Department for
Science, Innovation
& Technology

Government response to the Call for Views on the Cyber Security of AI

CP 1258

Government response to the Call for Views on the Cyber Security of AI

Presented to Parliament
by the Secretary of State for Science, Innovation and Technology by Command of His Majesty

January 2025

CP 1258

This publication is available at www.gov.uk/official-documents.

Any enquiries regarding this publication should be sent to us at AIcybersecurity@dsit.gov.uk

ISBN 978-1-5286-5409-8

E03283358 01-25

Printed on paper containing 40% recycled fibre content minimum

Printed in the UK by HH Associates Ltd. on behalf of the Controller of His Majesty's Stationery Office

Contents

1. Ministerial foreword

Artificial Intelligence (AI) is one of the most vital technologies of our lifetimes. It has incredible potential to improve our public services, boost productivity and rebuild our economy. However, to take full advantage and fully realise these benefits we need to build trust in these systems which are increasingly part of our day to day lives. We must protect end-users and address the very real security threats to AI systems and models. Organisations in the UK must be confident they can adopt AI, and security must be built in across the AI lifecycle as a key safeguard against misuse.

The voluntary Code of Practice on the cyber security of AI which is set out in this Government response will be used to inform the development of a global standard. The Code of Practice and the new implementation guide forms one-part of Government's wider work on AI and is aligned and contributing to the vital programme that DSIT is progressing on frontier AI to prepare the UK for future advanced AI models. As announced in the King's Speech this summer, we will deliver on our manifesto commitment by placing binding requirements on the handful of companies developing the most powerful AI systems. This highly targeted legislation will build on the voluntary commitments secured at the Bletchley and

Seoul AI Safety Summits and strengthen the role of the AI Safety Institute. This work on the cyber security of AI is also aligned with DSIT's other cyber security initiatives, such as the recently published draft Codes of Practice for Cyber Governance and Software Vendors which will both improve security practices, outcomes, and confidence for UK organisations.

I greatly appreciate all the responses we received to the Call for Views on the cyber security of AI and the many contributions from international partners and industry. My officials have analysed your responses, and I am pleased to now introduce the government's response to that Call for Views.

This government response outlines how we have taken your feedback on board. I am delighted by the scale of support for DSIT's approach and the technical feedback which has helped us to update the Code of Practice and create a brand-new implementation guide to support organisations in adopting it, particularly small and medium enterprises. We recognise it is vital that internationally agreed and aligned security requirements are developed and therefore my officials will be progressing with our plans to create a global standard.

We must ensure that all new and existing technologies are safely developed and deployed across the UK. The UK, as a world leader in securing technology, will continue to advocate the importance of cyber security and the need for a secure by design approach across all technologies.

This is another step to ensure we can all benefit from secure AI, and I look forward to continuing discussions on how the government, international partners, industry and civil society can collaborate to achieve this goal. Thank you again for your contributions to this generation-defining technology.

2. Executive summary

The UK is well positioned to take advantage of the range of benefits AI has to offer. However, through its recently published evidence base, the Government recognises there are clear risks to AI security which must be addressed so these benefits can be realised.[1] From the 15 May to 9 August 2024, DSIT held a Call for Views on the cyber security of artificial intelligence (AI). The Call for Views set out a proposed two-part intervention: the development of a voluntary Code of Practice, and to then use this as the basis for the development of a global standard focused on baseline cyber security requirements for AI models and systems.

DSIT received 123 responses to the Call for Views. Most responses were from organisations, including industry associations on behalf of their members, as opposed to individuals. This was a global Call for Views, and we welcome the views received from a wide range of international partners. We are satisfied that we have gathered the views of a considerable number of relevant stakeholders.

Most respondents to the Call for Views (80%) were supportive of DSIT's proposed two-part intervention. There was also overwhelming support for the inclusion of each of the 12 individual principles contained with the Code (ranging from 83% to 90%). There were several

1 Research on the cyber security of AI, Department for Science, Innovation & Technology, 2024.

recurring pieces of feedback received through the Call for Views. This included: the need for more detail or guidance on how to implement the Code; suggested changes to provisions within each principle of the Code; and suggestions for new provisions. Respondents also noted that the existing market might not provide the sufficient skills or capabilities need to implement the Code. The responses to open-text questions were rich in detail and varied widely, making it challenging to categorise the feedback into overarching themes. As a result, the themes identified are broad to ensure each theme captures feedback from multiple respondents.

We have taken this feedback and used it to update the Code of Practice and create a new implementation guide that supports the Code of Practice. The guide provides detail that supports organisations, particularly small and medium enterprises (SMEs), with implementing the Code. We have updated the Code and created a new principle on end of life that covers the transferring of ownership of the training data and/or a model as well as the decommissioning of a model and/or system. We have also updated the stakeholder groups to include “data custodian” and “affected entities”, while adding more clarity across the Code. The principles are now more contextualised to AI security risk, particularly where software requirements are referenced. We will be taking the updated Code of Practice and implementation guide into the European Telecommunications Standards Institute (ETSI) to develop a new global standard focused on baseline cyber security requirements for AI

models and systems. We will continue to advocate an international approach, pursue our goal of increasing the adoption of security principles domestically and internationally and provide clarity to organisations on how they should protect AI technologies.

This document provides a detailed overview of the feedback received from responses to the Call for Views. We have provided responses to each question to explain how feedback has been taken on board in updating the Code of Practice, and in determining the government's next steps in this space.

3. Background

AI continues to be one of this era's most defining and powerful technologies and is increasingly part of our daily lives both at home and at work. Across a range of areas such as technology, finance, transport, agriculture, and crime prevention, AI is changing the way we work and interact with data. The UK AI sector itself is strong and growing, generating £14.2 billion pounds in revenue in 2023 alone. Moreover, it is highly productive, contributing an estimated £5.8bn in Gross Value Added (GVA) to the economy in 2023 and employing over 64,000 people in the UK.[2]

The UK government's research[3] into the cyber security of AI found that there are clear and specific risks to the security of AI models and systems throughout the AI lifecycle. It is therefore imperative that these are addressed so that millions of consumers and organisations can safely benefit from AI technologies. A Call for Views on the government's proposed interventions was held from 15th May 2024 to 9th August

---

2 2023 AI Sector Study, Department for Science Innovation & Technology, 2024

3 Research on the cyber security of AI, Department for Science, Innovation & Technology, 2024.

2024.[4] The proposals included a two-part approach, comprised of a voluntary Code of Practice for the UK, which forms the basis for the second part, a global standard developed at an international standards body. Together, these will establish baseline security requirements that will help reduce the number and impact of successful cyber attacks and therefore protect users' data and the economy.[5]

This work is part of DSIT's wider technology security programme and its secure by design approach across all digital technologies, which places the responsibility on those that develop technology to build robust cyber security into their systems. Due to overlap between the technology areas and stakeholders, this work on AI is closely linked to the government's work on cyber governance and software security and resilience, as well as our other secure-by-design initiatives across consumer IoT, enterprise IoT, and App Stores. DSIT recently published a consultation response outlining approaches

---

4 This Call for Views is focused on addressing the cyber security risks to AI rather than wider issues relating to AI, such as safety or the cyber security risks that stem from AI. There is specific work on these areas being led by other parts of government.

5 Security is an essential component underpinning all types of AI. Therefore, the scope of the Call for Views as well as the voluntary Code of Practice and proposed technical standard, includes all AI technologies, including frontier AI.

on the Cyber Governance Code of Practice. In a recent Call for Views, DSIT sought feedback on a proposed Code of Practice for software vendors, a response to this Call for Views is expected to be published soon. All cyber security Codes of Practice produced by DSIT are part of the government's broader approach to improve baseline cyber security practices and increase cyber resilience in the economy. These Codes have been designed as part of a modular approach so that stakeholders can apply them in tandem depending on which technology areas are relevant to their business.[6]

This work also complements wider ongoing work across government to ensure the UK's economy will fully realise the benefits of AI. This includes the government's commitment to introduce highly targeted legislation for

---

6 The Codes of Practice provide guidance ranging from the development of baseline cyber security advice which all organisations should follow, moving progressively towards more product or domain-specific advice due to the increasing risk and evolving threat landscape. A modular approach has been developed to help organisations easily identify which Codes – and within those Codes, which provisions – are relevant to them according to both their business functions, and the types of technologies they either use or manufacture. More information is available here https://www.gov.uk/government/collections/cyber-security-codes-of-practice).

the handful of companies developing the most powerful AI models. These proposals will promote the safe development of AI, and support growth and innovation by reducing current regulatory uncertainty for AI developers[7], strengthening public trust and boosting business confidence.

