## Consultation questions

## DSIT: Call for views on the Cyber Security of AI

### About the Institution of Engineering and Technology (IET)

The IET is a trusted adviser of independent, impartial, evidence-based engineering and technology expertise. We are a registered charity and one of the world's leading professional societies for the engineering and technology community with over 155,000 members worldwide in 148 countries. Our strength is in working collaboratively with government, industry and academia to engineer solutions for our greatest societal challenges. We believe that professional guidance, especially in highly technological areas, is critical to good policy making. For further details on the evidence submitted, please contact policy@theiet.org.

### Questions

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

- Individual
- Organisation

**2.** [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:

**3.** [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): Professional Engineering Institution

**4.** [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)

**5.** [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)

**6.** [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)

**7.** 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.

Creating a global standard for cyber security requirements of AI systems is overall a good approach. Although, these proposals don't separate between advanced AI and basic AI, or between critical applications and consumer apps. This leads to a generalisation of AI, which may prove counterproductive as all branches of AI develop. Therefore, the proposed voluntary code would be better served being offered as a framework rather than a voluntary code.

The use of frameworks has already been proven to be beneficial in developing capacity in global arenas, for example the NIST Cyber Security Framework. Frameworks also enable the government to align all relevant standards, rather than the current proposed approach to align to one standard which does not appear to enable the industry fully. The government has the opportunity to create a framework which could drive momentum in this area.

Regardless of whether a code or framework is used, it is important that the UK aligns itself to international standards currently being developed by the British Standards Institution (BSI), although this may differ from the European Telecommunications Standards Institute (ETSI). Additionally, International Organization for Standardisation (ISO) standards exist under 'ISO/IEC TR 27563:2023 Security and privacy in artificial intelligence use cases — Best practices' which also outlines security and privacy practices for AI use cases. The code of practice makes sense as it is voluntary, however the wording alignment with additional standards may create more confusion.

Furthermore, development can take a considerable amount of time; as seen from the development of cyber security standards such as IEC 62443. The government should consider if interim requirements should be created until suitable global standards are available, given the pace of technology.

**8.** 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.

Defining roles, responsibilities and duties to define stakeholders is a sensible approach. This code covers all the key users across a business that would be involved, responsible and accountable for security within an organisation

developing AI. However, it is important to refer to enabling functions that are not directly involved, as this responsibility should be distributed.

We do also have some concerns surrounding how these stakeholders are defined, and the clarity of these roles in other areas of cybersecurity.

Regarding the titles of these stakeholders, ‘system operators’ are normally seen as ‘end users’, system operators will be better defined as system integrators, ‘data controller’ is a vague title and would be better labelled as ‘data governance’. Finally, it would be better to distinguish between developers and data creators, this will eradicate the potential for entire datasets being biased. Regardless of how the roles are labelled/defined, it should be made clear that these roles, responsibilities and duties shall not be transferred, along with the risks to other parties.

In addition, it is not clear how these roles will be allocated in the case of autonomous or semi-autonomous systems, such as vehicles. This leads to further questions around whether the end user would be the driver or the operator of the vehicle in the case of it being autonomous agricultural or industrial equipment. It would also be vague as to who would be the data controller if the autonomous system used unsupervised learning. The concerns we have outlined above need to be considered when using this approach as they may change the scope of this proposal.

**9.** 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.

Most of these codes of practice are cybersecurity, the next step is to apply these codes.

These actions clearly indicate what the stakeholders need to achieve to increase the cyber security of AI models. However, the challenge will be ensuring that the stakeholders understand the actions that are required to achieve the outcomes.

Data is very important in order to train the algorithm first and to validate it afterwards. There's a possibility that data manipulation would be the major cyber-attack that intend to trick AI algorithms into making the wrong decisions on purpose. Therefore, some stakeholders will need to be supported and a provided with clear examples of what represents a good outcome. Not providing the appropriate support/examples, could harm the usefulness of the guidance. Particularly as 31% of employers say that artificial intelligence / machine learning will be important to sector growth, but 50% of these employers say they don't have the necessary skills in this area (Source: Digital Skills Survey 2023, IET). It is therefore important that users are AI literate and understand when an output is 'wrong' because AI cannot 'know' the truth and may be biased.

**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.

Communicating the potential threats and risks around AI will increase awareness and is key to developing competence and creating the correct cyber security culture within an organisation.

Senior leaders and managers need to drive cultural change organisationally from the top. This principle should reflect this. Senior management should be responsible for making awareness and risk management a business priority.

For maximum effectiveness, this principle should be reviewed and updated in accordance with changing environments and at the minimum every six months to ensure that it is relevant and effective as technologies develop.

The wording in the code refers to terms such as ‘AI-Security awareness content’ is not entirely clear in its meaning. It could be enhanced by adjusting the words to include ‘conduct AI risk assessments around risk and threats and develop and communicate threat and risk scenarios which highlight risks with staff’.

**11.** 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
- Don’t know

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

We support the inclusion of this principle, however the principle should include a requirement to secure remote access via encryptions. There’s potential to apply previous IT design principles to add anything that is missing that can advance AI and cybersecurity.

Also, it may be clearer to call this principle “design your system for cybersecurity”, as “design your system for security” may be misunderstood to mean physical security.

**12.** 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.

We support the inclusion of this principle, however it should include a requirement for the risk assessment to be reviewed when new vulnerabilities are revealed.

