# Capabilities and Trends of Large Language Models: BCS' response to the House of Lords Communications and Digital Committee's call for evidence.

September 2023

Prepared by:

Arnoldis Nyamande

BCS Policy Manager

## Table of Contents

## Summary of BCS' position

BCS's position on developing Large Language Models (LLMs) over the next three years is grounded in a cautious yet optimistic outlook. This response builds on BCS' flagship paper entitled: 'Helping AI Grow Up Without Pressing Pause'. We acknowledge the incremental evolution of LLMs, characterised by a trend towards smaller, more accessible models that offer personalisation and integration into various applications. However, we emphasise the importance of addressing the environmental and economic challenges posed by the intensive training requirements of LLMs, which may limit their widespread adoption. BCS also underscores the need for responsible funding, monitoring, and ethical considerations

to navigate the risks associated with LLMs, including the potential for misinformation and biased responses. Ultimately, with careful management, LLMs can reshape various domains, offering opportunities and challenges that require ongoing research and ethical usage.

## Question 1: How will large language models develop over the next three years?

In the upcoming three-year period, the evolution of large language models (LLMs) is expected to follow a path marked by incremental changes, although the precise course remains uncertain. LLMs have been undergoing a reduction in size, while concurrently expanding in capabilities. The size of an LLM is significant as it affects both energy efficiency and processing speed. The trend towards smaller models is gaining traction due to their heightened accessibility, decreasing reliance on specific providers such as Microsoft or OpenAI[1].

These compact LLMs offer personalisation[2], allowing responses to be tailored to individual user requirements, and they can be seamlessly integrated into operating systems. Given that such pre-trained models are based on Machine Learning (ML), they possess the capacity to continue to learn from new data gathered about a user's preferences and interests. This potential for personalisation means smaller LLM models can be bundled in with an operating system (OS), increasing the capabilities of personal devices. It is worth noting, however, that while LLMs like those underpinning ChatGPT have been trained on a wide breadth of general-purpose knowledge sources, their capability is limited when delving into specific domains. For example, while they may be able to answer questions about general human biology, they would struggle to accurately understand, interpret and diagnose specific medical conditions or do non-trivial arithmetic[3]. This limitation gives rise to the possibility of specialised models focused on distinct topics, heralding an era where numerous models excel in specific areas, rather than relying on a single generalist model.

Despite their advantages, LLMs present challenges due to their demanding training requirements, which can be costly and create barriers to entry. LLMs are trained by feeding them vast amounts of data, teaching them to identify patterns and trends and use that information to reach a desired outcome. This process requires processing vast amounts of data in data centres, which comes at a great economic and environmental cost[4].

Only the most well-resourced actors can explore the boundaries of LLMs at any significant scale. One reason is that many organisations lack enough Graphical Processing Units (GPUs) – specialised computer chips which, among other things, help computers handle video games, cryptocurrency, mathematical and ML-based

[1] https://arxiv.org/abs/1503.02531

[2] https://arxiv.org/abs/2302.13971

[3] https://blog.research.google/2023/08/teaching-language-models-to-reason.html

[4] https://arxiv.org/abs/1906.02243

applications (like LLMs) more efficiently. There are 'GPU-Rich and GPU Poor[5]' organisations, which is a barrier to innovation.

Training algorithms, retaining data and the uncurbed growth of ML models like LLMs is already proven to be having detrimental impacts on the environment, and adverse impacts on the global effort to reduce the emission of greenhouse gases[6].

The development patterns with LLMs often see gradual progress, with an occasional leap that elevates their capabilities. LLMs build upon existing neural network architectures from 1958[7], with a big leap forward in 2017 with the invention of transformer models[8]. These are ML techniques that learn statistical patterns from data rather than explicit logic. There is also another branch of AI research, symbolic AI, which seeks to categorise and build interpretable rules and models based on knowledge and logic.

ChatGPT uses LLMs, lots of conventional software code, and apparently also some symbolic AI. The LLM recognises patterns from extensive examples but lacks any understanding of concepts founded in a knowledge base. The use of other technologies in combination with LLMs are the breakthrough represented by ChatGPT and similar systems.