---

7 We define AI developers as those organisations or individuals who design, build, train, adapt, or combine AI models and applications. In the context of the AI Cyber Security Code of Practice, this includes the companies and organisations, development teams, model engineers, data scientists, data engineers and AI designers who are responsible for creating a model and system.

4. Methodology

The Call for Views was open to the public and stakeholders were able to respond via an online survey, written mail, or via email. All responses were analysed using the same methodology, following the removal of any duplicates.

The Call for Views asked respondents 28 closed questions and 22 follow-up open text questions. The only mandatory questions in the survey required the respondent to provide demographic detail and to answer the initial question on whether they agree or not with DSIT's proposed approach set out in the Call for Views. For some questions, respondents were offered the opportunity to expand on answers and provide more detail with qualitative open text boxes. These open text boxes were not mandatory.

For open text response questions, all responses were reviewed and systematically analysed to identify common themes. Given the highly detailed and diverse nature of these responses, grouping them into overarching themes was often challenging. Consequently, the themes identified tend to be broad, and in many cases, are based on similar feedback from a relatively small number of respondents. When a particular theme emerged as the most frequently mentioned theme within a question, it has been highlighted in the summary below. If a theme was mentioned by 15% or more of respondents to a question, it has been categorised as a "frequently cited" theme.

Themes mentioned by fewer than 15% of respondents have been classified as “less commonly cited” in the summary.

The number of responses to the open text follow-up questions ranged from 27 to 65. An individual response to the open text questions could contain reference to more than one theme, where this has occurred all the themes from the response have been noted. Not all the open text responses to questions were relevant to the topic of the question. If they were relevant to other questions, then the response was considered and reflected in the analysis of that question.

Please note that some of the percentages in this write-up do not sum to 100% due to rounding.

5. Overview of responses

DSIT received 123 responses to the Call for Views. The majority (68%) of responses were from organisations, as opposed to individuals. The individuals who responded primarily identified themselves as cyber security/IT professionals (44% of individuals) or academics (20%).

Many of the responses came from trade or membership bodies which were predominantly based in the UK and US, but with members from other countries. These incorporated the views of multiple stakeholders simultaneously.

We are delighted that this was a global Call for Views, with responses received from the UK, the rest of Europe, the US, Japan, Singapore, the Republic of Korea, other parts of Asia and Oceania. Of the organisations that identified where they were from in their response, 54% were based in the UK, 26% in North America, 11% in Asia and 7% in Europe (excluding UK).[8]

We are satisfied that we have gathered the views of a considerable number of relevant global stakeholders through this Call for Views and are grateful to the stakeholders that responded as well as those that helped promote the Call for Views.[9]

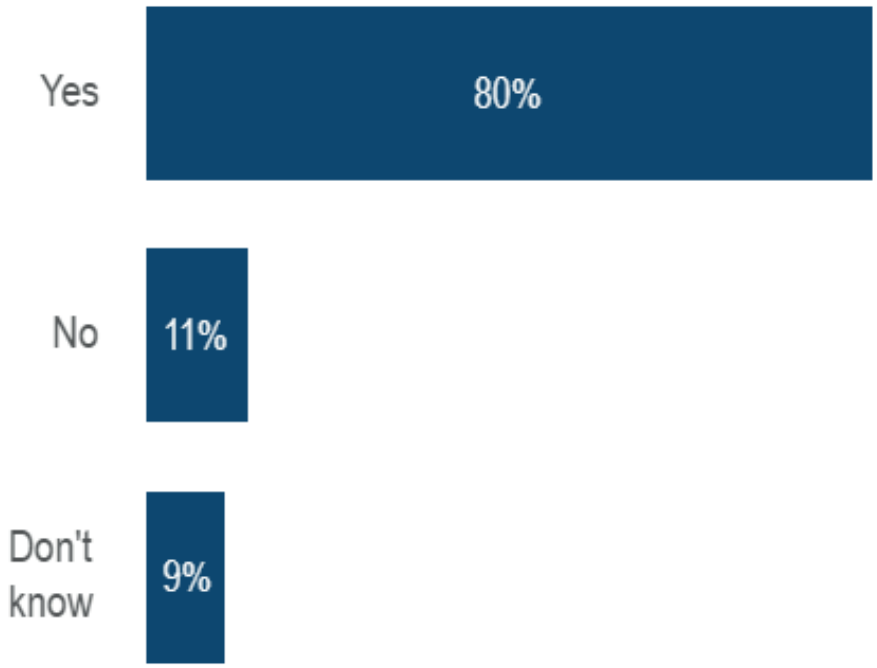
---

8 72 respondents disclosed the region where their organisation's headquarters are based.

9 Two responses to the Call for Views came from other parts of the UK government.

6. Section 1: Rationale & definitions

6.1 Question 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?



The vast majority (80%) of the 116 responses to this question showed agreement with the proposed approach, with 11% opposing and 9% responding with 'don't know'.

32 respondents provided additional evidence and reasons for their answer (both yes and no respondents), and some themes have been identified from these responses. However, these themes were only based on a small number of responses. A frequently cited theme from those that responded yes was the need to be aware of, and engage with, other international efforts (such as those by the US and EU). For those that responded no, the frequently sited themes were that a new standard is not needed and support for mandating security requirements immediately. Another frequently cited theme (from both yes and no respondents) was the need to provide more detail linked to the Code, such as sub-provisions and an implementation guide.

Other, less commonly cited themes included:

- Support for mandating requirements at some point in the future
- Confusion over the modular approach put forward by DSIT and the number of cyber security Codes of Practice
- The need for DSIT to conduct standards work in other standards development organisations.

6.2 Government Response

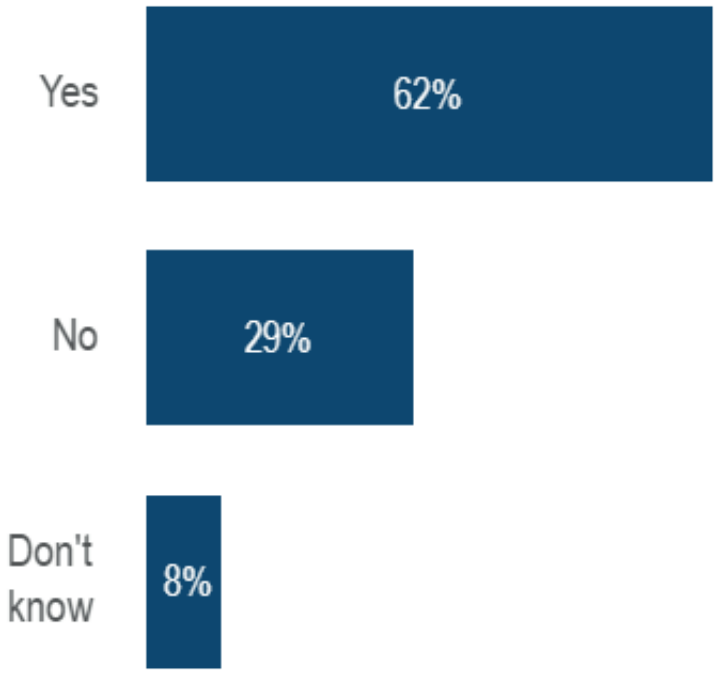
We welcome the overwhelming support for the Government's proposed two-part approach to address the cyber security risks to AI. We recognise that some stakeholders advocated that the Government should focus on other initiatives, such as those in other standards development organisations and work by the US and EU, rather than progress with a global standard in ETSI. We have been and will continue to actively participate in other standards development organisations so that internationally aligned security requirements are created for AI. We are also collaborating with the US and various European partners on this area. We believe ETSI is the most appropriate organisation for the development of a global standard because it enables industry to have a key role, the standards are free and the process is usually fast.

We note that a minority of stakeholders requested more detail in the Code in the context of sub-provisions and guidance on how to adhere to the Code's principles. DSIT therefore commissioned Kainos to create an implementation guide to support organisations. Each iteration was reviewed by DSIT and National Cyber Security Centre officials. This document has been published alongside the Code and Government Response. We decided not to add sub-provisions into the Code due to the level of support for its current level of detail (see Question 9 – 6.5 to 6.6) and because we did not want to make it prescriptive.

