Caution Your Blast Ltd—written evidence (LLM0077)

House of Lords Communications and Digital Select Committee inquiry: Large language models

About Caution Your Blast Ltd

1. We consult in efficient organisational design and operation across four fundamental information technology practice areas - 1. Transformation, 2. Research, 3. Design, and 4. Engineering.

2. Transformation brings Service, Product and Delivery Management together in a single practice area with an emphasis on working with an organisation to achieve a significant change in operational efficiency and service quality. Doing this by leading on team delivery and stakeholder engagement, being responsible for encouraging ambitions.

3. Our Research practice is world leading, creating new ways of conducting research and underpinning this with data. CYB has established research operations into our practice; promoting standardisation of tools, processes and technology that researchers need and driving innovative ways of working. This results in a complete understanding of users, providing operational technologies and building services on evidence.

4. Our Design practise works across all areas of design, including organisational operations, service, experience and content. We are experts in the Lean design process constantly pushing the boundaries of how we translate user needs into technology, through sustainable design systems.

5. Our Technology practice covers technology architecture, infrastructure, security, and full stack development of front-end and back-end services. We lead on innovative use of the latest technologies, including large language models, with a strong focus on open source where we both use and develop collaborative communities.

6. Our teams work in multi-disciplinary units applying our skills, experience and interest to solving real problems for people. We've set out how we work as part of multiple playbooks and studies, helping how governments and large multinational organisations deliver their operations and services.

7. We are an early adopter of LLM technology and are actively working on a number of initiatives that leverage LLMs for tasks such as question-answer responses and custom data set interrogation with a particular focus on data privacy and security.

Summary

8. This document provides Caution Your Blast Ltd's responses to the call for evidence requested by the Communications and Digital Committee. Responses to the questions in the capabilities and trends and domestic regulation sections are provided in this report.

9. We explore the development, opportunities, and risks associated with large language models (LLMs) over the next three years, with a focus on the UK context. Key insights include the rapid growth of LLMs, their potential to automate tasks and increase productivity, and the need for addressing the skills gap in AI practitioners. Additionally, the report evaluates the adequacy of existing regulatory approaches and proposes non-regulatory and regulatory options to manage LLM-related risks and opportunities. These options encompass ethical frameworks, compliance standards, sector-specific regulations, and international collaboration to ensure responsible LLM development and utilisation.

Responses

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?

10. The use of large language models (LLMs) will become widespread very quickly. This is already being seen across many sectors, particularly with the interest and applications of AI provoked since the release of OpenAI's GPT-3.5 model and other commercial platforms. There have also been a lot of open source models released recently[1] that will further accelerate exploration and adoption of the technology. This trend will continue and there will be new players entering both commercial and open source areas driving further innovation. Typically as in any new market there can be an expected explosion of new entrants followed by a rationalisation towards a consolidation of the market by players that have identified a viable business model and are in a position to acquire other players and control an industry sector.

    To understand and model this future projection better it is important to monitor if UK businesses and organisations are planning to leverage the technology and why they are attracted to use it.

11. LLMs require a lot of compute power and, hence, incur a relatively high financial and environmental (in terms of energy usage) cost burden. This can be prohibitive for both startups entering the space and organisations wishing to adopt the technology and we could see smaller firms on both fronts left behind. However, as LLMs develop we should see both the

[1] https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

scale (amount of parameters they're trained on and how much information they can process) and efficiency (improved the computational efficiency so they use less resources) increase, making them more easily leveraged.

12. The reducing cost of operating LLMs and the associated fall in cost could be modelled as per the predictable market model that moves products towards becoming commodities due to competition. AI should be recognised as being in the early stage of becoming a new commodity, like electricity or cloud infrastructure.

13. There will be more models trained on domain-specific data. This will allow for domain-specific models, like BloombergGPT[2] that recognises financial terms and is specialised for the finance domain. Specialist models can be used out of the box without the need for further fine-tuning and will provide more accurate specialist responses meaning applications can be built without the need to undergo expensive fine-tuning allowing teams to focus on application development. This highlights the impact of LLMs to professions and industry sectors, and hence suggests engagement with industry sectors and petitioning of reputable representative bodies (we used to leverage guilds) to describe and monitor the uses of AI that operate in and affect their sector.

