Call for views on the Cyber Security of AI

A public consultation issued by Department for Science, Innovation and Technology (DSIT), UK
Comments from ACCA to DSIT

25 July 2024

REF: TECH-CDR-2145

About ACCA:

ACCA (the Association of Chartered Certified Accountants) is the global professional body for professional accountants.

We're a thriving global community of **247,000** members and **526,000** future members based in **181** countries and regions, who work across a wide range of sectors and industries. We uphold the highest professional and ethical values.

We offer everyone everywhere the opportunity to experience a rewarding career in Accountancy, finance, and management. Our qualifications and learning opportunities develop strategic business leaders, forward-thinking professionals with the financial, business, and digital expertise essential for the creation of sustainable organisations and flourishing societies.

Since 1904, being a force for public good has been embedded in our purpose. We believe that Accountancy is a cornerstone profession of society and is vital in helping economies, organisations, and individuals to grow and prosper. It does this by creating robust trusted financial and business management, combating corruption, ensuring organisations are managed ethically, driving sustainability, and providing rewarding career opportunities.

And through our cutting-edge research, we lead the profession by answering today's questions and preparing for the future. We're a not-for-profit organisation. Find out more at accaglobal.com

For further enquiries please contact:

Glenn Collins
Head of Technical and Strategic Engagement, UK

glenn.collins@accaglobal.com

Jessica Bingham ACA FRSA
Regional Lead, Policy and insights (UK, Europe, Eurasia, Middle East & Americas)

jessica.bingham@accaglobal.com

GENERAL COMMENTS

ACCA welcomes the opportunity to comment on the open call for evidence issued by DSIT, UK. We support the importance and need for a trusted[1] eco-system for AI and commend this initiative particularly given the rapid developments in AI globally, and the heightened cybersecurity considerations linked to this.

In the section that follows this one, we provide responses to the specific survey questions that were posed. Our perspective on those responses is influenced by the following factors:

1. ACCA is a professional member body training accountancy and finance professionals. While some of our members as part of their work develop AI systems in addition to being trained in accountancy and finance, our members are most likely to be involved as system operators, data controllers, end-users, or assurance providers. The last of these refers to third-party, independent certification/verification of AI systems, particularly in relation to their deployment within organisations, as opposed to their development. The ACCA Qualification provides ACCA students and members with the opportunity to upskill in advancements in technology including AI, to enhance their professional skill set. A future integrated AI-driven learning and exam experience will enable ACCA to deliver personalised and tailored education support to help each and every learner through the ACCA journey. ACCA's current (and planned) use of AI across learning and assessment will have a profound impact on our partner network; improving the ability to do business with ACCA, improving partner learner outcomes by working closer with ACCA, and delivering finance professionals with the optimal experience and skill set for the modern workplace.

2. We support a principles-based approach as we feel that there are too many as-yet-unseen scenarios with AI, and consequently with the cybersecurity of AI. We are therefore supportive of the way that 'principles' have been given prominence in this Code, and this call for evidence.

3. We are a UK headquartered, and highly global body with offices in over 55 countries and members in 180+. In general, and across policy areas, we support global standards, and actively leverage our policy staff based around the world to advocate for consistent global standards and to draw attention to best practices advocated by the UK. We're therefore supportive of the government's stated approach of starting with the voluntary Code as a step towards a global standard.

4. We believe that given the very fast pace of change in AI, industry participants who are at the frontline of latest developments are best placed to manage constantly changing and newly emerging cyber risks. And the government is best placed to setup an overarching regulatory structure and principles, while giving space to industry experts to work within that. In that sense, philosophically, we see the value of a pro-innovation approach as explained in the government's AI whitepaper. However, this is provided it comes with appropriate safeguards and the ability to revisit requirements if needed, which is consistent with the government's proposed approach as per its response to the views received on the whitepaper.

