *The Government should provide more incentives to attract private sector investment in compute. These should be structured to maximise energy efficiency.***

147 Q 30
148 Q 14. An academic spinout is typically a company created by one or more academics or research staff with the aim of commercialising research.
149 Q 33
150 UK AI Council, *Draft Memo* (December 2022): https://mlatcl.github.io/papers/ai-council-llm-memo.pdf [accessed 8 January 2024]. The Government has a number of workstreams to support businesses, see for example DSIT, ' Secretary Michelle Donelan's speech at Plexal' (16 January 2024): https://www.gov.uk/government/speeches/science-innovation-and-technology-secretary-michelle-donelans-speech-at-plexal [accessed 19 January 2024].
151 Q 33. A recent independent review advocated further measures to support a self-sustaining spinout ecosystem. See DSIT, *Independent Review of University Spin-out Companies* (November 2023): https://assets.publishing.service.gov.uk/media/6549fcb23ff5770013a88131/independent_review_of_university_spin-out_companies.pdf [accessed 8 January 2024].
152 Written evidence from Professor David Barber (LLM0018). For details on the Centres for Doctoral Training see UK Research and Innovation, 'Centres for Doctoral Training (CDT)': https://www.ukri.org/what-we-do/developing-people-and-skills/nerc/nerc-studentships/directed-training/centres-for-doctoral-training-cdt/ [accessed 8 January 2024].
153 Written evidence from Professor David Barber (LLM0018)
154 'British universities are becoming dependent on China – and its military', *The Telegraph* (November 2023): https://www.telegraph.co.uk/news/2023/11/14/british-universities-dependent-china-military/ [accessed 8 January 2024] and 'Chinese money is pouring into British universities', *The Economist* (March 2022): https://www.economist.com/britain/2022/03/12/chinese-money-is-pouring-into-british-universities [accessed 8 January 2024]
155 Intelligence and Security Committee of Parliament, *China* (July 2023, HC 1605) and Cabinet Office, 'Government Response to the Intelligence and Security Committee of Parliament Report *China*' (September 2023): https://www.gov.uk/government/publications/government-response-to-the-isc-china-report/government-response-to-the-intelligence-and-security-committee-of-parliament-report-china-html [accessed 8 January 2024]

93. **Equitable access will be key.** ***UK Research and Innovation and DSIT must ensure that both researchers and SMEs are granted access to high-end computing facilities on fair terms to catalyse publicly beneficial research and commercial opportunity.***

94. **The Government should take better advantage of the UK's start-up potential.** ***It should work with industry to expand spin-out accelerator schemes. This could focus on areas of public benefit in the first instance. It should also remove barriers, for example by working with universities on providing attractive licensing and ownership terms, and unlocking funding across the business lifecycle to help start-ups grow and scale in the UK.***

95. ***The Government should also review UKRI's allocations for AI PhD funding, in light of concerns that the prospects for commercial spinouts are being negatively affected and foreign influence in funding strategic sectors may grow as a result.***

The case for sovereign capabilities

96. LLMs offer significant opportunities for the public sector if challenges around ethics, reliability, security and interpretability can be overcome.[156] LLMs could reduce general administrative burdens on office and frontline staff, while sector-specific tools could support education, intelligence analysis, healthcare processes and research, environmental and geospatial analyses, public engagement services, and more.[157] Public sector bodies are already starting to trial LLM-powered services.[158] Some countries are going further and establishing domestic capabilities.[159]

97. Our evidence suggested several options for developing a sovereign LLM capability. This might be an 'in-house model' used by Government and public sector bodies, or a wider facility available to researchers and industry.

98. We explored three main options for an in-house model. Purchasing an 'off the shelf' commercially available model would be quick and cheap, but carries risks around insufficient oversight of governance, safety guardrails, bias mitigations, and data privacy—as well as concerns around strategic dependence.[160] Developing a model from scratch would provide more

156 Ada Lovelace Institute, 'Foundation models in the public sector' (October 2023): https://www.adalovelaceinstitute.org/evidence-review/foundation-models-public-sector/ [accessed 8 January 2024]. See Chapter 5 for a discussion of the risks that need to be addressed.

157 Q 132, Adam C, Dr Richard Carter, 'Large Language Models and Intelligence Analysis': https://cetas.turing.ac.uk/publications/large-language-models-and-intelligence-analysis [accessed 21 December 2023] and Ada Lovelace Institute, 'Foundation models in the public sector' (October 2023): https://www.adalovelaceinstitute.org/evidence-review/foundation-models-public-sector [accessed 8 January 2024].

158 Cogstack, 'Unlock the power of healthcare data with CogStack': https://cogstack.org/ [accessed 21 December 2023]

159 For sample initiatives in Sweden, the United Arab Emirates, and Japan see Deloitte, *Large language models - a backgrounder* (September 2023): https://www2.deloitte.com/content/dam/Deloitte/in/Documents/Consulting/in-consulting-nasscom-deloitte-paper-large-language-models-LLMs-noexp.pdf [accessed 8 January 2023].

160 See for example NCSC, 'Exercise caution when building off LLMs' (30 August 2023): https://www.ncsc.gov.uk/blog-post/exercise-caution-building-off-llms [accessed 21 December 2023].

control—but would require a high-risk, high-tech and expensive in-house R&D effort to which the Government may be poorly suited.[161]

99. Commissioning an external developer to build a model which is deployed on secure Government infrastructure and UK-based data processing capabilities would provide a middle route.[162] The Government would set safety and ethical standards. The developer would provide the software and expertise for training and a licence for the Government to run the model in-house.[163] This would likely be lower risk, though not entirely risk-free.

100. Smaller in-house models could be built on top and fine-tuned for different departments. Dame Muffy noted the UK already had "fabulous resources in health data, ONS data, geospatial data, environmental data".[164] An in-house model might be used to try new safety or regulatory features, supporting Government aims to become an AI safety leader. A joint report by Lord Hague of Richmond and Sir Tony Blair argued that a domestic capability could underpin future public services, reduce strategic reliance on external providers for a critical technology, and help the Government respond with agility to fast-moving advances.[165]

101. The Government could also explore developing more widely accessible facilities. The Open Data Institute said sovereign capabilities could be used to support wider research and innovation, for example.[166] The AI Council has previously suggested the Government should develop a "proving ground" which offers world-class facilities and brings together researchers and practitioners to solve practical challenges that arise "when deploying AI models to address UK national priorities".[167]

102. Across all options, value for money would be key. EPOCH, a research initiative, estimated the current cost of building and maintaining LLM infrastructure at $300–600 million, while indicative costs to rent compute from the cloud to train a model might range from $40–100 million.[168] Costs may fall in time, while cheaper models requiring less compute may become more capable.[169]

---

161 The Government established an Advanced Research Innovation Agency in January 2023 to fund high-risk, high-reward scientific research. See Department for Business, Energy and Industrial Strategy, 'Research agency supporting high risk, high reward research formally established' (January 2023): https://www.gov.uk/government/news/research-agency-supporting-high-risk-high-reward-research-formally-established [accessed 8 January 2024]. The National Audit Office has in the past been critical of internal digital projects within Government. See for example National Audit Office, 'Digital transformation in government: addressing the barriers to efficiency' (March 2023): https://www.nao.org.uk/reports/digital-transformation-in-government-addressing-the-barriers/ [accessed 8 January 2024] and National Audit Office, 'Digital Transformation in Government (2017)' (March 2017): https://www.nao.org.uk/reports/digital-transformation-in-government/ [accessed 8 January 2024].

162 UK AI Council, *The UK Foundation Models Opportunity* (April 2023): https://mlatcl.github.io/papers/ai-council-foundation-models-policy-paper.pdf [accessed 14 December 2023]

163 See for example Sir Tony Blair and Lord Hague of Richmond, *A New National Purpose* (February 2023): https://www.williamhague.com/_files/ugd/067357_96e45c693747432e8bd21dca773fde28.pdf [accessed 3 January 2024].

164 Q 33

165 Sir Tony Blair and Lord Hague of Richmond, *A New National Purpose* (February 2023): https://www.williamhague.com/_files/ugd/067357_96e45c693747432e8bd21dca773fde28.pdf [accessed 3 January 2024]

166 Written evidence from the Open Data Institute (LLM0083)

167 UK AI Council, *The UK Foundation Models Opportunity* (April 2023): https://mlatcl.github.io/papers/ai-council-foundation-models-policy-paper.pdf [accessed 14 December 2023]

168 Written evidence from EPOCH (LLM002). Note the costs are indicative and it may not be feasible to rent such levels.

169 Written evidence from the Royal Statistical Society (LLM0055)

103. Ethics and reliability would also be vital. Professor Phil Blunsom, Chief Scientist at Cohere, highlighted the varying degrees of LLM reliability and thought any uses affecting life outcomes should be "heavily regulated".[170] The Committee on Standards in Public Life noted that the Government could learn lessons from abroad when considering the ethical use of public sector AI: Canada has compulsory ethics assessments for automated decision-making systems, for example.[171]

104. The Minister said he could see "in principle" the advantages of having a sovereign LLM but would "wait for the evidence" and further advice on next-generation model capabilities and uses, expected in early 2024.[172]

105. **A sovereign UK LLM capability could deliver substantial value if challenges around reliability, ethics, security and interpretability can be resolved. LLMs could in future benefit central departments and public services for example, though it remains too early to consider using LLMs in high-stakes applications such as critical national infrastructure or the legal system.**

106. **We do not recommend using an 'off the shelf' LLM or developing one from scratch: the former is too risky and the latter requires high-tech R&D efforts ill-suited to Government. But commissioning an LLM to high specifications and running it on internal secure facilities might strike the right balance. The Government might also make high-end facilities available to researchers and commercial partners to collaborate on applying LLM technology to national priorities.**

107. ***We recommend that the Government explores the options for and feasibility of acquiring a sovereign LLM capability. No option is risk free, though commissioning external developers might work best. Any public sector capability would need to be designed to the highest ethical and security standards, in line with the recommendations made in this report.***

170 Q 24
171 Written submission from the Committee on Standards in Public Life (LLM0052)
172 Q 132

CHAPTER 5: RISK

108. The nature, likelihood and impact of risks arising from large language models (LLMs) remains subject to much debate. The complexity stems in part from the extensive literature,[173] lack of agreed definitions, hype around rapid developments,[174] and the possibility that some organisations may have interests in emphasising or downplaying risk.[175]

109. This chapter examines a selection of security and societal risks.[176] We sought to distinguish hype from reality and provide some reference points to ground our review. We found credible evidence of both immediate and longer-term risks from LLMs to security, financial stability and societal values.

110. The first section of this chapter sets out our understanding of risk categories. The next section sets out near-term security risks that require immediate attention, followed by a discussion on longer-term concerns around catastrophic risk and then existential risk. Near-term societal risks such as bias and discrimination are discussed at the end of the chapter.

What are we talking about?

111. There are numerous frameworks for evaluating risk used by domestic and international authorities.[177] We found little consistency in terms or methods across the literature.[178] We adopt the framework from the Government's National Risk Register (NRR), set out in the table below, to help describe impacts of LLM-related security risks. Our categorisation is approximate only and we do not attempt to replicate the full National Security Risk

173 Our analysis draws on evidence submitted to this inquiry alongside Government publications, industry assessments, academic reviews and stakeholder engagements.

174 MIT Technology Review, 'AI hype is built on high test scores' (30 August 2023): https://www.technologyreview.com/2023/08/30/1078670/large-language-models-arent-people-lets-stop-testing-them-like-they-were/ [accessed 20 December 2023]

175 'How the UK's emphasis on apocalyptic AI risk helps business', *The Guardian* (31 October 2023): https://www.theguardian.com/technology/2023/oct/31/uk-ai-summit-tech-regulation [accessed 20 December 2023]

176 The distinction is made here for ease of analysis, noting that many of the risks and outcomes overlap. We describe bias as a societal risk, though a biased LLM used for defence-related decision-making might introduce security risks. Similarly a poorly calibrated LLM used in healthcare might result in fatalities. Our assessments are indicative only.

177 For a discussion on determining acceptable fatality rates see written evidence from Matthew Feeney (LLM047). For frameworks on risk see for example the US National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework* (January 2023): https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [accessed 20 December 2023] and European Commission, 'Regulatory framework proposal on artificial intelligence': https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai [accessed 20 December 2023]. See also National Cyber Security Centre, 'Guidelines for secure AI System development' (November 2023): https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development [accessed 8 January 2024].

178 See for example the AI Safety Summit discussion paper, alongside Annex A and Annex B, available at DSIT, 'Frontier AI' (25 October 2023): https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper [accessed 8 January 2024], 'The Bletchley Declaration by Countries Attending the AI Safety Summit' (1 November 2023): https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 [accessed 8 January 2024], 'Introducing the AI Safety Institute' (2 November 2023): https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute [accessed 8 January 2024], Department for Digital, Culture, Media and Sport, *National AI Strategy*, Cp 525 (September 2021): https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf [accessed 20 December 2023] and National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework* (January 2023): https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [accessed 8 January 2023].

Assessment process. It nevertheless provides a helpful yardstick to anchor discussion using a recognised framework.[179] This table does not cover existential risk, which we describe as a separate category later in this chapter.

Table 1: Risk categories

Risk Level	Fatalities	Casualties	Economic impact
Minor	1–8	1–17	£ millions
Limited	9–40	18–80	£ tens of millions
Moderate	41–200	81–400	£ hundreds of millions
Significant	201–1000	400–2000	£ billions
Catastrophic	More than 1,000	More than 2,000	£ tens of billions

Source: HM Government, National Risk Register (2023): https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1175834/2023_NATIONAL_RISK_REGISTER_NRR.pdf [accessed 20 December 2023]

112. There are also various ways of categorising societal risk and conducting impact assessments.[180] We draw on these to inform our review of societal risk, noting that the issues are highly context-dependent.

Threat models

113. Risks may arise from both open and closed models, for example through:

- inappropriate deployment (for example using a model to diagnose patients without proper procedures and safeguards);
- increasing the tools available to malicious actors (for example auto-generating phishing campaigns);
- poor performance or model malfunction (for example a safety feature failure);
- gradual over-reliance (for example handing increasingly important decisions and processes to machines with insufficient human engagement or accountability); or

179 Note the NRR evaluation timeframe is assessed over two years for malicious risks and five years for non-malicious risks. We acknowledge AI may be treated as both a chronic and acute risk.

180 See for example Cabinet Office, 'Ethics, Transparency and Accountability Framework for Automated Decision-Making' (November 2023): https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making [accessed 20 December 2023], Central Digital and Data Office, 'Data Ethics Framework' (September 2020): https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020 [accessed 20 December 2023], CDEI, 'Review into bias in algorithmic decision-making' (November 2020): https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making [accessed 20 December 2023], Information Commissioner's Office, 'Data protection impact assessments': https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/guide-to-accountability-and-governance/accountability-and-governance/data-protection-impact-assessments/ [accessed 20 December 2023] and House of Commons Library, 'The Public Sector Equality Duty and Equality Impact Assessments', Research Briefing SN06591, July 2020.

- loss of control (for example where a highly capable machine pursues its own objectives that may not be obvious to humans or aligned with our wellbeing).[181]

Near-term security risks

114. Our evidence was clear that LLMs will act as a force multiplier enhancing malicious capabilities in the first instance, rather than introducing qualitatively new risks.[182] Most models have some safeguards but these are not robust and can be circumvented.[183] We believe the most immediate security risks over the next three years are likely to include the (non-exhaustive) list below, with indicative impacts ranging from minor to moderate, rather than catastrophic.

115. Cyber: LLMs are likely to be of interest to hostile states, organised crime, and low-sophistication actors.[184] Some LLMs are reportedly being developed to create code for cyber attacks at increased scale and pace.[185] LLMs and multi-modal models will make it easier to create phishing campaigns, fraudulent websites and voice cloning to bypass security protocols.[186] Malicious actors may use prompt injection attacks to obtain sensitive information, or target models themselves to influence the outputs, poison training data or induce system malfunction.[187] Current security standards are unlikely to withstand attacks from sophisticated threat actors.[188]

116. Tools to mass produce high quality and openly available destructive cyber weapons appear limited at present. Chris Anley, Chief Scientist at the cyber security firm NCC Group, said LLMs currently provide efficiency and lower barriers to entry, rather than game-changing capability leaps.[189]

---

181 Written evidence from the Alan Turing Institute (LLM0081), Martin Hosken (LLM0009), Royal Academy of Engineering (LLM0063) and DSIT, 'Frontier AI' (25 October 2023): https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper [accessed 8 January 2024]

182 Q 27 (Professor Phil Blunsom), Q 24 (Chris Anley), Q 24 (Lyric Jain), written evidence from Ofcom (LLM0104), Competition and Markets Authority (LLM0100), Financial Conduct Authority (LLM0102), Open Data Institute (LLM0083), Alan Turing Institute (LLM0081) and HM Government, *Safety and Security Risks of Generative Artificial Intelligence to 2025* (2023): https://assets.publishing.service.gov.uk/media/653932db80884d0013f71b15/generative-ai-safety-security-risks-2025-annex-b.pdf [accessed 21 December 2023]

183 Q 26 (Lyric Jain) and 'GPT-4 gave advice on planning terrorist attacks when asked in Zulu', *New Scientist* (October 2023): https://www.newscientist.com/article/2398656-gpt-4-gave-advice-on-planning-terrorist-attacks-when-asked-in-zulu/ [accessed 20 December 2023]

184 Written evidence from NCC Group (LLM0014), Q 22 (Professor Phil Blunsom) and NCSC, 'Annual Review 2023' (2023): https://www.ncsc.gov.uk/collection/annual-review-2023/technology/case-study-cyber-security-ai [accessed 20 December 2023]

185 Check Point Research, 'OPWNAI: cyber criminals starting to use ChatGPT' (January 2023): https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/ [accessed 20 December 2023] and 'WormGPT: AI tool designed to help cybercriminals will let hackers develop attacks on large scale, experts warn', *Sky* (September 2023): https://news.sky.com/story/wormgpt-ai-tool-designed-to-help-cybercriminals-will-let-hackers-develop-attacks-on-large-scale-experts-warn-12964220 [accessed 20 December 2023]

186 Q 24 (Chris Anley)

187 A prompt injection involves entering a text prompt into an LLM which then enables the actor to bypass safety protocols. See written evidence from NCC Group (LLM0014), Q 24 (Chris Anley) and National Cyber Security Centre, 'Exercise caution when building off LLMs' (August 2023): https://www.ncsc.gov.uk/blog-post/exercise-caution-building-off-llms [accessed 20 December 2023].

