VE3’S RESPONSE TO DSIT'S AI CYBERSECURITY CODE OF PRACTICE CONSULTATION

Comprehensive Feedback and Recommendations on DSIT’s AI Cybersecurity Code of Practice

Abstract

This document outlines VE3's response to the Department for Science, Innovation and Technology (DSIT) consultation on the proposed Code of Practice for the cybersecurity of AI models and systems. VE3 strongly supports the Government’s initiative to establish a voluntary Code of Practice as part of the global standard for AI cybersecurity. The response provides detailed feedback on the proposed principles, including recommendations for improvement in areas such as the alignment of AI cybersecurity with existing standards, the inclusion of ethical considerations, and the practical implementation of the Code. VE3 also highlights additional principles that could be incorporated to address emerging risks, such as those linked to Frontier AI. Through this response, VE3 aims to contribute to the creation of a robust, comprehensive, and internationally harmonized cybersecurity framework for AI systems.

VE3

Introduction:

The rapidly evolving landscape of Artificial Intelligence (AI) presents both unprecedented opportunities and significant challenges. As AI systems become more integrated into various sectors, from healthcare to finance to critical infrastructure, ensuring their security has become a paramount concern. The Department for Science, Innovation and Technology (DSIT) has recognized this by proposing a Code of Practice focused on the cybersecurity of AI models and systems. This initiative is a crucial step towards establishing a robust framework that safeguards against the unique threats posed by AI technologies while fostering innovation.

VE3, as a leader in AI development and ethical AI implementation, welcomes the opportunity to provide comprehensive feedback on the proposed Code of Practice. With our extensive experience in deploying AI across various industries, we understand the critical importance of balancing security with innovation. Our response aims to address the specific provisions within the Code, offering insights and recommendations that we believe will strengthen its effectiveness and applicability.

This document outlines VE3's perspectives on the key principles and provisions proposed by DSIT, emphasizing the need for clarity, practical implementation guidance, and alignment with existing international standards. We also address potential gaps in the Code, such as the inclusion of emerging risks associated with Frontier AI and the importance of ethical considerations in AI security.

VE3 supports DSIT's efforts to create a voluntary Code of Practice that can serve as a global standard for AI cybersecurity. However, we believe that certain aspects of the Code require further refinement to ensure that it is both comprehensive and practical for organizations of all sizes. Our feedback is intended to contribute to the development of a robust, flexible, and forward-looking framework that not only mitigates the risks associated with AI but also promotes its responsible and ethical use.

Through this strategic response, VE3 aims to collaborate with DSIT and other stakeholders in shaping a Code of Practice that enhances the security and integrity of AI systems while enabling the continued growth and innovation of this transformative technology.

Company Introduction

About VE3

VE3 is a leading technology company specializing in the development and implementation of advanced artificial intelligence (AI) systems with a strong emphasis on ethical AI practices. Our unwavering commitment to cutting-edge AI research and development, combined with our strategic partnerships with industry leaders, positions VE3 as a trusted partner for organizations seeking to leverage AI responsibly and effectively. We are dedicated to delivering AI solutions that not only drive significant business value but also adhere to the highest standards of ethical and secure implementation.

VE3’s AI Expertise and Services

At VE3, we pride ourselves on our extensive expertise in AI technologies, which spans across key areas such as machine learning, deep learning, and natural language processing (NLP). Our services are tailored to meet the unique needs of our clients, ensuring that our AI solutions are not only innovative but also aligned with their strategic objectives. Our proven track record includes successful AI deployments across a wide range of sectors, including energy, healthcare, and public services.

Key Areas of Expertise:

- **Machine Learning and Deep Learning:** We utilize advanced algorithms to provide predictive analytics, enhance image and speech recognition, and more, empowering organizations to make data-driven decisions with confidence.
- **Ethical AI:** VE3 has developed robust frameworks and methodologies that ensure the responsible development and deployment of AI technologies, prioritizing transparency, fairness, and accountability.
- **AI Strategy and Governance:** We offer comprehensive advisory services on AI strategy, risk management, and compliance with industry standards, helping organizations navigate the complexities of AI adoption.

Ethical AI Maturity Framework

Understanding the critical importance of responsible AI, VE3 has developed an Ethical AI Maturity Framework that guides organizations in embedding ethical practices throughout the AI lifecycle. This framework is structured around five key dimensions: Strategy, Data, Technology, People, and Governance. It provides a clear roadmap for organizations to progress from initial exploration to transformative AI adoption, ensuring that ethical considerations are deeply integrated into every stage of AI development.

Framework Highlights:

- **Strategy:** Aligning AI initiatives with organizational goals while adhering to ethical standards.
- **Data:** Ensuring the highest levels of data quality, privacy, and security in AI projects.
- **Technology:** Leveraging cutting-edge AI technologies with a focus on maintaining transparency and accountability.
- **People:** Fostering a culture of ethical AI through continuous training and active stakeholder engagement.
- **Governance:** Establishing robust policies and oversight mechanisms to monitor and guide AI use.

Responsible AI Development

VE3 is committed to responsible AI development, emphasizing the principles of safety, security, and robustness in all our projects. Our lifecycle approach to AI development encompasses thorough planning, rigorous testing, and continuous monitoring. We integrate ethical reviews, conduct comprehensive risk assessments, and implement continuous feedback loops to ensure that our AI solutions are not only effective but also trustworthy and sustainable.

Development Approach:

- **Planning:** Identifying strategic AI opportunities and conducting detailed feasibility studies.
- **Testing:** Engaging in rigorous testing and validation processes to meet both ethical and performance standards.
- **Monitoring:** Continuously evaluating AI systems to identify potential risks and implement necessary mitigations.
- **Ethical Reviews:** Regularly conducting ethical reviews to align AI practices with organizational values and regulatory requirements.

Contributions to AI Policy and Standards

As a proactive contributor to the development of AI policies and standards, VE3 actively collaborates with industry bodies and participates in global initiatives. We are among the earliest members of the Coalition for Secure AI (CoSAI), alongside other industry giants such as Microsoft, Google, Nvidia, and IBM. Through our involvement with CoSAI, VE3 is advancing secure AI deployment and promoting best practices in the industry.

Key Contributions:

- **Policy Development:** Engaging with policymakers to influence the development of AI regulations and standards.
- **Research and Innovation:** Leading research initiatives that address emerging challenges in AI security and ethics.
- **Community Engagement:** Actively participating in forums and working groups to foster knowledge sharing and collaboration across the AI industry.

VE3's Commitment to AI Security and Ethical Practices

In recent consultations, such as our response to Ofgem's AI consultation, VE3 has consistently emphasized the importance of integrating ethical considerations and robust governance frameworks into AI deployments. Our insights highlight the potential of AI to enhance efficiency, reliability, and sustainability across various sectors, including energy. We advocate for a balanced approach that ensures the safe and effective use of AI technologies while maintaining a strong focus on ethical practices.

