Introducing the guidelines for secure AI

New guidelines will help developers make informed decisions about the design, development, deployment and operation of their AI systems.

Claire W

Artificial Intelligence (AI) systems have the potential to bring many benefits to society. However, for the opportunities of AI to be fully realised, it must be developed, deployed and operated in a secure and responsible way.

On 1st and 2nd November 2023, the UK hosted the first AI Safety Summit which brought together governments, leading technology organisations, academia and civil society to inform rapid national and international action at the frontier of AI development. The summit, building on events across the international community (such as the EU's AI Act and the G7 Hiroshima AI Process) agreed The Bletchley Declaration. The declaration acknowledges the need for inclusive and collaborative action to address risks around the most advanced and cutting edge 'frontier' AI.

On cyber security, the summit stressed the importance of a 'secure by design' approach to AI development, which is the key principle behind new Guidelines for secure AI system development, published today by the UK's National Cyber Security Centre (NCSC), US Cybersecurity and Infrastructure Security Agency (CISA), and 21 other international agencies (listed below).

Security as a core requirement of AI development

The guidelines, developed by the NCSC and CISA in partnership with international agencies, are for providers of any systems that use AI, whether those systems have been created from scratch, or built on top of tools and services provided by others. Implementing the guidelines will help providers to build AI systems that function as intended, are available when needed, and work without revealing sensitive data to unauthorised parties.

We've aimed the guidelines primarily at providers of AI systems who are using models hosted by an organisation (or are using external APIs), but we urge **all** stakeholders (including data scientists, developers, managers, decision-makers and risk owners) to read these guidelines to help them make informed decisions about the **design**, **development**, **deployment** and **operation** of their AI systems.

AI systems are subject to novel security vulnerabilities that need to be considered alongside standard cyber security threats. When the pace of development is high – as is the case with AI – security can often be a secondary consideration. Security must be a core requirement, not just in the development phase, but throughout the life cycle of the system. For this reason, the guidelines are broken down into four key areas within the AI system development life cycle:

- Secure design
- Secure development
- Secure deployment
- Secure operation and maintenance

For each section we suggest considerations and mitigations that will help reduce the overall risk to an organisational AI system development process.

These guidelines contribute to a growing body of work intended to support delivery of safe, secure and trustworthy AI. With its specific and in-depth focus on cyber security, the guidelines are complementary to the work of the G7 Hiroshima AI Process to develop a code of conduct for organisations developing advanced AI systems. They also build upon the US Voluntary AI Commitments (PDF) on 'Ensuring Safe, Secure and Trustworthy AI', set out initially in July 2023 and ratified by a number of leading AI companies and complement the Executive Order on Safe, Secure and Trustworthy Artificial Intelligence issued by President Biden in October which established new U.S. standards for AI safety and security.

Interested readers may also want to read a recent publication from the Australian Cyber Security Centre which provides approachable guidance on AI, and how to securely engage with it.

The NCSC and CISA commit to regularly reviewing these guidelines and we welcome feedback on how they have been implemented.

Claire W
NCSC Cyber Policy
The Guidelines for secure AI system development are published by the UK National Cyber Security Centre (NCSC), the US Cybersecurity and Infrastructure Security Agency (CISA), and the following international partners:

- National Security Agency (NSA)
- Federal Bureau of Investigations (FBI)
- Australian Signals Directorate's Australian Cyber Security Centre (ACSC)
- Canadian Centre for Cyber Security (CCCS)
- New Zealand National Cyber Security Centre (NCSC-NZ)
- Chile's Government CSIRT
- National Cyber and Information Security Agency of the Czech Republic (NUKIB)
- Information System Authority of Estonia (RIA)
- National Cyber Security Centre of Estonia (NCSC-EE)
- French Cybersecurity Agency (ANSSI)
- Germany's Federal Office for Information Security (BSI)
- Israeli National Cyber Directorate (INCD)
- Italian National Cybersecurity Agency (ACN)
- Japan's National center of Incident readiness and Strategy For Cybersecurity (NISC)
- Japan's Secretariat of Science, Technology and Innovation Policy, Cabinet Office (CSTI)
- Nigeria's National Information Technology Development Agency (NITDA)
- Norwegian National Cyber Security Centre (NCSC-NO)
- Poland Ministry of Digital Affairs
- Poland's NASK National Research Institute (NASK)
- Cyber Security Agency of Singapore (CSA)
- Republic of Korea National Intelligence Service (NIS)

- Cyber Security Agency of Singapore (CSA)

WRITTEN BY

Claire W

PUBLISHED

27 November 2023

WRITTEN FOR

Small & medium sized organisations

Large organisations

Public sector

Cyber security professionals

PART OF BLOG

NCSC publications