We acknowledge that several other stakeholders supported the mandating of the Code's security requirements. However, based on the scale of support for DSIT's approach, we plan to focus our efforts on the development of a global standard, while supporting the Government's overall approach to AI regulation. DSIT is also working closely with other government officials to ensure there is a consistent message for industry on the various Codes of Practice.

Based on the support for our approach, we will now take the updated Code of Practice and our newly created implementation guide into ETSI to form the basis of a new global standard focused on baseline cyber security requirements for AI models and systems. We will also continue to work closely with external stakeholders, including international partners to identify further avenues for collaboration with the ambition of creating international support for the security requirements.

6.3 Question 8 – In the proposed Code of Practice, we refer to and define four stakeholders that are primarily responsible for implementing the voluntary Code. These are Developers, System Operators, Data Controllers (and End-users). Do you agree with this approach?



The majority (62%) of the 109 responses to this question supported the approach, with 29% opposing and 8% responding with ‘don’t know’.

65 respondents to this question who agreed with our approach regarding the four stakeholder groups provided further detail. The most frequently cited theme among respondents was the need for additional stakeholders to be added. This was also the most frequently cited theme among the 27 respondents who stated “no” to this question.

Other, less commonly cited themes included:

- The need to update/change the definition used for end-user
- The need to update/change the definition used for data controller
- The need to update the terminology of the 4 stakeholders
- Disagreement (from respondents who answered ‘no’) with the 4 stakeholders as new terminology is needed

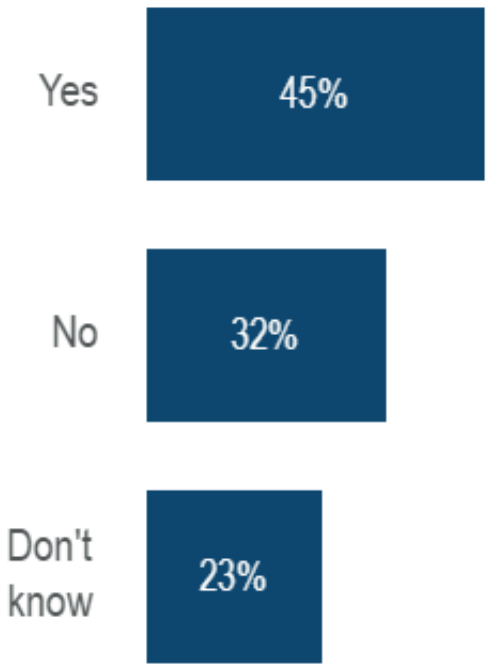
6.4 Government Response

There was considerable feedback that the term “data controller” should not be used because of its meaning in data protection law. The term has therefore been replaced with “data custodians”. The definition has been updated to note that this includes stakeholders who have responsibility for setting the policies for data use as well as the management of the data. When developing the voluntary Code of Practice, we consulted with the Information Commissioner’s Office (ICO) to provide consistency with ICO guidance relevant to compliance with data protection law, where applicable (this includes the term “data custodians”.

There was also feedback noting that the term “end-user” did not encompass a variety of circumstances, such as technology affected by AI, consumers that are impacted through the creation or use of an AI system, that would be relevant to this work. Additionally, some stakeholders noted that there are clear responsibilities for some end-users in the context of AI which needed to be highlighted. We have therefore created a new additional stakeholder group, “affected entities”, to capture individuals and technologies which are not directly affected by AI systems or decisions based on the output of AI systems. We have also modified the definition of end-user to align more closely with definitions used by international counterparts, such as NIST.

We have amended the definition of a “Developer” to also include those that are adapting an AI model and/or system to reflect the open-source market. There were also individual pieces of feedback that we wanted to address, such as the signposting that stakeholders can have multiple roles in the AI lifecycle. We have therefore created a new paragraph under “Audience” to provide added context on the stakeholder groups. We have also acknowledged that some of the requirements for Developers in the Code may not be applicable to open-source models / systems and that this nuance is further clarified within the Implementation Guide.

6.5 Question 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?



45% of the 107 responses to this question supported the current level of detail provided in the Code. 32% believed the Code does not provide enough detail, and 23% of respondents responded with ‘don’t know’.

48 respondents answered ‘yes’ to this question and provided further detail. The most frequently cited theme among these respondents was the need for additional detail including sub-provisions and guidance on how to implement the Code. This was also the most frequently cited theme among the 28 respondents who answered ‘no’ and provided further detail.

6.6 Government response

A common theme that emerged through this question and throughout the Call for Views was the need for more detail and guidance on how to implement the Code of Practice and its principles/provisions. However, we do recognise that a majority indicated that there was sufficient detail to enable an increase in the cyber security of AI models and systems.

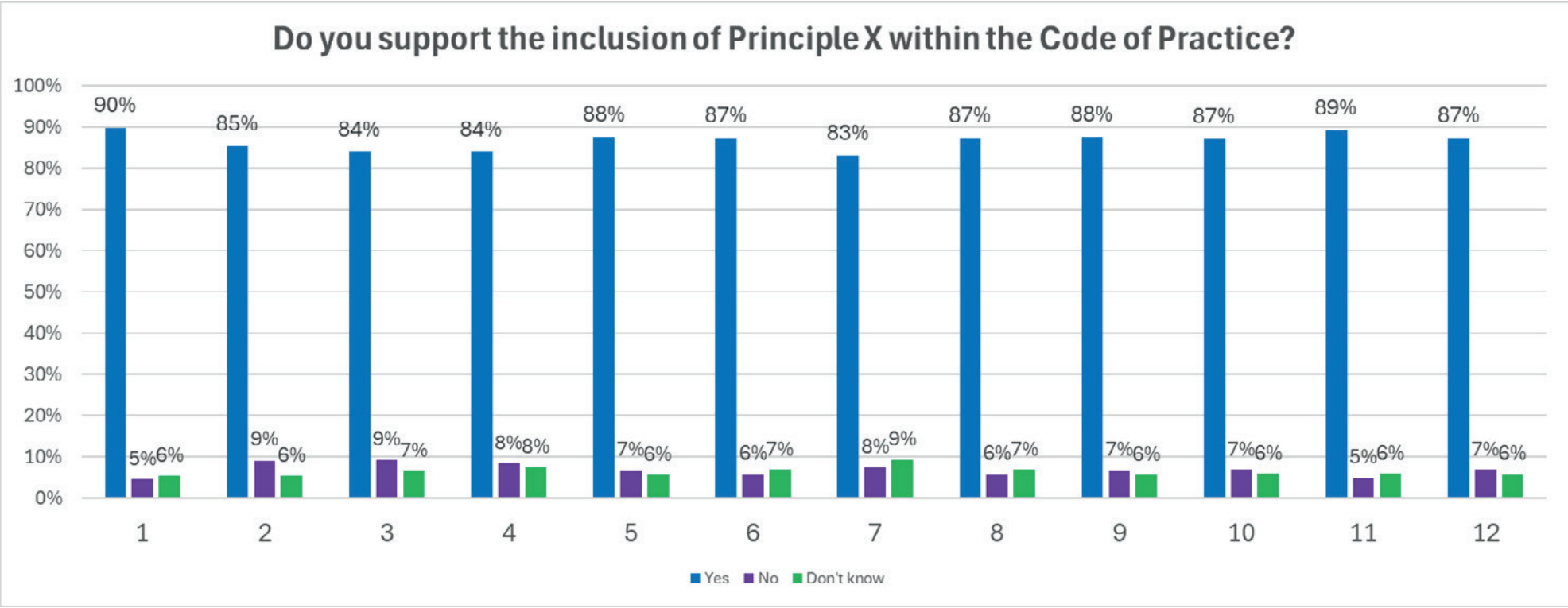
To address the feedback received, we have developed a new implementation guide that supports the Code of Practice. The guide provides detail that supports organisations, particularly small and medium enterprises (SMEs), with implementing the Code and future standard. We decided not to add sub-provisions into the Code because we did not want to make it prescriptive and due to the level of support for its current level of detail. Guidance on the various steps that could be taken for each provision / principle is provided in the Implementation Guide. Together, we believe the documents will ensure UK organisations have the guidance to immediately act to protect their infrastructure from security vulnerabilities associated with AI systems.