VE3 remains dedicated to driving innovation in AI while upholding the highest standards of ethics and security. Our comprehensive approach, grounded in our Ethical AI Maturity Framework and responsible development practices, ensures that our AI solutions are both cutting-edge and aligned with the broader goals of society. By partnering with organizations across various industries, VE3 is committed to advancing the responsible use of AI, contributing to the development of secure, transparent, and trustworthy AI ecosystems.

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

- Organisation

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

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

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

- Medium (50-499 employees)

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

- England

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

Do you agree with this proposed approach?

- Yes

VE3 acknowledges the Government's efforts to develop a voluntary Code of Practice for AI cybersecurity with the intention of aligning it with a global standard through ETSI. We agree with the need for international harmonization of standards to ensure consistent and robust cybersecurity practices across AI systems.

Several key points need addressing before the Code can effectively serve as the foundation for a global standard:

1. **Maturity of the Draft Code:** The current draft lacks sufficient practical implementation guidance, such as case studies and prescriptive recommendations. This makes it difficult for organizations to effectively implement the principles. Lessons from frameworks like the Cyber Assessment Framework (CAF) should be considered to enhance the Code's clarity and applicability.

2. **Overlap with Existing Standards:** Many of the principles in the draft Code overlap with existing software security practices and standards, such as the NIST Secure Software Development Framework and various ISO standards. Rather than creating a separate standard, the Government could focus on integrating AI-specific issues and controls into these existing frameworks. This approach would prevent redundancy and confusion while ensuring that AI systems benefit from established security practices.

3. **Modular Approach Concerns:** The Government's modular approach, layering one code of practice over another, risks becoming overly complex and difficult for organizations to navigate. It's essential to clarify how this AI Cybersecurity Code interacts with other codes, such as the Cyber Governance Code, and international frameworks like the NIST Cybersecurity Framework. Clear guidance is needed on how regulators will consider this Code in conjunction with others.

4. **Alignment with Upcoming Legislation:** There is also a need for more detail on how the AI Cybersecurity Code will align with the forthcoming AI Bill and the Cyber Security and Resilience

Bill. It's crucial to avoid overlapping or conflicting requirements for stakeholders who are subject to multiple regulations.

Given these considerations, VE3 supports the proposed approach and requests the Government to refine the draft Code, ensure it complements existing standards, and consider a more globally inclusive route for internationalization. Additionally, establishing a mechanism for ongoing consultation with industry would be beneficial in evolving the Code and ensuring it meets the needs of all stakeholders.

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

Do you agree with this approach?

- Yes

VE3 appreciates the Government's effort to clearly define the four stakeholder categories—Developers, System Operators, Data Controllers, and End-users—within the proposed Code of Practice, we believe that a more flexible responsibility model might be more beneficial.

Key Points:

1. **Responsibility Model:** A responsibility model that emphasizes the appropriate selection and use of AI tools based on their conformity to relevant codes, standards, and regulations might provide greater clarity and adaptability. This model would help organizations more effectively allocate responsibilities across different stages of the AI lifecycle, ensuring that security practices are appropriately tailored to the specific roles and activities being undertaken.
2. **Role Fluidity:** The distinction between Developers, System Operators, and Data Controllers may not always be clear-cut, especially given the fluid nature of roles in the AI value chain. For instance, data controllers may simultaneously be involved in developing and deploying AI systems, making the strict categorization potentially confusing. A more dynamic responsibility model could better accommodate these overlaps and ensure that all relevant security considerations are addressed, regardless of the stakeholder's specific designation.
3. **Focus on Conformity and Compliance:** By adopting a responsibility model that centres on the conformity of AI tools and systems to established standards and regulations, organizations can more effectively ensure compliance and security, regardless of their specific role in the AI lifecycle. This approach also aligns with VE3's emphasis on ethical AI practices, as it supports a comprehensive view of security and responsibility.

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

- No

While the actions outlined in the Code of Practice for Developers, System Operators, and Data Controllers provide a foundational framework, they may not offer sufficient detail to effectively support a comprehensive increase in the cybersecurity of AI models and systems.

Key Reasons:

1. **Lack of Practical Implementation Guidance:** The Code provides high-level principles, but it lacks specific, actionable guidance that organizations can easily implement. Practical examples, case studies, or detailed methodologies on how to apply these principles in real-world scenarios would greatly enhance the utility of the Code.

2. **Complexity of Roles:** The defined roles of Developers, System Operators, and Data Controllers overlap significantly, especially in complex AI environments where a single entity may perform multiple functions. The current actions do not adequately address these overlaps or provide clear instructions on how responsibilities should be managed when roles are fluid or shared.

3. **Insufficient Focus on AI-Specific Issues:** While the Code covers general cybersecurity practices, it does not sufficiently address AI-specific challenges, such as the unique vulnerabilities of machine learning models, the risks of data poisoning, or the complexities of managing AI-driven decision-making systems. More detailed guidance on these AI-specific issues would better support stakeholders in securing their AI systems.

4. **Integration with Existing Standards:** The Code does not clearly demonstrate how its actions align with or differ from existing cybersecurity standards and frameworks. This lack of integration could lead to confusion and inefficiencies, as stakeholders might struggle to reconcile the Code's actions with other established requirements they are already following.

5. **Need for a Responsibility Model:** As mentioned in response to Q8, a responsibility model that emphasizes the appropriate selection and use of AI tools based on their conformity to relevant codes and standards might provide a more effective framework. This would help ensure that the right actions are taken by the right stakeholders, according to their specific role in the AI lifecycle.

While the Code of Practice is a valuable step toward improving the cybersecurity of AI systems, it needs to be more detailed, practical, and aligned with existing standards to truly support stakeholders in enhancing their AI cybersecurity measures. Additional guidance and a more flexible responsibility model would make the Code more actionable and effective.

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

- Yes

VE3 supports the inclusion of Principle 1: "Raise staff awareness of threats and risks" in the Code of Practice. Ensuring that staff are knowledgeable about AI-specific threats is crucial for maintaining robust cybersecurity practices. However, we suggest the following changes to enhance the effectiveness and clarity of this principle:

1. **Integration with Broader Ethical AI Training:**
   - **Suggested Addition:** Consider adding a provision that encourages the integration of AI-specific security awareness into broader ethical AI training programs. This would help ensure that staff not only understand the cybersecurity risks but also the ethical implications of AI deployment.
   - **Example Wording:** "1.1.3 AI-specific security awareness training should be integrated with ethical AI training to provide a holistic understanding of both security, and ethical risks associated with AI systems."

2. **Frequency and Format of Updates:**
   - **Suggested Clarification:** While the principle mentions updating security awareness content every six months, it might be beneficial to specify that updates should be more frequent in fast-evolving threat landscapes. Additionally, providing more flexibility in how updates are delivered (e.g., interactive sessions, online modules) could enhance engagement.
   - **Example Wording:** "1.1.1 The AI-Security security awareness content shall be reviewed and updated at least every six months, with more frequent updates as necessary based on the evolving threat landscape. Updates should be delivered through a variety of formats to maximize engagement and retention."
3. **Secure Coding and Complexity Awareness:**
   - **Suggested Enhancement:** While the principle emphasizes secure coding and complexity awareness, it could be expanded to include continuous education on emerging AI threats and defensive strategies.
   - **Example Wording:** "1.3.2 Developers shall receive continuous education on emerging AI-specific threats and defensive coding strategies to ensure they are equipped to handle the latest security challenges."
4. **Tailored Training for Different Roles:**
   - **Suggested Addition:** Different roles within an organization face different security challenges. The principle could be strengthened by specifying that training should be tailored to the specific roles and responsibilities of the staff members.
   - **Example Wording:** "1.1.4 Security awareness training shall be tailored to the specific roles and responsibilities of staff members to ensure relevance and effectiveness."