14. As models continue to mature, more emphasis will be put on ensuring safeguards and prevention of bias, unintended and toxic responses. This will present itself in the form of inbuilt controls within the LLMs and also in a range of tooling[3] that is already emerging and will continue to develop further. Tooling will also be built to explain reasoning behind responses which could be an important factor in generating trust with generative AI technology, particularly around easing impending fears with copyright and content ownership.

    There is a natural causal driver for the evolution of trust within the development of LLMs as it is directly connected to an LLM's commercial viability. This is unlikely to need to be forecast but can of course be monitored.

15. While there is a lot of excitement and an urge to use generative-AI technology, understanding the return on investment is often an afterthought. There are already methods that can measure the effectiveness of LLMs but none are perfect and we would expect these techniques to develop further over the coming years and patterns of use emerging to eventually reach a consensus over standard approaches for different use-cases.

16. The types of business models being used to underpin LLM-based services should be studied and documented in a detailed form. To the degree that regulation may be required to investigate how a business model actually functions i.e. Who are the customers? What is the marketplace? How is

---

2 https://arxiv.org/abs/2303.17564

3 https://shreyar.github.io/guardrails/, https://github.com/NVIDIA/NeMo-Guardrails

revenue generated? What suppliers and commodities are required? What costs are incurred including those of ecological impact?

17. Generative AI not only covers text generation but also image, video, audio and others. Such models that generate more than one type of modality are known as multi-modality or multi-modal models. They are already capable of generating impressive content[4] and this capability will improve further. Multi-modal LLMs will be able to provide queries across multiple types of medium that could further accelerate adoption in areas like content generation, transport, education, entertainment and healthcare to name but a few.

18. Multi-modality is a function of available compute power and the cost of that commodity within a viable business model. By employing analysis of these 2 factors i.e. the forecast increase in available compute power and evolution of business models to pay for it, the trajectory of multi-modal solutions will be more easily forecast.

19. There is always a balance between introducing regulation and allowing innovation to prosper responsibly. Typically low regulation is required to foster research and innovation prior to public launch, and sensibly, regulation is required in the go-to-market stage - we note such regulation does not currently happen in any area of the world even though it is clearly the sensible and obvious approach to all new technologies.

20. Of course there are ways that the private sector can work with governments to inspire confidence that technology is being developed responsibly. A private environment would provide a better space for companies like Google and Meta, both of which have footprints in the UK, to report on ever evolving industry risks, developments and opportunities but it does not necessarily fall within their interests to reveal all risks. Going down the regulatory path, new legislation will be needed to mandate software service providers leveraging LLMs to implement a minimum bar of guardrails to ensure sufficient protection. At the very least, working groups can be setup to develop an industry standard such as the OWASP top 10[5] used for managing cyber security risks and threats for web applications. This is explored further below.

21. The trajectory as it relates to the area of regulation is best understood by partnering with other governments and seeing the commercialisation of LLM capabilities as a global opportunity and a global threat. The foresight awarded through what's happening elsewhere is a helpful tool for what might be happening at home under one's own eyes, but going unseen.

---

4 https://elevenlabs.io/, https://runwayml.com/, https://www.midjourney.com/

5 https://owasp.org/www-project-top-ten/

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

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

22. LLMs will be used across many sectors to automate mundane tasks and increase productivity where properly implemented. Initially at least, the technology will assist staff rather than replace them allowing them to be used for more important tasks that require human reasoning. There is a risk that this could change as the technology improves to a level of reasoning and understanding that can generate extremely accurate responses equaling or even surpassing those of humans resulting in the reduction in demand for skilled knowledge-based workers.

23. Of course, in order for this technology to be used there needs to be enough skilled practitioners capable of building the tools and products. Without an approach to produce many skilled practitioners to implement these solutions, the UK risks falling behind. Regardless, more people will need to become "data and technology literate" as job roles change. Investment in education in schools and universities to produce more skilled practitioners as well as introduce schemes to help businesses utilise these practitioners to build solutions that significantly increase productivity would be one approach to harnessing the opportunity this technology offers. In the short-term, however, immigration schemes could be introduced to make it easier to employ skilled workers from outside the UK that could plug the skills shortage gap. There is an opportunity to further educate and reward UK people as part of the continuing evolution of information technology and its growing role in how we organise, interact, and live our lives.

