Machine learning security principles updated

Revised principles will help people make the right security decisions when developing systems with AI/ML components.

Martin R

The NCSC's 'Principles for the security of machine learning' were originally published in August 2022. Since then, a huge amount has happened in the world of artificial intelligence (AI) and machine learning (ML), and it's time for an update.

Over the last two years, AI/ML tools have become increasingly integrated into different industries and applications, from the smart in our smartphones, to the industrial control systems running our critical national infrastructure.

Most of us are exposed to ML every day, whether we realise it or not, in products like social media photo tagging and product recommendations. Large language models (LLMs) like ChatGPT and Google Gemini have captured public attention and become some of the fastest-growing technology applications in history.

This explosion of interest has made the security of ML systems more important than ever. Earlier this week, the UK co-hosted the AI Seoul Summit 2024, with a session chaired by the Prime Minister. The summit focussed on AI safety and built upon the Bletchley Declaration from 2023, which itself led to the publication of Guidelines for secure AI system development by the NCSC, CISA, and 20 other partner agencies from around the world.

The principles for the security of ML continue to underpin those higher-level AI guidelines.

As we said when the principles were originally published, *'machine learning is important, but its security is hard'*. There are inherent weaknesses in how AI/ML works, meaning that to secure systems containing AI/ML components requires effort over and above normal cyber security best practice. In this iteration they have been brought up-to-date to cover recent developments, and now include:

- risks to LLM systems
- updates that reinforce the importance of supply chain security and life cycle management
- more focus on ‘security by design’ (the idea that AI/ML tools, like any software system, should be developed in a way that treats security as a core business priority)

We have also taken the opportunity to simplify the structure of the guidance, making it easy for users to navigate and put into action.

We recognise that AI and ML can bring huge benefits to society, and we want to ensure those benefits are realised safely and securely. **Using our principles will help people make the right security decisions when developing systems with AI/ML components**.

Please use the "Was this article helpful?" box on the guidance pages if you would like to give any feedback.

Martin R
Data Science Researcher, NCSC

WRITTEN BY

Martin R
Data Science Researcher, NCSC

PUBLISHED

22 May 2024

WRITTEN FOR

Cyber security professionals

Large organisations

Public sector

PART OF BLOG

NCSC publications