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

- Yes

Yes, VE3 supports the inclusion of Principle 2: "Design your system for security as well as functionality and performance" within the Code of Practice. This principle aligns with our core values of responsible AI development, emphasizing the integration of security considerations from the initial design phase.

We propose the following changes to enhance the wording and clarity of the provisions within Principle 2:

2.1: As part of deciding whether to create an AI system, a System Operator shall conduct a thorough assessment that includes determining and documenting the business requirements and/or problem they are seeking to address, along with potential security risks and mitigation strategies.

2.1.1: Data controllers shall actively participate in the design and development process, providing expertise on data sensitivity, privacy considerations, and access controls throughout the AI system lifecycle.

2.2: To support the process of preparing data for an AI system, Developers shall maintain comprehensive documentation and an auditable trail of the entire data lifecycle, including data collection, preprocessing, labelling, and storage, in addition to the creation, operation, and life cycle management of models and prompts incorporated into the system.

2.3 and 2.7: Consolidate 2.3 and 2.7 and strengthen the language: If a Developer and/or System Operator decides to use an external component such as an Application Programming Interface (API) or library, they shall conduct a thorough risk assessment and due diligence process. This assessment should include evaluating the security practices, vulnerability management, and incident response capabilities of the external provider. Additionally, appropriate controls, such as input validation, data encryption, and access restrictions, should be implemented to protect data sent to or received from external services.

2.4: Data controllers shall ensure that the intended usage of the system is commensurate with the sensitivity of the data it was trained on, and that robust controls are in place to ensure the confidentiality, integrity, and availability of the data throughout its lifecycle.

2.5: Where the AI system will be interacting with other systems (internal or external), Developers and System Operators shall adhere to the principle of least privilege, granting the AI system only the minimum necessary permissions required for its intended functionality. All permissions should be subject to regular risk assessments and reviews.

General:

- **Explicitly address the balance between security, functionality, and performance:** The principle could include guidance on how to make informed trade-offs when these considerations conflict, emphasizing the importance of prioritizing security without unduly compromising functionality or performance.
- **Promote transparency and explainability**: Encourage the use of AI models and techniques that are inherently more interpretable and explainable, facilitating understanding of their decision-making processes and potential biases.

By incorporating these suggested changes, we believe Principle 2 can be further strengthened, promoting a more secure and responsible approach to AI system design.

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

- Yes

VE3 supports the inclusion of Principle 3: “Model the threats to your system” within the Code of Practice. Threat modelling is a crucial part of the risk management process and is essential for identifying and mitigating potential security risks in AI systems. However, we suggest the following changes to further enhance the clarity and applicability of this principle:

1. **Clarification of Threat Modelling Process:**
    - **Suggested Change:** The principle could provide more specific guidance on the steps involved in the threat modelling process. This would help ensure that all relevant aspects are considered during the process.
    - **Example Wording:** "3.1 Developers and System Operators shall undertake comprehensive threat modelling as part of their risk management process. This should include identifying potential threats, evaluating their impact, and determining the likelihood of occurrence.

The process should be iterative, with regular updates as new information or technologies emerge."

2. **Expansion on AI-Specific Threats:**
   - **Suggested Enhancement:** While the principle mentions AI-specific attacks and failure modes, it could provide examples or scenarios to illustrate these risks more concretely.
   - **Example Wording:** "3.1 The threat modelling process shall specifically address AI-specific attacks such as adversarial attacks, data poisoning, model inversion, and membership inference. Scenarios should be developed to assess the potential impacts of these threats on the system and its users."
3. **Inclusion of Ethical Considerations:**
   - **Suggested Addition:** It would be beneficial to include considerations of ethical risks in the threat modelling process, particularly when AI systems could have significant societal impacts.
   - **Example Wording:** "3.1.4 Developers and System Operators shall include ethical considerations in their threat modelling process, particularly where AI systems could lead to unintended societal impacts, such as bias or discrimination."
4. **Explicit Communication of Unresolved Threats:**
   - **Suggested Enhancement:** Strengthen the requirement for communicating unresolved threats to ensure that all relevant stakeholders are adequately informed.
   - **Example Wording:** "3.3 Where threats are identified that cannot be fully mitigated by Developers, they shall be clearly communicated to System Operators and End-users. This communication should include detailed descriptions of the risks, potential impacts, and recommended actions to address or monitor these threats."
5. **Risk Tolerance and Continuous Monitoring:**
   - **Suggested Enhancement:** Emphasize the importance of aligning risk tolerance with corporate governance and ensuring continuous monitoring.
   - **Example Wording:** "3.6 Developers and System Operators shall recognize that some level of risk will always remain, even after controls are applied. Continuous monitoring and regular reviews of the system infrastructure should be conducted in line with the organization's risk appetite and corporate governance policies."

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

- Yes

VE3 supports the inclusion of Principle 3: "Model the threats to your system" within the Code of Practice. This principle is essential for proactively identifying and mitigating potential security risks in AI systems, aligning with industry best practices and our own commitment to responsible AI development.

Suggested Changes:

- **3.1**: Developers and System Operators shall conduct regular and comprehensive threat modelling exercises throughout the AI system lifecycle, incorporating AI-specific threats, traditional IT system attacks, and potential social impacts.
- **3.1.1**: The risk management process shall be continuous and iterative, ensuring that any changes to the AI system, including new settings, configurations, or updates, are thoroughly assessed for potential security implications and addressed accordingly.
- **3.1.2**: As part of this process, Developers shall maintain a living document that outlines potential adversarial motivations, attack vectors, and corresponding mitigation strategies. This document should be regularly updated to reflect the evolving threat landscape.
- **3.1.3**: Developers shall conduct thorough risk assessments when utilizing models with multiple functionalities, ensuring that unused or partially utilized capabilities are appropriately secured and do not introduce unnecessary vulnerabilities.
- **3.2**: Data controllers shall conduct data protection impact assessments (DPIAs) whenever processing personal data in the context of AI systems, as mandated by UK data protection regulations, to identify and mitigate potential privacy risks.
- **3.3, 3.4, 3.5**: These provisions appropriately emphasize communication, collaboration, and risk mitigation across the AI supply chain. We support their inclusion without any suggested changes.
- **3.6**: Developers and System Operators shall acknowledge the inherent residual risk in AI systems, even with robust security controls. They shall implement continuous monitoring and evaluation processes to identify and address emerging threats and vulnerabilities, adjusting their security posture in line with their evolving risk appetite.

