Written Evidence Submitted by NCC Group
(GAI0040)

Introduction

NCC Group welcomes the opportunity to respond to the Science and Technology Committee's call for views and offer our expertise as a UK headquartered, globally-operating cyber security and software resilience business.

NCC Group's mission is to make the world safer and more secure. We are trusted by more than 14,000 customers worldwide to help protect their operations from ever-changing cyber threats. Recognising the increasing convergence of cyber security and safety in the connected world including in the application of artificial intelligence (AI), we recently announced the acquisition of Adelard – a well-established UK computer system safety advisory business – extending our risk management service offering into the field of safety critical systems. To ensure we match the rapidly evolving and complex technological environment, we continually invest in research and development as an intrinsic part of our business model. **We have many years' experience[1] researching AI and machine learning (ML) to understand the risks and opportunities these technologies present.** Most recently, NCC Group's Chief Scientist, Chris Anley, published a Whitepaper[2] collating details of practical attacks on ML systems for use by security practitioners to enable more effective security auditing and security-focused code review of ML systems.

Through our work and research, we are acutely aware of the rapidly evolving use of ML and AI across the economy, and the risks and opportunities this presents. We are therefore delighted that the Committee is taking the time to review the way in which we govern and regulate AI systems in the UK. **We believe that it is crucially important that security implications are considered from the outset, embedding 'secure by design' principles.** We are keen to ensure that security and safety considerations are not seen as a blocker or a cost, but as an enabler of future-proof systems that, by their design, avoid mistakes that are expensive and otherwise costly to fix later. It is through this lens that we look to offer our input to the call for evidence.

Definitions

We believe that some confusion has arisen around the terms AI and ML. So, to aid clarity, we define these terms below:

- ***Artificial Intelligence (AI)*** is an overarching term for systems that employ computer intelligence. This includes, for example, systems that can play games against humans, or systems that automate creative processes such as legal writing.
- ***Machine Learning (ML)***, for us, is a subfield of AI and computer science that provides computers with the ability to learn, without being explicitly programmed, when exposed to new data. This is done through the study and construction of algorithms that produce models from training data which are then used to make predictions on further data. In that context, supervised learning entails an algorithm being trained with labelled data, such as using natural language processing (NLP), to extract relevant textual data from legal documents. Unsupervised learning entails the algorithm making its own decisions

[1] Offensive Security & Artificial Intelligence – NCC Group Research
[2] Whitepaper – Practical Attacks on Machine Learning Systems – NCC Group Research

and inferences, such as arbitration or contract-negotiation mobile phone apps. Reinforcement learning entails data being presented as a dynamic environment, such as in autonomous vehicles.

The security and safety landscape

There are two primary security risks we perceive with the adoption of AI. For one, **AI and ML algorithms are, by design, susceptible to influence and change based on their inputs and lifetime.** This presents the opportunity for significant security risks, particularly from adversarial ML[3] attacks. We also believe that it is **inevitable that attackers will start using AI and ML for offensive operations** to aid their own efficacy (e.g. the use of AI to augment the discovery of security vulnerabilities, or the use of deepfakes for fraud/misinformation). In our experience, research and tools to support both scenarios are becoming more accessible, datasets are becoming larger, and skills are becoming more widespread. We believe that once criminals or maligned state actors decide that it is economically rational or valuable to use AI and ML in their attacks, they will.

The democratisation of technology and its widespread availability risks inadvertent consequences too. As a result of a growing number of openly accessible AI/ML frameworks becoming available to software developers that abstract data science and algorithmic details, **developers may deploy ML and AI systems without necessarily understanding their underlying mathematics and associated operations, leading to potentially poor outputs**.

In addition, as technologies like AI become increasingly ubiquitous across society and the economy, **the potential for bias exists**. There have been reports[4] of AI-based facial recognition tools repeatedly falsely identifying minority groups and genders – where individuals with multiple minority characteristics are particularly at risk. To build a safer and more secure future for all, removing or reducing inherent existing biases while balancing data privacy needs and taking steps to ensure that social issues are not exacerbated, will be crucial.

A regulatory and governance framework

Against this backdrop, we broadly support the UK Government's endeavours[5] to create a common framework for AI governance that is **risk-based and delivers on the Government's ambitions to promote innovation, while keeping the UK and its allies safe and secure**. We share wider industry views that such a framework will provide much needed clarity and consistency across sectors, ensuring a **level-playing field** for organisations developing and deploying AI and ML systems. We believe, however, that the **Government's plans could be strengthened in the following ways** to build trust in AI and ML technologies and cement the UK's position as a global leader:

- **The UK's risk appetite should be clearly defined.** AI will never be zero risk if we wish to pioneer its development and, crucially, its deployment. As such, the UK should define its risk appetite so that red lines with regards to AI and ML systems and their security, safety and resilience are known.