24. We will see many opportunities in new developments across all industries as a result of LLM deployment at scale, such as in the healthcare sector where for example we'll see the ability to use LLMs to provide personalised health plans according to a patient's genetic profile, discovery of new medicines and expanding the use and breadth of scientific data and literature. And in the shorter term we will see the development of solutions to address more laborious administrative tasks like taking transcriptions and medical results to produce medical reports or analysing medical scans and images to detect anomalies.

25. LLMs are very good at processing vast amounts of data to derive valuable insights. There could be an opportunity for them to play a role in developing new climate models that can predict trends or impacts of climate change by processing masses of climate data and research papers along with satellite imagery. This could in turn help governments make more informed policy decisions that impact on climate change. On the same topic, as more models are used widespread, this will also lead to a tremendous increase in compute and energy requirements and could have a detrimental impact on climate goals. Introducing policies to encourage businesses and public cloud providers like AWS, Google and Microsoft to use renewable-only-powered data centres could help mitigate this.

26. The risk of all this is as the technology develops, there will be more opportunities for criminal actors to utilise the capabilities of LLMs. For example, they can imitate another person's voice[6] and socially engineer unknowing parties like a support agent into revealing additional information about them. Combining with a corpus of publicly available social media data, a person's identity could easily be stolen. The risk of this applies beyond the personal, however, as government officials and industry leaders are prime targets for nefarious actors.

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?

27. The AI White Paper presents a strong foundation for ethical AI regulation. The approach to have sector-specific regulators implement the framework with support from central functions vs a centralised regulatory body providing regulation across all sectors inevitably comes with trade-offs. How the new central functions are implemented and operate in practice will determine the effectiveness of this approach. The suggestion to create industry standards and guidelines around the use of AI models will help businesses adopt the technology responsibly. While the principles and spirit are well-intended there is room for improvement.

28. Industry-specific expertise can be harnessed and applied to best use but could present problems across the regulatory landscape in the way different regulators interpret the framework. This could lead to a fragmented and incoherent regulatory framework that leads to confusion and compliance challenges for businesses. The cross-regulator collaboration mentioned in the White Paper will be a key part of implementing the framework and it will be vital for the UK government to support this by creating forums to encourage a consistent joined-up approach to implementing the framework across all sectors.

29. The rapid pace of technological change in AI will require regulators to keep updated with the latest developments and potentially the framework. Such developments may be outside their specialised domains and could hinder regulators from being able to provide effective oversight. Regulators will need to build out this function which may require hiring in experts from outside their domain or upskilling existing staff.

30. Applying the framework and setting up the central functions to support regulators will introduce financial burden that the government needs to fully cost. The investment needed should not be underestimated in order for it to be effective. Inadequate investment puts it at risk of being an expensive, toothless fanfare with no tangible benefits, undermining public trust in the technology and stifling the opportunities of adopting the technology.

---

6 https://pca.st/kl34tchs

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

31. There are advantages and disadvantages associated with the proliferation of open-source LLMs. The biggest benefits of open sources are accessibility to the technology that will drive experimentation and innovation. For those that have the skills, they can integrate the technology into businesses to automate administrative tasks, increase efficiency and productivity gains it promises. This would have previously been limited to academia and large tech companies with large budgets available to fund such projects.

32. Open-source models can also help foster public trust in the technology as the exact data sets used to train the models can be accessed. Businesses can also apply more control over the technology by deploying open-source models into their own infrastructure and ensuring their own customers' data does not end up used by third party service providers like OpenAI[7].

33. On the other hand, open-source models also allow for bad actors to use the technology. Misinformation and disinformation are an obvious application that we will undoubtedly see more of over the coming years. Automated campaigns could be much more effective at targeting would-be victims, increasing the chance of producing skewed election results or shifting public opinion on a certain subject.

34. We may see more advanced adversarial cyber attacks against AI systems themselves if malicious actors are able to access the underlying code of LLMs. It will be important for new adopters to ensure there are safeguards in place to control access to the LLM to avoid exacerbating the impact of an LLM exploit. Similarly, LLMs could be used by malicious actors to find security issues with any publicly accessible open-source software. Coupled with the knowledge of the organisations that use such software, the businesses could be targeted to exploit the newly found vulnerabilities.