By adopting these suggestions, we believe the Code of Practice can provide even stronger guidance on threat modelling and risk management, contributing to the development of more secure and resilient AI systems.

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

- Yes

VE3 supports the inclusion of Principle 4: "Ensure decisions on user interactions are informed by AI-specific risks" within the Code of Practice. This principle aligns with our focus on responsible AI development, emphasizing the need to consider and address the unique risks associated with AI systems when designing user interactions.

Suggested Changes

- **4.1**: Developers and System Operators shall implement robust safeguards and controls, including human-in-the-loop processes and explainability mechanisms, to ensure that AI system outputs are reliable, accurate, and aligned with ethical and safety standards.
- **4.2**: This provision is well-articulated and emphasizes the collaboration between developers and data controllers. We support its inclusion without any suggested changes.
- **4.3**: Developers shall implement rate limiting and resource management mechanisms to protect AI systems from excessive or malicious usage, preventing denial-of-service attacks and ensuring fair access for all users.

- **4.4 & 4.5**: Combine 4.4 & 4.5 and make the language stronger: Developers and System Operators shall provide clear and accessible documentation to end-users, outlining the intended use cases, limitations, potential failure modes, and prohibited uses of the AI system. This information should promote transparency, manage expectations, and prevent overreliance on the system.
- **4.6**: If a Developer offers an API to external customers or collaborators, they shall implement robust security measures, such as authentication, authorization, input validation, and encryption, to protect the AI system from unauthorized access and malicious attacks through the API.

Overall, Principle 4 is crucial in ensuring that user interactions with AI systems are safe, secure, and aligned with ethical considerations. By incorporating our suggested changes, the Code of Practice can further enhance its clarity and effectiveness in addressing the unique risks posed by AI in user interactions.

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

- Yes

VE3 supports the inclusion of Principle 5: "Identify, track, and protect your assets" within the Code of Practice. This principle is fundamental to any robust cybersecurity framework, and its application to AI systems is critical given the sensitivity and potential value of the data and models involved.

Suggested Changes:

- **5.1:** Developers, Data Controllers and System Operators shall maintain a comprehensive inventory of their AI assets, including their physical and logical locations, and conduct regular risk assessments to identify and address any evolving security threats.
- **5.2:** Developers, Data Controllers and System Operators shall implement robust processes and utilize appropriate tools to track, authenticate, manage version control, and secure their AI assets throughout their lifecycle.
- **5.3:** System Operators shall implement backup and recovery mechanisms to enable the restoration of AI systems to a known secure state in the event of a compromise or data loss.

Overall, Principle 5 provides a strong foundation for asset management and protection in the context of AI systems. By incorporating these suggested changes, the Code of Practice can further enhance its clarity, comprehensiveness, and effectiveness in ensuring the security of AI assets.

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

- Yes

VE3 supports the inclusion of Principle 6: "Secure your infrastructure" within the Code of Practice. This principle is fundamental to protecting AI systems from unauthorized access, data breaches, and other cybersecurity threats, and aligns with our commitment to building robust and secure AI solutions.

Suggested Changes:

- **6.1**: In addition to implementing foundational cybersecurity practices for system infrastructure, Developers and System Operators shall adopt a zero-trust security model, employing robust access controls, authentication mechanisms, and encryption protocols to protect their APIs, models, data, and training/processing pipelines.

- **6.2 and 6.2.1 & 6.2.2**: These provisions are well-structured and emphasize the importance of data segregation and isolation for security. We support their inclusion without any suggested changes.
- **6.3**: Developers and System Operators shall establish and maintain a clear and accessible vulnerability disclosure process, encouraging responsible reporting of security vulnerabilities and ensuring timely remediation. This process should promote transparency and collaboration with the security research community.
- **6.4**: Developers and System Operators shall develop and regularly test a comprehensive incident response plan that outlines procedures for identifying, containing, and recovering from security incidents. This plan should include clear communication protocols and escalation procedures.

Additional Considerations

- **Regular Security Audits and Penetration Testing**: While not explicitly mentioned in the principle, we recommend adding a provision emphasizing the importance of conducting regular security audits and penetration testing to proactively identify and address vulnerabilities in the AI system infrastructure.
- **Secure Configuration Management**: The principle could also benefit from including guidance on secure configuration management practices, ensuring that AI systems are deployed and maintained with secure configurations and settings.

Overall, Principle 6 establishes a strong foundation for securing AI infrastructure. By incorporating these suggested changes and additional considerations, the Code of Practice can further enhance its effectiveness in promoting robust security practices and protecting AI systems from cyber threats.

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

- Yes

VE3 supports the inclusion of Principle 7: "Secure your supply chain" within the Code of Practice. This principle is of paramount importance in today's interconnected world, where AI systems often rely on a complex network of third-party components and data sources. Ensuring supply chain security is vital for maintaining the overall integrity and resilience of AI systems.

Suggested Changes:

- **7.1**: Developers and System Operators shall establish clear security requirements and contractual obligations for all suppliers and third-party providers involved in the AI supply chain. These requirements should be aligned with the organization's risk management policies and regularly reviewed and updated to address evolving threats.
- **7.2 and 7.2.1**: Consolidate 7.2 and 7.2.1 and strengthen the language: Developers and System Operators shall prioritize the use of well-secured and well-documented hardware and software components from trusted sources. In cases where the use of components with limited documentation or security assurances is unavoidable, a thorough risk assessment should be conducted, and clear justifications should be documented and communicated to relevant stakeholders, including end-users and System Operators.

- **7.3**: For mission-critical AI systems, Developers and System Operators shall establish contingency plans and redundancy measures, including failover mechanisms to alternative solutions, to ensure continuity of operations in the event of a supply chain compromise or security breach.
- **7.3.1 and 7.3.2**: When utilizing publicly available data for training AI models, Developers and Data Controllers shall implement rigorous validation and sanitization procedures to ensure the data's integrity and security. Data Controllers shall also establish continuous monitoring mechanisms to detect any changes or vulnerabilities in the publicly available data sources that could impact the security of AI models.

By incorporating these suggested changes, the Code of Practice can provide more explicit and actionable guidance on securing the AI supply chain, contributing to the development of more resilient and trustworthy AI systems.

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

- Yes

VE3 supports the inclusion of Principle 8, recognizing the crucial role of documentation in ensuring transparency, accountability, and the effective management of AI systems. We believe that clear and comprehensive documentation is essential for understanding AI systems, facilitating audits, and supporting incident response efforts.

Suggested Changes:

We propose the following changes to enhance clarity and reduce documentation burden:

1. **Streamlining Documentation (8.1 and 8.1.1)**: We suggest consolidating 8.1 and 8.1.1 and adjusting the language to emphasize concise but comprehensive documentation that focuses on key security-relevant aspects, including:
   - Core model design principles and rationale
   - Key stages of development, training, and deployment processes
   - Post-deployment maintenance and monitoring plans
   - Security-relevant information, such as data sources (with appropriate anonymization or aggregation), intended scope and limitations, key guardrails, retention policies, and potential failure modes
2. **Clarifying Complexity Documentation (8.1.2)**: Developers should document key areas of model and system complexity that could introduce security vulnerabilities. This should include high-level information about software dependencies and configurations, striking a balance between transparency and the protection of proprietary information.
3. **Strengthening Output Sanitization (8.2)**: We recommend reinforcing the language to emphasize the implementation of output sanitization and filtering mechanisms to prevent data leakage or exposure of sensitive metadata.