188 DSIT, *Capabilities and risks from frontier AI* (October 2023), p 18: https://assets.publishing.service.gov.uk/media/65395abae6c968000daa9b25/frontier-ai-capabilities-risks-report.pdf [accessed 20 December 2023]

189 Q 24 (Chris Anley)

Even moderate gains could however prove costly when deployed against under-prepared systems, as previous attacks on the NHS have shown.[190]

117. A reasonable worst case scenario might involve malicious actors using LLMs to produce attacks achieving higher cyber infection rates in critical public services or national infrastructure.[191]

118. Terrorism: A recent report by Europol found that LLM capabilities are useful for terrorism and propaganda.[192] Options include generating and automating multilingual translation of propaganda, and instructions for committing acts of terror.[193] In future, openly available models might be fine-tuned to provide more specific hate speech or terrorist content capabilities, perhaps using archives of propaganda and instruction manuals.[194] The leak of Meta's model (called LLaMa) on 4chan, a controversial online platform, is instructive. Users reportedly customised it within two weeks to produce hate speech chatbots, and evaded take-down notices.[195]

119. National Statistics data show 93 victim deaths due to terrorism in England and Wales between April 2003 and 31 March 2021.[196] A reasonable worst case scenario might involve a rise in attacks directly attributable to LLM-generated propaganda or made possible through LLM-generated instructions for building weapons.[197]

120. Synthetic child sexual abuse material: Image generation models are already being used to generate realistic child sexual abuse material (CSAM).[198] The Stanford Internet Observatory predicts that in under a year "it will become significantly easier to generate adult images that are indistinguishable from actual images".[199] The Internet Watch Foundation has confirmed this is "happening right now",[200] and stated legal software can be downloaded and

190 The 2017 WannaCry cyber-attack for example affected 30 per cent of NHS Trusts, costing £92 million. See 'Cost of WannaCry cyber-attack to the NHS revealed', *Sky*, 11 October 2018: https://news.sky.com/story/cost-of-wannacry-cyber-attack-to-the-nhs-revealed-11523784 [accessed 20 December 2023].

191 Cabinet Office, 'National Risk Register' (2023), p 15: https://www.gov.uk/government/publications/national-risk-register-2023 [accessed 20 December 2023]

192 EUROPOL, *ChatGPT—The impact of Large Language Models on Law Enforcement* (March 2023): https://www.europol.europa.eu/cms/sites/default/files/documents/Tech%20Watch%20Flash%20-%20The%20Impact%20of%20Large%20Language%20Models%20on%20Law%20Enforcement.pdf [accessed 20 December 2023]

193 Tech Against Terrorism, 'Early Terrorist Adoption of Generative AI' (November 2023): https://techagainstterrorism.org/news/early-terrorist-adoption-of-generative-ai [accessed 20 December 2023]

194 Global Network on Extremism and Technology, ''RedPilled AI': A New Weapon for Online Radicalisation on 4chan' (June 2023): https://gnet-research.org/2023/06/07/redpilled-ai-a-new-weapon-for-online-radicalisation-on-4chan/ [accessed 20 December 2023]

195 *Ibid.*

196 House of Commons Library, 'Terrorism in Great Britain: the statistics', Research Briefing, CBP7613, 19 July 2022

197 HM Government, *National Risk Register 2023 Edition* (2023): https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1175834/2023_NATIONAL_RISK_REGISTER_NRR.pdf [accessed 20 December 2023]. See section on terrorism pp 30–54.

198 Q 10 (Ian Hogarth)

199 David Thiel, Melissa Stroebel and Rebecca Portnoff, *Generative ML and CSAM: Implications and Mitigations* (June 2023): https://stacks.stanford.edu/file/druid:jv206yg3793/20230624-sio-cg-csam-report.pdf [accessed 21 December 2023]

200 Matt O'Brien and Haleluya Hadero, 'AI-generated child sexual abuse images could flood the internet', *AP* (October 2023): https://apnews.com/article/ai-artificial-intelligence-child-sexual-abuse-c8f17de56d41f05f55286eb6177138d2 [accessed 21 December 2023]

used offline to produce illegal content "with no opportunity for detection".[201] This suggests more abuse imagery will be in circulation, law enforcement agencies may find it more difficult to identify and help real-world victims, and opportunities to groom and coerce vulnerable individuals will grow.[202]

121. AI CSAM currently represents a small proportion of the total amount of CSAM (reportedly 255,000 webpages last year with potentially millions of images).[203] A reasonable worst case scenario might involve widespread availability of illegal materials which overwhelms law enforcement ability to respond.[204]

122. Mis/disinformation: LLMs are well placed to generate text-based disinformation at previously unfeasible scale, while multi-modal models can create audio and visual deepfakes which even experts find increasingly difficult to identify.[205] LLMs' propensity to hallucinate also means they can unintentionally misinform users.[206] The National Cyber Security Centre assesses that large language models will "almost certainly be used to generate fabricated content; that hyper-realistic bots will make the spread of disinformation easier; and that deepfake campaigns are likely to become more advanced in the run up to the next nationwide vote, scheduled to take place by January 2025".[207]

123. Professor Dame Angela McLean, Government Chief Scientific Adviser, said she was "extremely worried" and called for a public awareness campaign.[208] Dr Jean Innes, CEO of the Alan Turing Institute, similarly warned about "mass disinformation".[209] Professor Phil Blunsom, Chief Scientist at Cohere, likewise highlighted "disinformation [and] election security" as issues of concern.[210]

124. Lyric Jain, CEO of the counter-disinformation firm Logically, said one of the main impacts of generative AI was increased efficiency and lower costs. He estimated the Internet Research Agency's disinformation campaign targeting the US 2016 election cost at least $10 million,[211] whereas generating comparable disinformation materials could now be done for $1,000 by private

---

201 Internet Watch Foundation, *How AI is being abused to create child sexual abuse imagery* (October 2023): https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf [accessed 21 December 2023)

202 David Thiel, Melissa Stroebel and Rebecca Portnoff, *Generative ML and CSAM: Implications and Mitigations* (June 2023): https://stacks.stanford.edu/file/druid:jv206yg3793/20230624-sio-cg-csam-report.pdf [accessed 21 December 2023]

203 Internet Watch Foundation, *How AI is being abused to create child sexual abuse imagery* (October 2023): https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf [accessed 21 December 2023]

204 *Ibid.*

205 Written evidence from the Alan Turing Institute (LLM0081), Logically AI (LLM0062), Dr Jeffrey Howard et al (LLM0049) and Full Fact (LLM0058)

206 Written evidence from the Surrey Institute for People-Centred Artificial Intelligence (LLM0060)

207 NCSC, 'NCSC warns of enduring and significant threat to UK's critical infrastructure' (14 November 2023): https://www.ncsc.gov.uk/news/ncsc-warns-enduring-significant-threat-to-uks-critical-infrastructure [accessed 21 December 2023]

208 Q 119

209 Q 3

210 Q 24

211 For details of the US intelligence community assessment of activities conducted by the Russian Federation see The Director of National Intelligence, 'Assessing Russian Activities and Intentions in Recent US Elections' (January 2017): https://www.dni.gov/files/documents/ICA_2017_01.pdf [accessed 20 December 2023].

individuals. He further noted that model safeguards were preventing only 15 per cent of disinformation-related prompts.[212]

125. A reasonable worst case scenario might involve state and non-state interference undermining confidence in the integrity of a national election, and long-term disagreement about the validity of the result.[213]

Mitigations

126. A range of mitigation work is underway across Government and industry. The main issue remains scale and speed: malicious actors enjoy first-mover advantages whereas it will take time to upgrade public and private sector mitigations, including public awareness.[214] And as the Government's AI Safety Summit paper noted, there are limited market incentives to provide safety guardrails and no standardised safety benchmarks.[215]

127. We wrote to the Government seeking more information. It declined to provide details on whether mitigations were being expanded. But it did confirm workstreams included:

- Cyber: Research from the AI Safety Institute and DSIT's new AI central risk function; delivery of the National Cyber Strategy; and Cabinet Office work on AI cyber risks.
- Counter-terror: delivery of the CONTEST strategy, and monitoring the early-stage experimentation of generative AI for terrorist purposes.
- CSAM: Measures under the Online Safety Act; delivery of the 2021 Child Sexual Abuse Strategy; international partnerships; monitoring technology developments; investments in the National Crime Agency, GCHQ and policing; and setting up a "new central strategic function" looking at emerging technology.
- Disinformation: Measures under the Defending Democracy Taskforce, National Security Online Information Team and Election Cell; implementation of the Online Safety Act; media literacy; and international partnerships.[216]

128. **The most immediate security concerns from LLMs come from making existing malicious activities easier, rather than qualitatively new risks. *The Government should work with industry at pace to scale existing mitigations in the areas of cyber security (including systems vulnerable to voice cloning), child sexual abuse material,***

212 QQ 24–25

213 For further details of disinformation affecting elections and other Government priorities see HM Government, *National Risk Register 2020 Edition* (2020): https://assets.publishing.service.gov.uk/media/6001b2688fa8f55f6978561a/6.6920_CO_CCS_s_National_Risk_Register_2020_11-1-21-FINAL.pdf [accessed 21 December 2023].

214 Written evidence from NCC Group (LLM0014) and letter from Viscount Camrose, Parliamentary Under Secretary of State Department for Science, Innovation & Technology to Baroness Stowell of Beeston, Chair of the Communications and Digital Committee (16 January 2024): https://committees.parliament.uk/work/7827/large-language-models/publications/3/correspondence/

215 DSIT, 'Frontier AI' (25 October 2023): https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper [accessed 8 January 2024]

216 Letter from Viscount Camrose, Parliamentary Under Secretary of State Department for Science, Innovation & Technology to Baroness Stowell of Beeston, Chair of the Communications and Digital Committee (16 January 2023): https://committees.parliament.uk/work/7827/large-language-models/publications/3/correspondence/

counter-terror, and counter-disinformation. It should set out progress and future plans in response to this report, with a particular focus on disinformation in the context of upcoming elections.

129. **The Government has made welcome progress on understanding AI risks and catalysing international co-operation. There is however no publicly agreed assessment framework and shared terminology is limited. It is therefore difficult to judge the magnitude of the issues and priorities.** ***The Government should publish an AI risk taxonomy and risk register. It would be helpful for this to be aligned with the National Security Risk Assessment.***

Catastrophic risk

130. Catastrophic risks might arise from the deployment of a model with highly advanced capabilities without sufficient safeguards.[217] As outlined in the previous table, indicative impacts might involve over 1,000 fatalities, 2,000 casualties and/or financial damages exceeding £10 billion.

131. There are threat models of varying plausibility.[218] The majority of our evidence suggests these are less likely within the next three years but should not be ruled out—particularly as the capabilities of next-generation models become clearer and open access models more widespread.[219] We outline some of the most plausible risks below.

132. Biological or chemical release: A model might be used to lower the barriers to malicious actors creating and releasing a chemical or biological agent. There is evidence that LLMs can already identify pandemic-class pathogens, explain how to engineer them, and even suggest suppliers who are unlikely to raise security alerts.[220] Such capabilities may be attractive to sophisticated terror groups, non-state armed groups, and hostile states. This scenario would still require a degree of expertise, access to requisite materials and, probably, sophisticated facilities.[221]

133. Destructive cyber tools: Next generation LLMs and more extensive fine tuning may yield models capable of much more advanced malicious activity.[222] These may be integrated into systems capable of autonomous self-improvement and a degree of replication.[223] Such advances would raise the possibility of

217 HM Government, *Safety and Security Risks of Generative Artificial Intelligence to 2025* (2023): https://assets.publishing.service.gov.uk/media/653932db80884d0013f71b15/generative-ai-safety-security-risks-2025-annex-b.pdf [accessed 21 December 2023]

218 Center for AI Safety, 'An overview of catastrophic AI risks': https://www.safe.ai/ai-risk [accessed 20 December 2023]

219 QQ 22–23, written evidence from Royal Academy of Engineering (LLM0063), Microsoft (LLM0087), Google and Google DeepMind (LLM0095), OpenAI (LLM0013) and DSIT (LLM0079)

220 Kevin Esvelt et al, 'Can large language models democratize access to dual-use biotechnology?' (June 2023): https://arxiv.org/abs/2306.03809[accessed 21 December 2023]

221 Andrew D White et al, 'ChemCrow: Augmenting large-language models with chemistry tools' (April 2023): https://arxiv.org/abs/2304.05376 [accessed 8 January 2024] and Nuclear Threat Initiative, *The Convergence of Artificial Intelligence and the Life Sciences* (October 2023): https://www.nti.org/wp-content/uploads/2023/10/NTIBIO_AI_FINAL.pdf [accessed 21 December 2023]

222 Effective Altruism Forum, ' Possible OpenAI's Q* breakthrough and DeepMind's AlphaGo-type systems plus LLMs' (November 2023): https://forum.effectivealtruism.org/posts/3diski3inLfPrWsDz/possible-openai-s-q-breakthrough-and-deepmind-s-alphago-type [accessed 21 December 2023]

223 Note that the Government assesses generative AI is unlikely to fully automate computer hacking by 2025. See HM Government, *Safety and Security Risks of Generative Artificial Intelligence to 2025* (2023): https://assets.publishing.service.gov.uk/media/653932db80884d0013f71b15/generative-ai-safety-security-risks-2025-annex-b.pdf [accessed 21 December 2023].

advanced language model agents navigating the internet semi-autonomously, performing sophisticated exploits, using resources such as payment systems, and generating snowball effects created by self-improvement techniques.[224] Recent research suggests such capabilities do not yet exist, though progress on the component parts of such tools is already underway and capability leaps cannot be ruled out.[225]

134. Critical infrastructure failure: Models may in time be linked to systems powering critical national infrastructure (CNI) such as water, gas and electricity transmission, or security platforms (for example in military planning or intelligence analysis systems). This might occur either through direct integration of models with the infrastructure platform itself, or through software used in the supply chain.[226] In the absence of safeguards, a sudden model failure may trigger a CNI outage or sudden security lapse, and could be extremely difficult to rectify given the black-box nature of LLM processes.

Mitigations

135. Professor Dame Angela McLean, Government Chief Scientific Adviser, confirmed that there were no agreed warning indicators for catastrophic risk. She said warning indicators for pandemics and similar were well understood, but:

> "we do not have that spelled out for the more catastrophic versions of these risks. That is part of the work of the AI Safety Institute: to make better descriptions of things that might go wrong, and scientific descriptions of how we would measure that."[227]

136. OpenAI told us work was underway to evaluate "dangerous capabilities" and appropriate safety features but noted "science-based measurements of frontier system risks … are still nascent".[228]

137. Professor John McDermid OBE, Professor of Safety-Critical Systems at the University of York, said industries like civil aviation designed software with fault-detection in mind so that sudden failures could be fixed with speed and confidence.[229] He did not believe such safety-critical system analysis was possible yet for LLMs and believed it should be a research priority.[230]

138. Professor Stuart Russell OBE, Professor of Computer Science at the University of California, Berkeley, was sceptical that the biggest safety challenges could be addressed without fundamental design changes. He noted that high-stakes industries like nuclear power had to show the likelihood

224 Megan Kinniment et al, *Evaluating Language-Model Agents on Realistic Autonomous Tasks*: https://evals.alignment.org/Evaluating_LMAs_Realistic_Tasks.pdf [accessed 21 December 2023]
225 *Ibid.*, written evidence from the Alan Turing Institute (LLM0081)
226 See for example Adam C, Dr Richard J. Carter, 'Large Language Models and Intelligence Analysis': https://cetas.turing.ac.uk/publications/large-language-models-and-intelligence-analysis [accessed 21 December 2023], War on the Rocks, 'How large language models can revolutinise military planning (12 April 2023): https://warontherocks.com/2023/04/how-large-language-models-can-revolutionize-military-planning/ [accessed 9 January 2024] and National Cyber Security Centre, 'NCSC CAF guidance': https://www.ncsc.gov.uk/collection/caf/cni-introduction [accessed 21 December 2023].
227 Q 118 (Professor Dame Angela McLean)
228 Written evidence from OpenAI (LLM0113)
229 The bug responsible for the 2014 UK air traffic control failure was found within 45 minutes, for example. See 'Flights disrupted after computer failure at UK control centre', *BBC* (12 December 2014): https://www.bbc.co.uk/news/uk-30454240 [accessed 20 December 2023].
230 Q 70

of sudden catastrophic failure rates, which LLM developers could not. He also noted it was straightforward to bypass a model's safety guardrails by prefixing a harmful question with something unintelligible to confuse it, and maintained that:

> "The security methods that exist are ineffective and they come from an approach that is basically trying to make AI systems safe as opposed to trying to make safe AI systems. It just does not work to do it after the fact".[231]

139. Ian Hogarth, Chair of the (then) Frontier AI Taskforce, told us that the Government took catastrophic risk very seriously. Viscount Camrose, Minister for AI and Intellectual Property, said the AI Safety Institute was focusing on frontier AI safety and driving "foundational" research.[232]

140. **Catastrophic risks resulting in thousands of UK fatalities and tens of billions in financial damages are not likely within three years, though this cannot be ruled out as next generation capabilities become clearer and open access models more widespread.**

141. **There are however no warning indicators for a rapid and uncontrollable escalation of capabilities resulting in catastrophic risk. There is no cause for panic, but the implications of this intelligence blind spot deserve sober consideration.**