Elements of symbolic AI are likely to have been integrated to improve ethical responses, including reducing bias. This means that it has been given rules that, for example, make it refrain from giving answers that perpetuate racism, sexism and other forms of discrimination, despite the fact that the data used to train it will have undoubtedly contained elements of these forms of bias. Reinforcement learning is also incorporated to improve an LLM's capabilities based on feedback in the form of human responses.

The future development trajectory remains uncertain, and refining the techniques employed by LLMs could lead to more accurate responses and enhanced technology.

## Improving Understanding of Future Trajectories

### Question 1a: Given the inherent uncertainty of forecasts in this area, what can be done to improve understanding of and confidence in future trajectories?

Given the inherent uncertainty associated with predicting developments, enhancing our comprehension of future trajectories calls for focused actions. Streamlining funding processes for AI research and more technical support like increasing access to GPUs could empower researchers to operate more efficiently. Additionally, investing in local technology talent and providing tech-specific funding for AI and

[5] https://www.semianalysis.com/p/google-gemini-eats-the-world-gemini
[6] https://onlinelibrary.wiley.com/doi/10.1002/advs.202100707
[7] https://en.wikipedia.org/wiki/Perceptron
[8] https://blog.research.google/2017/08/transformer-novel-neural-network.html?m=1

LLMs can nurture innovative businesses and reduce dependency on foreign sources like the US and China.

Efficient monitoring of LLM and AI development, along with responsible funding practices, is essential. To foster meaningful discussions, establishing a platform for scientifically informed discourse rather than anecdotal exchanges is vital. Acknowledging the emergence of abilities in LLMs and redefining success metrics can provide deeper insights into their progress. While leaps in development are unpredictable, supporting continuous incremental improvements can contribute to a better understanding of future trajectories.

## Greatest Opportunities and Risks

### Question 2: What are the greatest opportunities and risks over the next three years?

Anticipated over the next three years are substantial opportunities and risks within the realm of LLMs. One significant risk is the potential for "hallucination", which is when LLMs generate false information, incorrectly combining data sources, or generating randomised tokens that lead to factual inaccuracies. This phenomenon, observed during domain-specific tasks like generating legal cases[9] as well as general enquiries about Australian mayors[10] or American law professors[11], underscores the challenges posed by the sequential nature of automated generation.

Dependence on major tech companies and presumed reliability of said companies can also pose risks, as demonstrated by the lawyer case cited above. Caution is advised in anthropomorphising scenarios, as LLMs are trained to defer to human expertise, potentially reinforcing user biases. Concerns about data breaches and inadequate regulation, exemplified by the Italian ban on data storage, further highlight the need for careful consideration of data handling.

The rise of mis- and disinformation and content generated by LLMs poses significant societal risks, potentially exploited by malicious actors to manipulate public opinion. The potential harms arising from misuse remain largely uncharted territory. Furthermore, the integration of LLMs across all educational settings raises questions about assessment methods and authentic evaluation.

A lack of understanding of AI risks scaremongering, ‘fake news’, and poor decision-making. On the other hand, there is an opportunity to roll out a programme of education in AI literacy for everyone, rather than just computer scientists as has been the norm until now. Such a programme could cover how to use AI, understanding how AI works, and the ethical and societal implications of AI.

[9] https://www.bbc.co.uk/news/world-us-canada-65735769

[10] https://www.theguardian.com/technology/2023/apr/06/australian-mayor-prepares-worlds-first-defamation-lawsuit-over-chatgpt-content

[11] https://www.independent.co.uk/tech/chatgpt-sexual-harassment-law-professor-b2315160.html

## Balancing Risk Considerations

### Question 2a: How should we think about risk in this context?

Considering risks within their context offers insights into potential outcomes. LLMs hold promise in enhancing efficiency and automating routine tasks. However, the impact on employment remains uncertain and dependent on existing expert knowledge to effectively utilise these technologies.

The emergence of new businesses and job opportunities is possible as LLMs streamline processes. Rather than erasing jobs, LLMs could reshape roles by automating routine tasks and enabling humans to focus on more complex aspects of their work. Nurturing innovation and staying ahead of LLM advancements is advised to avoid playing catch-up.

“Automation Bias”, or the tendency of humans to believe the machine, is a major risk of LLMs, and we have to insist that information we rely on must be backed by verified citations (as the legal case showed).