Additional Feedback:

- **Proprietary Information**: Documentation requirements should be balanced with the need to protect legitimate trade secrets. The Code of Practice should clarify that sensitive details can be anonymized or aggregated to maintain confidentiality.
- **Cryptographic Hashes or Signatures**: The Code should clarify the use of cryptographic hashes or signatures for verifying the integrity and authenticity of AI models and data.

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

- Yes

VE3 supports the inclusion of Principle 9: "Conduct appropriate testing and evaluation" within the Code of Practice. However, there are some areas where additional clarity and adjustments could be beneficial, particularly regarding the sharing of proprietary information, the role of benchmarking, and the scope of post-deployment testing.

1. **Clarity on Sharing Proprietary Information:**
    - **Suggested Change:** It's essential to provide guidance on the extent to which companies are expected to share proprietary information about their AI models during testing and evaluation, particularly in relation to independent evaluations and collaboration with System Operators.
    - **Example Wording:** "9.5 Developers shall be transparent about the security and performance aspects of their AI models during testing and evaluation. However, the sharing of proprietary information shall be limited to what is necessary to ensure the integrity and security of the evaluation process, with appropriate confidentiality agreements in place to protect intellectual property."
2. **Clarification on Benchmarking Reference:**
    - **Suggested Change:** Ensure consistency between principles by referencing benchmarking in Principle 2 or clarifying its role within Principle 9.
    - **Example Wording (in Principle 2):** "Developers should perform benchmarking as part of the design and development process, ensuring that AI systems meet established performance and security standards. This benchmarking process should continue throughout the AI lifecycle as part of ongoing risk management (see Principle 9 for more detail)."
3. **Scope of Post-Deployment Testing (9.2.1):**
    - **Suggested Clarification:** Clarify whether 9.2.1 applies only when the Developer is also the System Operator, or if it applies more broadly. If the latter, consider specifying the conditions under which post-deployment testing by Developers is necessary to avoid undue burden.
    - **Example Wording:** "9.2.1 Developers shall work closely with System Operators for post-deployment testing where the Developer is also responsible for the operation of the AI system. In cases where the Developer is not the System Operator, the responsibility for post-deployment testing should be clearly defined in the contractual agreement, ensuring

that System Operators have the necessary tools and guidance to conduct these tests independently."

While Principle 9 is crucial for ensuring that AI systems are robustly tested and evaluated, the suggested changes provide necessary clarity on proprietary information sharing, the role of benchmarking, and the scope of post-deployment testing. These adjustments will help ensure that the principle is practical for developers to implement while maintaining the security and integrity of AI systems.

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

- Yes

VE3 supports the inclusion of Principle 10: "Communication and processes associated with end-users" within the Code of Practice. This principle highlights the importance of transparency, user education, and support, which are critical for building trust and ensuring the responsible and safe use of AI systems.

Suggested Changes:

- **10.1**: Developers and System Operators shall provide clear and accessible information to end-users regarding their role in maintaining security, including best practices for data protection and secure usage of the AI system. They shall also be transparent about how user data may be used, accessed, or stored, ensuring compliance with relevant privacy regulations.
- **10.2**: Developers and System Operators shall establish and communicate clear incident response and support procedures to end-users, outlining the steps to be taken in the event of a cybersecurity incident. These procedures should be documented and readily available to affected parties.
- **10.3 and 10.3.1**: Combine 10.3 and 10.3.1 and strengthen the language: Developers shall provide end-users with comprehensive and user-friendly guidance on the secure and responsible use of the AI system. This guidance should cover proper configuration, integration with other systems, and understanding the system's limitations, potential biases, and failure modes.
- **10.3.2:** Developers shall proactively inform end-users of any significant updates or changes to the AI model's functionality, providing clear explanations and offering opt-out options where appropriate, respecting user autonomy and control over their data and interactions with the system.

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

- Yes

VE3 supports the inclusion of Principle 11 within the Code of Practice. Regular security updates are vital for maintaining the security and integrity of AI systems over time. To enhance the clarity and effectiveness of this principle, we offer the following suggestions:

1. **Clarification on Security Audits:**
   - We recommend providing more detailed guidance on what should be included in security audits. This includes specifying the scope of the audits, such as evaluating compliance

with security standards, identifying new vulnerabilities, and assessing the effectiveness of implemented security controls. Additionally, it would be beneficial to emphasize the importance of conducting these audits regularly, particularly in dynamic AI environments where new risks may emerge frequently.

2. **Guidance on Security Updates and Notifications:**
   - Clearer guidance on the process for delivering security updates and notifying System Operators and End-users is essential. Best practices should be outlined to ensure timely and effective communication of updates, including instructions on how to apply them, the risks of not updating, and a summary of the issues addressed by the update. Providing a timeline for when updates need to be applied to maintain security would also be helpful.
3. **Mechanisms for Handling Unpatched Vulnerabilities:**
   - It is important to expand on the mechanisms for addressing vulnerabilities that cannot be patched. This should include strategies for mitigating risks associated with these vulnerabilities and fostering collaboration with the wider community to develop workarounds or alternative solutions. Clear guidance on how to communicate these issues to stakeholders, including publishing security bulletins, should also be included.
4. **New Version Testing Requirements:**
   - The principle should reinforce the importance of treating major system updates as new versions, requiring a full testing and validation process before release. This process should ensure that updates do not introduce new vulnerabilities or negatively impact existing functionalities. Documentation of the testing process and communication of significant findings to System Operators and End-users should be emphasized.
5. **Support for System Operators:**
   - Additional guidance should be provided on how Developers can support System Operators in evaluating and responding to model changes. This could include offering tools, documentation, and training to help System Operators effectively manage and deploy updates. Providing preview access through beta-testing and versioned APIs would also support this process.

These suggested enhancements are designed to ensure that Principle 11 effectively supports the ongoing security of AI models and systems through regular updates and maintenance. By providing clearer guidance on audits, update processes, and support mechanisms, this principle will help organizations maintain robust security practices throughout the AI system lifecycle, aligning with VE3’s commitment to responsible and secure AI development.

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

- Yes

VE3 supports the inclusion of Principle 12: "Monitor your system's behaviour and inputs" within the Code of Practice. Continuous monitoring and analysis of AI system behaviour and inputs are fundamental for detecting anomalies, potential security breaches, and performance degradation. It enables proactive identification and mitigation of risks, contributing to the overall resilience and trustworthiness of AI systems.