5. As an education body that trains accountancy and finance professionals, we think deeply about skills. And are acutely aware that there is an urgent need for upskilling in the AI space, particularly for members like ours who are not technology experts. This is a consideration which will apply to the vast majority of staff in organisations

[1] https://www.accaglobal.com/gb/en/professional-insights/technology/trusted-artificial-intelligence.html

across the country. We see opportunities for the Apprenticeship Levy to be expanded for instance to a 'Growth and Skills Levy' that is more flexible and can be used to fund shorter-term accredited training programmes that upskill and reskill workers on the cybersecurity of AI. Companies should also be able to increase the proportion of their unspent levy funds to their supply chains – we'd suggest from 25% to 40%. This could unlock millions of pounds to develop AI skills. Ultimately cybersecurity issues linked to AI need staff to be trained on current and emerging risks – absent a focus on this aspect, the standards and frameworks will fail to achieve impact.

SURVEY RESPONSES

Demographics

Q1. Are you responding as an individual or on behalf of an organisation?

~~Individual~~

Organisation

~~Q2. [if individual] Which of the following statements best describes you?~~

~~Cyber security/IT professional~~

~~Developer of AI components~~

~~Software engineer~~

~~Data scientist~~

~~Data engineer~~

~~Senior leader in a company~~

~~Consumer expert~~

~~Academic~~

~~Interested member of the public~~

~~Government official (including regulator)~~

~~Other (please specify)~~

Q3. [if organisation/business] Which of the following statements describes your organisation? Select all that apply.

Organisation/Business that develops AI for internal use only

Organisation/Business that develops AI for consumer and/or enterprise use

Organisation/Business that does not develop AI, but has adopted AI

Organisation/Business that plans to adopt AI in the future

Organisation/Business that has no plans to adopt AI

A cyber security provider

An educational institution

A consumer organisation

A charity

Government

Other (please specify)

Q4. [if organisation], What is the size of your organisation?

Micro (fewer than 10 employees)

Small (10-49 employees)

Medium (50-499 employees)

Large (500+ employees) ; as well as 100k members and 70k students in the UK

~~Q5. [if individual], Where are you based?~~

~~England~~

~~Scotland~~

~~Wales~~

~~Northern Ireland~~

~~Europe (excluding England, Scotland, Wales and Northern Ireland)~~

~~North America~~

~~South America~~

~~Africa~~

~~Asia~~

~~Oceania~~

~~Other (please specify)~~

Q6. [if organisation], Where is your organisation headquartered?

England

Scotland

Wales

Northern Ireland

Europe (excluding England, Scotland, Wales and Northern Ireland)

North America

South America

Africa

Asia

Oceania

Other (please specify)

Call for Views Questions

Question 7:

Q7. In the Call for Views document, the Government has set out our rationale for why we advocate for a two-part intervention involving the development of a voluntary Code of Practice as part of our efforts to create a global standard focused on baseline cyber security requirements for AI models and systems. The Government intends to align the wording of the voluntary Code's content with the future standard developed in the European Telecommunications Standards Institute (ETSI).

Do you agree with this proposed approach?

Yes

No

Don't know

[If no], please provide evidence (if possible) and reasons for your answer.

Q8. In the proposed Code of Practice, we refer to and define four stakeholders that are primarily responsible for implementing the Code. These are Developers, System Operators, Data Controllers (and End-users).

Do you agree with this approach?

Yes

No

Don't know

Please outline the reasons for your answer.

On this question, while we agree with the basic approach, we would highlight two aspects:

- **Assurance providers**: we anticipate utility from such a code for those providing assurance or third-party verification of AI systems. This is an important category of stakeholders who will have a key role to play in creating a trusted AI eco-system to supplement the regulatory and legal direction from policy makers. We do not anticipate this group to be subject to the requirements of the code itself, but assurance requires checks against a well-defined, and ideally, publicly available

standard - which this code could provide. And cyber risks are a part of what the assurance of an AI system may need to check for. Therefore, those providing assurance would find such a cyber code and associated standards helpful. We would suggest for consideration, that they are added within the parentheses at the end of the 4 categories - ie '(and End users, AI Assurance providers)'. In addition, our general starting point on assurance is a preference for global standards, which is similar to the stated aim for this cyber code, ie use it as a stepping stone to a global standard.