We have also rewritten the background section of the Code of Practice. The introduction section has been updated to explain the rationale for the Code and support for DSIT's proposed approach. We have also included a scope section and a glossary of key terms to more closely align with the structure of a standard and to support the reader. There are also sections explaining the implementation guide and purpose of the document for different audiences reading the Code of Practice.

7. Section 2: Code of Practice principles

This section looked at gathering views on the 12 principles presented in the draft Code of Practice. Questions in this section presented each principle in full and asked whether the respondent supported the inclusion of the principle within the Code of Practice.[10] Support for the inclusion of each principle in the Code of Practice was very positive across each principle, ranging from 83-90% of respondents.

10 As part of this, respondents also had the opportunity to provide feedback on whether the requirements were shall, should or could/can. For clarification, standards development organisations define shall to mean that it is a requirement, should is a recommendation and could/can indicates where something is possible. Stakeholders that seek to adhere to the Code are expected to at least adhere to all the shall requirements.


Do you support the inclusion of Principle X within the Code of Practice?
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
90%
5%
6%
85%
9%
6%
84%
9%
7%
84%
8%
8%
88%
7%
6%
87%
6%
7%
83%
8%
9%
87%
6%
7%
88%
7%
6%
87%
7%
6%
89%
5%
6%
87%
7%
6%
1
2
3
4
5
6
7
8
9
10
11
12
Yes
No
Don't know

Respondents were then asked follow-up open-text questions:

Where respondents indicated ‘yes’, they were asked whether they had any suggestions on wording of any specific provisions in the principle in question.

For respondents who answered ‘no’, they were asked to provide the reasons for their answer

The additional feedback received via the open-text responses to both follow-up questions could be grouped into similar consistent themes across all 12 principles.[11] There were multiple recommendations for changes to the wording of provisions and requests to remove certain provisions. This feedback has been captured under the theme ‘suggestions to changes to specific ‘provisions’. Also, there was feedback on the overall approach or focus of principles as a whole and these types of suggestions have been captured under the theme ‘suggestion of changes to the framing of the principle’. The full list of themes across the questions in this section were:

- Suggestions of changes to specific provisions
- Suggestions of changes to the title of the principle
- The need for more guidance on implementing the principle/provision
- Suggestion of changes to the framing of the principle

11 Not all the themes were cited in each ’principle’ question

- Suggestion of new provision(s)
- Principle is too burdensome/not practical

Secure design

7.1 Question 10 – Principle 1: Raise staff awareness of threats and risks – Do you support the inclusion of Principle 1 within the Code of Practice?

The vast majority (90%) of the 107 responses to this question supported the inclusion of Principle 1 in the Code, with only 5% opposed and 6% responding with 'don't know'.

58 of the respondents who answered 'yes' and 6 of the respondents who answered 'no' provided an answer to the respective open text follow-up questions. The most frequently cited theme among responses was the suggestion of the inclusion of new provisions within the principle. Another frequently cited theme was the suggestion of changes to provision 1.1.

Other, less commonly cited themes included:

Suggestion of changes to specific provisions including 1.1.2, 1.2, 1.2.1, 1.3, 1.3.1 and 1.4

Suggestion of changes to the framing of the principle

7.2 Government response

We have added a short section at the start of each principle to explain other relevant cyber security practices and international standards for that area following some feedback. Stakeholders are encouraged to read the Implementation Guide for clarity on the actions required (and recommended) for adhering to each provision in the Code.

We have deleted the previous provision 1.1 to reflect feedback. We've also sought to place the AI requirements in the context of an organisation's wider staff security training programme to clarify that we're not proposing an entirely separate regime for AI staff training. We've removed any specific time periods linked to requirements so that the provisions are not overly prescriptive. Provision 1.1.2 has been amended to reflect that training needs to be tailored to the specific roles and responsibilities of staff. Provision 1.2 is now clearer on an organisations' expectations for their staff in the context of raising awareness of threats and risks. The previous provision 1.4 has been deleted to ensure the Code remains relevant to technological changes. The content from this provision is now incorporated within the implementation guide. Provision 1.3.1 is now provision 1.2.2 and has been amended for additional clarity.

7.3 Question 11 – Principle 2: Design your system for security as well as functionality and performance – Do you support the inclusion of Principle 2 within the Code of Practice?

The vast majority (85%) of the 108 responses to this question supported the inclusion of Principle 2 in the Code, with only 9% opposed and 6% responding with 'don't know'.

58 of the respondents who answered 'yes' and 8 of the respondents who answered 'no' provided an answer to the respective open text follow-up questions. The most frequently cited theme among responses was the suggestion of the inclusion of new provisions within the principle. Other frequently cited themes included the suggestion of changes to provision 2.2 and 2.7.

Other, less commonly cited themes included:

Suggestion of changes to specific provisions including 2.1, 2.1.1, 2.3, 2.4, 2.5 and 2.6

Suggestion of changes to the framing of the principle

7.4 Government response

There was some feedback that noted the need for the principle to apply to Developers as well as System Operators, this has been incorporated into principle 2.

We have amended provision 2.1 to reflect that is also applicable to Developers and feedback that an organisation should undertake an assessment to help with determining and documenting the business requirements for AI. We have also noted that this process should include the potential security risks and mitigations strategies. Provision 2.1.1 has been amended to focus solely on Data Custodians who are part of a Developer because the original wording inferred requirements around involving potential third-party organisations. The section from NCSC's Guidelines has been moved to the Implementation Guide to ensure a consistent level of detail in the Code.

There is a new provision 2.2 to reflect proposed wording on AI system design that was suggested by a few stakeholders. The previous provision 2.2 is now provision 2.3 and has been expanded to clarify why Developers need to document and audit various areas. The new provision 2.4 represents the merger of previous provisions 2.3 and 2.7. Responders noted that the wording could be broadened to provide clarity on what is required if a Developer or System Operator decides to use an external component.

The previous provision 2.4 is now provision 2.5 and the word “safety” has been replaced by “security” for additional clarity on the scope of the provision. Based on feedback, a new “should” sub-provision has been added (2.5.1) to note an organisations’ role in enabling employees to report and identify potential security risks in AI systems whilst ensuring safeguards are in place. The previous provision 2.5 is now 2.6 and further wording has been added to clarify the scope of the requirements. Provision 2.6 is now 2.7 and we’ve clarified that the requirement could be applicable to both Developers or System Operators and added that external providers “should” adhere to the Code of Practice.

7.5 Question 12 – Principle 3: Model the threats to your system – Do you support the inclusion of Principle 3 within the Code of Practice?

The vast majority (84%) of the 106 responses to this question supported the inclusion of Principle 3 in the Code, with only 9% not supporting its inclusion and 7% responding with ‘don’t know’.

57 of the respondents who answered ‘yes’ and 9 of the respondents who answered ‘no’ provided an answer to the respective open text follow-up questions. The most frequently cited theme among responses was the suggestion of the inclusion of new provisions within the principle.

Other, less commonly cited themes:

- Suggestion of changes to specific provisions including 3.1, 3.1.1, 3.1.2, 3.2, 3.3, 3.4, and 3.5.
- Suggestion of changes to the title of the principle
- Suggestion of changes to the framing of the principle

7.6 Government response

The title of this principle has been amended to reflect feedback that managing the risks to your system is a key part of the design phase of the AI lifecycle. Provision 3.1 has been expanded to clarify what could constitute threat modelling to provide clarity to Developers and System Operators. Provision 3.1.1 has been broadened to note that threat modelling sits alongside a risk management process and the importance of this process being carried out when a setting or configuration is updated (as well as implemented). To ensure the Code is not too prescriptive, the previous provision 3.1.2 has been incorporated within the implementation guide. Provision 3.1.3 is therefore now 3.1.2 and we have clarified the wording to make clear that the provision is focused on security risks and superfluous functionalities.

The previous provision 3.2 has been removed from the Code based on responders' feedback that there shouldn't be data protection regulation requirements in the principle within the Code. When developing the voluntary Code of Practice, we consulted with the ICO to provide consistency with ICO guidance relevant to compliance with data protection law, where applicable. Provision 3.3 is now 3.2 and has been expanded to note that additional actions (both for Developers and System Operators) are needed if a security threat can't be resolved based on concerns from responders on the previous wording. Provision 3.4 is now 3.3 and we have changed "third-party organisations" to "external entity", so consistent language is used in the Code. Provision 3.5 is now 3.4 and has been made a "shall" rather than a "should" provision based on feedback. Provision 3.6 is now 3.1.3 due to its link to the analysis required in provision 3.1 and we've added some minor content to help contextualise the provision.