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

- Yes

VE3 believes that the proposed Code of Practice is comprehensive but could be further strengthened by including additional principles and provisions that address emerging concerns and ensure a more robust approach to AI security and ethics. We suggest the inclusion of the following principles and provisions:

1. **Principle: Ethical AI Usage and Fairness:**
   - **Rationale:** As AI systems are increasingly being deployed in sensitive areas such as healthcare, law enforcement, and finance, it is crucial to ensure that these systems are used ethically and do not perpetuate biases or discrimination. A principle focused on ethical AI usage would emphasize the importance of fairness, transparency, and accountability in AI systems.
   - **Suggested Provision:** "Developers and System Operators shall ensure that AI systems are designed and deployed in ways that uphold ethical standards, including fairness, transparency, and accountability. This includes implementing measures to prevent bias in AI models, providing clear explanations of AI decisions, and ensuring that AI systems do not unfairly discriminate against any individuals or groups."
2. **Principle: Robustness and Resilience**
   - **Rationale:** AI systems should be designed and implemented to withstand adversarial attacks, unexpected inputs, and system failures. This principle would emphasize the importance of building resilience into AI systems to maintain their functionality and security even in the face of disruptions.
   - **Suggested Provisions:**
     1. Developers and System Operators shall implement measures to enhance the robustness of AI models and systems, such as input validation, error handling, and graceful degradation.
     2. AI systems should be designed to detect and recover from failures, ensuring minimal disruption to operations and user experience.
     3. Regular stress testing and fault injection exercises should be conducted to evaluate the system's resilience under various adverse conditions.
3. **Principle: Explainability and Interpretability**
   - **Rationale:** The ability to understand and interpret AI system decisions is crucial for building trust, ensuring fairness, and identifying potential biases or vulnerabilities. This principle would encourage the use of explainable AI techniques and tools.
   - **Provisions:**
     1. Developers should prioritize the use of inherently interpretable models or implement explainability techniques to provide insights into AI system decision-making processes.
     2. System Operators should be able to provide explanations for AI-generated outputs, especially in high-stakes or sensitive applications.

3. Documentation should include information on the factors influencing AI system decisions and any known limitations or biases.

4. **Principle: Continuous Learning and Adaptation**

   - **Rationale:** AI systems often operate in dynamic environments and are exposed to new data and challenges over time. This principle would emphasize the need for ongoing monitoring, evaluation, and adaptation to maintain security and performance.

   - **Provisions:**

     1. Developers and System Operators should implement mechanisms for continuous learning and adaptation of AI models, ensuring they remain effective and secure in the face of evolving threats and data distributions.

     2. Regular retraining and updates should be performed based on new data and insights.

     3. Monitoring and evaluation processes should be in place to track model performance, detect drifts, and identify potential biases or vulnerabilities.

5. **Principle on Environmental Impact of AI Systems:**

   - **Rationale:** The environmental impact of AI systems, particularly those requiring significant computational resources, is becoming an increasingly important consideration. A principle addressing the sustainability of AI systems would encourage developers to consider the environmental footprint of their models and take steps to minimize energy consumption and carbon emissions.

   - **Suggested Provision:** "Developers shall consider the environmental impact of AI systems, particularly in terms of energy consumption and carbon emissions. Where possible, developers should optimize models for efficiency, implement energy-saving measures, and explore the use of renewable energy sources to power AI systems."

Reasoning:

These additional principles and provisions would enhance the Code of Practice by:

- **Addressing emerging threats**: The AI landscape is constantly evolving, and new security challenges are emerging. These additional principles would help ensure that the Code remains relevant and effective in addressing these challenges.

- **Promoting responsible AI use**: By emphasizing robustness, explainability, and continuous learning, the Code would encourage the development and deployment of AI systems that are not only secure but also transparent, fair, and adaptable to changing circumstances.

- **Building user trust**: Addressing these additional aspects would help build greater trust in AI systems by demonstrating a commitment to security, transparency, and ethical considerations.

VE3 believes that incorporating these additional principles and provisions would further strengthen the Code of Practice, promoting a more holistic and responsible approach to AI cybersecurity.

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

- Yes

As VE3 is a technology company specializing in advanced AI systems and ethical AI implementations, there would likely be several financial implications and other impacts associated with implementing the baseline requirements outlined in the proposed Code of Practice.

Financial implications:

- **Increased personnel costs:** The Code emphasizes the need for staff training and awareness programs on AI-specific security risks. This would require investments in training resources, potentially hiring additional cybersecurity experts, and allocating staff time for training sessions.
- **Enhanced security infrastructure**: Provisions related to securing infrastructure, data protection, and supply chain security would necessitate investments in advanced security tools, technologies, and processes. This might involve upgrading existing infrastructure, implementing new security solutions, and conducting regular security audits.
- **Documentation and compliance efforts**: The Code mandates extensive documentation and audit trails throughout the AI lifecycle. This would require additional resources for documentation, record-keeping, and ensuring compliance with the Code's requirements.

Other impacts:

- **Operational adjustments:** Implementing the Code's provisions might require changes to existing workflows and processes, potentially impacting productivity and efficiency in the short term.
- **Potential delays in AI development and deployment**: The emphasis on thorough testing, evaluation, and risk assessments could lead to longer development and deployment timelines, especially for complex AI systems.
- **Increased collaboration and communication**: The Code encourages close collaboration and communication between Developers, System Operators, and Data Controllers. This could require adjustments to existing organizational structures and communication channels.

Data to quantify the impact:

Providing precise financial estimates is challenging without a detailed implementation plan. However, based on our experience and industry benchmarks, we anticipate the following potential costs:

- **Staff training and awareness**: Estimated cost of £50,000 - £100,000 per year, depending on the size of the organization and the complexity of AI systems.
- **Security infrastructure enhancements**: Estimated cost of £100,000 - £500,000 or more, depending on the existing infrastructure and the specific security solutions implemented.
- **Documentation and compliance**: Estimated cost of £20,000 - £50,000 per year, depending on the complexity of AI projects and the level of documentation required.

It is important to note that these are rough estimates and actual costs may vary depending on various factors. However, this information can provide a general understanding of the potential financial implications of implementing the Code of Practice.

While there may be financial and operational challenges in the short term, VE3 believes that the long-term benefits of implementing the Code of Practice, such as enhanced security, increased user trust, and a more resilient AI ecosystem, far outweigh the initial costs.

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

- Yes

VE3 agrees with DSIT's analysis of alternative actions that the Government could take to address the cybersecurity of AI, as set out in Annex E of the Call for Views document. The approach taken by DSIT appears to be well-considered, taking into account the need to balance effectiveness with practicality, ensuring that interventions do not impose undue burdens on stakeholders or stifle innovation.

The rationale provided for prioritizing a voluntary Code of Practice, while keeping regulation and certification schemes under consideration, aligns with the need to foster international cooperation and avoid fragmented standards that could complicate compliance efforts for AI developers. Additionally, the decision to avoid burdening organizations with excessive documentation or certification requirements until a broader consensus on baseline security requirements is reached seems prudent.

Moreover, DSIT's emphasis on leveraging existing guidance and tools, such as those developed by NCSC, rather than creating new, potentially redundant, resources is a sensible approach. This strategy helps to avoid confusion and ensures that organizations can focus on meeting clear and consistent expectations.