142. ***The AI Safety Institute should publish an assessment of engineering pathways to catastrophic risk and warning indicators as an immediate priority. It should then set out plans for developing scalable mitigations. (We set out recommendations on powers and take-down requirements in Chapter 7). The Institute should further set out options for encouraging developers to build systems that are safe by design, rather than focusing on retrospective guardrails.***

Uncontrollable proliferation

143. There is a clear trend towards faster development, release and customisation of increasingly capable open access models.[233] Some can already be trained in just 6 hours and cost a few hundred dollars on public cloud computing platforms.[234]

144. We heard widespread concern about the ease of customisation leading to a rapid and uncontrollable proliferation of models which may be exploited by malicious actors, or contain safety defects affecting businesses and service users.[235]

145. Google DeepMind told us that that "once a model is openly available, it is possible to circumvent any safeguards, and the proliferation of capabilities is

231 Q 24
232 Q 134
233 Written evidence from Hugging Face (LLM0019), Advertising Association (LLM0056) and Meta (LLM0093)
234 Xinyang Geng et al, 'Koala: A Dialogue Model for Academic Research' (April 2023): https://bair.berkeley.edu/blog/2023/04/03/koala/ [accessed 21 December 2023]
235 Q 10 (Ian Hogarth), written evidence from British Copyright Council (LLM0043), Dr Baoli Zhao (LLM0008), Google DeepMind (LLM0095) and IEEE, 'Protesters Decry Meta's "Irreversible Proliferation" of AI' (October 2023): https://spectrum.ieee.org/meta-ai [accessed 21 December 2023]

irreversible."[236] There is no 'undo' function if major safety or legal compliance issues subsequently emerge,[237] and no central registry to determine model provenance once released. It may be possible to embed identifying features in models to help track them, though such research remains at an early stage.[238] The Royal Academy of Engineering emphasised that many models will be hosted overseas, posing major challenges to oversight and regulation.[239]

146. As we set out in Chapter 3, open access models can provide speedy community-led improvements, including to security issues, but those same characteristics can also drive proliferation in malicious use.[240]

147. Closed models are not a security panacea, however. Previous breaches from hack and leak operations, espionage and disgruntled employees suggest that even well-protected systems may not remain closed forever.[241] The Minister said the AI Safety Institute was working on the issues but believed the risks around open access proliferation remained an "extremely complex problem".[242]

148. **There is a credible security risk from the rapid and uncontrollable proliferation of highly capable openly available models which may be misused or malfunction. Banning them entirely would be disproportionate and likely ineffective. But a concerted effort is needed to monitor and mitigate the cumulative impacts. *The AI Safety Institute should develop new ways to identify and track models once released, standardise expectations of documentation, and review the extent to which it is safe for some types of model to publish the underlying software code, weights and training data.***

Existential risk

149. The threat model for existential risk remains highly disputed. A baseline scenario involves the gradual integration of hyper intelligent AI into high-impact systems to achieve political, economic or military advantage, followed by loss of human control. This might occur because humans gradually hand over control to highly capable systems that vastly exceed our understanding; and/or the AI system pursues goals which are not aligned with human welfare and reduce human agency.[243] Humans might also increasingly rely on AI evaluations in high-stakes areas such as nuclear strategy, for example.[244]

---

236 Written submission from Google and Google DeepMind (LLM0095)

237 Centre for the Governance of AI, *Open-Sourcing Highly Capable Foundation Models*: https://cdn.governance.ai/Open-Sourcing_Highly_Capable_Foundation_Models_2023_GovAI.pdf [accessed 21 December 2023]

238 See for example C2PA, *Guidance for Artificial Intelligence and Machine Learning*: https://c2pa.org/specifications/specifications/1.3/ai-ml/ai_ml.html#_attestation_for_ai_ml_models [accessed 21 December 2023].

239 Written evidence from the Royal Academy of Engineering (LLM0063)

240 Q 10 and Q 75

241 See for example Foreign, Commonwealth and Development Office, 'Russia: UK exposes Russian involvement in SolarWinds cyber compromise' (April 2021): https://www.gov.uk/government/news/russia-uk-exposes-russian-involvement-in-solarwinds-cyber-compromise [accessed 8 January 2023].

242 Q 141

243 Q 22 (Professor Stuart Russell) and DSIT, *Capabilities and risks from frontier AI* (October 2023): https://assets.publishing.service.gov.uk/media/65395abae6c968000daa9b25/frontier-ai-capabilities-risks-report.pdf [accessed 21 December 2023]

244 AI in Weapon Systems Committee, *Proceed with Caution: Artificial Intelligence in Weapon Systems* (Report of Session 2023–24, HL Paper 16), paras 157–158

150. Long-term indicative impacts have been compared to outcomes in other fields, including pandemics and nuclear.[245] At the most extreme end, the first- and second-order consequences of uncontrolled nuclear exchange between superpowers have been variously estimated at 2–5 billion fatalities.[246] A biosecurity extinction event might involve above 7 billion fatalities.[247]

151. Systems capable of posing such risks do not yet exist and there is no consensus about their long-term likelihood. Professor Phil Blunsom, Chief Scientist at the LLM firm Cohere, did "not see a strong existential risk from large language models".[248]

152. Professor Stuart Russell OBE argued that "large language models are not on the direct path to the super intelligent system … but they are a piece of the puzzle". He maintained current systems lacked features including "the ability to construct and execute long-term plans, which seems to be a prerequisite" to overcome human resistance, but "could not say with any certainty that it will take more than 20 years" for researchers to address those shortcomings.[249]

153. Some surveys of industry respondents predict a 10 per cent chance of human-level intelligence by 2035, while others say such developments are not likely and do not believe it is a concern.[250] Researchers at the Oxford Internet Institute emphasised that current capabilities were "meaningfully different" to those required for existential risk.[251] Owen Larter, Director of Public Policy at Microsoft's Office for Responsible AI, anticipated a "further maturation of AI safety" in the coming years.[252]

154. This indicates a non-zero likelihood (remote chance) of existential risks materialising, though it is almost certain that these will not occur within the next three years and it seems highly likely that they will not materialise within the next decade. We note the possibility and (longer-term) timing remains a matter of debate and concern for some in the expert community.[253] Several stakeholders suggested concerns about existential risk were distracting from

245 Center for AI Safety, 'Statement on AI risk': https://www.safe.ai/statement-on-ai-risk [accessed 25 January 2024]

246 See 'Global food insecurity and famine from reduced crop, marine fishery and livestock production due to climate disruption from nuclear war soot injection' *Nature Food* (August 2022): https://www.nature.com/articles/s43016–022-00573-0 [accessed 23 December 2023], Cold War estimates of deaths in nuclear conflict', *Bulletin of the Atomic Scientists* (January 2023): https://thebulletin.org/2023/01/cold-war-estimates-of-deaths-in-nuclear-conflict/ [accessed 21 December 2023] and Department of Homeland Security, 'Nuclear Attack' : https://www.dhs.gov/publication/nuclear-attack-fact-sheet [accessed 8 January 2024].

247 Piers Millett et al, 'Existential Risk and Cost-Effective Biosecurity', *Health Security* (August 2017): https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5576214/ [accessed 8 January 2023]

248 Q 22

249 *Ibid.*

250 DSIT, 'Frontier AI: capabilities and risks—discussion paper' (October 2023): https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper/frontier-ai-capabilities-and-risks-discussion-paper [accessed 21 December 2023].

251 Written evidence from the Oxford Internet Institute (LLM0074)

252 Q 74

253 DSIT, *Capabilities and risks from frontier AI* (October 2023): https://assets.publishing.service.gov.uk/media/65395abae6c968000daa9b25/frontier-ai-capabilities-risks-report.pdf [accessed 21 December 2023] and Reuters, 'AI pioneer says its threat to world may be 'more urgent' than climate change' (9 May 2023): https://www.reuters.com/technology/ai-pioneer-says-its-threat-world-may-be-more-urgent-than-climate-change-2023–05-05/ [accessed 24 January 2024]

efforts to address limited but more immediate risks,[254] as well as from the opportunities LLMs may provide.[255]

155. **It is almost certain existential risks will not manifest within three years and highly likely not within the next decade. As our understanding of this technology grows and responsible development increases, we hope concerns about existential risk will decline. The Government retains a duty to monitor all eventualities. But this must not distract it from capitalising on opportunities and addressing more limited immediate risks.**

Societal risks

156. LLMs may amplify any number of existing societal problems, including inequality, environmental harm, declining human agency and routes for redress, digital divides, loss of privacy, economic displacement, and growing concentrations of power.[256]

Bias and discrimination

157. Bias and discrimination are particular concerns, as LLM training data is likely to reflect either direct biases or underlying inequalities.[257] Depending on the use, this might entrench discrimination (for example in recruitment practices, credit scoring or predictive policing); sway political opinion (if using a system to identify and rank news stories); or lead to casualties (if AI systematically misdiagnoses healthcare patients from minority groups).[258] Professor Neil Lawrence cautioned that emergent societal risks could arise in unforeseen ways from mass deployment, as has been the case with social media.[259]

158. Such issues predate LLMs but, as Sense About Science warned, economic logic is driving competition for early adoption of LLMs before adequate guardrails are in place.[260] The Post Office Horizon scandal provides a cautionary tale about the risks of relying on faulty technology systems.[261]

159. We heard that longstanding recommendations remain pertinent: educate developers and users, and embed explainability, transparency, accuracy and accountability throughout the AI lifecycle.[262] This appears particularly difficult for LLMs. They are very complex and poorly understood; operate black-box decision-making; datasets are so large that meaningful

254 Q 55 (Arnav Joshi) and written evidence from Andreessen Horowitz (LLM0114)
255 Written evidence from Kairoi Ltd (LLM0110)
256 See for example Emily M Bender et al, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' (March 2021): https://dl.acm.org/doi/pdf/10.1145/3442188.3445922 [accessed 21 December 2023] and House of Lords Library, 'Artificial intelligence: Development, risks and regulation' (July 2023): https://lordslibrary.parliament.uk/artificial-intelligence-development-risks-and-regulation/ [accessed 8 January 2024].
257 Emily M Bender et al, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' (March 2021): https://dl.acm.org/doi/pdf/10.1145/3442188.3445922 [accessed 21 December 2023]
258 Written evidence from Sense about Science (LLM0046), the Advertising Association (LLM0056), Dr Jeffrey Howard (LLM0049), Society of Authors (LLM0044) and British Copyright Council (LLM0043)
259 Q 3
260 Written evidence from Sense about Science (LLM0046)
261 BBC, 'Post Office scandal explained' (16 January 2024): https://www.bbc.co.uk/news/business-56718036 [accessed 18 January 2024]
262 Written evidence from the Committee on Standards in Public Life (LLM0052), Copyright Clearance Center (LLM0018), Cambridge Language Sciences (LLM0053), DMG Media (LLM0068), Guardian Media Group (LLM0108)

transparency is difficult; hallucinations are common;[263] and accountability remains highly disputed.[264]

160. Irene Solaiman, Head of Global Policy at Hugging Face, said efforts to improve model design and post-deployment practices were underway, but emphasised "how difficult, and frankly impossible, complex social issues are to quantify or to robustly evaluate".[265] Dr Koshiyama, CEO of the audit firm Holistic AI, noted there were limited market incentives to prioritise ethics, and said many earlier AI systems had well-known bias problems but remained in widespread use.[266] Some jurisdictions are introducing mandatory ethics impact assessments.[267] Sam Cannicott, Deputy Director of AI Enablers and Institutions at DSIT, said the AI Safety Institute would examine "societal harms" and would engage professional ethicists in its work.[268]

161. **LLMs may amplify numerous existing societal problems and are particularly prone to discrimination and bias. The economic impetus to use them before adequate guardrails have been developed risks deepening inequality.**

162. ***The AI Safety Institute should develop robust techniques to identify and mitigate societal risks. The Government's AI risk register should include a range of societal risks, developed in consultation with civil society. DSIT should also use its White Paper response to propose market-oriented measures which incentivise ethical development from the outset, rather than retrospective guardrails. Options include using Government procurement and accredited standards, as set out in Chapter 7.***

Data protection

163. LLMs may have personal data in their training sets, drawn from proprietary sources or information online. Safeguards to prevent inappropriate regurgitation are being developed but are not robust.[269]

164. Arnav Joshi, Senior Associate at Clifford Chance, did not believe there was currently widespread non-compliance with data protection legislation but thought "that might happen [… without] sufficient guardrails".[270] He said the General Data Protection Regulation (GDPR) provided "an incredibly powerful tool" to guide responsible innovation, but noted measures in the Data Protection and Digital Information Bill would, if enacted, have

263 Hallucinations refer to the phenomenon of LLMs producing plausible-sounding but inaccurate responses.
264 Written evidence from the Alan Turing Institute (LLM0081) and Royal Society of Statisticians (LLM0055)
265 Q 68
266 Q 67
267 Written evidence from the Committee on Standards in Public Life (LLM0052) and Oxford Internet Institute (LLM0074)
268 Q 136
269 Haoran Li et al, 'Privacy in Large Language Models: Attacks, Defenses and Future Directions' (October 2023): https://arxiv.org/abs/2310.10383 [accessed 8 January 2024]
270 Q 55

a "dilutive effect on rightsholders", for example around rights to contest decisions made by AI.[271]

165. Data protection in healthcare will attract particular scrutiny. Some firms are already using the technology on NHS data, which may yield major benefits.[272] But equally, models cannot easily unlearn data, including protected personal data.[273] There may be concerns about these businesses being acquired by large overseas corporations involved in related areas, for example insurance or credit scoring.[274]

166. Stephen Almond, Executive Director at the Information Commissioner's Office, told us data protection was complex and much depended on who was doing the processing, why, how and where. He said the ICO would "clarify our rules on this and our interpretation of the law to ensure that it is crystal clear".[275]

167. **Further clarity on data protection law is needed.** ***The Information Commissioner's Office should work with DSIT to provide clear guidance on how data protection law applies to the complexity of LLM processes, including the extent to which individuals can seek redress if a model has already been trained on their data and released.***

168. ***The Department for Health and Social Care should work with NHS bodies to ensure future proof data protection provisions are embedded in licensing terms. This would help reassure patients given the possibility of LLM businesses working with NHS data being acquired by overseas corporations.***

271 Written evidence from Arnav Joshi (LLM0112). We noted further concerns from the Public Law Project about the Bill's proposals to "weaken" protections around automated decision-making, as well as uncertainty around the extent to which models 'hold' personal data and hence how far data protection duties apply. See for example Public Law Project, 'How the new Data Bill waters down protections' (November 2023): https://publiclawproject.org.uk/resources/how-the-new-data-bill-waters-down-protections/ [accessed 21 December 2023], and Q 56.

272 Cogstack, 'Unlock the power of healthcare data with CogStack': https://cogstack.org/ [accessed 21 December 2023]

273 Written evidence from the Creators' Rights Alliance (LLM0039)

274 See recent debates on related topics, for example 'Palantir NHS contract doubted by public for data privacy', *The Times* (November 2023): https://www.thetimes.co.uk/article/palantir-nhs-contract-doubted-by-public-for-data-privacy-q9sccsmln [accessed 8 January 2024].

275 Q 86. The ICO already provides extensive guidance on data protection. See for example: Information Commissioner's Office, 'Generative AI: eight questions that developers and users need to ask' (April 2023): https://ico.org.uk/about-the-ico/media-centre/blog-generative-ai-eight-questions-that-developers-and-users-need-to-ask/ [accessed 21 December 2023].

CHAPTER 6: INTERNATIONAL CONTEXT AND LESSONS

International context

169. We examined the extent to which the UK should replicate regulatory approaches adopted by the most influential actors in AI: the US, EU and China.