7.7 Question 13 – Principle 4: Ensure decisions on user interactions are informed by AI-specific risks – Do you support the inclusion of Principle 4 within the Code of Practice?

The vast majority (84%) of the 106 responses to this question supported the inclusion of Principle 4 in the Code, with only 8% not supporting its inclusion and 8% responding with 'don't know'.

47 of the respondents who answered ‘yes’ and 8 of the respondents who answered ‘no’ provided an answer to the respective open text follow-up questions. The most frequently cited theme among responses was the suggestion of the inclusion of new provisions within the principle. Other frequently cited themes included the suggestion of changes to provision 4.4 and 4.5, and changes to the framing of the principle.

Other, less commonly cited themes included:

- Suggestion of changes to specific provisions including 4.1, 4.2, 4.3 and 4.6
- Suggestion of changes to the title of the principle
- The need for more guidance on implementing the principle
- Suggestion of changes to the framing of the principle
- Principle is too burdensome

7.8 Government response

Principle 4 has undergone various changes because feedback indicating that some stakeholders had misinterpreted the focus of the principles and others were concerned by the extensiveness and thus the burden placed by the requirements. The title has therefore changed to make clear that this principle centres around enabling human responsibility for AI systems. The previous provision 4.1 has been rewritten to reflect how designing an AI system should involve enabling human oversight. The previous provision 4.2 is now provision 4.4 and we've added the need for Developers to verify (and validate) that the controls specified by Data Custodians have been built into the system.

A new provision 4.2 has been created to cover how AI systems should be designed by Developers to support human involvement. The previous provision 4.3 has been removed and has been replaced with a requirement that sets out actions to be taken where human oversight is a risk control. Provision 4.4 is now provision 4.5 and we've modified the wording to be clear on its scope. Provision 4.6 has been moved to Principle 6 (provision 6.2).

Secure development

7.9 Question 14 – Principle 5: Identify, track and protect your assets – Do you support the inclusion of Principle 5 within the Code of Practice?

The vast majority (88%) of the 104 responses to this question supported the inclusion of Principle 5 in the Code, with only 7% not supporting its inclusion and 6% responding with ‘don’t know’.

50 of the respondents who answered ‘yes’ and 6 of the respondents who answered ‘no’ provided an answer to the respective open text follow-up questions. The most frequently cited theme among responses was the suggestion of changes to provision 5.1. Another frequently cited theme was the suggestion of the inclusion of new provisions within the principle. Finally, another frequently cited theme was the suggestion of changes to provision 5.3.

Other, less commonly cited themes included:

- Suggestion of changes to specific provisions including 5.2, 5.4 and 5.4.1.
- Suggestion of changes to the framing of the principle

7.10 Government response

Provision 5.1 has been amended to clarify what is required by the stakeholder groups in relation to their assets, i.e. the maintaining of a comprehensive inventory. Provision 5.2 has been further contextualised. Provision 5.3 has been rewritten following feedback that it should call out the importance of disaster recovery plans. The latter part of the original requirement on ensuring a known good state of the system can be restored has now formed the new provision 5.3.1. This provision has been changed from "shall" to "should" to reflect feedback. We have kept 5.4 and 5.4.1 because they set out important requirements for protecting different types of data. A new provision 5.4.2 has been added to recognise the potential confidentiality of training data and model weights.

7.11 Question 15 – Principle 6: Secure your infrastructure – Do you support the inclusion of Principle 6 within the Code of Practice?

The vast majority (87%) of the 102 responses to this question supported the inclusion of Principle 6 in the Code, with only 6% not supporting its inclusion and 7% responding with 'don't know'.

49 of the respondents who answered 'yes' and 6 of the respondents who answered 'no' provided an answer to the respective open text follow-up questions. A frequently cited theme among responses was the suggestion of the inclusion of new provisions within the principle. Another

frequently cited theme was the suggestion of changes to provision 6.4.

Other, less commonly cited themes included:

Suggestion of changes to specific provisions including 6.1, 6.2, 6.2.1, 6.2.2 and 6.3

7.12 Government response

Provision 6.1 has been reduced so that the focus of the requirement is more contextualised on the AI ecosystem. As noted in the Government's response section 7.8, provision 6.2 was previously provision 4.6. We have added an additional line to explain the provision's importance in the context of specific AI security risks. Provisions 6.2, 6.2.1 and 6.2.2 have been merged into the new provision 6.3. This has been undertaken based on feedback on the need to avoid repetition and to provide clarity, such as through changing segregated environments to dedicated environments and explaining why it is necessary for AI. The previous provision 6.3 is now 6.4. Provision 6.4, (now 6.5), has been expanded based on the feedback that Developers and Operators should create an incident management plan as well as a AI system recovery plan and that it needs to be tested and maintained. We have created a new provision 6.6 following feedback that cloud service operators will play an important role in helping Developers and System Operators to deliver the requirements in principle 6.

7.13 Question 16 – Principle 7: Secure your supply chain – Do you support the inclusion of Principle 7 within the Code of Practice?

The vast majority (83%) of the 106 responses to this question supported the inclusion of Principle 7 in the Code, with only 8% not supporting its inclusion and 9% responding with ‘don’t know’.

55 of the respondents who answered ‘yes’ and 7 of the respondents who answered ‘no’ provided an answer to the respective open text follow-up questions. A frequently cited theme among responses was the suggestion of the inclusion of new provisions within the principle. Other frequently cited themes included the suggestion of changes to provision 7.1, 7.2 and 7.2.1 and the need for more guidance on implementing the principle.

Other, less commonly cited themes included:

Suggestion of changes to specific provisions including 7.3, 7.3.1 and 7.3.2.

Suggestion of changes to the framing of the principle

The principle is too burdensome

7.14 Government response

Provision 7.1 has been scaled back to ensure the wording on secure software supply chain processes is consistent with other frameworks. Provision 7.2 and 7.2.1 have been merged for simplicity and to avoid repetition. Based on feedback received, a new provision 7.2.1 has been created which sets out the need for mitigating controls and the undertaking of a risk assessment linked to the use of other models and components. A component of the original provision 7.2.1 on the need for information to be shared with end-users forms the new provision 7.2.2. A new provision 7.3 has been created from the previous 7.3.1 and 7.3.2 with a focus on requiring Developers to document aspects of the training data used to create a model. A new provision 7.3.1 has been created to provide further clarity on key aspects of the training data that needs to be documented. Following Feedback from responders, a new provision 7.4 is included so that evaluations are re-run on released models and 7.5 has been added so that end-users are made aware of upcoming changes to models.

7.15 Question 17 – Principle 8: Document your data, models and prompts – Do you support the inclusion of Principle 8 within the Code of Practice?

The vast majority (87%) of the 102 responses to this question supported the inclusion of Principle 8 in the

Code, with only 6% not supporting its inclusion and 7% responding with ‘don’t know’.

41 of the respondents who answered ‘yes’ and 5 of the respondents who answered ‘no’ provided an answer to the respective open text follow-up questions. A frequently cited theme among responses was the suggestion of the changes to provision 8.1.1. Another frequently cited theme was the suggestion of the inclusion of new provisions within the principle. Finally, another frequently cited theme was the suggestion of changes to the framing of the principle.

Other, less commonly cited themes included:

- Suggestion of changes to specific provisions including 8.1, 8.1.2, and 8.2.
- Suggestion of changes to the title of the principle
- The need for more guidance on implementing the principle
- The principle is too burdensome

7.16 Government response

Provision 8.1 has been expanded to clarify that the documentation they are creating and maintaining for an audit trail of their system design should be made available to downstream System Operators and Data Custodians. The latter is a “should” rather than a “shall” requirement due to the potential complexities that some Developers may face with this activity. The wording “cryptographic hashes or signatures” has been removed from Provision 8.1.1 due to stakeholder feedback that it needed to be repurposed around the releasing of said hashes to help verify the authenticity of components. This now forms the basis of the new provision 8.1.2 which has changed from “should” to “shall” following responders’ feedback. Provision 8.2 has also been replaced with new wording following feedback from stakeholders on the need for it to be focused on the need for Developers to have an audit log of changes to system prompts or other model configuration that affect the underlying working of the systems. Following feedback, we have also noted the contextual relevance of data poisoning to highlight the provision’s importance for AI stakeholders. Provision 8.3 has been amended to more clearly express the requirements linked to changes to system prompts or other model configuration.