---

[3] Adversarial ML describes an attack whereby ML models are manipulated – usually by manipulating data inputs – with the objective of causing the model to make incorrect assessments.
[4] For example: Many Facial-Recognition Systems Are Biased, Says U.S. Study - The New York Times (nytimes.com)
[5] As set out in its recent policy paper: Establishing a pro-innovation approach to regulating AI - GOV.UK (www.gov.uk)

- There needs to be a shift away from the current reliance on advice, guidance and voluntary measures to secure the UK's digital systems towards **more stringent, forward-looking regulatory intervention and mandating of security requirements**, prioritising higher-risk applications.
- **Flexibility, agility and periodic regulatory and legislative reviews** should be built in from the outset to keep pace with technological and societal developments. This could include requirements for regulators and policymakers to engage regularly with innovation centres and industry experts, drawing from a wide range of backgrounds (including safety and security) and generational perspectives.
- To achieve genuinely forward-looking outcomes, the Government should **invest in coordinating and improving horizon-scanning**. At present, there is a myriad of horizon scanning activity and initiatives across government, the private sector and academia, as well as multiple government bodies and advisers whose remit involves considering future risks and opportunities[6]. This can lead to overlap and duplication of effort, with no central coordination and collation of data. We were therefore pleased when the Government committed[7] to consolidating and joining up horizon scanning across the public sector and note that the Digital Regulation Cooperation Forum will play a key role in this regard.
- Efforts should be made to move away from a 'tick-box' compliance approach to security regulation to one where there is a **true understanding of cyber threats, greater 'buy in' and accountability at an organisational level**. This should include regular and independent assessments of real-world resilience at the organisational and system level.
- In assuming a greater role in regulating the use of AI and ML, **digital and sectoral regulators should be strengthened in their powers, resources and capabilities**.
- There remains a significant shortage of the skills we need to develop AI frameworks, and assure systems' safety, security and privacy. If the UK wants to be a global leader in AI, the Government must focus investment on **developing the skills we need** to make its regime a success.
- **The drafting, approval and implementation of technical standards** that underpins the UK's regulatory framework will be critical. We know that the UK is already taking steps to be at the forefront of developing world-leading AI standards, and this is something we firmly support. We ask that, in the forthcoming AI White Paper, a clear route map for the development of technical standards is laid out, detailing where those standards are cross-sectoral and where sector-specific standards are required (e.g. medical devices).

Response to questions

How effective is current governance of AI in the UK? What are the current strengths and weaknesses of current arrangements, including for research?

In our experience from operating within the world of cyber security, industry does not always learn from the lessons of others. Indeed, despite daily publicised data breaches, many organisations continue to make the same mistakes that eventually result in their own data breach or cyber incident. Incidents that could have been avoided by following industry best practice and learning from the mistakes of others. To avoid this happening with AI and ML technologies, assurance and testing, backed up by investment in research, will play an

[6] For example: Regulatory Horizons Council, Departmental Chief Scientific Advisers, Science Advisory Councils, UK Research and Investment (UKRI) etc
[7] https://committees.parliament.uk/publications/9464/documents/161530/default/

important role in ensuring organisations involved in the development and deployment of AI and ML are taking the right steps and learning from the mistakes of the past. Such activities need to be undertaken on a continuous basis to ensure vulnerabilities are addressed and the latest threat landscape is understood and acted upon. In addition, where the risk profile necessitates, we believe independent, third-party product validation should be mandated. In our experience, many claims made by AI and ML product vendors, predominantly about products' effectiveness in detecting threats, can be unproven or lack independent verification.

What measures could make the use of AI more transparent and explainable to the public?

Most AI-based products in use today are 'black box' appliances that are placed onto networks and configured to consume data, process it and output decisions without humans having much knowledge of what's happening. This means understanding and explaining why an AI-based system reached a certain decision can be very difficult. One area of evolving research and development that could resolve this issue is 'explainable AI'. Explainable AI tools help technical experts to understand how and why an AI reached an outcome. It is an important area of research that we believe should be prioritised for investment.

In addition, we also see potential problems with the use of predictive AI and ML, where correlation may be confused with causation. For example, recidivism scores – used to assess whether an individual convicted of a crime is likely to reoffend – can be based on statistical correlations, such as low income, rather than causations. Some have argued that this could result in people from low-income households being automatically assigned a high recidivism score, and, as a result, would be more likely to receive a prison sentence[8]. 'Causal AI' – which can help identify the precise relationships of cause and effect – could have a greater role to play, alongside explainable AI, in deepening developers' and users' understanding of the root causes of outcomes and ensuring correlation is not mistaken for causation.

However, the need to be transparent and explainable (and by extension ensure the effectiveness and safety of systems) must also be balanced with privacy rights. The Government's approach to enabling transparency should reflect this.

How should the use of AI be regulated, and which body or bodies should provide regulatory oversight?