170. The EU reached initial agreement on its AI Act in December 2023.[276] Supporters believe the legislation will set a global standard for a tiered mitigation of risks, preserving consumer rights and upholding democratic principles. Detractors said it is too prescriptive and risks becoming obsolete as general purpose systems continue to evolve.[277]

171. The US is pursuing a market-driven approach. Dr Mark MacCarthy, Senior Fellow at the Institute for Technology Law and Policy at Georgetown Law, said the US would likely go "beyond voluntary commitments". In his view, this would involve government-stipulated requirements enforced via a "supplemental approach of giving existing regulators more authority".[278]

172. China's approach may be characterised as 'security first'. Paul Triolo, Senior Associate with the Trustee Chair in Chinese Business and Economics at the Center for Strategic and International Studies, said China took a positive attitude to technological progress and had recently shifted regulatory oversight into "overdrive" to ensure generative AI delivered against the Chinese Communist Party's strategic objectives. This included rapid iterative measures (for example on watermarking, the quality of data inputs and accuracy of model outputs) to provide businesses with initial direction, followed by stricter codified rules.[279]

173. The Government could learn lessons from the US vision for context-specific regulation, the EU's objectives to mitigate high-impact risks, and China's positive attitude to technological adoption while addressing its societal and security concerns at pace.[280] But wholesale replication of their regulatory approaches appeared unwise: the UK lacks the distinctive features that shape the their positions—such as the EU's customer base and appetite for regulatory heft; American market power; and China's political objectives.[281]

174. Katherine Holden of techUK said the UK should continue to pursue its own regulatory pathway which is "proportionate, risk-based and outcomes-focused".[282] Many others such as the Startup Coalition and Google DeepMind offered similar views. [283] As the Alan Turing Institute emphasised, being proactive in delivering this "middle of the road" approach would mean the UK is "better placed to advocate for those policies globally,

276 Council of the EU, 'Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world' (December 2023): https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/ [accessed 8 January 2024]
277 Written evidence from the Startup Coalition (LLM0089), AGENCY (LLM0028) and Q 50
278 Q 48
279 Q 49 and written evidence from Dr Xuechen Chen (LLM0031)
280 Q 31, Q 50, written evidence from the Open Data Institute (LLM0083), Matthew Feeney (LLM0047) and AGENCY (LLM0028)
281 Q 50, written evidence from the Alan Turing Institute (LLM0081) and Startup Coalition (LLM0089)
282 Q 38
283 Written evidence from Google and Google DeepMind (LLM0095)

which will in turn generate further credibility and support for the UK's domestic AI ecosystem".[284]

175. ***The UK should continue to forge its own path on AI regulation, balancing rather than copying the EU, US or Chinese approaches. In doing so the UK can strengthen its position in technology diplomacy and set an example to other countries facing similar decisions and challenges.***

176. International co-ordination will be key, but difficult. We found substantial support for the Government's work to convene global stakeholders, including China,[285] and for its efforts to create a shared approach to risks.[286] Competing priorities, agendas and forums suggest however that global regulatory divergence is more likely than convergence in the short- to medium-term.[287]

177. We found support for further international co-ordination,[288] perhaps involving a convening body modelled on other sectors like nuclear or aviation.[289] Professor McDermid thought greater co-ordination would be valuable but warned that the UK would fall "far behind the curve" if it waited for international consensus without progressing domestic action first.[290]

178. **International regulatory co-ordination will be key, but difficult and probably slow. Divergence appears more likely in the immediate future. We support the Government's efforts to boost international co-operation, but it must not delay domestic action in the meantime.**

Lessons for regulation

179. We further explored the case for comprehensive primary legislation relating specifically to foundation models. (Wider legislation on AI governance was beyond the scope of this inquiry).[291]

180. Professor Anu Bradford, Professor of Law and International Organisation at Columbia Law School, advocated starting early, arguing that developers should not have a "free pass". She acknowledged challenges around regulating fast-moving technical issues, but said medical and airline regulations showed

284 Written evidence from the Alan Turing Institute (LLM0081)
285 Q 50 (Professor Bradford)
286 Written evidence from Google and Google DeepMind (LLM0095) and Alan Turing Institute (LLM0081)
287 Written evidence from Dr Xuechen Chen, Dr Xinchuchu Gao and Dr Lingpeng Kong (LLM0031), Oxford Internet Institute (LLM0074), Open Data Institute (LLM0083) and Q 70 (Professor John McDermid)
288 Written evidence from Google and Google DeepMind (LLM0095), Microsoft (LLM0087) and Alan Turing Institute (LLM0081)
289 Q 26 (Professor Stuart Russell OBE). See also 'Is it possible to regulate artificial intelligence', *BBC* (September 2023): https://www.bbc.co.uk/news/business-66853057 [accessed 21 December 2023]. The Government has committed to supporting a 'State of the Science' report on AI, see for example DSIT, 'State of the Science report' (2 November 2023): https://www.gov.uk/government/publications/ai-safety-summit-2023-chairs-statement-state-of-the-science-2-november/state-of-the-science-report-to-understand-capabilities-and-risks-of-frontier-ai-statement-by-the-chair-2-november-2023 [accessed 21 December 2023].
290 Q 70
291 For a review of wider AI governance see Artificial Intelligence Committee, *AI in the UK: ready, willing and able* (Report of Session 2017–2019, HL Paper 100) and Science, Innovation and Technology Committee, *The governance of artificial intelligence: interim report* (Ninth Report, Session 2022–23, HC 1769).

it was possible.[292] Arnav Joshi of Clifford Chance noted the EU's work on legislation had begun in 2019 and would not take effect until around 2025.[293]

181. Owen Larter, Director of Public Policy at Microsoft's Office for Responsible AI, advocated tiered regulation with different requirements for each layer of the technology stack.[294] The Glenlead Centre supported legislation, arguing that its absence would make the UK a "rule-taker" as businesses comply with more stringent rules set by other countries.[295]

182. Others were more cautious. Mind Foundry, a software firm, warned that "ill-conceived and strict regulation" would hamper opportunities.[296] The Oxford Internet Institute identified some areas where primary legislation would help, but noted greater clarity on standards and regulatory gaps was needed.[297]

183. Rachel Coldicutt OBE of Careful Industries thought getting regulation right would be difficult: moving quickly risks poor rules which lead to chilling effects, while waiting for harms to emerge and legislating retrospectively may involve years-long processes to develop an overly complex regime that attempts to unpick entrenched business models.[298] She cited the progress of the Online Safety Act as a cautionary tale, and advocated instead stronger Government-led strategic direction backed up by forward-looking measures to prevent harm and incentivise responsible innovation.[299]

184. We noted numerous other lessons to inform LLM oversight, though no system could be replicated wholesale. Medicine has a robust system of phased discovery trials and closely supervised release,[300] though the Government Chief Scientific Adviser said we did not yet have AI tests that would approximate even first-stage trials.[301] Dr Koshiyama pointed to the financial sector's ongoing self-assessments against clear benchmarks as a helpful yardstick.[302]

185. Professor McDermid said aviation showed that high-stakes software can be made in ways that are safe, interpretable and internationally co-ordinated.[303] Data protection law has shown the viability of tiered penalties, as well as the risks of 'one-size-fits-all' approaches disproportionately burdening small businesses.[304] Health and safety laws have proved remarkably durable.[305] Digital markets show the value of acting ahead of time before damaging practices become normalised.[306]

186. The Government told us that legislation had not been ruled out.[307] The Minister had no "philosophical objection" and anticipated "binding

292 Q 47
293 Written evidence from Arnav Joshi (LLM0112)
294 Q 76
295 Written evidence from the Glenlead Centre (LLM0051)
296 Written evidence from Mind Foundry (LLM0030)
297 Written evidence from the Oxford Internet Institute (LLM0074)
298 Written evidence from Careful Industries (LLM0041)
299 *Ibid.*
300 See for example The Medicines for Human Use (Clinical Trials) Regulations 2004 (SI 2004/1031).
301 Q 128
302 Q 72 (Dr Adriano Koshiyama)
303 Q 70 (Professor John McDermid)
304 Q 47 (Professor Anu Bradford) and QQ 55–57 (Arnav Joshi)
305 Written evidence from Carnegie UK (LLM0096)
306 Written evidence from Stability AI (LLM0078)
307 Q 139 (Lizzie Greenhalgh)

requirements" at some point in future, but emphasised the Government's current focus on a non-statutory approach to enable flexible and reactive progress.[308]

187. **Extensive primary legislation aimed solely at LLMs is not currently appropriate: the technology is too new, the uncertainties too high and the risk of inadvertently stifling innovation too great. Broader legislation on AI governance may emerge in future, though this was outside the scope of our inquiry.** ***Setting the strategic direction for LLMs and developing enforceable, pro-innovation regulatory frameworks at pace should remain the Government's immediate priority.***

308 Q 143

CHAPTER 7: MAKING THE WHITE PAPER WORK

188. The Government's White Paper aims to bring "clarity and coherence" to AI regulation. It relies substantially on existing regulators to deliver this complex task, rather than establishing a new overarching AI regulator.[309] Many stakeholders have raised concerns about a patchwork of disjointed rules, gaps, definitions, overlapping remits, and inconsistent enforcement emerging from the UK's 90 or so regulators of varying size, heft and expertise.[310]

189. The White Paper committed to setting up Government-led "central functions" to provide support, co-ordination and coherence. It said many stakeholders preferred this to a new AI regulator. Key areas for the central functions include:

- monitoring, assessment and feedback;
- supporting coherent implementation of the principles;
- cross-sector risk assessment;
- horizon scanning;
- supporting innovators (including testbeds and sandboxes);
- education and awareness; and
- international interoperability.[311]

Where are the central functions?

190. Regulators will need to navigate issues of immense complexity, uncertainty and importance with technologies developing at an unprecedented rate. The Ada Lovelace Institute and techUK emphasised that the central function teams were key to the White Paper's success and believed it was critical for them to be well resourced.[312] Dr Florian Ostmann, Head of AI Governance and Regulatory Innovation at the Alan Turing Institute, said the central function co-ordination teams were particularly important to ensure challenging issues did not fall between gaps in regulators' remits.[313]

191. Speed will be key. Numerous contributors emphasised the importance of providing clear guidelines quickly and iteratively.[314] This would encourage good practice early on, prevent harmful business models from becoming

309 DSIT, ' A pro-innovation approach to AI regulation' (August 2023): https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [accessed 8 January 2024]

310 See for example Public Law Project, *Public Law Project response to the AI White Paper consultation* (June 2023): https://publiclawproject.org.uk/content/uploads/2023/06/Public-Law-Project-AI-white-paper-consultation-response.pdf [accessed 8 January 2024], Taylor Wessing, 'The UK's approach to regulating AI': (May 2023): https://www.taylorwessing.com/en/interface/2023/ai---are-we-getting-the-balance-between-regulation-and-innovation-right/the-uks-approach-to-regulating-ai [accessed 8 January 2024] and Ada Lovelace Institute, 'Regulating AI in the UK: three tests for the Government's plans' (June 2023): https://www.adalovelaceinstitute.org/blog/regulating-ai-uk-three-tests/ [accessed 8 January 2024].

311 DSIT, ' A pro-innovation approach to AI regulation' (August 2023): https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [accessed 8 January 2024]

312 Q 41

313 *Ibid.*

314 Written evidence from Royal Academy of Engineering (LLM0063), Alan Turing Institute (LLM0081), Startup Coalition (LLM0089) and Carnegie UK (LLM0096)

entrenched and minimise longer-term disputes about the subsequent cost of retrospective compliance.[315]

192. However, progress in Government seems slow. Regulators in our evidence session in November 2023 did not appear to know what was happening with the teams proposed in the March White Paper to provide cross-regulator co-ordination and support. Dr Yih-Choung Teh, Group Director of Strategy and Research at Ofcom, remained unclear what "shape that will take". Stephen Almond, Executive Director of Regulatory Risk at the Information Commissioner's Office, suggested regulators were "keen to see progress".[316]

193. Our review of ten regulators' staffing suggests significant variation in technical expertise,[317] which further underscores the need for support from the Government's central functions:

Table 2: Indicative staffing overview

Regulator	Specialised staff (full time equivalent)	Future plans (full time equivalent)
Office of Communications (Ofcom)	60 data scientists and machine learning experts 0 dedicated AI governance staff, though current related work draws on 20+ staff	Currently recruiting
Information Commissioner's Office (ICO)	9 on AI governance, with "a much larger number" working on "AI-related issues"	Under review
Equality and Human Rights Commission (EHRC)	0 AI governance specialists 0 data scientists	Desire for internal data science capacity but limited funding to do so
Competition & Markets Authority (CMA)	9 data scientists and 3 data engineers, supported by 20 technologists 0 AI governance specialists but numerous staff involved in AI initiatives	3 further AI specialists, 5 further data scientitsts and 3 data engineers

315 Written evidence from Careful Industries (LLM0041), Carnegie UK (LLM0096). Some guidance is emerging already. See for example Medicines and Healthcare products Regulatory Agency, 'Large Language Models and software as a medical device' (3 March 2023): https://medregs.blog.gov.uk/2023/03/03/large-language-models-and-software-as-a-medical-device/ [accessed 26 January 2023] and Information Commissioner's Office, 'Generative AI: eight questions that developers and users need to ask' (3 April 2023): https://ico.org.uk/about-the-ico/media-centre/blog-generative-ai-eight-questions-that-developers-and-users-need-to-ask/ [accessed 26 January 2023].

316 QQ 93–94

317 See correspondence from regulators, available at Communications and Digital Committee, 'Correspondence': https://committees.parliament.uk/work/7827/large-language-models/publications/3/correspondence/.

Medicines and Healthcare products Regulatory Agency (MHRA)	1.5 AI governance specialists and 2 data scientists	3 further AI specialists and 14 roles in Digital & Technology
Office of Qualifications and Examinations Regulation (Ofqual)	0 AI governance specialists, but a range of data experts	Under review
Bank of England (BoE), Prudential Regulation Authority PRA)	1.5 on AI regulation, supported by a large working group 82 data scientists (mostly in Monetary Policy and in the PRA), plus additional machine learning experts	No plans
Financial Conduct Authority (FCA)	75+ data scientists, 3 staff in the AI Lab, supported by others from other sectors 9 staff on regulatory and digital sandboxes which has an increasing AI focus. 5 staff on emerging technology	No plans
Solicitors Regulation Authority (SRA)	0 AI governance staff, 3 data scientists	Under review
Advertising Standards Agency (ASA)	0 AI governance specialists, 5 data scientists	2 data scientists in 2024

194. In response to our request for further details, the Department said that the central functions totalled 23 staff, of which 10 were dedicated to evaluating risk and 13 to AI analysis, regulatory co-ordination and delivery. The minister said this work was complemented by the Centre for Data, Ethics and Innovation and the AI Standards Hub.[318]

195. **We support the overall White Paper approach. But the pace of delivering the central support functions is inadequate. The regulatory support and co-ordination teams proposed in the March 2023 White Paper underpin its entire success. By the end of November 2023, regulators were unaware of the central function's status and how it would operate. This slowness reflects prioritisation choices and undermines confidence in the Government's commitment to the regulatory structures needed to ensure responsible innovation.**

196. ***DSIT should prioritise resourcing the teams responsible for regulatory support and co-ordination, and publish an update on staffing and policy progress in response to this report.***

318 Q 135

Do the regulators have what it takes?

197. We wrote to ten regulators seeking information on their level of preparedness to deliver on the White Paper objectives.[319] We found a significant variation in remits, powers, resource and expertise.

198. Some, notably Ofcom and the ICO, acknowledged the scale of the challenge and appeared relatively well resourced to respond. The Medicines and Healthcare products Regulatory Agency (MHRA) said it had growing capacity gaps relative to the scale of demand. The Equality and Human Rights Commission is expected to face mounting difficulties around bias issues, but has no AI governance experts and insufficient funding to pursue legal remedies. The lack of expertise to conduct technical audits was a recurring theme across regulators, as were gaps in powers to gather information from developers and interrogate AI in its working context.[320]

199. We also found significant variation in regulators' sanctioning powers, suggesting enforcement on similar types of problems caused by LLMs could vary considerably across sectors. The National Union of Journalists believed there was insufficient focus on ensuring regulatory requirements are backed up by meaningful sanctions to deter wrongdoing.[321]

200. The Royal Academy of Engineering said numerous cross-cutting LLM-related issues were not the direct responsibility of any regulator, for example accuracy, interpretability, and bias. It suggested forthcoming sector-specific codes should be accompanied by cross-cutting guidelines too.[322] We further noted there were numerous different co-ordination forums involving different regulators, suggesting there would be some value in further coherence brought by the central function co-ordination team.

201. **Relying on existing regulators to ensure good outcomes from AI will only work if they are properly resourced and empowered.** ***The Government should introduce standardised powers for the main regulators who are expected to lead on AI oversight to ensure they can gather information relating to AI processes and conduct technical, empirical and governance audits. It should also ensure there are meaningful sanctions to provide credible deterrents against egregious wrongdoing.***

202. ***The Government's central support functions should work with regulators at pace to publish cross-sector guidance on AI issues that fall outside individual sector remits.***

Liability

203. We heard conflicting views on the extent to which regulators could and should be able to hold upstream developers to account. The Alan Turing Institute outlined the "many hands" problem, where the number of parties involved in LLMs and extent of possible uses makes liability attribution difficult.[323]

319 See Communications and Digital Committee, 'Correspondence': https://committees.parliament.uk/work/7827/large-language-models/publications/3/correspondence/.
320 Written evidence from the Solicitors Regulation Authority (LLM0106)
321 Written evidence from the National Union of Journalists (LLM0007)
322 Written evidence from the Royal Academy of Engineering (LLM0063)
323 Written evidence from the Alan Turing Institute (LLM0081)

The number of actors involved with open access models introduces further complexity.[324]

204. Upstream developers have greatest insight into and control over the base model, and typically specify acceptable use policies.[325] Dr Nathan Benaich of Air Street Capital said their responsibility for subsequent downstream use remained a "grey zone", particularly if models were customised in inappropriate ways.[326] Rob Sherman of Meta believed there had to be responsibility at "every level of the chain".[327] Microsoft said developers would "not be in a position to mitigate the risks of the many different downstream use cases of which they will have little visibility".[328]

205. Downstream actors may however lack sufficient information to be confident of their liabilities. Dr Zoë Webster, Director of Data and AI Solutions at BT, said she was concerned that:

> "we will be held accountable … for issues with a foundation model where we have no idea what data it was trained on, how it was tested and what the limitations are on how and when it can be used. There are open questions and that is a limiting factor on adoption".[329]

206. Professor McDermid noted that liability ultimately lies with the manufacturer in safety-critical industries like aviation, unless a downstream customer has erred (for example through faulty maintenance).[330] He thought the issue with LLMs was not directly comparable, though it remained "far too complex to transfer liability to the user". He suggested the complexities around mid-tier customisation of models should be referred to the Law Commission for an authoritative review.[331]

207. Michael Birtwistle of the Ada Lovelace Institute said the White Paper focused on AI use rather than development, and that regulators had limited capacity to address the source of problems in upstream developers.[332] Poor data labelling by developers may create downstream bias issues, for example.[333]

208. Our discussion with regulators suggested the issue remained complex, context-dependent and in many cases unclear.[334] The ICO believed they could operate across the "entirety of the value chain".[335] The EHRC thought likewise. (In practice this might involve attempting to obtain information on the base model via an intermediary service provider and it remains unclear how successful this would be).[336] Ofcom said it focused more on downstream services.[337] The Minister said liability was "one of the areas that the [AI

---

324 Written evidence from the Oxford Internet Institute (LLM0074)
325 Written evidence from BT Group (LLM0090)
326 Q 15
327 Q 76
328 Written evidence from Microsoft (LLM0087)
329 Q 17
330 Q 70
331 *Ibid.*
332 Q 38
333 Written evidence from the Alan Turing Institute (LLM0081)
334 Q 86
335 *Ibid.*
336 See for example Equality Act 2010, section 29.
337 Q 86

Safety Institute] is looking into to give us the evidence and opinion to guide our approach".[338]

209. **Model developers bear some responsibility for the products they are building—particularly given the foreseeable risk of harm from misuse and the limited information available to customers about how the base model works. But how far such liability extends remains unclear. *The Government should ask the Law Commission to review legal liability across the LLM value chain, including open access models. The Government should provide an initial position, and a timeline for establishing further legal clarity, in its White Paper response.***

High-risk high-impact testing

210. In June 2023 the Prime Minister said that Google DeepMind, OpenAI and Anthropic had agreed to provide early access to their models "for research and safety purposes".[339] This was followed by pledges at the AI Safety Summit for "increased emphasis on AI safety testing and research", led in the UK by the AI Safety Institute.[340]

211. In October the White House published an executive order on AI safety for the US, which moved from voluntary commitments to mandatory requirements for sharing safety testing information before "the most powerful AI systems" are made public.[341]

212. Many calls for further action on testing regimes have been led by large tech firms themselves,[342] though others have highlighted options for going further too. The Law Society has argued for a regime that combines adaptable regulations with firmer requirements "focusing on inherently high-risk contexts and dangerous capabilities".[343] Dr Baoli Zhao, an AI entrepreneur, said the White Paper should have given greater consideration to "mandatory compliance testing".[344] Martin Hosken, an industry expert, cautioned against the types of "over-regulation [that] makes people's lives more difficult" but highlighted mandatory impact assessments and heightened auditing as options to consider.[345]

---

338 Q 140

339 Prime Minister Rishi Sunak, speech given at London Tech Week, 12 June 2023: https://www.gov.uk/government/speeches/pm-london-tech-week-speech-12-june-2023 [accessed 8 January 2024]

340 DSIT, 'Safety Testing: Chair's Statement of Session Outcomes, 2 November 2023' (November 2023): https://www.gov.uk/government/publications/ai-safety-summit-2023-chairs-statement-safety-testing-2-november/safety-testing-chairs-statement-of-session-outcomes-2-november-2023 [accessed 8 January 2024]

341 The White House, 'Fact sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence' (October 2023): https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ [accessed 8 January 2024]

342 Q 76 (Owen Larter), Bloomberg, 'OpenAI backs idea of requiring licences for advanced AI systems' (20 July 2023): https://www.bloomberg.com/news/articles/2023–07-20/internal-policy-memo-shows-how-openai-is-willing-to-be-regulated [accessed 8 January 2024], see also written evidence from Microsoft (LLM0087) and Google and Google DeepMind (LLM0095).