7.17 Question 18 – Principle 9: Conduct appropriate testing and evaluation – Do you support the inclusion of Principle 9 within the Code of Practice?

The vast majority (88%) of the 104 responses to this question supported the inclusion of Principle 9 in the Code, with only 7% not supporting its inclusion and 6% responding with ‘don’t know’.

50 of the respondents who answered ‘yes’ and 6 of the respondents who answered ‘no’ provided an answer to the respective open text follow-up questions. A frequently cited theme among responses was the suggestion of changes to the framing of the principle. Another frequently cited theme included the suggestion of changes to provision 9.2.2.

Other, less commonly cited themes included:

- Suggestion of changes to specific provisions including 9.2, 9.2.1, 9.2.3 and 9.3
- The suggestion of new provisions within the principle

7.18 Government response

To reflect the fact that many of the requirements in this principle are assigned to System Operators, this stakeholder group is now referenced in the “primarily applies to” section. Provision 9.1 has been clarified in scope to make it clear we’re referring to models, applications and systems that are released to System Operators and/or End-users. There were concerns from some stakeholders that the requirements in provision 9.2 were too prescriptive / detailed and that this should be considered alongside the fact that there isn’t an agreed international framework for red teaming. Therefore, previous provisions 9.2, 9.2.1, 9.2.2 and 9.2.3 have been scaled back to the new 9.2 and 9.2.1 which focus on the need for System Operators to conduct security testing of their systems and that independent security testers should be used. The previous provision 9.3 has been removed due to the high-level nature of the requirement and he previous provision 9.4 is now 9.3. Two new provisions, 9.4 and 9.4.1, have been added based on recommendations from some responders on the need for requirements tied to evaluating model outputs.

Secure deployment

7.19 Question 19 – Principle 10: Communication and processes associated with end-users – Do you support the inclusion of Principle 10 within the Code of Practice?

The vast majority (87%) of the 101 responses to this question supported the inclusion of Principle 10 in the Code, with only 7% not supporting its inclusion and 6% responding with ‘don’t know’.

48 of the respondents who answered ‘yes’ and 5 of the respondents who answered ‘no’ provided an answer to the respective open text follow-up questions. Frequently cited themes among responses were the suggestion of new provisions within the principle, and the suggestion of changes to provision 10.1,10.2 and 10.3.2.

Other, less commonly cited themes included:

Suggestion of changes to specific provisions including 10.3 and 10.3.1

The need for more guidance on implementing the principle

Suggestion of changes to the framing of the principle

7.20 Government response

The content of Provision 10.1 has been rephrased for clarity as multiple stakeholders interpreted its purpose differently. There are now specific requirements for System Operators and Developers. The previous provision 10.2 has been moved to 10.3 and we’ve amended it to note obligations are on both Developers and System Operators to support End-users and Affected Entities in the event of a cyber security incident.

The previous provision 10.3 is now 10.2 due to its links with 10.1. We have clarified the provision, including by noting that the guidance provided to End-users for AI systems needs to be accessible. We've split the requirement to express the actions that need to be taken by System Operators as well as Developers. The requirement has also been made a "shall" rather than a "should" requirement based on stakeholder feedback. The previous provision 10.3.1 is now 10.2.1 and we've clarified that System Operators are responsible for implementing this provision and changed the requirement from a "should" to a "shall" following stakeholder feedback. Provision 10.3.2 is now 10.2.2 and focuses on updates rather than model functionality based on feedback that in the open-source environment, the requirement would have been difficult to implement.

Secure maintenance

7.21 Question 20 – Principle 11: Maintain regular security updates for AI model and systems – Do you support the inclusion of Principle 11 within the Code of Practice?

The vast majority (89%) of the 101 responses to this question supported the inclusion of Principle 11 in the Code, with only 5% not supporting its inclusion and 6% responding with 'don't know'.

46 of the respondents who answered ‘yes’ and 4 of the respondents who answered ‘no’ provided an answer to the respective open text follow-up questions. Frequently cited themes among responses were the suggestion of changes to the framing of the principle and the suggestion of new provisions within the principle.

Other, less commonly cited themes included:

Suggestion of changes to specific provisions including 11.1, 11.2, 11.2.1 and 11.3

7.22 Government response

Principle 11 has been scaled back slightly because other areas of the Code focus on security updates and we did not want to repeat content in the document. Provision 11.1 was therefore deleted. The previous provision 11.2 is now 11.1 and we’ve set out separate requirements for Developers and System Operators for updates and patches. Provision 11.2.1 is now 11.1.1. The latter provision was refined to provide further clarity on what is expected from Developers if an update can’t be provided for AI systems. The previous provision 11.3 is now 11.2 and we have clarified that the new testing and evaluation process for a new version of a model should be focused on security. The previous provision 11.4 is now 11.3.

7.23 Question 21 – Principle 12: Monitor your system’s behaviour – Do you support the inclusion of Principle 12 within the Code of Practice?

The vast majority (87%) of the 102 responses to this question supported the inclusion of Principle 12 in the Code, with only 7% not supporting its inclusion and 6% responding with ‘don’t know’.

47 of the respondents who answered ‘yes’ and 6 of the respondents who answered ‘no’ provided an answer to the respective open text follow-up questions. A frequently cited theme among responses was the suggestion of changes to the framing of the principle.

Other, less commonly cited themes included:

Suggestion of changes to specific provisions including 12.1, 12.2, 12.3 and 12.4

Suggestion of new provisions within the principle

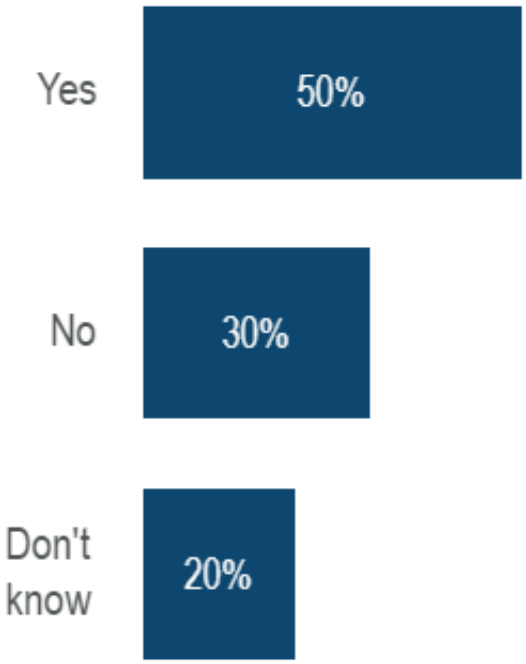
The principle is too burdensome

7.24 Government response

The layout of the provisions in this principle has changed so that the requirements align with the actions that System Operators and Developers would likely take when maintaining their AI systems. Additionally, Provision 12.1 has been slightly amended to due to some confusion around the wording of “inputs and outputs” in relation to logging. Provision 12.2 is now 12.3 and provision 12.3 is now provision 12.4. Provision 12.4 is now 12.2 and has been expanded to reflect other areas that should be considered through the analysing of logs.

8. Section 3: Further questions

8.1 Question 22 – Are there any principles and/or provisions that are currently not in the proposed Code of Practice that should be included?



50% of the 100 responses to this question believed that there were other provisions that should be included within the code, with 30% believing there were not and 20% responding with ‘don’t know’.

48 respondents who answered ‘yes’ to this question provided further detail. There were several less commonly cited themes, including:

- The Code should include requirements focused on ethics
- The Code should include requirements focused on data security

The Code should include requirements for areas or topics outside of security

The Code should provide more detail or guidance

The Code should include requirements on end of life/ cessation

8.2 Government response

We appreciate the various recommendations provided for additional principles and provisions for the Code of Practice. A significant amount of this feedback focused on areas (ethics, bias, safety etc) which are outside the scope of the Code of Practice (which seeks to address the security risks to AI). These recommendations were therefore not taken forward.