- **End-users**: the importance of ensuring that the rights and protections of end-users in relation to cyber safety for AI systems is given particular thought. There is a significant information and comprehension asymmetry between developers and end-users: certainly, in relation to AI systems and in many instances, in relation to cyber security more generally as well. So, end-user protection should be a key consideration (even in the face of competing commercial considerations) in all aspects of the design and deployment of the code. We would particularly highlight the risks and impact to end users in the small and medium sized enterprises. ACCA has a significant number of members operating in this segment, and we are acutely aware of the greater challenges faced by this group of stakeholders on cyber readiness – across both skills and budgets.

Q9. Do the actions for Developers, System Operators and Data Controllers within the Code of Practice provide stakeholders with enough detail to support an increase in the cyber security of AI models and systems?

Yes

No

Don't know

Please outline the reasons for your answer.

The Code addresses areas we would ordinarily expect for such an endeavour. Therefore, rather than commenting on individual line items, we would like to draw attention to two points of broad significance which the Code would benefit from addressing as explicitly as possible:

- **Handovers between Developers and System Operators**: clarity of what aspects of cyber risk are the responsibility of Developers and which ones are of System Operators. This is particularly the case for AI systems because it often involves linkages to other systems across the organisation and outside of it - such as via APIs to link to external sources of data as part of the design choice for the AI system. So, there could for example, be an AI system with its Developer, System Operator, and interface with another system (which may or may not be AI) that has its own separate developer. The nature of this means the need for clarity and vigilance for cyber responsibilities to not fall in between the cracks of the System Operator and various developers.

- **Data Controllers and Training data**: It has not been unusual for AI systems to rely on data sets drawn from the data of end-users for model training. Therefore, Data Controllers have a key role to ensure the appropriate contractual arrangements are in

place to ensure legal access to training data, as well as adherence to privacy considerations on this data. Having said this, with the advent of Generative AI models, the notion of training data has somewhat altered with applications not always requiring custom data sets for training at the start, as was common with earlier machine learning models. If the AI application is sitting on top of, and referencing, an underlying foundation model (say Chat GPT), the Data Controller may need further clarity or a revised Data Protection Impact Assessment. For eg what data was scraped from websites by the application developer specifically for the use case being deployed by the System Operator, and what data was pulled into the application via the underlying, supporting foundation model. It is likely that data obligations regarding the latter (eg copyright) may be covered within the obligations of the foundation model provider.

The next questions are going to ask you specifically about the Code of Practice that has been designed and proposed by DSIT. There will be a question on whether you support the inclusion of each principle in the Code of Practice and whether you have any feedback on the provisions in each principle.

Q.10 Do you support the inclusion of Principle 1: "Raise staff awareness of threats and risks within the Code of Practice?"

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

The code makes explicit reference to AI model and code related aspects of training. It would be helpful to also emphasize the 'management' aspects of AI cybersecurity explicitly. In other words, training on the required people and organisational process aspects, to enable cyber security. Breaches are often driven by human error, and faulty processes (including things as simple as inadequate documentation or version control for example). Therefore, training that uses case studies to highlight different scenarios where breaches could happen due to these types of lapses would be helpful and practical. Members of professional bodies like ACCA are subject to a code of ethics based on fundamental principles set by a global ethics standard setting body - which they have to reaffirm every year to maintain membership. These include[2] integrity, objectivity, professional competence and due care, confidentiality and professional behaviour. It may be helpful for this code to assess if some/all of these may be relevant to inform aspects of the wording here as well.

[If No], please provide the reasons for your answer.

Q11. Do you support the inclusion of Principle 2: "Design your system for security as well as functionality and performance" within the Code of Practice?

Yes

No

[2] https://www.ethicsboard.org/consultations-projects/revised-code-ethics-completed

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

No comment.

[If No], please provide the reasons for your answer.

Q12. Do you support the inclusion of Principle 3: "Model the threats to your system" within the Code of Practice?

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

No comment.

[If No], please provide the reasons for your answer.