4. Do the UK's regulators have sufficient expertise and resources to respond to large language models?[5] If not, what should be done to address this?

35. Addressing the adequacy of UK regulators in responding to LLMs is crucial to ensure responsible governance of this technology. While some regulators may have a baseline level of expertise and resources to deal with AI and LLMs, others may not and there is room for improvement and adaptation to the evolving landscape.

36. Firstly, regulators should focus on enhancing their expertise. This can be achieved through training programs and collaborations with AI research institutions. Hiring experts in AI ethics, fairness, and safety can further bolster their capabilities. To effectively address LLM-related challenges, regulators could benefit from forming cross-disciplinary teams with their

7 https://openai.com/

newly hired experts. These teams would comprise technologists, ethicists, legal experts, and domain specialists, providing a well-rounded perspective on the multifaceted issues posed by LLMs.

37. Partnerships with industry, academia, and civil society organisations are also vital. Such collaborations can facilitate knowledge exchange and resource sharing, allowing regulators to tap into external expertise and resources. Such a forum could be facilitated by one of the central functions referenced in the White Paper. Regulators should actively participate in research efforts focused on LLMs' societal and industry-specific impacts. This includes studying bias and fairness issues, understanding the technology's capabilities, and assessing its implications for different sectors.

38. Ensuring regulators have access to adequate technical resources, such as computational infrastructure and tools for analysing LLMs, is essential. These resources enable them to effectively evaluate AI systems. This cost should not be underestimated and should be factored as part of the investment into the framework mentioned earlier.

39. Regulators must not work in a vacuum but connect with innovative thinkers within the UK and globally. The regulation of AI needs to be global not local. It is akin to regulating carbon emissions in that no matter which country we are in, the pollution is atmospheric by default. All countries need to agree and conform.

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

40. Non-regulatory options could include sponsoring research and development into AI in general. This includes understanding their capabilities, limitations, and potential societal impacts as technologies and opinions further develop. Public and private investments in AI research can also lead to advancements in LLM technology and help shape the direction of travel. Investment could also promote connecting professionals and interested parties to support the reinforcement of the UK as a hub for AI innovation (see the success of the Data Lab[8] in Edinburgh in fostering data science in the region).

41. Non-regulatory options also include the education of the general public into what AI is, how it works, what it delivers and promises and how to use it. It is important to encourage mature conversations and debate about all new technologies at a societal level. This approach will yield significant insights and opportunities that would otherwise be non-starters.

42. The development of standards and guidelines for the ethical design, implementation, deployment and management of applications using LLMs can help ensure they include fairness, transparency and accountability as foundational principles. They will also guide businesses in understanding

8 https://thedatalab.com/

the impact and risks associated with deploying AI solutions such as identifying potential biases and such ethical concerns.

43. Areas for regulation could include ethical principles, compliance standards and penalties for non-compliance. As the technology is adopted, there may be a need for a similar model to how GDPR was introduced with businesses appointing individuals responsible for ensuring compliance.

44. To support the regulation and standards, auditing and certification could also be introduced. Businesses and individuals would gain a level of confidence in knowing they are dealing with an ethical and responsible party that gains certification. High standards should be met but such auditing and certification should not be so bureaucratic and cost prohibitive as to close out smaller businesses from being able to gain it.

45. Sector-specific regulations may be required for specialist areas like healthcare or finance. Essentially, in domains that could have far reaching societal impacts or risk to life. Regulations can enforce guardrails to protect against such circumstances. As touched on earlier, a shortage of skilled software engineers, data scientists and associated practitioners will inhibit the adoption of this technology. Identifying industries that have most to gain and removing barriers to hire in those sectors will address the short-term skills shortage. Investment in education and training right through all age groups will help ensure the UK is well equipped in the longer term.

46. More radically it may be pertinent to globally regulate the operation of any commercial LLM or AI instance such that the supplier of the commodity service e.g. Amazon Web Services, is required by law to identify the party using the capability, and that the party needs to explicitly state the details of their business model and where the AI/LLM use fits in, including the details of risky elements such as the source of any proprietary data used for the LLM. As this relates to Open Source solutions it is indeed possible to legislate that these too cannot be executed without a legal identity on record with disclosure of purpose and relevant details of the solution.

