Department for Science, Innovation & Technology

Rt Hon Michelle Donelan MP
Secretary of State for Science, Innovation and Technology
3rd Floor
100 Parliament Street
London SW1A 2BQ

The Rt Hon. Greg Clark MP

Chair, Science, Innovation and Technology Committee

House of Commons

London

SW1A 0AA

13 May 2024

PUBLICATION OF THE CALL FOR VIEWS ON THE CYBER SECURITY OF AI

Dear Greg,

I am writing to the Science, Innovation and Technology Committee to inform you of the Call for Views on the cyber security of Artificial Intelligence (AI), which my department will publish on the 15 of May 2024. Please see enclosed a copy of the Call for Views which will run for 8 weeks until 10 July 2024. The Call for Views sets out evidence on the risk landscape of the cyber security of AI models and systems, and seeks feedback on our proposed interventions, which include a proposal to introduce an AI cyber security Code of Practice and seek feedback on how government should implement the Code as the basis for a new global standard on the security of AI models.

Background

AI has become an integral part of our daily lives and is used by organisations and individuals as a powerful tool to enhance the way we work and interact with data. AI developers are pushing the boundaries of what AI models and systems can achieve, which has the potential to bring many benefits for society and the economy. However, proliferation of AI systems and models also creates opportunities for cyber attacks through the manipulation of models and exploitation of vulnerabilities. To unlock the benefits of AI, we need to ensure that AI systems and models are designed securely, and users are protected.

The need to secure AI systems has been recognised in the 2022 National Cyber Strategy.[1] This was reflected through the AI Safety Summit in November 2023, where the cyber security

[1] National Cyber Strategy, Cabinet Office, 2022.

of AI was a key theme. The momentum from the AI Safety Summit continued through the publication of the National Cyber Security Centre (NCSC) Guidance on Secure AI System Development alongside the United States Cybersecurity Infrastructure and Security Agency, co-sealed with 18 other countries.[2]

The government believes that the risks to AI models originate from the design and development of the models and systems. As such, the Department of Science, Innovation, and Technology (DSIT) has developed a programme of work that will focus on the cyber security of AI models and systems. This work will align to the work of the AI Safety Institute and AI regulation, which focusses on the safety and security risk that stems from the use of AI. The Code of Practice included in the Call for Views incorporates a secure by design approach to AI design and deployment with the aim of providing assurance to AI users and ensuring that the responsibility is on AI developers and system operators to develop and use secure AI systems and models.

Alongside the Call for Views on the Cyber Security of AI, DSIT will simultaneously launch a Call for Views on the Code of Practice for Software Vendors. The Call for Views on software outlines a proposed Code of Practice containing actions that senior leadership can take to ensure that software is securely developed, built, distributed, and maintained, and that vendors are effectively communicating with their software customers. Outputs from both these Calls for Views will have a significant impact in driving improved security behaviour across the software supply chain and providing increased confidence to organisations adopting digital products and services.

The Code of Practice and future global standard

The main intervention that will be tested through the Call for Views is the AI cyber security Code of Practice and subsequent work to develop a global standard based on the Code. Building on the NCSC's Guidelines for Secure AI System Development and taking account of evidence from research commissioned by DSIT, the Code of Practice sets out 12 principles to improve the cyber security of AI systems and models. The Code of Practice provides technical requirements for AI developers and system operators on what is expected from them to ensure they are meeting baseline requirements. This includes providing assurance to users of AI and supporting the UK market in preparing for further enhancements of AI.

It is our ambition the Code of Practice, as set out in the Call for Views, will ultimately protect users and ensure the economy can securely benefit. Through doing this, we will be building security requirements that fosters international alignment and continue to promote the UK as a world-leader for AI.

Next steps

We will seek feedback from stakeholders over the next 8 weeks. This will be through formal written feedback submitted via gov.uk, but also by engaging with stakeholders directly, including several events that my officials and I are attending in the coming weeks. Notably, I will be launching this Call for Views at CyberUK on 15th May 2024. This feedback received will help formulate an official response following the closure of the Call for Views. Where relevant, we will make appropriate amendments to our approach and proposed interventions, such as

[2] Guidelines for secure AI system development, National Cyber Security Centre, 2023.

the Code of Practice, to ensure that they are appropriate and implementable as we look to take it forward into a standards development organisation.

I will place a copy of this letter, the Call for Views, and the official government response in the Libraries of the House.

Yours sincerely,

Rt Hon Michelle Donelan MP
Secretary of State for Science, Innovation and Technology