**13.** 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.

We support the inclusion of this principle. The government should consider mandating that developers and system operators inform users of the residual risks that are hard to restrain associated with the use of the AI, such as: image generation, ethics risks, data poisoning and AI enabled password hacking.

It should also be assessed dependant on the AI model itself. The most advanced models are advanced precisely because they work in a different way from rule-based systems. At times, the best way to monitor AI is for AI to monitor AI.

**14.** 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.

Whilst we support the inclusion of this principle, the term “assets” is too vague as anything can deemed as an “asset”. This would be better called “assets driven by AI” We would suggest renaming this proposal to “assets driven by AI” to make it clear what Principle 5 is referring to.

**15.** 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.

As stated in paragraph 6.4, the response should include an incident management plan. If the infrastructure lies with a third party, then the response plan should be agreed with that party.

The terms “secure” and “infrastructure” are also vague as there are several ‘infrastructures’ that are not directly related to this consultation, for example, the railways. The principle would be clearer if it was called “Cybersecure your IT and Digital Infrastructure”.

**16.** 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.

We broadly agree this the inclusion of this principle, however, there is a concern with the difficulty of adhering to this in practice – particularly if you’re a smaller firm. Small and Medium Enterprises (SME’s) should be supported to achieve this in practice.

Paragraph 7.1 should include a duty to assure that the supply chain adheres to the same security expectations and requirements. The requirements in this principle should include suppliers of hardware platforms as well as software assets.

It is unclear whether this principle is to secure the supply chain or securing the organisation from the supply chain. Calling this “Secure against your digital supply chain” would help to differentiate what is being asked of those taking on these principles.

This principle also does not appear to consider that some of the foundational AI technology (particularly Gen AI) is Open Source developed and relies on training data that may or may not have been validated. This is a challenge, whilst 7.2.1 touches on this as a principle, there are some behaviours or conduct that can be expected, such as: engaging with legal counsel and due diligence on the models they are using, which ensures that all has been done to maintain the integrity of their systems.

**17.** 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.

Yes, we support the inclusion of this principle. However, it will be difficult to monitor the best AI models in the same way we would have monitored technologies in the past, therefore, paragraph 8.1.2 should require that the complexity of the system is sufficiently detailed to allow for adequate regression testing due to modifications or the correction of defects.

Also, there is an element of ambiguity within the title of Principle 8. There’s a chance that a large number of people will not know what “prompt” is. This may extend to SME’s who may not have the similar resources to that of bigger companies. It would be beneficial to include further explanation of what this principle means for organisations.

**18.** 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.

The principle should include a requirement to ensure that testing is sufficiently documented to enable any tests to be repeated by a third party if required. On top of post deployment testing and red teaming, there should be research into the behaviours of the users, recognising that how they interact with a product or system also creates vulnerabilities.

**19.** 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 principle should include the end user being supplied with details of any risks from the AI they may be exposed to because of the use of their data. The risks associated with the AI should be clearly communicated to the end user so that they can make an informed decision.

**20.** 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.

Yes, however, security updates should be treated as a dual risk. These updates need to be tested and certified too, as they can contain malicious codes. One example of this would be if there was a cyber-attack against a software development company that added a backdoor to implement a forced update to a legitimate third-party user.

**21.** 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.

Although we support this inclusion of this principle, there are questions to be asked regarding privacy and confidentiality. It will be difficult for a developer monitor a system’s behaviour without compromising security and privacy. This needs to be considered before affirming the principle.

**22.** 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.

There are some provisions, and detail around the ethics of cybersecurity in AI that is missing. For example, how developers work to understand that biases in AI could result in unfair profiling. There are discussions around transparency and explanation, but not necessarily enough coverage of how Developers, system operators can ensure that they are transparent and conveying key concepts in a way that is explainable.

The department should also consider including an additional principle defining the requirement for a management system and for periodic audit and review.

**23.** [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.

**24.** 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.]

There should be consideration within business guidance for the impacts different types of AI applications might require. A sales business that is business-to-customer or business-to-business, may require additional stakeholders mentioned in the code of practice, including sales development specialists.

There are also Explainable AI or Algorithms Audits that confirm how the algorithm makes decisions and learns. These must not be overlooked.

**25.** 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.

Terms such as ‘bias’, ‘diversity’ and ‘inclusion’ need to be included.

**26.** 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.

It is positive that DSIT are working with ETSI on AI security and it is key to collaborate with other key stakeholders in this field, such as BSI, ISO and CEN, and Centre for emerging technology and security (CeTas).

For the best outcomes, all the existing standards that apply should be used and this should only address the gaps posed by AI, not all encompassing. This will make it more impactful.

**27.** 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.

The risk assessment did not include examples relating to cyber-physical systems such as autonomous vehicles. These systems will be susceptible to cyber security risks and could represent significant hazards.

In addition, there is a challenge with systems that use unsupervised learning in order to localise their responses according to their operating environment. The peculiarities of cyber-physical systems do not appear to be fully addressed within the risk assessment. For example: There's a risk of AI swapping the learning source for the solution to make wrong decisions and humans believing the AI is performing at it should be. This further justifies the need for operators to be AI literate when assessing performance.

**28.** 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.

The Code is a sensible mechanism for improving the cyber security of AI systems. The code appears to be most applicable to applications that are purely software based.

However, the requirements of AI systems that interface and direct physical systems do not appear to be fully considered. Further information and guidance should be given describing how this Code and the principles would be applied to autonomous cyber physical systems, particularly how the stakeholder roles would be allocated and how relevant principles would be applied to applications that adapted during use with un-supervised learning.