VE3 supports the criteria used to assess the different interventions, particularly the focus on benefit vs. cost, likely effectiveness, barriers to implementation, and consistency with international approaches. The analysis reflects a balanced approach that takes into consideration the complexities of AI security while striving to promote innovation and protect users.

In conclusion, VE3 agrees with DSIT's analysis and believes that the proposed path forward is well-aligned with the needs of the industry and the broader goal of enhancing AI cybersecurity.

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

- Yes

While VE3 acknowledges that the Government's proposed interventions are comprehensive, there are additional policy interventions that could further strengthen the approach to addressing cybersecurity risks in AI. These suggestions aim to enhance existing efforts and ensure a more resilient AI ecosystem:

1. **Public-Private Partnerships for AI Security Innovation:**
    - **Proposal:** The Government could consider establishing public-private partnerships (PPPs) focused on AI security innovation. These partnerships could bring together government agencies, academic institutions, and private sector companies to collaborate on developing advanced AI security technologies, share threat intelligence, and create new standards. This would help in accelerating the development and adoption of cutting-edge security solutions tailored to the unique challenges of AI.
    - **Reasoning:** Public-private partnerships have proven effective in other areas of cybersecurity by pooling resources, expertise, and knowledge from diverse stakeholders. In the context of AI, such collaborations could drive innovation, improve response times to emerging threats, and facilitate the creation of more robust security frameworks.
2. **Incentivizing AI Security Research and Development:**

- **Proposal:** The Government could introduce incentives, such as grants, tax breaks, or funding programs, specifically aimed at encouraging research and development (R&D) in AI cybersecurity. This could include developing new methodologies for securing AI systems, creating tools for testing and validation, and exploring the ethical implications of AI security.

- **Reasoning:** By incentivizing R&D in AI security, the Government can stimulate innovation and help UK companies stay at the forefront of global AI security practices. This would also support the development of more advanced tools and techniques for securing AI systems, contributing to overall resilience in the AI ecosystem.

3. **Establishment of an AI Security Certification and Training Program:**

   - **Proposal:** The Government could establish a certification and training program for AI security professionals. This program could provide specialized training on the latest AI security practices and certify individuals who meet certain standards of expertise. This could be done in collaboration with industry bodies, universities, and professional organizations.

   - **Reasoning:** As AI systems become more widespread, the demand for skilled professionals who understand the nuances of AI security will grow. A certification program would help build a skilled workforce capable of addressing the unique security challenges posed by AI, thereby enhancing the overall security posture of AI systems in the UK.

4. **Regular Review and Update of AI Security Regulations:**

   - **Proposal:** The Government could commit to a regular review and update cycle for AI security regulations and guidelines. This would ensure that the regulatory framework remains relevant and effective in the face of rapid technological advancements and evolving cyber threats.

   - **Reasoning:** AI technology and cyber threats are constantly evolving. Regularly reviewing and updating regulations would help ensure that they remain aligned with the latest developments and continue to effectively mitigate risks without stifling innovation.

5. **Cross-Border Collaboration on AI Security Standards:**

   - **Proposal:** The Government could take a more proactive role in fostering cross-border collaboration on AI security standards. This could involve working with international bodies to harmonize AI security regulations and practices, making it easier for companies to comply with global standards and reducing the risk of conflicting regulations.

   - **Reasoning:** Given the global nature of AI development and deployment, harmonized international standards are crucial for ensuring consistent security practices across borders. Cross-border collaboration would help UK companies operate more effectively in the global market while maintaining high security standards.

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

- Yes

VE3 recommends that the Government engage with the Coalition for Secure AI (CoSAI) as part of its program of work on the cybersecurity of AI. CoSAI is a collaborative effort initiated by industry leaders, including VE3, Microsoft, Google, and IBM, to address the pressing security challenges posed by AI implementations.

Reasons for Engaging with CoSAI:

1. **Industry Leadership and Expertise:**
   - CoSAI is composed of some of the most influential companies and thought leaders in the AI industry. Engaging with CoSAI would allow the Government to tap into cutting-edge research, best practices, and innovative solutions developed by these leaders. This collaboration would enhance the Government's efforts to create robust and relevant AI security standards.
2. **Access to Cutting-Edge Standards and Research:**
   - CoSAI is at the forefront of developing standardized approaches to mitigate AI-related cybersecurity risks. Engaging with this coalition would provide the Government with early access to cutting-edge research, tools, and methodologies that have been developed by some of the world's leading technology companies. This would enable the Government to stay ahead in the rapidly evolving field of AI security.
3. **Standardization and Best Practices:**
   - CoSAI is dedicated to developing standardized approaches for mitigating AI-related cybersecurity risks. By participating in CoSAI, the Government can contribute to and benefit from the development of global standards that ensure AI systems are secure by design. This alignment would also help harmonize UK standards with international practices, making it easier for UK companies to comply with global regulations.
4. **Strategic Alignment with Global Leaders:**
   - Given the global influence of CoSAI's founding members, including VE3, Microsoft, Google, and IBM, the Government's involvement would help ensure that the UK's AI security standards are aligned with those of other major technology leaders. This alignment is crucial for fostering international cooperation, reducing compliance burdens for UK companies, and ensuring that AI systems are secure on a global scale.
5. **Facilitating Collaboration:**
   - As an active member of CoSAI, VE3 can facilitate the Government's engagement with the coalition. Our Managing Director, Manish Garg, who serves on the CoSAI Governing Board, can provide direct insights and connections, helping to align the Government's initiatives with the coalition's ongoing work. This collaboration could lead to more effective policy-making and a stronger international presence for the UK in AI security discussions.
6. **Reinforcement of the UK's Leadership in AI Security:**
   - By partnering with CoSAI, the Government would signal its commitment to leading in AI security. This partnership would reinforce the UK's position as a key player in the global AI security landscape, fostering trust in AI technologies and ensuring that UK-developed AI systems meet the highest security standards.

Conclusion: Engaging with CoSAI offers the Government a unique opportunity to collaborate with leading AI and cybersecurity experts, contribute to the development of global standards, and ensure that the UK remains at the forefront of AI security. VE3 is ready to support this engagement and facilitate connections between the Government and CoSAI to advance the shared goal of creating a secure and trustworthy AI ecosystem.

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

- Yes

As a contributing member of the OWASP AI Exchange, VE3 recognizes the evolving and intricate landscape of cyber security risks associated with AI, particularly in the realm of **Frontier AI** systems. These systems, representing the cutting edge of AI technology, introduce unique and potentially severe risks that extend beyond the vulnerabilities outlined in the DSIT's commissioned risk assessment and the Call for Views publication document. Below, we highlight several additional risks that merit attention:

1. Autonomous Decision-Making and Critical System Risks:

- **Risk:** Frontier AI systems are increasingly being integrated into critical decision-making processes, including autonomous vehicles, financial trading systems, and military applications. The compromise of these systems through cyber-attacks could lead to catastrophic outcomes, including threats to public safety, financial markets, and national security.

- **Rationale:** Autonomous systems rely on AI to make real-time decisions in complex environments. If these systems are attacked or manipulated, the consequences could be immediate and severe, surpassing traditional cyber security threats. This is a significant concern for systems operating in high-stakes environments where failure is not an option.