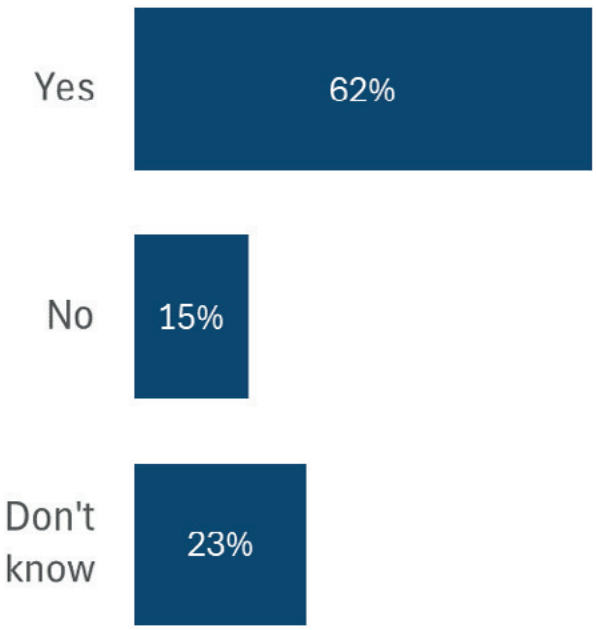
For context, ethics, bias and safety are covered by other parts of DSIT and are subsequently out of scope for this work on the cyber security of AI. The Responsible Technology Adoption Unit (RTA) leads the government's work to build trust in AI across the UK by championing responsible innovation. On bias specifically, the RTA is currently running the Fairness Innovation Challenge alongside Innovate UK, a grant challenge that has given over £465,000 of government funding to support the development of socio-technical solutions to address bias and discrimination in AI systems. Regarding the impact of AI on safety, in many cases, harmful AI content is already regulated in the UK and we are taking steps to tackle the malicious use of AI technologies, whilst ensuring young people can benefit from the opportunities AI brings. AI generated content is regulated by the Online Safety Act where it is shared on an in-scope service (user to user services, search services or service providers which publish pornographic content) and constitutes either illegal content or content which is harmful to children. Additionally, for the largest in-scope services, AI generated content is captured where it contravenes terms of service.

The Government has also committed to placing new binding requirements on the developers of the most powerful AI Models. These proposals intend to build on the voluntary commitments secured at the Bletchley and Seoul AI Safety Summits, and place the AI Safety Institute (AISI) on a statutory footing. AISI is building a world-leading technical organisation to tackle the key issues of AI safety: understanding what frontier AI risks are and will be, and how the UK and our partners overseas should deal with them. It has recruited a team of technical experts, world class researchers and engineers to evaluate publicly available frontier AI models and conduct pre-deployment testing.

AISI's research primarily focuses on AI capabilities' contribution to the most critical risks facing the UK and humanity. This includes severe catastrophic risk, cyber misuse, and the capacity for systems to act autonomously and evade human oversight. In addition, AISI has tested the robustness of system safeguards, and conducted research on the broader societal impacts from frontier AI deployment and use. AISI will pursue a route to impact which prioritises the provision of government with a continuous understanding of frontier risks, the development of AI safety tooling, and the creation of best practice approaches around which the wider international ecosystem can cohere.

We have included a new principle 13 focused on end of life, that covers the transferring or sharing of ownership of the training data and/or a model as well as the decommissioning of a model and/or system. This is based on the feedback received, the UK's support for the Council of Europe's declaration that references OECD's definition that the AI lifecycle can include end of life and our willingness to align with other international efforts. As mentioned earlier, we have published alongside this response and the updated Code, an implementation guide to address feedback that more detail and guidance is needed on the Code's principles.

8.3 Question 23 – Where applicable, would there be any financial implications, as well as other impacts, for your organisation to implement the baseline requirements?



This question was only presented for respondents who identified themselves as responding on behalf of an organisation. The majority (62%) of the 65 respondents to this question responded with ‘yes’, with 15% responding with ‘no’ and 23% responding with ‘don’t know’.

35 respondents who answered ‘yes’ to this question provided further detail. This follow-up question specifically asked respondents to provide data to support their response, however, very few responses provided this data. The most frequently cited theme among responses was the recognition that there would be costs or financial implications for their organisation to implement the baseline requirements. Another frequently cited theme was that implementing the baseline requirements would have capacity and capability implications for organisations implementing the Code.

A less commonly cited theme included that there could be a skill gap in this space and that finding expertise will be challenging.

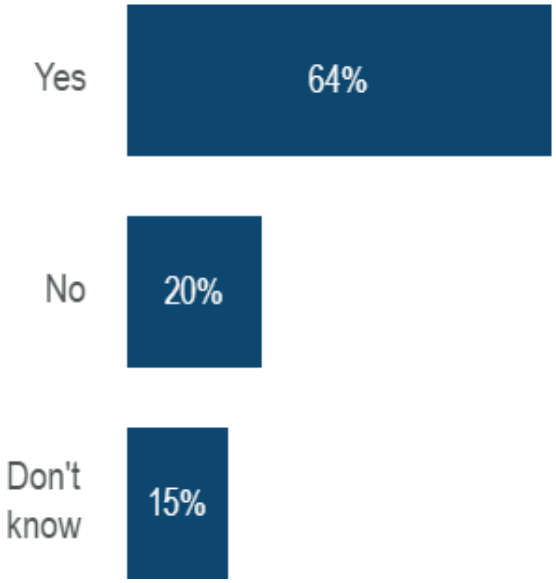
8.4 Government response

It was noticeable that very few stakeholders pointed to specific costs that would arise from implementing the Code and instead focused on resources and a broader financial impact. We recognise that implementing a new set of security requirements will bring additional costs and other challenges to stakeholders. However, there are crucial benefits that it will bring. Firstly, it will reduce the likelihood of cyber attacks and the resultant loss of money and data as well as any reputational damage that may stem from such an attack. The costs to implement good security, we believe, are outweighed by the impact of a successful cyber attack. We want organisations across the UK and abroad to be able to exploit the economic opportunities that AI can offer to improve services. Secondly, it will enable organisations to demonstrate that they are complying with a set of security requirements that have been brought together from an extensive list of international frameworks and standards. For some entities offering AI services, this will provide the organisations with an opportunity to positively differentiate themselves from their competitors whilst ensuring safeguards are in place for their employees and customers.

We have therefore created the updated Code and implementation guide with consideration of the costs for stakeholders across the AI supply chain. The Government believes that by creating a new global standard based on this updated Code of Practice at ETSI, which publishes their standards for free, many early adopters of the UK Code of Practice will already be at least partially compliant with the standard.

We also recognise the concerns around a skill gap in this area. The government is supporting the UK Cyber Security Council as the body responsible for setting the standards and pathways for the cyber security profession. We are interested to see how AI impacts what is required of cyber practitioners and how that informs the required skillsets. Additionally, the 2025 publication of DSIT's upcoming Cyber security skills in the UK labour market survey will include questions on AI cyber security that will be used to inform future interventions in this area.

8.5 Question 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?



The majority (64%) of the 98 respondents to this question agreed with the list of alternative actions, with 20% opposing and 15% answering with ‘don’t know’.

28 respondents who answered ‘no’ to this question provided further details. A frequently cited theme among these responses was support to mandate the security requirements.

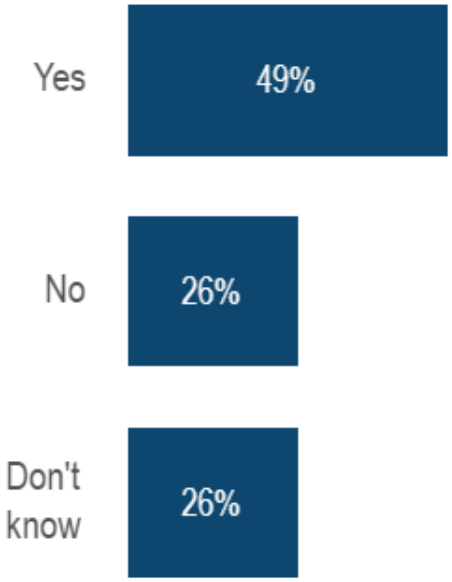
Other, less commonly cited themes included:

- Support for guidance for industry stakeholders
- Agreement with DSIT’s analysis that regulation would be burdensome, particularly for smaller companies

8.6 Government response

We welcome the support for our analysis of alternative proposals and our rationale for not progressing with them currently. We recognise a minority of respondents supported mandating elements of the security requirements within the Code of Practice immediately. However, as indicated in Question 7, there was clear support for the Government to progress ahead with the two-part intervention. As noted above, the UK will continue to work with international partners to build international consensus for baseline security requirements in this area. Our priority for now will be to socialise the updated Code of Practice, implementation guide, and develop a global standard within ETSI.