343 The Law Society, 'A pro-innovation approach to AI regulation – Law Society response' (June 2023): https://www.lawsociety.org.uk/campaigns/consultation-responses/a-pro-innovation-approach-to-ai-regulation [accessed 8 January 2024]

344 Written evidence from Dr Baoli Zhao (LLM0008)

345 Written evidence from Martin Hosken (LLM0009)

213. The risk profile of the most powerful models suggests further safeguards may indeed be needed as next-generation capabilities come online.[346] Advanced capabilities to plan and execute tasks autonomously through external tools might be a particular concern.[347] We welcomed the Government's initial progress on engaging tech firms but were not convinced voluntary agreements would suffice in the long-term. The recent furore around OpenAI's governance showed that the tech leaders with whom the Government strikes deals can change overnight,[348] and their successors may not be likeminded. The scale of controversy and litigation in technology around the world over the past 25 years suggests the current period of constructive engagement between governments and tech firms is unlikely to last forever.[349]

214. It would also be naïve to assume that high-risk high-impact models will be developed only in countries like the US, where the UK can draw on goodwill and longstanding relationships. The Minister acknowledged there were no safety testing agreements with Chinese firms, for example, though that country is likely to produce highly capable models.[350]

215. Further, there does not appear to be a clear set of tools and powers to compel a business to comply with Government recommendations on pre-release safety requirements. What happens if a highly risky model is released (including in open access format) remains unclear. The Minister suggested developers might break an existing rule and trigger some form of sanction.[351] However, the current absence of benchmarks with legal standing and lack of clarity on liability suggests there are limited options to issue market recall directives to the developer, or platform take-down notices at websites hosting dangerous open access models.[352] Some bodies have comparable powers (for example the Health and Safety Executive) but none appears designed to address the scale and cross-cutting nature of LLMs.[353] The Minister noted any gaps would be "a piece of evidence" supporting further regulatory action.[354]

---

346 'OpenAI chief seeks new Microsoft funds to build 'superintelligence'', *Financial Times* (November 2023): https://www.ft.com/content/dd9ba2f6-f509-42f0-8e97-4271c7b84ded [accessed 8 January 2024], DSIT, 'The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023' (November 2023): https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 [accessed 8 January 2024] and Humza Naveed et al, *A Comprehensive Overview of Large Language Models* (July 2023): https://arxiv.org/pdf/2307.06435.pdf [accessed 27 December 2024]. See also Chapter 5 on risk.

347 Centre for Security and Emerging Technology, *Skating to where the puck is going* (October 2023): https://cset.georgetown.edu/wp-content/uploads/Frontier-AI-Roundtable-Paper-Final-2023CA004-v2.pdf [accessed 5 January 2024]

348 Roberto Tallarita, Harvard Business Review, 'AI Is Testing the Limits of Corporate Governance' (December 2023): https://hbr.org/2023/12/ai-is-testing-the-limits-of-corporate-governance [accessed 8 January 2024]

349 See for example 'As Google Turns 25, It Faces The Biggest Tech Antitrust Trial Of A Generation', *Forbes* (September 2023): https://www.forbes.com/sites/richardnieva/2023/09/11/google-antitrust-trail-25th-birthday/?sh=502ac98910e4 [accessed 8 January 2024] and 'Why it is becoming easier to sue Big Tech in the UK', *BBC News* (January 2023): https://www.bbc.co.uk/news/technology-64210531 [accessed 8 January 2024].

350 Q 141

351 Q 139

352 Written evidence from Reset (LLM0042)

353 Health and Safety Executive, 'HSE's role as a market surveillance authority': https://www.hse.gov.uk/work-equipment-machinery/hse-role-market-surveillance-authority.htm [accessed 8 January 2024] and Medicines & Healthcare products Regulatory Agency, 'Homepage': https://www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency [accessed 8 January 2024]

354 Q 139

216. Defining the criteria for what counts as a high-risk high-impact model will be difficult, as will deciding what an acceptable boundary is for passing any tests. Avoiding onerous red tape and market barriers would be key. Scope could be determined by model size, compute power, cost, general capability or risk-specific capability. None is a perfect predictor and capability is likely the key (if most challenging) metric.[355] A combination of factors which evolves in line with technology may prove best.[356]

217. Dr Adriano Koshiyama said that agreeing the pass or fail rate for safety tests would be challenging,[357] particularly if the skills to create safeguards lie in upstream developers but the societal and legal liability costs are largely borne by downstream users.[358] OpenAI said that safety benchmarks and guardrails were among its research priorities.[359] Bringing a wide range of actors including civil society into such discussions will be important in ensuring the benchmarks are fair and avoid the concerns around regulatory capture outlined in earlier chapters.[360]

218. **We welcome the commitments from model developers to engage with the Government on safety. But it would be naïve to believe voluntary agreements will suffice in the long-term as increasingly powerful models proliferate across the world, including in states which already pose a threat to UK security objectives.**

219. ***The Government should develop mandatory safety tests for high-risk high-impact models. This must include an expectation that the results will be shared with the Government (and regulators if appropriate), and clearly defined powers to require compliance with safety recommendations, suspend model release, and issue market recall or platform take-down notices in the event of a credible threat to public safety.***

220. ***The scope and benchmarks for high-risk high-impact testing should involve a combination of metrics that can adapt to fast-moving changes. They should be developed by the AI Safety Institute through engagement with industry, regulators and civil society. It is imperative that these metrics do not impose undue market barriers, particularly to open access providers.***

Accredited standards and auditing practices

221. A clear pathway to better standards and auditing practices is crucial. These will underpin much of the work needed to incentivise, stipulate and (where necessary) enforce good practice across very different types of business

355 There are numerous ways of evaluating capability already, and extensive work is ongoing. See for example Dan Hendrycks et al, 'Measuring Massive Multitask Language Understanding' (January 2021): https://arxiv.org/abs/2009.03300 [accessed 8 January 2024], Papers With Code, 'Arithmetic Reasoning on GSM8K: https://paperswithcode.com/sota/arithmetic-reasoning-on-gsm8k [accessed 8 January 2024] and Papers With Code, 'HumanEval': https://paperswithcode.com/dataset/humaneval [accessed 8 January 2024].

356 Written evidence from the Oxford Internet Institute (LLM0074)

357 Q 70

358 *Ibid.*

359 Written evidence from OpenAI (LLM0113)

360 Written evidence from Andreessen Horowitz (LLM0114)

model across upstream developers and downstream service providers.[361] The Royal Academy of Engineering said the UK had a "major" opportunity to lead the way.[362]

222. We heard it would be impractical and undesirable for sector regulators to directly evaluate all models and uses. Equally, LLM technology is developing at an unprecedented rate and the lack of ongoing assessment carries safety and societal risks. Grey areas will also dent business confidence.[363] An accredited system of technical and regulatory standards would clarify what good looks like, while accredited auditing practices would enable businesses to check and showcase their good practice.[364] Potential benefits of progress in this space include:

    - business confidence: businesses would have a clear set of guidelines to follow and reduce legal risk of trialling new products. They would also be able to demonstrate good practice certification when bidding for contracts in high-stakes industries, while non-technical clients would have more confidence in what they are getting;

    - incentives: the Government could use its procurement market to encourage good practice by requiring high standards. This could help de-risk public sector use while simultaneously shaping good business practices;

    - a new commercial sector: a 2021 review published by the Centre for Data Ethics and Innovation found that the AI assurance market was likely to grow significantly and that the UK should take advantage, drawing on its strengths in tech, legal and professional services;[365]

    - regulatory enforcement: a common set of good auditing practices would provide regulators with the toolkit to investigate and address malpractice with confidence;

    - liability: accredited standards would help determine expectations and assign liability across complex value chains; and

    - public trust: the adoption of LLMs in some industries is likely to follow the speed of public trust. Demonstrating that LLMs can be built, used

361 Business models and practices vary significantly across LLMs, including how they monetise the models, whether the data used is proprietary or scraped from the internet, and what the model may be used for. Clear but nuanced guidance will be key. See the CDEI Innovation, *Industry Temperature Check* (December 2022): https://assets.publishing.service.gov.uk/media/638f3af78fa8f569f7745ab5/Industry_Temperature_Check_-_Barriers_and_Enablers_to_AI_Assurance.pdf [accessed 20 December 2023].

362 Written evidence from the Royal Academy of Engineering (LLM0063)

363 Written evidence from Hugging Face (LLM0019) and Bright Initiative (LLM0033)

364 Written evidence from the British Standards Institution (LLM0111)

365 CDEI, 'The roadmap to an effective AI assurance ecosystem' (December 2021): https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem/the-roadmap-to-an-effective-ai-assurance-ecosystem [accessed 8 January 2024]

and audited in ways familiar to other products would help alleviate concern about the proliferation of increasingly powerful tools.[366]

223. Much of the initial groundwork is in place. The Centre for Data Ethics and Innovation has an AI Assurance Programme, which recently highlighted the need for better ways to navigate the complex standards landscape and industry "desire for certification or accreditation schemes".[367] Regulators are already expected to develop codes and expand sandboxes to support innovators.[368] Work is underway to apply existing standards to LLMs, and work out what new ones would look like.[369] These might cover governance, data provenance and protection, bias, security, incident reporting, watermarking, interpretability and appropriate use.[370] BT Group suggested requiring a standardised model card which summarises relevant information to help deployers understand how to use the base model appropriately.[371]

224. The UK's AI Standards Hub provides a forum for bringing standards together, while bodies such as the British Standards Institute and UK Accreditation Service provide assurance on the quality of standards and accreditation pathways, including for regulator-led schemes like the ICO's age appropriate design framework.[372] We noted the importance of working at pace given the complexity of issues and number of actors involved (see Figure 9 below).

366 CDEI, 'The roadmap to an effective AI assurance ecosystem' (December 2021): https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem/the-roadmap-to-an-effective-ai-assurance-ecosystem [accessed 8 January 2024], CDEI and the DSIT, 'CDEI portfolio of AI assurance techniques' (June 2023): https://www.gov.uk/guidance/cdei-portfolio-of-ai-assurance-techniques [accessed 8 January 2024], Q 116, written evidence from IEEE Standards Association (LLM0072), BT Group (LLM0090), Arnav Joshi (LLM0112), Local Government Association (LLM0048) and Policy Connect (LLM0065)

367 CDEI, *Industry Temperature Check*

368 Regulatory sandboxes enable innovators to test products with close supervision and access to regulatory expertise. See Department of Science, Innovation and Technology, 'New advisory service to help businesses launch AI and digital innovations' (September 2023): https://www.gov.uk/government/news/new-advisory-service-to-help-businesses-launch-ai-and-digital-innovations [accessed 8 January 2024].

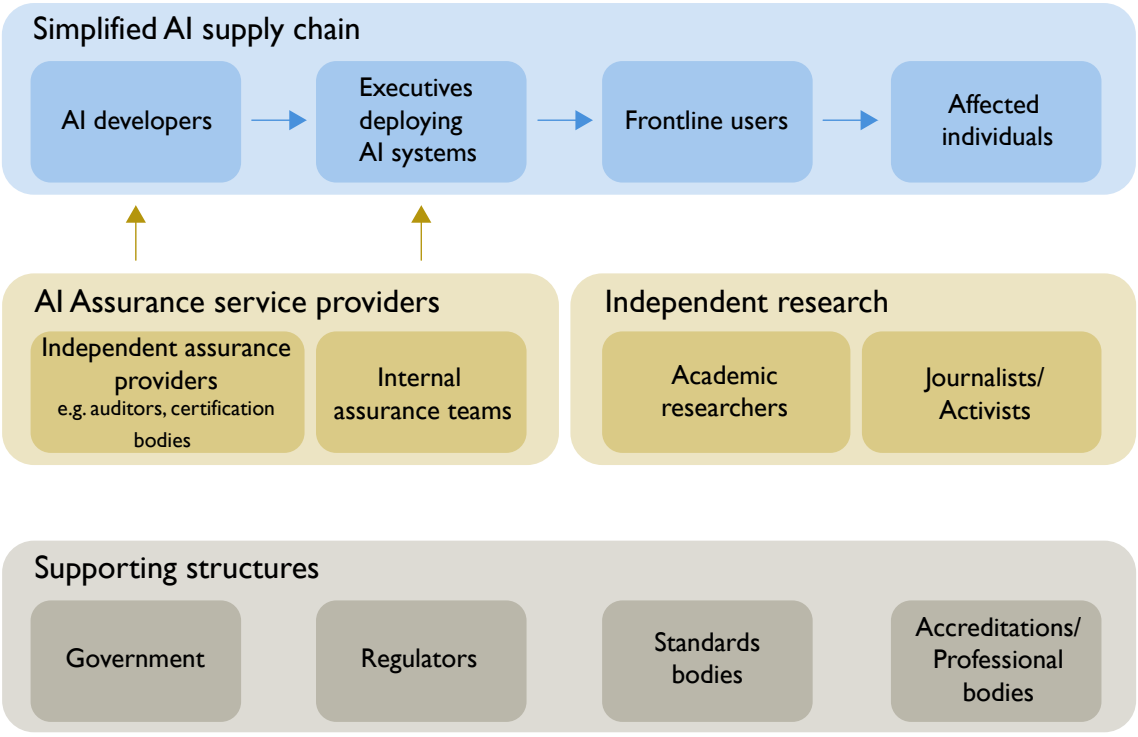
369 Written evidence from the Royal Academy of Engineering (LLM0063) and IEEE (LLM0072), See also the Ethical Black Box standard.

370 Written evidence from the IEEE (LLM0072), RAE (LLM0063), British Standards Institution (LLM0111), Hugging Face (LLM0019) and BT (LLM0090)

371 Written evidence from BT Group (LLM0090). Model cards are a type of documentation used in AI to provide information about a model. See for example Hugging Face, 'Model Cards': https://huggingface.co/docs/hub/model-cards [accessed 8 January 2024].

372 UKAS, 'Digital Sector Accreditation': https://www.ukas.com/accreditation/sectors/digital/ [accessed 8 January 2024], UKAS, 'Homepage': https://www.ukas.com/ [accessed 8 January 2024] and ICO, 'Age Appropriate Design Certification Scheme' (July 2021): https://ico.org.uk/for-organisations/advice-and-services/certification-schemes/certification-scheme-register/age-appropriate-design-certification-scheme-aadcs/ [accessed 8 January 2024]

Figure 9: Key actors in the AI assurance ecosystem



Source: Centre for Data Ehics and Innovation, The roadmap to an effective AI assurance ecosystem (8 December 2021): https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem/the-roadmap-to-an-effective-ai-assurance-ecosystem [accessed 21 December 2023]

225. Progress on standards will help inform decisions on what audits of LLMs should cover and how they should be conducted.[373] Accredited private sector auditors could provide AI assurance in ways similar to the financial sector. This would also deepen the pool of experts available for regulators to draw on too.[374] Hayley Fletcher, a Director at the Competition and Markets Authority, highlighted the importance of audits led both by regulators and third parties, and said the Digital Regulation Co-operation Forum was making progress on auditing practices.[375]

226. **Accredited standards and auditing practices are key. They would help catalyse a domestic AI assurance industry, support business clarity and empower regulators. *We urge the Government and regulators to work with partners at pace on developing accredited standards and auditing practices for LLMs (noting that these must not be tick-box exercises). A consistent approach to publishing key information on model cards would also be helpful.***

227. ***The Government should then use the public sector procurement market to encourage responsible AI practices by requiring bidders to demonstrate compliance with high standards when awarding relevant contracts.***

373 Digital Regulation Cooperation Forum, 'Auditing algorithms: the existing landscape, role of regulators and future outlook' (September 2023): https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook [accessed 8 January 2024]

374 Q 72 and written evidence from Holistic AI (LLM0010)

375 Q 89

CHAPTER 8: COPYRIGHT

228. Many contributors to our inquiry contended that LLM developers were acting unethically and unlawfully by using copyrighted data to train models without permission.[376] Developers disagreed, citing the societal value of their products and the legal exemptions. We examined the balance of evidence and ways forward.