We support a context-driven, outcomes-focused, proportionate approach to regulation, which understands and reflects the potential harm-benefit profile of the use of AI. For example, we are increasingly seeing the development and deployment of cyber-physical systems[9] that are underpinned by AI, such as autonomous vehicles, medical devices and unmanned air systems. In these instances, the security of the AI is critical to the physical safety of the systems. The safety risk will, however, differ depending on the application of the system and its kinetic effects, and this should be reflected as part of a proportionate risk management approach. To this end, we propose that the UK Government, working with regulators, identifies "high risk" sectors or safety-critical applications, and that more stringent

[8] The Case for Causal AI (ssir.org)

[9] We note that BEIS is undertaking a review into how it promotes and regulates cyber-physical systems. Given the widespread application of AI and ML systems in the cyber-physical world, it's critical that the review is aligned with the forthcoming AI White Paper.

requirements are applied in these circumstances. This should align, where possible, with similar frameworks being developed in other jurisdictions, including the EU.

In addition, the UK's regulatory framework must promote best practice data management measurements. There are two core components to AI and ML-based applications: (1) the algorithms themselves and (2) the data they use. In the end, any application is only as good and fair as the quality of data used to train it. To ensure AI and ML-based applications can consistently produce reliable outputs, it is essential that steps are taken to ensure the data is up to date, secure and, as far as possible, free from bias. These steps should include:

- Establishing clear processes and mechanisms through which applications can be carefully vetted and their respective data supply chains sanitised, particularly where data originates from untrusted sources, such as the Internet and end-users.
- Where proportionate to do so, updating and retraining applications with the latest available data.
- Analysing datasets to ensure they are representative and appropriate for the jurisdiction in which they are used. This should take into account the diversity of the development team responsible for sourcing the datasets, as this may result in unconscious biases. Creating synthetic data (i.e. information that is artificially manufactured rather than generated by real-world events) that is representative could be a future solution to this.
- A multidisciplinary approach to reviewing the decision criteria used for automated decisions - which offers a legal, policy and operational perspective in addition to a technical review – should be taken to reduce bias wherever possible.
- Responsibility for issues of bias shouldn't end when products or systems have been released. There should be a clear reporting process that allows organisations to receive and act on information about potential biases in a system. Lessons can be learnt from the security industry where there are established protocols for disclosing vulnerabilities in a system. In addition, the Government should consider establishing an Ombudsman, or similar regulatory body, that allows individuals and organisations to appeal any automated decisions that they believe to be unfair.

Notwithstanding the need for these safeguards, the experience of the global pandemic has shown us all that the use of data, at scale, can save lives. The Government's regulatory framework should therefore recognise that data, applied, used and shared in a responsible way, can lead to good societal outcomes.

The effective regulation and governance of AI will also require individuals with the right skillsets, including across the following areas:

- **Within regulators:** It will be important equip sectoral regulators with the resources and skills to effectively enforce the adoption of safety and security standards. While some regulators are more advanced in their understanding of and ability to regulate AI, in our experience working in the sector, there remains a significant skills gap across authorities which will only widen as technologies and applications evolve. In the short term, the Government should consider whether drawing in external expertise is needed to plug the skills gap. In the long term, in addition to skills investment within the civil service, a requirement could be established within all regulatory frameworks to regularly and systematically engage with the AI ecosystem. This includes academics, incubators and accelerators, disruptors, Catapults and other innovation centres to understand technological developments, in order to exchange views with industry experts and technologists (including digital natives). This engagement could take a number of forms,

including: secondment models as pioneered through the National Cyber Security Centre's Industry100 (i100) scheme; formal and informal government consultations and calls for evidence; and regular sounding board mechanisms such as advisory groups and councils, or the Department for Digital, Culture, Media and Sport's 'College of Experts'.

- **The development of AI frameworks:** Focused skills investment is required to ensure genuine UK leadership in AI and ML which we would define as producing core AI frameworks, as opposed to using AI frameworks developed by others. Indeed, AI/ML are closely linked to data science and are very mathematical subjects. While there are many AI frameworks available for use that abstract away from the low-level minutiae/ mathematics of AI, there is likely a major skills shortage of people with deep technical understanding of AI and its algorithms. **There is therefore a danger that, as a nation, the UK will be using AI frameworks developed by other nations**, reliant on the assurances that they provide regarding the security of those frameworks. We strongly believe that this is a much less desirable outcome to being in a position where the UK is the producer of the core AI frameworks (that others might then use).
- **Assurance:** Assurance, as we outlined above, will play a core role in determining compliance with the new regulatory regime. However, there remains a distinct lack of people in the AI assurance sector with the experience and/or qualifications to undertake assessments, particularly assessments of cyber security risks that are unique to AI systems. We believe that policymakers should consider how post-16 education can be more appropriately geared toward developing educational programmes that bridge AI and related disciplines such as cyber security.

What lessons, if any, can the UK learn from other countries on AI governance?

While we do not have views on other regimes the UK should look to learn from, we would emphasise that when it comes to the digital sphere, no country is an island. International regulatory cooperation should be front and centre of policymakers' minds when developing the UK's approach to AI. In aligning the UK's domestic and international approach, we recommend that the Government:

- Utilises existing successful partnerships, including the 'Five Eyes' alliance;
- Invests time in developing practical outcomes with other governments, that go deeper than high-level principles; and,
- Ensures that civil society and industry – who will play a central role in delivering governments' objectives - are involved in discussions from the outset.

(November 2022)