Q13. Do you support the inclusion of Principle 4: "Ensure decisions on user interactions are informed by AI-specific risks" within the Code of Practice?

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

In relation to 4.5 (Developers and System Operators should be transparent with end-users about known limitations or potential failure modes to protect against overreliance.) '

It may be helpful to give a basic explanation of what happens 'under the hood' of the AI model and the impact of design choices on the limitations described. For example, Generative AI models could give different answers to the same question to a greater/lesser extent at different points in time depending on where the model setting is in the spectrum between deterministic and probabilistic.

[If No], please provide the reasons for your answer.

Q14. Do you support the inclusion of Principle 5: "Identify, track and protect your assets" within the Code of Practice?

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

No comment

[If No], please provide the reasons for your answer.

Q15. Do you support the inclusion of Principle 6: "Secure your infrastructure" within the Code of Practice?

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

No comment

[If No], please provide the reasons for your answer.

Q16. Do you support the inclusion of Principle 7 "Secure your supply chain" within the Code of Practice?

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

There may be benefits in explicitly citing procurement and vendor management guidelines and policies as part of this principle. This is to ensure a joined up, holistic and coherent approach across the organisation to the checks conducted when using AI systems.

[If No], please provide the reasons for your answer.

Q17. Do you support the inclusion of Principle 8: "Document your data, models and prompts" within the Code of Practice?

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

We would draw particular attention to the importance of version control as models and the data they use undergo changes over time. We would also emphasize the importance of documentation that is easily understandable by non-experts to mitigate against key person dependency risk. This is specifically for System Operators where the person familiar with the system details leaves the organisation taking the 'tacit knowledge' on the model with them. Related to this is the point of transparency for users – this is an important aspect of ensuring a trustworthy AI eco-system.

[If No], please provide the reasons for your answer.

Q18. Do you support the inclusion of Principle 9: "Conduct appropriate testing and evaluation" within the Code of Practice?

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

No comment

[If No], please provide the reasons for your answer.

Q19. Do you support the inclusion of Principle 10: "Communication and processes associated with end-users" within the Code of Practice?

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

The Code focuses on communication from the Developer/System Operator to the end-user. However, there may be scenarios where the communication is in the other direction. For instance, where the end-user has queries, particularly in the aftermath of an actual/potential cyber incident. For example, say a bank uses an AI system sourced from a third-party developer to identify fraudulent transactions in relation to the bank's credit card. And an individual who has a legitimate credit card transaction receives an email informing them of card cancellation. They would need to contact the System Operator (bank) to resolve the situation. Overall, the end user communication and requirements on the end user should be simple, transparent and where incorrect; quick remedies need to be put in place.

[If No], please provide the reasons for your answer.

Q20. Do you support the inclusion of Principle 11: "Maintain regular security updates for AI models and systems" within the Code of Practice?

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

No comment

[If No], please provide the reasons for your answer.

Q21. Do you support the inclusion of Principle 12: "Monitor your system's behaviour and inputs" within the Code of Practice?

Yes

No

Don't know

[If Yes], please set out any changes you would suggest on the wording of any provisions in the principle.

ACCA is in general supportive of a principles-based approach. In our experience in a fast-changing world, accountancy and finance professionals most often add value by interpreting principles to specific use cases rather than blindly following overly prescriptive rules that are disproportionate to the requirement. In that spirit, we are very supportive of this Principle, and believe it is at the heart of recognising the cyber challenge with AI systems. Namely, that they are dynamic, and a point-in-time view that can become outdated very quickly. For example, data characteristics can evolve over time thus causing the model to drift away from optimal outcomes, provide confusing signals to System Controllers, and issues for Data Controllers who might find it challenging to differentiate normal model drift from deliberate data poisoning - all happening even though no changes may have been made to the model.

[If No], please provide the reasons for your answer.

Q22. Are there any principles and/or provisions that are currently not in the proposed Code of practice that should be included?

Yes

No

Don't know

[If Yes], please provide details of these principles and/or provisions, alongside your reasoning.