Background on data mining

229. Text and data mining (TDM) involves accessing and analysing large datasets to identify patterns and trends to train AI. Obtaining permission for this typically involves acquiring a licence or relying on an exception. Non-commercial research is permitted. In 2022 the Intellectual Property Office (IPO) proposed to change this system to allow any form of commercial mining. Our report on the creative industries noted the £108 billion sector relied on copyright protections and criticised the IPO's plans for undercutting business models.[377] The Government's response confirmed it would no longer pursue a "broad copyright exception" and set up a working group to develop a new code of practice by "summer" 2023.[378] A separate creative industries strategy published in June 2023 emphasised the Government's continued commitment "to promote and reward investment in creativity" and ensure rightsholder content is "appropriately protected" while also supporting AI innovation.[379]

Using rightsholder data

230. Many LLM developers have used extensive amounts of human-generated content to train their models. We heard that much of this had taken place without permission from or compensation for rightsholders. Many felt that allowing such practices was morally unfair and economically short sighted.[380]

231. The Society of Authors noted that AI systems "would simply collapse" if they did not have access to creators' works for training and believed tech firms should reward creators fairly.[381] The Copyright Licensing Agency argued that current LLM practices "severely undermine not only the economic value of

376 Written evidence from the British Copyright Council (LLM0043), Publishers Licensing Services (LLM0082) and Creators Rights Alliance (LLM0039)

377 Communications and Digital Committee, *At risk: our creative future* (2nd Report, Session 2022–23, HL Paper 125), para53. The £108 billion figure refers to a more recent update from the Government. See Department for Culture, Media and Sport, 'Ambitious plans to grow the economy and boost creative industries' (June 2023): https://www.gov.uk/government/news/ambitious-plans-to-grow-the-economy-and-boost-creative-industries [accessed 8 January 2024].

378 Department for Culture, Media and Sport, Government response to *At risk: our creative future* (18 April 2023): https://committees.parliament.uk/publications/39303/documents/192860/default/

379 Department for Culture, Media and Sport, *Creative Industries Sector Vision*, CP 863 (June 2023): https://assets.publishing.service.gov.uk/media/64898de2b32b9e000ca96712/Creative_Industries_Sector_Vision__accessible_version_.pdf [accessed 8 January 2024]

380 Written evidence from Publishers' Licensing Services (LLM0082), British Copyright Council (LLM0043), Authors' Licensing and Collecting Society (LLM0092), British Equity Collecting Society (LLM0085), British Recorded Music Industry (LLM0084), Creators' Rights Alliance (LLM0039), PRS for Music (LLM0071), Ivors Academy of Music Creators (LLM0071), Publishers Association (LLM0067), RELX (LLM0064), Getty Images (LLM0054), DACS (LLM0045), Society of Authors (LLM0044), Association of Illustrators (LLM0036), Copyright Licensing Agency (LLM0026), Alliance for Intellectual Property (LLM0022) and Copyright Clearance Center (LLM0018). Note that we refer to 'rightsholders' as a shorthand for stakeholders critical of LLM developers' use of copyrighted works. We recognise that both parties are rightsholders and should not be seen as entirely separate groups.

381 Written evidence from the Society of Authors (LLM0044)

the creative industries but the UK's internationally respected 'gold-standard' copyright framework".[382]

232. The Financial Times said there were "legal routes to access our content which the developers … have chosen not to take".[383] DMG Media said its news content was being used to train models and fact check outputs, and believed the resulting AI tools "could make it impossible to produce independent, commercially funded journalism".[384] The Guardian Media Group said current practices represented a "one sided bargain … without giving any value back" to rightsholders, and warned that openly available high quality news would be "hollow[ed] out" as a result.[385]

233. We heard further concern that the debate on innovation and copyright was too often presented as a mutually exclusive choice. Richard Mollet, Head of European Government Affairs at the information business RELX, noted that RELX was managing to "innovate while at the same time preserving all the things we want to preserve about copyright".[386]

234. OpenAI told us however that it "respect[ed] the rights of content creators and owners" and that its tools helped creative professionals innovate. It noted it had already established "partnership deals with publishers like the Associated Press", though maintained it was "impossible to train today's leading AI models without using copyrighted materials" and attempting to do so "would not provide AI systems that meet the needs of today's citizens".[387] Meta, Stability AI and Microsoft similarly said that limiting access to data risked leading to poorly performing or biased models and less benefit for users.[388]

Legal compliance

235. We heard further disagreement about the extent to which the methods used by LLM developers to acquire and use data are lawful. Dan Conway, CEO of the Publishers' Association, argued that LLMs "are infringing copyrighted content on an absolutely massive scale … when they collect the information, how they store the information and how they handle it." He said there was clear evidence from model outputs that developers had used pirated content from the Books3 database, and alleged they were "not currently compliant" with UK law.[389]

236. Microsoft argued in contrast that conducting TDM on "publicly available and legally accessed works should not require a licence" and was "not copyright infringement".[390] It cited international copyright conventions[391] suggesting copyright should "not extend to ideas … Everyone should have the right to read, learn and understand these works, and copyright law in

382 Written evidence from the Copyright Licensing Agency (LLM0026)
383 Written evidence from the Financial Times (LLM0034)
384 Written evidence from DMG Media (LLM0068)
385 Written evidence from the Guardian Media Group (LLM0108)
386 Q 61
387 Written evidence from OpenAI (LLM0113)
388 Q 4 (Ben Brooks), Q 78 (Rob Sherman) and written evidence from Microsoft (LLM0087)
389 Q 52
390 Written evidence from Microsoft (LLM0087)
391 TRIPS is an international agreement among World Trade Organization members, see World Trade Organisation, 'Frequently asked questions about TRIPS [trade-related aspects of intellectual property rights] in the WTO': https://www.wto.org/english/tratop_e/trips_e/tripfq_e.htm [accessed 8 January 2023].

the UK includes exceptions that allow for the use of technology as a tool to enable this".[392]

237. OpenAI said it complied with "all applicable laws" and believed that, in its view, "copyright law does not forbid training".[393] Stability AI said its activities were "protected by fair use doctrine in jurisdictions such as the United States".[394] Professor Zoubin Ghahramani of Google DeepMind said that if models were to directly reproduce works then rightsholder concerns would be "very valid … We try to take measures so that does not happen."[395]

Technical complexity

238. A large language model may not necessarily 'hold' a set of copyrighted works itself. As Dr Andres Guadamuz has noted, the text from books and articles is converted into billions of sequences (called tokens).[396] The final model contains only statistical representations of the original training data.[397] Jonas Andrulis, CEO of Aleph Alpha, said it was "technically not possible to trace the origin of a certain word or sentence down to one or even a handful of sources".[398]

239. The process for extracting data from websites and transferring it to processing platforms may however involve some form of temporary copy. There is disagreement as to whether such usage is exempt from the Copyright, Designs and Patents Act 1988.[399]

240. Dr Hayleigh Bosher, Reader in Intellectual Property Law and Associate Dean at Brunel University London, said the Act covered the reproduction or "storing the work in any medium by electronic means".[400] She argued that the exceptions allowing transient or incidental copies were narrow and did not apply to LLMs.[401] Dan Conway, CEO of the Publishers Association, agreed.[402] This issue may be a focus of future legal action.[403]

241. Dr Bosher further argued that it was more helpful to consider the underlying purpose and principles of copyright law. She noted that metaphors comparing LLMs to people reading books were misleading, because the intent behind LLM development was clearly commercial whereas reading a book for

---

392 Written evidence from Microsoft (LLM0087)
393 Written evidence from OpenAI (LLM0113)
394 Q 4
395 Q 110
396 Dr Andres Guadamuz, 'A scanner darkly' (February 2023): https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4371204 [accessed 8 January 2024], OpenAI blog, 'What are tokens and how to count them?' (2023): https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them [accessed 8 January 2024];
397 Dr Andres Guadamuz, 'A scanner darkly' (February 2023): https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4371204 [accessed 8 January 2024]
398 Q 108
399 Alec Radford et al, 'Language Models are Unsupervised Multitask Learners', OpenAI Research Paper (2018): https://bit.ly/3mfceXg [accessed 8 January 2024], Dr Andres Guadamuz, 'A scanner darkly' (February 2023): https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4371204 [accessed 8 January 2024] and Q 54 (Dan Conway)
400 Written evidence from Dr Hayleigh Bosher (LLM0109)
401 *Ibid.*
402 Q 54
403 Dr Andres Guadamuz, *A scanner darkly* (February 2023): https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4371204 [accessed 8 January 2024]. Some legal action is underway already. See for example BBC, 'New York Times sues Microsoft and OpenAI for "billions"' (27 December 2023): https://www.bbc.co.uk/news/technology-67826601 [accessed 8 January 2024].

interest was not. She said the application of copyright law should be future proof and not overly specific to how a particular technology works:

> "because it is not the point. It does not matter how you do it; it is why you are doing it."[404]

Reviewing the Government's position

242. We were disappointed that the Government could not articulate its current legal understanding. The Minister said the issues were context dependent and he "worr[ied] about committing … because of the uses and the context in which these potential infringements are occurring". We heard the Government was "waiting for the courts' interpretation of these necessarily complex matters".[405]

243. We were not convinced that waiting for the courts to provide clarity is practical.[406] Rob Sherman of Meta thought it would take "a decade or more for this to work through the court system",[407] and cases may be decided on narrow grounds or settled out of court. In the meantime rightsholders would lose out and contested business practices would become normalised.[408]

244. We welcomed the Minister's acknowledgement of the challenges however. He did "not believe that infringing the rights of copyright holders is a necessary precondition for developing successful AI".[409] And he was clear that AI:

> "can copy an awful lot of information quickly, inexpensively and in new ways that have not been available to copyright infringers before. So it is the same risk of copyright infringement, but it is happening many millions of times faster, which is why it is more complex. It is quite straightforward for someone who intends to infringe copyright to train their model in a different jurisdiction".[410]

245. **LLMs may offer immense value to society. But that does not warrant the violation of copyright law or its underpinning principles. We do not believe it is fair for tech firms to use rightsholder data for commercial purposes without permission or compensation, and to gain vast financial rewards in the process. There is compelling evidence that the UK benefits economically, politically and societally from upholding a globally respected copyright regime.**

246. **The application of the law to LLM processes is complex, but the principles remain clear. The point of copyright is to reward creators for their efforts, prevent others from using works without permission, and incentivise innovation. The current legal framework is failing to ensure these outcomes occur and the Government has a duty to act. It cannot sit on its hands for the next decade until sufficient case law has emerged.**

247. ***In response to this report the Government should publish its view on whether copyright law provides sufficient protections to***

404 Q 54
405 Q 143
406 Q 61 (Dan Conway)
407 Q 77
408 Written evidence from the Authors' Licensing and Collecting Society (LLM0092)
409 Q 142
410 *Ibid.*

rightsholders, given recent advances in LLMs. If this identifies major uncertainty the Government should set out options for updating legislation to ensure copyright principles remain future proof and technologically neutral.

Ways forward

248. Viscount Camrose, Minister for AI and Intellectual Property, said he "had hoped" the IPO-convened working group could develop a voluntary code for AI and copyright by the end of 2023. If talks failed he would consider "other means, which may include legislation".[411] Dan Conway said he still supported the IPO's efforts but believed they would fail without an explicit acknowledgement from the Government and tech firms about the application of copyright and IP law. He said a "legislative handbrake" was needed "if the voluntary conversations fall apart".[412]

249. **The voluntary IPO-led process is welcome and valuable. But debate cannot continue indefinitely.** ***If the process remains unresolved by Spring 2024 the Government must set out options and prepare to resolve the dispute definitively, including legislative changes if necessary.***

250. We heard there were difficult decisions over whether access to and payment for data should be conducted on an 'opt-in' or 'opt-out' basis. Stability AI said it already operated an 'opt-out' system and believed requirements to obtain licenses before conducting TDM would "stifle AI development" and encourage activity to shift to more permissive jurisdictions.[413] OpenAI, Google DeepMind and Aleph Alpha also supported opt-out approaches.[414] Richard Mollett of RELX noted the EU already has an "opt-in/opt-out regime … [which] operates tolerably well".[415]

251. Getty Images argued that "ask for forgiveness later" opt-out mechanisms were "contrary to fundamental principles of copyright law, which requires permission to be secured in advance".[416] The Publishers' Licensing Services said an opt-out approach would also be "impractical" because models could not easily unlearn data they had already been trained on.[417] DMG Media noted that opt-outs could also be commercially damaging, as it is not always clear whether web crawlers are being used for internet search services (which contribute significantly to publishers' revenue) or for AI training. The uncertainty means that publishers have been reluctant to block bots from some large tech firms.[418]

252. ***The IPO code must ensure creators are fully empowered to exercise their rights, whether on an opt-in or opt-out basis. Developers should make it clear whether their web crawlers are being used to acquire data for generative AI training or for other purposes. This would***

411 Q 142
412 Q 58
413 Stability AI highlighted the EU's tiered approach which allowed greater opt-out options, and licensing regimes in the US and Japan. See written evidence from Stability AI (LLM0078).
414 Q 106, Q 109 and written evidence from OpenAI (LLM0113)
415 Q 60
416 Written evidence from Getty Images (LLM0054)
417 Written submission from PLS (LLM0028)
418 Written evidence from DMG Media (LLM0068)

help rightsholders make informed decisions, and reduce risks of large firms exploiting adjacent market dominance.

Better licensing options

253. The Copyright Licensing Agency said that there were already collective licensing mechanisms providing a "practical" system for developers to access data responsibly.[419] Work is underway to develop further licensing options specifically for generative AI.[420] LLMs require vast amounts of data however. The IP Federation believed that a licensing framework was "not feasible for large scale AI".[421]

254. Expanding existing licensing systems and developing new, commercially attractive curated datasets may help address concerns about the viability of licensing agreements and about AI activity shifting to more permissive jurisdictions.[422] Reaching the scale required by LLM developers may be challenging, though some content aggregators already run businesses which reportedly offer access to trillions of words.[423]

255. BT said the Government should boost access to publicly held data and invest in large curated datasets.[424] Jisc, an education and technology firm, likewise thought the UK could play a leading role in this space.[425] The Copyright Clearance Center suggested the Government should use its leverage over public sector technology use and procurement to restrict the use of "products built upon infringement of UK creators' rights".[426]

256. ***The Government should encourage good practice by working with licensing agencies and data repository owners to create expanded, high quality data sources at the scales needed for LLM training. The Government should also use its procurement market to encourage good practice.***

New powers to assert rights

257. We heard that copyright holders are often unable to exercise their rights because they cannot access the training data to check if their works have been used without permission. The British Copyright Council said the IPO should be "empowered" to oversee and enforce copyright issues relating to AI models.[427] RELX called for a transparency mechanism which "requires developers to maintain records, which can be accessed by rightsholders".[428] Dan Conway suggested a searchable repository of citations and metadata would be helpful.[429]

---

419 Written evidence from the CLA (LLM0026)
420 CLA, *Friend or Foe? Attitudes to Generative Artificial Intelligence Among the Creative Community* (4 December 2023): https://assets.cla.co.uk/media/2023/12/ai-research-report.pdf [accessed 8 January 2024]
421 Written evidence from the IP Federation (LLM0057)
422 Written evidence from Human Native AI (LLM0119)
423 See SyndiGate, 'Global content solutions': https://www.syndigate.info/ [accessed 21 December 2023].
424 Written evidence from BT (LLM0090)
425 Written evidence from Jisc (LLM025)
426 Written evidence from the Copyright Clearance Center (LLM0018)
427 Written evidence from the British Copyright Council (LLM0043)
428 Written evidence from RELX (LLM0064)
429 Q 59

258. Google DeepMind said such schemes would be technically "challenging".[430] PRS for Music argued however that it was:

> "insufficient for AI developers to say that the scale of ingestion prevents licensing, record keeping, good data stewardship and disclosure. They have designed and built the product; the ability to meet these fundamental expectations should be built in from the start."[431]

259. ***The IPO code should include a mechanism for rightsholders to check training data. This would provide assurance about the level of compliance with copyright law.***

430 Q 109
431 Written evidence from PRS for Music (LLM0071)

SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS

Future trends

1. Large language models (LLMs) will have impacts comparable to the invention of the internet. (Paragraph 28)

2. *The UK must prepare for a period of heightened technological turbulence as it seeks to take advantage of the opportunities.* (Paragraph 28)

Open or closed

3. Fair market competition is key to ensuring UK businesses are not squeezed out of the race to shape the fast-growing LLM industry. The UK has particular strengths in mid-tier businesses and will benefit most from a combination of open and closed source technologies. (Paragraph 40)

4. *The Government should make market competition an explicit policy objective. This does not mean backing open models at the expense of closed, or vice versa. But it does mean ensuring regulatory interventions do not stifle low-risk open access model providers.* (Paragraph 41)

5. *The Government should work with the Competition and Markets Authority to keep the state of competition in foundation models under close review.* (Paragraph 42)

6. The risk of regulatory capture is real and growing. External AI expertise is becoming increasingly important to regulators and Government, and industry links should be encouraged. But this must be accompanied by stronger governance safeguards. (Paragraph 48)

7. *We recommend enhanced governance measures in DSIT and regulators to mitigate the risks of inadvertent regulatory capture and groupthink. This should apply to internal policy work, industry engagements and decisions to commission external advice. Options include metrics to evaluate the impact of new policies and standards on competition; embedding red teaming, systematic challenge and external critique in policy processes; more training for officials to improve technical know-how; and ensuring proposals for technical standards or benchmarks are published for consultation.* (Paragraph 49)

8. The perception of conflicts of interest risks undermining confidence in the integrity of Government work on AI. Addressing this will become increasingly important as the Government brings more private sector expertise into policymaking. Some conflicts of interest are inevitable and we commend private sector leaders engaging in public service, which often involves incurring financial loss. But their appointment to powerful Government positions must be done in ways that uphold public confidence. (Paragraph 56)

9. *We recommend the Government should implement greater transparency measures for high-profile roles in AI. This should include further high-level information about the types of mitigations being arranged, and a public statement within six months of appointment to confirm these mitigations have been completed.* (Paragraph 57)