2. Dual-Use Technology and Misuse:

- **Risk:** The dual-use nature of Frontier AI technologies poses significant risks if these technologies are misused. This could involve their application in disinformation campaigns, cyber warfare, or the creation of autonomous weapons systems, which could have devastating impacts.

- **Rationale:** Frontier AI models, such as large language models or advanced generative systems, can be weaponized in ways that are difficult to predict or control. The potential for misuse by malicious actors is high, and the consequences could be far-reaching, affecting everything from individual privacy to global stability.

3. Model and Data Poisoning at Scale:

- **Risk:** Large-scale AI models, particularly those trained on vast, diverse datasets, are vulnerable to data poisoning attacks. Adversaries can subtly introduce biased or malicious data into the training process, leading to compromised AI systems that behave unpredictably or unethically.

- **Rationale:** Given the complexity and opacity of many Frontier AI models, it is challenging to detect and mitigate data poisoning. The impact of such attacks can be profound, particularly if the AI system is deployed in sensitive or critical areas.

4. Supply Chain Vulnerabilities:

- **Risk:** Frontier AI systems are often dependent on complex global supply chains, including specialized hardware and open-source software. These supply chains are vulnerable to compromises that can introduce security risks at multiple points, from hardware tampering to the inclusion of malicious code in software libraries.
- **Rationale:** The security of the AI supply chain is crucial for ensuring the integrity and reliability of AI systems. The SolarWinds incident underscores how a supply chain attack can have widespread and long-lasting effects, a scenario that could be catastrophic if applied to AI systems at the frontier of innovation.

5. Ethical and Governance Challenges:

- **Risk:** The rapid advancement of Frontier AI technologies has outpaced the development of comprehensive ethical frameworks and governance structures. This creates significant risks related to the deployment and use of AI systems in ways that may not align with societal norms or legal standards.
- **Rationale:** Without robust governance and ethical guidelines, there is a risk that Frontier AI systems could be deployed in ways that are harmful or unjust, leading to long-term negative consequences for society. This is particularly concerning in areas such as surveillance, autonomous weapons, and AI-driven decision-making in justice systems.

6. Adversarial AI and the Arms Race:

- **Risk:** The development of Frontier AI has led to an escalating arms race between AI developers and adversaries, with increasingly sophisticated attacks and defences. This arms race heightens the risk of AI systems being targeted by highly advanced and persistent threats.
- **Rationale:** The growing capabilities of adversarial AI, including techniques to bypass AI-driven security measures, pose a significant challenge to the security and reliability of AI systems. As AI becomes more advanced, the stakes of this arms race increase, requiring continuous innovation in defensive measures.

VE3, as a contributor to the OWASP AI Exchange, stresses the importance of addressing these additional cyber security risks associated with Frontier AI. The complexity and potential impact of these risks necessitate a proactive and collaborative approach, leveraging the collective expertise of the cyber security and AI communities. It is imperative that we continue to develop robust security frameworks and strategies that can keep pace with the rapid evolution of AI technologies, ensuring that their benefits are realized while minimizing the potential for harm.

By highlighting these additional risks, VE3 aims to contribute to a more comprehensive understanding of the cyber security landscape as it pertains to Frontier AI, and to encourage the development of targeted solutions that address these emerging challenges.

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

- Yes

VE3 appreciates the opportunity to provide feedback on the proposed Code of Practice and the broader initiatives outlined by DSIT. We would like to offer a few additional thoughts that we believe could further enhance the effectiveness of the Government's efforts in securing AI systems:

1. **Continuous Engagement and Iterative Improvement:**
   - We encourage the Government to adopt a continuous engagement model with industry stakeholders, academia, and international partners. AI and cybersecurity are rapidly evolving fields, and the Code of Practice should be seen as a living document that is regularly updated based on new insights, technological advancements, and emerging threats. This iterative approach will ensure that the Code remains relevant and effective over time.
2. **Emphasizing the Role of Ethics in AI Security:**
   - While the focus on cybersecurity is critical, we suggest that the Government also emphasizes the ethical implications of AI security practices. Ensuring that AI systems are not only secure but also used ethically and responsibly is essential for maintaining public trust and achieving long-term success in AI deployment. Incorporating ethical guidelines into the Code of Practice or as a complementary document would be a valuable addition.
3. **Global Leadership and Standardization:**
   - The UK has the opportunity to lead globally in the development of AI security standards. We encourage the Government to continue working closely with international bodies to harmonize AI security standards and practices. This global leadership can help set a high bar for AI security worldwide and ensure that UK-developed AI systems are recognized and trusted globally.
4. **Support for Small and Medium Enterprises (SMEs):**
   - Implementing the baseline requirements of the Code of Practice may be more challenging for SMEs due to limited resources and expertise. We recommend that the Government consider providing additional support for SMEs, such as grants, training programs, or access to shared security tools, to help them comply with the Code and protect their AI systems effectively.
5. **Education and Public Awareness:**
   - Building a secure AI ecosystem requires not only the participation of developers and system operators but also an informed public. We suggest that the Government invest in public education campaigns to raise awareness about AI security risks and best practices. This will help create a more knowledgeable user base that can better protect itself and contribute to the overall security of AI systems.
6. **Future-Proofing AI Security:**
   - Finally, we recommend that the Government take a forward-looking approach by considering future trends in AI and cybersecurity, such as the potential impact of quantum computing, advances in AI autonomy, and the increasing integration of AI into critical infrastructure. By anticipating these trends, the Government can develop proactive strategies to address emerging risks before they become critical issues.

VE3 is committed to supporting the Government’s efforts in securing AI systems and contributing to the development of a robust, ethical, and globally recognized AI security framework. We appreciate the opportunity to provide feedback and look forward to continued collaboration in this important area.

Concluding Note

In conclusion, VE3 fully endorses the initiatives and principles outlined in the consultation on the cybersecurity of AI. We believe that the proposed Code of Practice represents a significant step forward in establishing a robust framework that ensures the secure, ethical, and responsible development and deployment of AI systems. VE3 appreciates the opportunity to contribute our expertise and perspectives to this important discussion, and we are committed to supporting the ongoing refinement and implementation of these guidelines.

Our feedback highlights the need for a holistic approach that not only addresses the technical aspects of AI security but also emphasizes the importance of ethical considerations, transparency, and continuous improvement. By integrating these elements into the Code of Practice, we can create a comprehensive framework that not only mitigates risks but also fosters innovation and trust in AI technologies.

VE3 looks forward to continued collaboration with the government, industry stakeholders, and the broader AI community to ensure that the Code of Practice evolves to meet the challenges and opportunities of the rapidly advancing AI landscape. We are dedicated to leveraging our expertise to support the development of AI systems that are secure, transparent, and beneficial to all of society.

As we move forward, VE3 remains committed to upholding the highest standards of AI development and to playing an active role in shaping the future of AI governance. Together, we can achieve a balanced approach that promotes the safe, responsible, and ethical use of AI, ensuring that its benefits are realized across all sectors while minimizing potential risks.