8.7 Question 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?



48% of the 101 respondents to this question believed there were other policy interventions not included in Annex E, with 27% answering with ‘no’ and 26% answering with ‘don’t know’.

47 respondents who answered ‘yes’ to this question provided further detail. A frequently cited theme identified among responses was support for investment in developing skills, such as to implement AI cyber security.

Other less commonly cited themes included:

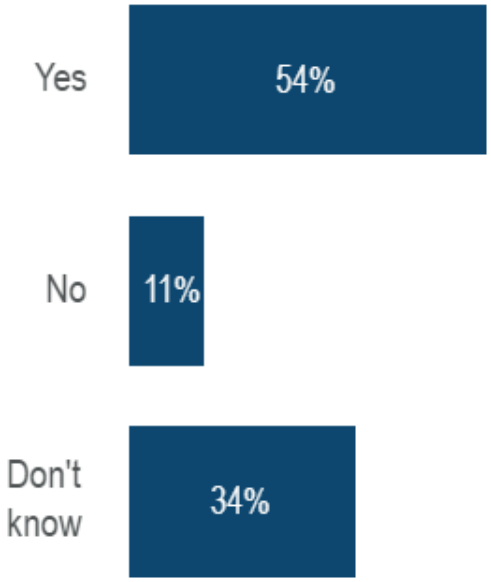
- The need to collaborate with stakeholders to ensure consistent development of policy
- Support for the creation of an implementation guide
- Support for a certification scheme to prove adherence to the Code
- Support for UK government investment in more AI security research
- Support for government to utilise procurement frameworks/processes to promote the Code

8.8 Government response

We recognise that there was significant support for our rationale for not taking forward various other interventions set out in the Call for Views document at this time based on responses to Q24. However, we note that Q25 identified some further areas, particularly different viewpoints on a potential AI security skills gap. We have therefore undertaken a study to evaluate what AI cyber security services are being offered in the UK market. This has highlighted that there are a significant number that offer various services that map across to the Code of Practice. We have also created the implementation guide to further to support stakeholders, particularly SMEs, who may lack technical expertise on AI security. Lastly, we are working with colleagues who are leading cyber skills policy to support their various initiatives, including the CyberFirst programme.

In the context of the other less commonly cited themes, we are working with government colleagues and regulators on other interlinked areas, including software, AI policy, data protection and procurement to ensure this work is aligned. We are continuing discussions with the assurance/certification sector to encourage involvement from the sector and wider industry to contribute to the development of the global standard.

8.9 Question 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?



The majority (54%) of the 96 respondents to this question believed there were other initiatives or forums the government should be engaging with, with 11% answering ‘no’ and 34% answering ‘don’t know’.

55 respondents who answered ‘yes’ provided further detail. The most frequently cited theme was the need to collaborate with other government entities. Another frequently cited theme was the need to engage with other standards development organisations.

Other, less commonly cited groups that were suggested included:

- Not for profit organisations
- International organisations

Membership/community interest bodies

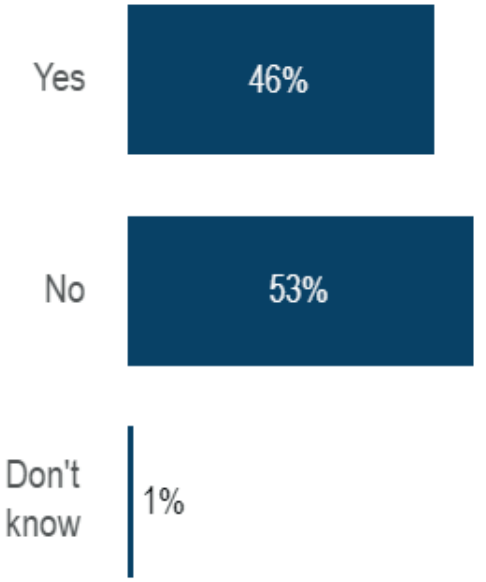
8.10 Government response

Prior to the Call for Views, we had been engaging with various government entities, standards development organisations as well as stakeholders that form the other groups highlighted by responses. We plan to increase this engagement and utilise the insights and recommendations provided by responders to the Call for Views to further promote this work area. This is vital because we believe a global approach is needed for this area.

While we intend to create a global standard in ETSI, we fully recognise the importance of work being undertaken across multiple standards development organisations such as ISO, CEN-CENELEC, and ITU. We will continue to monitor work being undertaken through these organisations to support standardisation efforts on AI cyber security and ensure we are internationally aligned in our approach.

8.11 <u>Question 27</u> – 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-commissioned risk assessment (which has published alongside the Call for Views document)? Risk is defined here as "The potential for harm or adverse

consequences arising from cyber security threats and vulnerabilities associated with AI systems".



The majority (53%) of the 99 respondents to this question answered with 'no', with 46% answering with 'yes' and 1% with 'don't' know'.

43 respondents who answered 'yes' provided further details. Responses to this question were very detailed and varied. Frequently cited themes to this question were risks linked to data and risks specific to AI cyber security risks, such as those associated with frontier AI.

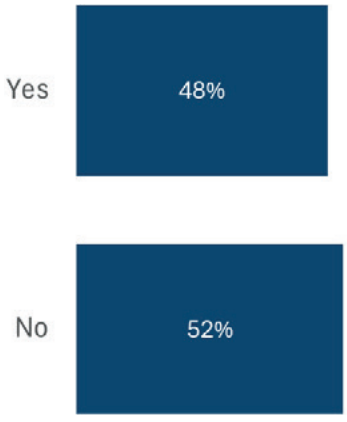
Other, less commonly cited themes included:

- Risks linked to supply chains and employees
- Crossover risks associated with different technologies

8.12 Government response

We appreciate the extensive feedback provided for this question. It has helped shape the new introduction section for the Code of Practice to highlight the distinct cyber security risks to AI. Additionally, through workshops with NCSC, we have used the responses to ensure that the Code addresses the various risks faced in the AI ecosystem. Importantly, many of the risks highlighted from responses had previously been captured in DSIT's risk assessment which was published alongside the Call for Views document. Moreover, quite a few of the risks that were signposted were outside the scope of the Code/ this work area. Lastly, we wanted to thank several stakeholders who signposted examples of cyber attacks which occurred as a result of vulnerabilities in specific AI systems. This data has been very helpful in developing the updated Code and informing DSIT's future work in this area.

8.13 Question 28 – Is there any other feedback that you wish to share?



The majority (52%) of the 99 respondents to this question did not wish to provide any further feedback, with 48% responding with yes.

48 respondents who answered ‘yes’ to this question provided further written feedback. Naturally responses to this question were varied, and many respondents thanked the government for its work in this area. One frequently cited theme was the need for the government to establish a mechanism for consulting industry on a continual basis.

Other less commonly cited themes included:

- Lack of clarity on how the Code aligns with other HMG publications
- Offers to further collaborate with DSIT
- More detail needed in the Code
- The Code needs to be regularly updated/reviewed
- Further changes suggested to the Code
- Support for mandating security requirements
- Requests that encouraged international standards and alignment

8.14 Government response

As illustrated in section 2, we have taken onboard a variety of feedback to the Code of Practice via the Call for Views. This updated Code was tested with a variety of close stakeholders within industry and across government. Equally, we believe the implementation guide provides that additional detail for organisations, particularly SMEs, with implementing the Code and future standard, without overburdening the Code itself.

We now intend to conduct a variety of external stakeholder engagement on these products and will explore how we can best engage with industry for future updates on this work as well as DSIT's wider cyber security and AI activities. We also want to ensure that stakeholders have full clarity on how our work aligns with other HMG publications and will consider how we can best socialise and take our modular approach (as described within the Call for Views) forward.

We welcome further engagement and dialogue on this topic and will collaborate, support and share information with the global community as we all look to ensure we extract the best from AI and realise its full potential.

E03283358
978-1-5286-5409-8