A pro-innovation strategy

10. Large language models have significant potential to benefit the economy and society if they are developed and deployed responsibly. The UK must not lose out on these opportunities. (Paragraph 65)

11. Some labour market disruption looks likely. Imminent and widespread cross-sector unemployment is not plausible, but there will inevitably be those who lose out. The pace of change also underscores the need for a credible strategy to address digital exclusion and help all sectors of society benefit from technological change. (Paragraph 66)

12. *We reiterate the findings from our reports on the creative industries and digital exclusion: those most exposed to disruption from AI must be better supported to transition. The Department for Education and DSIT should work with industry to expand programmes to upskill and re-skill workers, and improve public awareness of the opportunities and implications of AI for employment.* (Paragraph 67)

13. The Government is not striking the right balance between innovation and risk. We appreciate that recent advances have required rapid security evaluations and we commend the AI Safety Summit as a significant achievement. But Government attention is shifting too far towards a narrow view of high-stakes AI safety. On its own, this will not drive the kind of widespread responsible innovation needed to benefit our society and economy. The Government must also recognise that long-term global leadership on AI safety requires a thriving commercial and academic sector to attract, develop and retain technical experts. (Paragraph 80)

14. The Government should set out a more positive vision for LLMs and rebalance towards the ambitions set out in the National AI Strategy and AI White Paper. It otherwise risks falling behind international competitors and becoming strategically dependent on a small number of overseas tech firms. The Government must recalibrate its political rhetoric and attention, provide more prominent progress updates on the ten-year National AI Strategy, and prioritise funding decisions to support responsible innovation and socially beneficial deployment. (Paragraph 81)

15. A diverse set of skills and people is key to striking the right balance on AI. *We advocate expanded systems of secondments from industry, academia and civil society to support the work of officials—with appropriate guardrails as set out in Chapter 3. We also urge the Government to appoint a balanced cadre of advisers to the AI Safety Institute with expertise beyond security, including ethicists and social scientists.* (Paragraph 82)

16. Recent Government investments in advanced computing facilities are welcome, but more is needed and the Government will struggle to afford the scale required to keep pace with cutting edge international competitors. *The Government should provide more incentives to attract private sector investment in compute. These should be structured to maximise energy efficiency.* (Paragraph 92)

17. Equitable access will be key. *UK Research and Innovation and DSIT must ensure that both researchers and SMEs are granted access to high-end computing facilities on fair terms to catalyse publicly beneficial research and commercial opportunity.* (Paragraph 93)

18. The Government should take better advantage of the UK's start-up potential. *It should work with industry to expand spin-out accelerator schemes. This could focus on areas of public benefit in the first instance. It should also remove barriers, for example by working with universities on providing attractive licensing and ownership terms, and unlocking funding across the business lifecycle to help start-ups grow and scale in the UK.* (Paragraph 94)

19. *The Government should also review UKRI's allocations for AI PhD funding, in light of concerns that the prospects for commercial spinouts are being negatively affected and foreign influence in funding strategic sectors may grow as a result.* (Paragraph 95)

20. A sovereign UK LLM capability could deliver substantial value if challenges around reliability, ethics, security and interpretability can be resolved. LLMs could in future benefit central departments and public services for example, though it remains too early to consider using LLMs in high-stakes applications such as critical national infrastructure or the legal system. (Paragraph 105)

21. We do not recommend using an 'off the shelf' LLM or developing one from scratch: the former is too risky and the latter requires high-tech R&D efforts ill-suited to Government. But commissioning an LLM to high specifications and running it on internal secure facilities might strike the right balance. The Government might also make high-end facilities available to researchers and commercial partners to collaborate on applying LLM technology to national priorities. (Paragraph 106)

22. *We recommend that the Government explores the options for and feasibility of acquiring a sovereign LLM capability. No option is risk free, though commissioning external developers might work best. Any public sector capability would need to be designed to the highest ethical and security standards, in line with the recommendations made in this report.* (Paragraph 107)

Risk

23. The most immediate security concerns from LLMs come from making existing malicious activities easier, rather than qualitatively new risks. (Paragraph 128)

24. *The Government should work with industry at pace to scale existing mitigations in the areas of cyber security (including systems vulnerable to voice cloning), child sexual abuse material, counter-terror, and counter-disinformation. It should set out progress and future plans in response to this report, with a particular focus on disinformation in the context of upcoming elections.* (Paragraph 128)

25. The Government has made welcome progress on understanding AI risks and catalysing international co-operation. There is however no publicly agreed assessment framework and shared terminology is limited. It is therefore difficult to judge the magnitude of the issues and priorities. (Paragraph 129)

26. *The Government should publish an AI risk taxonomy and risk register. It would be helpful for this to be aligned with the National Security Risk Assessment.* (Paragraph 129)

27. Catastrophic risks resulting in thousands of UK fatalities and tens of billions in financial damages are not likely within three years, though this cannot be ruled out as next generation capabilities become clearer and open access models more widespread. (Paragraph 140)

28. There are however no warning indicators for a rapid and uncontrollable escalation of capabilities resulting in catastrophic risk. There is no cause for panic, but the implications of this intelligence blind spot deserve sober consideration. (Paragraph 141)

29. *The AI Safety Institute should publish an assessment of engineering pathways to catastrophic risk and warning indicators as an immediate priority. It should then set out plans for developing scalable mitigations. (We set out recommendations on powers and take-down requirements in Chapter 7). The Institute should further set out options for encouraging developers to build systems that are safe by design, rather than focusing on retrospective guardrails.* (Paragraph 142)

30. There is a credible security risk from the rapid and uncontrollable proliferation of highly capable openly available models which may be misused or malfunction. Banning them entirely would be disproportionate and likely ineffective. But a concerted effort is needed to monitor and mitigate the cumulative impacts. (Paragraph 148)

31. *The AI Safety Institute should develop new ways to identify and track models once released, standardise expectations of documentation, and review the extent to which it is safe for some types of model to publish the underlying software code, weights and training data.* (Paragraph 148)

32. It is almost certain existential risks will not manifest within three years and highly likely not within the next decade. As our understanding of this technology grows and responsible development increases, we hope concerns about existential risk will decline. The Government retains a duty to monitor all eventualities. But this must not distract it from capitalising on opportunities and addressing more limited immediate risks. (Paragraph 155)

33. LLMs may amplify numerous existing societal problems and are particularly prone to discrimination and bias. The economic impetus to use them before adequate guardrails have been developed risks deepening inequality. (Paragraph 161)

34. *The AI Safety Institute should develop robust techniques to identify and mitigate societal risks. The Government's AI risk register should include a range of societal risks, developed in consultation with civil society. DSIT should also use its White Paper response to propose market-oriented measures which incentivise ethical development from the outset, rather than retrospective guardrails. Options include using Government procurement and accredited standards, as set out in Chapter 7.* (Paragraph 162)

35. Further clarity on data protection law is needed. *The Information Commissioner's Office should work with DSIT to provide clear guidance on how data protection law applies to the complexity of LLM processes, including the extent to which individuals can seek redress if a model has already been trained on their data and released.* (Paragraph 167)

36. *The Department for Health and Social Care should work with NHS bodies to ensure future proof data protection provisions are embedded in licensing terms. This would help reassure patients given the possibility of LLM businesses working with NHS data being acquired by overseas corporations.* (Paragraph 168)

International context and lessons

37. *The UK should continue to forge its own path on AI regulation, balancing rather than copying the EU, US or Chinese approaches. In doing so the UK can strengthen its position in technology diplomacy and set an example to other countries facing similar decisions and challenges.* (Paragraph 175)

38. International regulatory co-ordination will be key, but difficult and probably slow. Divergence appears more likely in the immediate future. We support the Government's efforts to boost international co-operation, but it must not delay domestic action in the meantime. (Paragraph 178)

39. Extensive primary legislation aimed solely at LLMs is not currently appropriate: the technology is too new, the uncertainties too high and the risk of inadvertently stifling innovation too great. Broader legislation on AI governance may emerge in future, though this was outside the scope of our inquiry. (Paragraph 187)

40. *Setting the strategic direction for LLMs and developing enforceable, pro-innovation regulatory frameworks at pace should remain the Government's immediate priority.* (Paragraph 187)

Making the White Paper work

41. We support the overall White Paper approach. But the pace of delivering the central support functions is inadequate. The regulatory support and co-ordination teams proposed in the March 2023 White Paper underpin its entire success. By the end of November 2023, regulators were unaware of the central function's status and how it would operate. This slowness reflects prioritisation choices and undermines confidence in the Government's commitment to the regulatory structures needed to ensure responsible innovation. (Paragraph 195)

42. *DSIT should prioritise resourcing the teams responsible for regulatory support and co-ordination, and publish an update on staffing and policy progress in response to this report.* (Paragraph 196)

43. Relying on existing regulators to ensure good outcomes from AI will only work if they are properly resourced and empowered. (Paragraph 201)

44. *The Government should introduce standardised powers for the main regulators who are expected to lead on AI oversight to ensure they can gather information relating to AI processes and conduct technical, empirical and governance audits. It should also ensure there are meaningful sanctions to provide credible deterrents against egregious wrongdoing.* (Paragraph 201)

45. *The Government's central support functions should work with regulators at pace to publish cross-sector guidance on AI issues that fall outside individual sector remits.* (Paragraph 202)

46. Model developers bear some responsibility for the products they are building—particularly given the foreseeable risk of harm from misuse and the limited information available to customers about how the base model works. But how far such liability extends remains unclear. (Paragraph 209)

47. *The Government should ask the Law Commission to review legal liability across the LLM value chain, including open access models. The Government should provide an initial position, and a timeline for establishing further legal clarity, in its White Paper response.* (Paragraph 209)

48. We welcome the commitments from model developers to engage with the Government on safety. But it would be naïve to believe voluntary agreements will suffice in the long-term as increasingly powerful models proliferate

across the world, including in states which already pose a threat to UK security objectives. (Paragraph 218)

49. *The Government should develop mandatory safety tests for high-risk high-impact models. This must include an expectation that the results will be shared with the Government (and regulators if appropriate), and clearly defined powers to require compliance with safety recommendations, suspend model release, and issue market recall or platform take-down notices in the event of a credible threat to public safety.* (Paragraph 219)

50. *The scope and benchmarks for high-risk high-impact testing should involve a combination of metrics that can adapt to fast-moving changes. They should be developed by the AI Safety Institute through engagement with industry, regulators and civil society. It is imperative that these metrics do not impose undue market barriers, particularly to open access providers.* (Paragraph 220)

51. Accredited standards and auditing practices are key. They would help catalyse a domestic AI assurance industry, support business clarity and empower regulators. (Paragraph 226)

52. *We urge the Government and regulators to work with partners at pace on developing accredited standards and auditing practices for LLMs (noting that these must not be tick-box exercises). A consistent approach to publishing key information on model cards would also be helpful.* (Paragraph 226)

53. *The Government should then use the public sector procurement market to encourage responsible AI practices by requiring bidders to demonstrate compliance with high standards when awarding relevant contracts.* (Paragraph 227)

Copyright

54. LLMs may offer immense value to society. But that does not warrant the violation of copyright law or its underpinning principles. We do not believe it is fair for tech firms to use rightsholder data for commercial purposes without permission or compensation, and to gain vast financial rewards in the process. There is compelling evidence that the UK benefits economically, politically and societally from upholding a globally respected copyright regime. (Paragraph 245)

55. The application of the law to LLM processes is complex, but the principles remain clear. The point of copyright is to reward creators for their efforts, prevent others from using works without permission, and incentivise innovation. The current legal framework is failing to ensure these outcomes occur and the Government has a duty to act. It cannot sit on its hands for the next decade until sufficient case law has emerged. (Paragraph 246)

56. *In response to this report the Government should publish its view on whether copyright law provides sufficient protections to rightsholders, given recent advances in LLMs. If this identifies major uncertainty the Government should set out options for updating legislation to ensure copyright principles remain future proof and technologically neutral.* (Paragraph 247)

57. The voluntary IPO-led process is welcome and valuable. But debate cannot continue indefinitely. (Paragraph 249)

58. *If the process remains unresolved by Spring 2024 the Government must set out options and prepare to resolve the dispute definitively, including legislative changes if necessary.* (Paragraph 249)

59. *The IPO code must ensure creators are fully empowered to exercise their rights, whether on an opt-in or opt-out basis. Developers should make it clear whether their web crawlers are being used to acquire data for generative AI training or for other purposes. This would help rightsholders make informed decisions, and reduce risks of large firms exploiting adjacent market dominance.* (Paragraph 252)

60. *The Government should encourage good practice by working with licensing agencies and data repository owners to create expanded, high quality data sources at the scales needed for LLM training. The Government should also use its procurement market to encourage good practice.* (Paragraph 256)

61. *The IPO code should include a mechanism for rightsholders to check training data. This would provide assurance about the level of compliance with copyright law.* (Paragraph 259)

APPENDIX 1: LIST OF MEMBERS AND DECLARATIONS OF INTEREST

Members

Baroness Featherstone
Lord Foster of Bath
Baroness Fraser of Craigmaddie
Lord Griffiths of Burry Port
Lord Hall of Birkenhead
Baroness Harding of Winscombe
Baroness Healy of Primrose Hill
Lord Kamall
The Lord Bishop of Leeds
Lord Lipsey
Baroness Stowell of Beeston (Chair)
Baroness Wheatcroft
Lord Young of Norwood Green

Declarations of interest

Baroness Featherstone

Former career in creative industries

Lord Foster of Bath

No relevant interests declared

Baroness Fraser of Craigmaddie

Board Member, Creative Scotland

Board Member, British Library (which houses the Alan Turing Institute)

Lord Griffiths of Burry Port

No relevant interests declared

Lord Hall of Birkenhead

Chairman, City of Birmingham Symphony Orchestra

Chairman, Harder Than You Think Ltd (start-up documentary producer)

Member, Advisory Board, Qwilt (edge cloud application developer)

Trustee, National Trust

Trustee, Natural History Museum

Trustee, Oxford Philharmonic Orchestra

Trustee, Paul Hamlyn Foundation (independent grant-making organisation focusing on the arts)

Baroness Harding of Winscombe

Fellow, Royal Society of Arts

Baroness Healy of Primrose Hill

No relevant interests declared

Lord Kamall

Former member, Tech UK Brexit advisory committee (unpaid)

Member, Advisory Board, Startup Coalition (unpaid)

Consultant to two think tanks (IEA and Politeia) that have published on AI

Non-Executive Director, Department for Business and Trade

The Lord Bishop of Leeds

Trustee, Reading Culture (Bradford Literature Festival)

Lord Lipsey

Chair, Premier Greyhound Racing

Trustee, Mid Wales Music Trust (formerly Cambrian Music Trust)

Baroness Stowell of Beeston (Chair)

No relevant interests declared

Baroness Wheatcroft

Chair, Financial Times appointments and oversight committee

Lord Young of Norwood Green

Former professional engagement with BT

A full list of Members' interests can be found in the Register of Lords' Interests: https://members.parliament.uk/members/lords/interests/register-of-lords-interests

Specialist Adviser

Professor Michael Wooldridge

Scientific advisor for RocketPhone, a startup

Scientific advisory board for Mind Foundry and Aioi Nissay Dowa Insurance and Aioi Nissay Dowa Europe

Royal Institution Christmas Lectures

Director of Foundational AI Research, Alan Turing Institute. Professor Wooldridge agreed to a series of mitigations with the Alan Turing Institute to mitigate potential conflicts of interest. These included agreements to avoid disclosure of information relating to the inquiry, and the avoidance of policy influence work for the duration of the inquiry relating to its core topics. Professor Wooldridge's work for the Committee was primarily in the area of providing expert technical advice and relevant declarations were made to the Committee throughout the inquiry.

Professor Wooldridge holds a variety of additional posts and academic engagements: https://www.cs.ox.ac.uk/people/michael.wooldridge

APPENDIX 2: LIST OF WITNESSES

Evidence is published online at https://committees.parliament.uk/committee/170/communications-and-digital-committee/publications/ and available for inspection at the Parliamentary Archives (020 7219 3074).

Evidence received by the Committee is listed below in chronological order of oral evidence session and in alphabetical order. Those witnesses marked with ** gave both oral evidence and written evidence. Those marked with * gave oral evidence and did not submit any written evidence. All other witnesses submitted written evidence only.