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

BCS, along with The Royal Statistics Society, National Physics Laboratory, the Alan Turing Institute, The Operational Research Society and the Institute for Mathematics and its Applications is part of the Alliance for Data Science Professionals. This means that registration, professional standards and responsible computing form the basis of our approach to engage with AI safely in general.[12]

The government’s UK AI Regulation White Paper highlights five principles for UK regulators to achieve responsible AI in their specific sectors. Section 4 recognises a central role for tools in trustworthy AI, including technical standards and assurance techniques to help implement the principles.

The technical and operational standards provide management systems, processes and measurement methods to support implementation of the five principles. Organisations should align themselves with these best practices and ensure that their AI technology providers also adopt them as and when they are published.

While significant advances have clearly been made in LLMs, it is not entirely clear that this rapid pace of development will continue, although we expect that LLMs will become more integrated with other tools and become increasingly multi-modal (for instance, consuming and generating images and sound in addition to text).

The concerns from most experts are not that AI is too powerful, but that basic guard rails are not in place to ensure AI is deployed responsibly. It should be noted that the

[12] Helping AI grow up without pressing pause https://www.bcs.org/articles-opinion-and-research/helping-ai-grow-up-without-pressing-pause/

guard rails that need to be put in place are governance-based rather than being technical innovations.

The Institute broadly welcomes the government's regulatory proposals but with caveats.[13] While the light touch approach is positive in that it enables innovation, there are areas that need more consideration to ensure the proposals maximise the public benefit of AI. The proposed cross sectorial principles are appropriate and useful, but should be extended further.

For example, all AI systems must have appropriate safeguards to ensure they remain technically sound and are used ethically under reasonably foreseeable exceptional circumstances, as well as under normal circumstances. Organisations must show they have properly explored and mitigated against reasonably foreseeable unintended consequences of AI systems.

In a sense, we view ChatGPT (and other LLMs) as a distraction in relation to increasing the reasoning power of AI systems. While LLMs are AI systems that can generally respond to human queries well, they do not represent a significant improvement in reasoning ability. We acknowledge that LLMs do appear to be able to reason better on first glance, but that is because they have illusory anthropomorphic qualities.

At this point it is important to keep monitoring the consequences and impact of LLMs, however as this is an emerging technology it is too early to say if we need tailored regulatory approach for this specific use of AI.

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

As with most emerging technologies, there is a steep learning curve. We would suggest that all stakeholders have a basic understanding of the ethical considerations around AI and LLMs, and the challenges and opportunities these technologies bring.

As the professional body for information technology a major part of BCS' remit is the development of digital skills and literacy, an ethical understanding of technology, and the value of competent professionals from diverse backgrounds to deliver change.

We would welcome the opportunity to discuss further with the government how we can help.

## Long-Term Implications and Responsible Usage

The development of large language models in the next three years remains uncertain yet promising. Their evolution holds the potential to reshape various

[13] Light touch approach to AI regulation welcomed by IT industry body https://www.bcs.org/articles-opinion-and-research/light-touch-approach-to-ai-regulation-welcomed-by-it-industry-body/

domains, offering both opportunities and challenges that require careful consideration, ethical usage, and ongoing research.

## Thanks to the following for their contributions to this consultation:

Dr. Matthew Shardlow, Senior Lecturer at Manchester Metropolitan University; Professor Adrian Hopgood FBCS (BCS AI Specialist Group); Adam Leon Smith FBCS - Chair of BCS, The Chartered Institute for IT Fellows Advisory Group (F-TAG) and CTO Dragonfly; Professor James Davenport FBCS FIMA CITP CMath

## About BCS

The purpose of BCS, as defined by its Royal Charter, is to promote and advance the education and practice of computing for the benefit of the public. We bring together industry, academics, practitioners, and governments to share knowledge, promote new thinking, inform the design of new curricula, shape public policy and inform the public.

As the professional membership and accreditation body for Information Technology, we serve 70,000 members, including practitioners, businesses, academics, and students, in the UK and internationally. We accredit the computing degree courses in ninety-eight universities around the UK and as a leading IT qualification body, we offer a range of widely recognised professional and end-user qualifications.