Q23. [If you are responding on behalf of an organisation] Where applicable, would there be any financial implications, as well as other impacts, for your organisation to implement the baseline requirements?

Yes

No

Don't know

[If yes], please provide any data to explain this. This will help the Government to quantify the impact of the Code and its requirements on different types of organisations.

Inevitably, doing anything rigorously comes with costs. In our case many of these costs may be indirect (particularly staff time and effort on supporting adherence with the Code) rather than direct. Examples of the latter may be where we bring in an advisor for advisory on setup and review of our processes. There may be some areas where we effectively pay for the costs through the payment to our suppliers/vendors who may be responsible for adherence with some of the relevant principles. We also anticipate costs in monitoring and enforcement of the requirements of the code; as well as promotion to raise awareness of the requirements it places among our stakeholders.

Q24. Do you agree with DSIT's analysis of alternative actions the Government could take to address the cyber security of AI, which is set out in Annex E within the Call for Views document?

Yes

No

Don't know

[If no], please provide further details to support your answer.

Q25. Are there any other policy interventions not included in the list in Annex E of the Call for Views document that the Government should take forward to address the cyber security risks to AI?

Yes

No

Don't know

[If yes], please provide further details to support your answer.

Annex E mentions 'Creating a certification scheme based on the security requirements for AI companies'.

We would be happy to conduct exploratory discussions with government on the creation of a certification scheme with a particular emphasis on the needs of those deploying (as opposed to developing) AI systems. This could be relevant for System Operators and Data Controllers (as well as have value for end-users) - who could then say they are following best practice in relation to the cybersecurity of AI, having been certified via the scheme. We would also be happy to support the awareness campaigns mentioned in Annex E through our extensive network in the UK and if relevant, internationally.

Q26. Are there any other initiatives or forums, such as in the standards or multilateral landscape, that that the Government should be engaging with as part of its programme of work on the cyber security of AI?

Yes

No

Don't know

[If yes], please provide evidence (if possible) and reasons for your answer.

The government is presumably already well aware of this, so mentioning for completeness. Alignment with internal standards would benefit from linkage with ISOs in this and related areas such as ISO/IEC 27001:2022 - Information security, cybersecurity, and privacy protection.

Q27. Are there any additional cyber security risks to AI, such as those linked to Frontier AI, that you would like to raise separate from those in the Call for Views publication document and DSIT's commissioned risk assessment. Risk is defined here as "The potential for harm or adverse consequences arising from cyber security threats and vulnerabilities associated with AI systems".

Yes

No

[If yes], please provide evidence (if possible) and reasons for your answer.

Again, these are implicitly/explicitly covered but worth reiterating given their indirect, but real connection to cyber risk:

- **Concentration risk**: A small number of foundation models sit behind most Generative AI applications. So, a hacker with specialist expertise and insider access, who has spotted an emerging bug in one of these models could affect an inordinately high number of front-end applications.

- **Energy consumption**: There are extremely high compute costs of foundation models, and a premium on sourcing CPU power. This pressure could create a risk of reducing model 'quality' to manage costs - and a question of whether the priority will be to measure quality in terms of maintaining model performance, while sacrificing model security. The recent exit of the Chief Scientist of a leading Generative AI provider is because of his view that not enough emphasis was being placed on safety. While this call for evidence focuses specifically on cyber security rather than wider safety issues and the impact of AI on society – there is a core point here that remains valid. Namely that wider pressures to deliver quickly to market, against a backdrop of astronomical compute costs is real, not going away and may impact emphasis on security aspects if not properly prioritised and emphasized.

  While the above point on costs speaks directly to cyber, it's important not to lose sight of the impact of energy consumption from an environmental perspective – and of the sustainability-related challenges that come with further AI development. ACCA is actively exploring and articulating emerging thinking on the reporting of sustainability related disclosures. And greater, and higher quality disclosures, do ultimately also contribute to a more trustworthy eco-system which reduces risks, including cyber risks.

Q28. Thank you for taking the time to complete the survey. We really appreciate your time. Is there any other feedback that you wish to share?

Yes

No

[If yes], Please set out your additional feedback.