Oral evidence in chronological order

*	Ian Hogarth, Chair, Frontier AI Taskforce	QQ 1–11
*	Dr Jean Innes, Chief Executive Officer, Alan Turing Institute	
*	Professor Neil Lawrence, DeepMind Professor of Machine Learning, University of Cambridge	
*	Ben Brooks, Head of Public Policy, Stability AI	
*	Dr Peter Waggett, UK Director of Research, IBM	QQ 12–20
**	Dr Zoë Webster, Director of Data and AI Solutions, BT	
*	Francesco Marconi, Co-Founder, Applied XL	
*	Dr Nathan Benaich, Founder, Air Street Capital	
*	Professor Stuart Russell OBE, Professor of Computer Science, University of California, Berkeley	QQ 21–28
*	Professor Phil Blunsom, Chief Scientist, Cohere	
**	Lyric Jain, Founder and Chief Executive Officer, Logically	
**	Chris Anley, Chief Scientist, NCC Group	
*	Professor Dame Wendy Hall, Regius Professor of Computer Science, University of Southampton	QQ 29–36
*	Professor Dame Muffy Calder, Vice-Principal and Head of College, University of Glasgow	
*	Dr Jeremy Silver, Chief Executive Officer, Digital Catapult	
**	Dr Florian Ostmann, Head of AI Governance and Regulatory Innovation, Alan Turing Institute	QQ 37–45
**	Michael Birtwistle, Associate Director (Law & Policy), Ada Lovelace Institute	
*	Katherine Holden, Head of Data Analytics, AI and Digital ID, techUK	
*	Professor Anu Bradford, Professor of Law and International Organisation, Columbia Law School	QQ 46–50

*	Dr Mark MacCarthy, Senior Fellow, Institute for Technology Law and Policy, Georgetown Law	
*	Paul Triolo, Senior Associate with the Trustee Chair in Chinese Business and Economics, Center for Strategic and International Studies	
**	Dan Conway, Chief Executive Officer, Publishers Association	QQ 51–63
**	Arnav Joshi, Senior Associate, Clifford Chance	
**	Richard Mollet, Head of European Government Affairs, RELX	
**	Dr Hayleigh Bosher, Associate Dean and Reader in Intellectual Property Law, Brunel Law School	
*	Dr Moez Draief, Managing Director, Mozilla.ai	QQ 64–72
**	Irene Solaiman, Head of Global Policy, Hugging Face	
*	Professor John McDermid OBE, Chairman, Rapita Systems, and Professor of Safety-Critical Systems, University of York	
**	Dr Adriano Koshiyama, Co-Chief Executive Officer, Holistic AI	
**	Owen Larter, Director of Public Policy, Office for Responsible AI, Microsoft	QQ 73–82
**	Rob Sherman, Vice President and Deputy Chief Privacy Officer for Policy, Meta	
**	Hayley Fletcher, Director, Competition and Markets Authority	QQ 83–94
**	Dr Yih-Choung Teh, Group Director of Strategy and Research, Ofcom	
**	Stephen Almond, Executive Director, Regulatory Risk, Information Commissioner's Office	
**	Anna Boaden Director of Policy and Human Rights, Equality and Human Rights Commission	
*	Jonas Andrulis, Founder and Chief Executive Officer, Aleph Alpha	QQ 95–111
**	Professor Zoubin Ghahramani, Vice-President of Research, Google DeepMind	
*	Professor Dame Angela McLean DBE FRS, Government Chief Scientific Adviser, Government Office for Science	QQ 112–129
**	Viscount Camrose, Minister for AI and Intellectual Property, HM Government—Department for Science, Innovation and Technology	QQ 130–145
**	Lizzie Greenhalgh, Deputy Director of AI Regulation, AI Policy Directorate, HM Government—Department for Science, Innovation and Technology	

**	Sam Cannicott, Deputy Director of AI Enablers and Institutions, AI policy directorate, HM Government—Department for Science, Innovation and Technology	

Alphabetical list of all witnesses

	Dr Elena Abrusci, Senior Lecturer in Law, Brunel University London (joint submission)	LLM0061
	Dr Alberto Acerbi, Assistant Professor, University of Trento (joint submission)	LLM0024
**	Ada Lovelace Institute (QQ 37–45)	
	The Advertising Association	LLM0056
	AGENCY	LLM0028
	AI Governance	LLM0013
**	The Alan Turing Institute (QQ 1–11, QQ 37–45)	LLM0081
*	Aleph Alpha (QQ 95–111)	
	Alliance for Intellectual Property	LLM0022
	Andreessen Horowitz	LLM0114
	Dr. Plamen P. Angelov, Professor of Intelligent Systems, ELSA project Work Package 3 leader, Lancaster University (joint submission)	LLM0032
	ASA System	LLM0098
	The Association of Illustrators	LLM0036
	Authors' Licensing and Collecting Society (ALCS)	LLM0092
	Dr Brian Ball, Associate Professor of Philosophy, Northeastern University—London (joint submission)	LLM0038
	Sir Jon Cunliffe, Deputy Governor, Financial Stability, Bank of England (joint submission)	LLM0099
	Professor David Barber, Professor of Machine Learning, Department of Computer Science, University College London)	LLM0118
	BCS, The Chartered Institute for IT	LLM0094
	Professor Mark Beer OBE (joint submission)	LLM0040
*	Dr Nathan Benaich (QQ 12–20)	
*	Professor Phil Blunsom (QQ 21–28)	
*	Dr Hayleigh Bosher (QQ 51–63) (joint submission)	LLM0061 LLM0109
	BPI (British Recorded Music Industry)	LLM0084
*	Professor Anu Bradford (QQ 46–50)	
	The Bright Initiative	LLM0033
	British Copyright Council	LLM0043

	The British Equity Collecting Society (BECS)	LLM0085
	British Screen Forum	LLM0097
	British Standards Institution	LLM0111
**	BT Group (QQ 12–20)	LLM0090
*	Professor Dame Muffy Calder (QQ 29–36)	
	Cambridge Language Sciences	LLM0053
	Careful Industries	LLM0041
	Carnegie UK	LLM0096
	Caution Your Blast	LLM0077
*	Center for Strategic and International Studies (QQ 46–50)	
	Dr Xuechen Chen, Assistant Professor in Politics and International Relations and Head of Digital Governance Research Cluster, Northeastern University—London (joint submission)	LLM0031
**	Clifford Chance (QQ 51–63)	LLM0112
	Committee on Standards in Public Life	LLM0052
**	Competition and Markets Authority (QQ 83–94)	LLM0100
	Confederation of British Industry	LLM0069
	Connected by Data	LLM0066
	Copyright Clearance Center	LLM0018
	The Copyright Licensing Agency	LLM0026
	Melissa Coutino	LLM0059
	Creators' Rights Alliance	LLM0039
	DACS	LLM0045
	Dr Rishi Das-Gupta, Chief Executive Officer, Health Innovation Network South London	LLM0037
	Deep Learning Partnership	LLM0005
	Digital Regulation Cooperation Forum	LLM0086
	DMG Media	LLM0068
	EPOCH	LLM0002
**	Equality and Human Rights Commission (QQ 83–94)	LLM0101
	Matthew Farmer (joint submission)	LLM0040
	Matthew Feeney, Head of Technology and Innovation, Centre for Policy Studies	LLM0047
	Financial Conduct Authority	LLM0091
		LLM0102
	Financial Times	LLM0034

	Professor Mario Fritz, Professor, ELSA project coordinator, CISPA, Germany (joint submission)	LLM0032
	Full Fact	LLM0058
	Dr Xinchuchu Gao, Lecturer in International Relations, University of Lincoln (joint submission)	LLM0031
	Getty Images (UK)	LLM0054
	The Glenlead Centre	LLM0051
**	Google DeepMind (QQ 95–111)	LLM0095
*	Government Office for Science (QQ 112–129)	
	Guardian Media Group	LLM0108
*	Professor Dame Wendy Hall (QQ 29–36)	
	Dr Alice Helliwell, Assistant Professor of Philosophy, Northeastern University—London	LLM0038
	Professor Ali Hessami, Director of R&D and Innovation, Vega Systems (joint submission)	LLM0075
**	HM Government—Department for Science, Innovation and Technology (QQ 130–145)	LLM0079 LLM0116 LLM0120
*	Ian Hogarth (QQ 1–11)	
**	Holistic AI (QQ 64–72)	LLM0010
	Martin Hosken, Chief Technologist, Cloud for VMware EMEA, VMware	LLM0009
	Dr Jeffrey Howard, Associate Professor of Political Philosophy & Public Policy, and Principal Investigator of the Digital Speech Lab, University College London (joint submission)	LLM0049
**	Hugging Face (QQ 64–72)	LLM0019
	Human Native AI	LLM0119
*	IBM (QQ 12–20)	
	IEEE Standards Association	LLM0072
	Dr Sam Illingworth, Associate Professor, Edinburgh Napier University	LLM0003
**	Information Commissioner's Office (QQ 83–94)	LLM0006 LLM0103
	IP Federation	LLM0057
	The Ivors Academy of Music Creators	LLM0070
	Dr Karen Jeffrey, Postdoctoral Data Analyst, Medical Informatics, University of Edinbugh	LLM0001
	JISC	LLM0025

	Kairoi	LLM0110
	Dr Dmitry Kangin, Senior Research Associate, ELSA project, Lancaster University (joint submission)	LLM0032
	Michael Karanicolas, Executive Director, UCLA Institute for Technology, Law & Policy (joint submission)	LLM0020
	Dr Dimosthenis Karatzas, Associate Professor, ELSA Board member, Computer Vision Center (CVC), Barcelona (joint submission)	LLM0032
	Dr Beatriz Kira, Lecturer in Law, University of Sussex (joint submission)	LLM0049
	Dr Lingpeng Kong, Assistant Professor, Department of Computer Science, University of Hong Kong (joint submission)	LLM0031
*	Professor Neil Lawrence (QQ 1–11)	
	Local Government Association (joint submission)	LLM0048
**	Logically AI (QQ 21–28)	LLM0062
*	Dr Mark MacCarthy (QQ 46–50)	
*	Professor John McDermid OBE (QQ 64–72)	
	Dr Dan McQuillan, Lecturer in Creative and Social Computing, Goldsmiths, University of London	LLM0015
*	Francesco Marconi (QQ 12–20)	
	Market Research Society	LLM0088
	Medicines and Healthcare products Regulatory Agency	LLM0107
**	Meta (QQ 73–82)	LLM0093
**	Microsoft (QQ 73–82)	LLM0087
	Mind Foundry	LLM0030
	Dr Alina Miron, Lecturer in Computer Science, Brunel University London (joint submission)	LLM0061
*	Mozilla.ai (QQ 64–72)	
	National Union of Journalists	LLM0007
**	NCC Group (QQ 21–28)	LLM0014
	The News Media Association	LLM0029
	NquiringMinds	LLM0073
	Oaklin Consulting	LLM0035
**	Ofcom (QQ 83–94)	LLM0080
		LLM0104
	Ofqual	LLM0105
	OpenAI	LLM0113

	Open Data Institute	LLM0083
	OpenUK	LLM0115
	Oxford Internet Institute, University of Oxford	LLM0074
	Pact	LLM0011
	Policy Connect	LLM0065
	Professional Publishers Association	LLM0017
	PRS for Music	LLM0071
	Sam Woods, Deputy Governor, Prudential Regulation and Chief Executive Officer, Prudential Regulation Authority (joint submission)	LLM0099
**	Publishers Association (QQ 51–63)	LLM0067
		LLM0117
	Publishers' Licensing Services	LLM0082
**	RELX (QQ 51–63)	LLM0064
	Reset	LLM0042
	The Royal Academy of Engineering	LLM0063
	Royal Statistical Society	LLM0055
*	Professor Stuart Russell OBE (QQ 21–28)	
	Nadim Sadek, Founder and Chief Executive Officer, Shimmr AI	LLM0021
	Sense about Science	LLM0046
	Patricia Shaw, Chief Executive Officer, Beyond Reach Consulting (joint submission)	LLM0075
*	Dr Jeremy Silver (QQ 29–36)	
	Dr Martin Smith, Visiting Fellow in Creative Industries, Goldsmiths, University of London	LLM0004
	The Society for Innovation, Technology and Modernisation (Socitm) (joint submission)	LLM0048
	The Society of Authors	LLM0044
	The Society of Local Authority Chief Executive (Solace) (joint submission)	LLM0048
	Solicitors Regulation Authority	LLM0106
**	Stability AI (QQ 1–11)	LLM0078
	Startup Coalition	LLM0089
	Professor Marc Stears, Director, UCL Policy Lab, University College London (joint submission)	LLM0049
	Dr Joseph Stubbersfield, Lecturer, University of Winchester (joint submission)	LLM0024
	Surrey Institute for People-Centred AI	LLM0060

*	techUK (QQ 37–45)	
	Trustworthy Autonomous Systems Hub (TAS Hub), University of Southampton	LLM0027
	UCL Institute of Health Informatics	LLM0076
	Warner Music Group	LLM0023
	Eleanor Watson, President, European Responsible AI Office (joint submission)	LLM0075
	WITNESS	LLM0050
	Writers' Guild of Great Britain	LLM0016
	Dr Baoli Zhao, Founder and Managing Director, Vieunite Ltd	LLM0008
	Alessia Zornetta, Research Assistant/Doctoral Candidate, UCLA Institute for Technology, Law & Policy (joint submission)	LLM0020

APPENDIX 3: CALL FOR EVIDENCE

Large language models (LLMs) are a type of generative AI, which have attracted significant interest for their ability to produce human-like text, code and translations. There have been several recent advances, notably OpenAI's GPT-3 and GPT-4 models. Many experts say these developments represent a step change in capability. Smaller and cheaper open-source models are set to proliferate.

Governments, businesses and individuals are all experimenting with this technology's potential. The opportunities could be extensive. Goldman Sachs has estimated generative AI could add $7 trillion (roughly £5.5 trillion) to the global economy over 10 years. Some degree of economic disruption seems likely: the same report estimated 300 million jobs could be exposed to automation, though many roles could also be created in the process.[432]

The speed of development and lack of understanding about these models' capabilities has led some experts to warn of a credible and growing risk of harm. Several industry figures have been calling for urgent reviews or pausing new release plans. Large models can generate contradictory or fictious answers, meaning their use in some industries could be dangerous without proper safeguards. Training datasets can contain biased or harmful content. Intellectual property rights over the use of training data are uncertain. The 'black box' nature of machine learning algorithms makes it difficult to understand why a model follows a course of action, what data were used to generate an output, and what the model might be able to do next, or do without supervision. Some models might develop counterintuitive or perverse ways of achieving aims. And the proliferation of these tools will make easier undesirable practices, such as spreading disinformation, hacking, fraud and scams.

This all presents challenges for the safe, ethical and trusted development of large language models, and undermines opportunities to capitalise on the benefits they could provide.

Regulation

There are growing calls to improve safeguards, standards and regulatory approaches that promote innovation whilst managing risks. Many experts say this is increasingly urgent. The UK Government released its AI White Paper in March 2023. It highlights the importance of a "pro-innovation framework designed to give consumers the confidence to use AI products and services, and provide businesses the clarity they need to invest in AI and innovate responsibly".[433] Regulators are expected to address key issues using existing powers. The Prime Minister's Office has expressed an interest in the UK becoming a world-leading centre for AI safety.

Inquiry objectives

The Communications and Digital Committee will examine what needs to happen over the next 1–3 years to ensure the UK can respond to the opportunities and risks posed by large language models. [434] This will include evaluating the work

432 Goldman Sachs, 'Generative AI Could raise global GDP by 7 per cent' (5 April 2023): https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html [accessed 11 January 2024]

433 Department for Science, Innovation & Technology and Office for Artificial Intelligence, 'A pro-innovation approach to AI regulation' (29 March 2023): https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [accessed 11 January 2024]

434 The main focus of this inquiry will be on large language models. The Committee will also examine wider generative AI capabilities, though in less depth.

of Government and regulators, examining how well this addresses current and future technological capabilities, and reviewing the implications of approaches taken elsewhere in the world.

Questions

Capabilities and trends

1. How will large language models develop over the next three years?

    (a) Given the inherent uncertainty of forecasts in this area, what can be done to improve understanding of and confidence in future trajectories?

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

    (a) How should we think about risk in this context?

Domestic regulation

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

    (a) What are the implications of open-source models proliferating?

4. Do the UK's regulators have sufficient expertise and resources to respond to large language models?[435] If not, what should be done to address this?

5. What are the non-regulatory and regulatory options to address risks and capitalise on opportunities?

    (a) How would such options work in practice and what are the barriers to implementing them?

    (b) At what stage of the AI life cycle will interventions be most effective?

    (c) How can the risk of unintended consequences be addressed?

International context

6. How does the UK's approach compare with that of other jurisdictions, notably the EU, US and China?

    (a) To what extent does wider strategic international competition affect the way large language models should be regulated?

    (b) What is the likelihood of regulatory divergence? What would be its consequences?

435 The Committee will be focusing in particular on the members of the Digital Regulation Co-operation Forum (Ofcom, the Competition and Markets Authority, the Information Commissioner's Office and the Financial Conduct Authority).

APPENDIX 4: VISITS

Committee visit to Intuit

On 5 December 2023 the Committee held a visit to Intuit's offices in London. In attendance were Baroness Stowell of Beeston, Baroness Featherstone, Lord Foster of Bath, Baroness Fraser of Craigmaddie Lord Griffiths of Burry Port, Lord Hall of Birkenhead, Baroness Healy of Primrose Hill, Lord Lipsey, and Lord Young of Norwood Green.

The purpose of the visit was to develop a better understanding of how small and medium enterprises (SMEs) are making use of AI, current and future opportunities, concerns and barriers to wider adoption.

Intuit is an American business software company. The Committee heard from representatives from Mailchimp about the use of AI in its business areas, followed by a roundtable with small business owners and providers of AI-driven services. Topics included the value of AI in speeding up rote tasks and customising services, alongside a recognition that large language models were just the latest in a series of AI developments. The discussion also covered issues relating to digital exclusion and the importance of ensuring all sectors of the public have sufficient skills to use new digital tools.

We are grateful to all those who took part in the discussions.

Committee visit to Google Health and UCL Centre for Artificial Intelligence

On 12 December 2023, the Committee visited Google and University College London (UCL) Centre for Artificial Intelligence. In attendance were Baroness Stowell of Beeston, Lord Foster of Bath, Baroness Fraser of Craigmaddie, Lord Hall of Birkenhead, Baroness Harding of Winscombe, Baroness Healy of Primrose Hill, Lord Bishop of Leeds, Lord Lipsey and Lord Young of Norwood Green.

The purpose was to understand how AI products are being developed and applied within healthcare, opportunities for commercialising academic research, and barriers to progress.

The visit to Google involved talks from members of the Google Health team followed by demonstrations of large language model tools and a question-and-answer session. The discussion topics included the opportunities provided by applying AI to healthcare, existing partnerships and the value these can deliver, challenges and mitigations (including around data protection and accuracy), and future trends.

The subsequent engagement at UCL involved a roundtable discussion with staff from the institution including Dr Anne Lane, CEO of UCL Business and Professor David Barber, Director of the Centre for Artificial Intelligence; academic staff involved in AI research as well as commercial enterprises; and representatives from spinouts including CogStack and Humanloop.

Discussions focused on Centre's work in providing guidance and support in bringing ideas to market, and helping develop opportunities for translating research into commercial applications. The Committee heard that the way public funding is allocated to PhDs, mainly through Centres for Doctoral Training, was having an adverse impact on the number of relevant PhD places and the prospects for institutions with a good track record of producing academic excellence and

commercial value in AI. The Committee also heard about the needs of AI start-ups, the limited level of funding in the UK and the attraction of scaling opportunities in the US.

We are grateful to all those who took part in the discussions.