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

47. The UK government could allocate funding and resources to support AI research and development, particularly in the field of LLMs. Collaborations between academic institutions, businesses, and government-backed research initiatives would be fostered, potentially similar to how Innovate UK operates. However, barriers may include limited public funding availability and international competition for top AI talent.

48. One of the central functions could collaborate with industry and academia to develop a set of standards and guidelines for the practical application of LLMs. Once a baseline has been created, industry-specific extensions could be developed, for example in healthcare. Further collaboration also involving auditors can then lead to the development of assessment frameworks and certifications. Awarding the certifications could be granted by an awarding body such as UKAS and recorded in a centralised

registry similar to ISO 27001 certifications and would be valid for a set period. Continued assessment would be required to ensure ongoing compliance.

49. As an alternative approach or a second-tier option, a self-certified assessment could be applied, similar to how SWIFT introduced the Customer Security Controls Framework. An independent internal audit department separate to the technology or product team managing the AI application would assess compliance with the framework. Results could then be logged in a central registry and areas of concern followed up by awarding bodies or regulators. This approach could allow smaller businesses to achieve certification without having to incur potentially expensive auditors fees.

50. Regulation could be applied to businesses above a certain size (people, revenue, value, etc) that would likely include the most widespread applications of LLMs. Where there are more critical consequences of misuse of this technology, like loss of life or harm to individuals, stiff penalties need to be introduced and enforceable by regulators.

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

51. Ideally regulation will cover all stages of the AI life cycle. However, the most effective areas are provided below:

52. Research, testing and validation: Regulation in this stage can be highly effective because it helps ensure that AI systems are thoroughly tested for the protection of human rights, safety, fairness, and reliability before deployment. It can establish standardised testing procedures and criteria, reducing the risk of errors and unintended consequences.

53. Explainability and accountability: Regulations that mandate transparency and accountability measures are crucial for holding organisations responsible for the actions of their AI systems prior to them being deployed. They can promote transparency in AI decision making, making it easier to understand and audit the system's behaviour once in wider operation as well as contest judgements where a party is wronged.

54. Deployment and commercial operation: Regulations are most essential here ensuring that AI systems are used responsibly within societal standards and that organisations continuously monitor their performance. They can address issues related to system misuse, bias, and the ongoing maintenance of AI systems to prevent harmful impacts.

55. Retirement and decommissioning: Effective regulations in this area are essential to ensure that AI systems are safely retired when they become outdated, unsafe, or are no longer needed. Proper decommissioning procedures and data handling during retirement are critical to protect sensitive information and minimise risks.

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

56. A multifaceted approach can be taken to mitigate these risks effectively through a combination of proactive measures, transparency initiatives, research efforts, and ongoing evaluation. First and foremost, robust ethical guidelines and regulations must be established. These frameworks should emphasise transparency, fairness, accountability, and non-discrimination. It's imperative to ensure that human oversight is integrated into AI driven decision making processes, particularly in critical applications such as healthcare, finance, and legal systems.

57. Accountability and liability frameworks need to be clarified. Legal responsibility should be defined for unintended consequences arising from LLM use, distinguishing between developers, users, and other stakeholders. Exploring the need for AI-specific liability insurance and compensation mechanisms for individuals harmed by LLM-generated decisions is also essential.

58. The guidelines and compliance standards referenced earlier also play a pivotal role in risk mitigation. These assessments should encompass bias and fairness evaluations, aiming to identify and rectify biases in LLMs that could perpetuate discrimination or reinforce stereotypes. Additionally, conducting thorough risk analyses is crucial to evaluate potential consequences, including the propagation of misinformation and data privacy breaches.

59. Encouraging the development of highly explainable LLMs through regulation and auditing frameworks makes their decision-making processes more understandable to users and regulators. Moreover, promoting third-party audits of LLM algorithms to assess fairness, transparency, and compliance with regulations can enhance trust in their use.

60. International collaboration is crucial. Collaborating with global partners to develop common standards for LLM development and use ensures consistent ethical practices. Sharing information about AI best practices and lessons learned on an international scale promotes responsible AI innovation.

5 September 2023