Cyber Shield: The path to an agentic AI future for cyber defence

Why the UK is pioneering an initiative to develop a national scale, sovereign defence capability

In her recent inaugural GCHQ Annual Lecture at Bletchley Park, Director GCHQ announced that:

> *we need to reimagine cyber security in the AI world. In the past few months, GCHQ has developed the blueprint for a new national cyber defence capability that will hardwire cutting-edge agentic AI into machine speed cyber defence.*
>
> Anne Keast-Butler, Bletchley Park, 27 May 2026

The NCSC and the Department for Science, Innovation and Technology (DSIT) are developing this blueprint, which we are calling **Cyber Shield**. The objective of Cyber Shield is to **build a national-scale, collaborative approach to agentic cyber defence, using frontier AI to identify, reduce and resolve our national cyber risk**.

In this blog, we:

- set out why we need a new approach to cyber security, and outline our aim and vision for Cyber Shield – as well as the challenges
- invite debate and engagement from academia, Critical National Infrastructure (CNI) organisations, frontier labs, the cyber defence sector and others to help collectively solve the challenges and develop the blueprint

Why the UK needs a step change in cyber defence

The UK faces a **cyber threat** that is **growing in scale, speed and sophistication**. Attacks from hostile states, organised crime and others are increasingly disrupting services, harming businesses and exposing sensitive data. Frontier AI is accelerating this trend, with the potential to shift the balance in favour of attackers – and with serious implications for defenders.

Securing our critical technology systems

We need to keep our critical technology systems secure against both existing and emergent cyber threats.

Existing threats

Today's challenge is that there are **many preventable weaknesses.** We know that a large proportion of critical systems do not fully meet the aims set out in the Cyber Assessment Framework (CAF).

Many attacks still succeed because of basic vulnerabilities, including outdated or unsupported systems, delays in applying security updates, and weak controls over access to systems and data. These are well-understood risks, but they remain widespread, leaving the UK exposed to attacks that are often avoidable. This makes getting cyber security fundamentals in place now more important than ever.

Emerging threats

At the same time, the emerging challenge is that **AI is changing how attacks are carried out, increasing their scale and speed**.

AI is already helping attackers to conduct elements of offensive cyber activity, such as vulnerability discovery and reconnaissance at a much **greater scale and faster pace**. As a result, activities that once took weeks can now take minutes, reducing the time available for defenders to respond, detect and contain them. This increases the likelihood of successful attacks.

Fix today's challenges

Organisations must take urgent tactical action to ensure critical systems are well defended.

Act now to strengthen the fundamentals. Fundamental cyber security remains essential and organisations should prioritise:

- rapid patching of vulnerabilities
- reducing reliance on legacy systems
- adopting secure-by-design technologies

Organisations should also **start to use AI in cyber defence,** to stay ahead of the attacker**.** This means:

- using agentic AI to identify exposed vulnerabilities autonomously ('blue' team capability)
- using AI to detect and contain security incidents
- working to address the challenge of safely automating mitigation

Be ready for tomorrow's challenges

As well as getting their cyber fundamentals in place, organisations must be prepared for the future challenges.

The shift towards full lifecycle automation of attacks

While some stages of a cyber attack can already be automated, we have not yet seen fully autonomous attacks operating across the complete intrusion lifecycle in real-world systems. In practice, the complexity of these environments still requires human judgement and oversight.

However, frontier AI models are likely to become capable of operating across the full lifecycle – from initial access through to actions on objectives. This could allow attackers to move at machine speed and greater scale, reducing opportunities for detection and response. This has the potential to overwhelm traditional defences and increase the risk of advantage shifting towards the attacker.

Developing viable solutions that scale and execute at the pace we need in the modern era is the remit of the Cyber Shield.

The evolution of AI-enabled cyber capabilities

AI-enabled offensive cyber capabilities are developing rapidly, alongside a growing commercial market for such tools. Together, these trends are accelerating the pace and expanding the scale of cyber threats, and enabling more actors to exploit such capabilities. This evolution presents strategic challenges for the UK.

The responsible use of dual-use capabilities

Many advanced defensive capabilities are inherently dual-use, meaning they could be repurposed for offensive or hostile activity.

Organisations developing these capabilities must act responsibly. Identifying and fixing vulnerabilities is essential, but not sufficient on its own. Engaging with initiatives such as Cyber Shield will help ensure that these technologies deliver a net benefit to cyber security.

The Cyber Shield vision

Cyber Shield supports the ambitions of the UK government to build national-scale, AI-powered defensive capabilities that can operate at speed and scale.

How AI-powered cyber defence could work

In the near future, we envisage a world where cyber defence is supported by 'red' and 'blue' agents which identify weaknesses in systems ('red') and defend against threats in real time ('blue'). These AI systems would:

- initially identify vulnerabilities and threats at machine speed, before progressing toward automated remediation
- generate and share insight whilst detecting and containing breaches
- work under the control and authority of their owners across government and non-government institutions
- collaborate seamlessly across organisational boundaries
- contribute to improving the national security of the UK

These agents are enabled by strong foundations of data, identity, reliability, cyber security, and regulatory compliance.

Our approach: test, iterate, scale

We will initially partner with network defenders across government and critical UK sectors to test and deploy newly-researched capabilities. This will accelerate learning and improve resilience where it will have the greatest impact.

Our aim is to transition to commercially-scalable solutions to deliver a level of national resilience which is ready for the future threat.

The UK will pioneer this approach and provide a case study to the world on how to successfully engineer and deliver the future of active cyber defence in the AI era, in a safe and secure manner, consistent with our values and policies.

Core capabilities required

We will need a **number of functions** to deliver a national-scale, sovereign cyber shield capability, in association or partnership with leading frontier AI capabilities, cyber defence organisations and academia. We recognise that some of these areas present challenges which will need significant progress in research to unlock. The functions include:

1 **Reliable and explainable AI for cyber security**

This means our AI systems can be:

- used confidently in production environments at scale
- authorised by system owners to make safe, reliable and significant real-time changes in support of cyber defence, in a predictable manner

2 **Federated agents**

Agents will be federated with underpinning trust infrastructure. These agents will:

- run **national-level operations** on behalf of the country
- run under the control and authority of **individual organisations**
- secure the means to **identify, trust and communicate** between themselves, to allow co-operation

3 **Vulnerability discovery and mitigation**

Harnessing cutting-edge UK research into agentic red/blue team functions, the aim is to:

- develop and demonstrate **automated discovery of network vulnerabilities**
- develop fully **automated vulnerability mitigation** workflows

This will allow cyber defenders to operate at beyond human scale, helping them keep pace with attacker capability. These functions will be focussed on critical networks but able to operate at national scale.

4 **Co-ordinated detection and response**

Building on the red/blue team functions, the aim is to:

- provide a means for real-time sharing of insights across organisation boundaries
- get agents to leverage these insights to detect and contain adversary behaviour

5 **National-level scanning**

This means:

- automated scanning of critical UK IP ranges for exposed vulnerabilities

- analysis of aggregated data to understand national level exposure

6 **National-level mitigation**

As HMG and/or with major service providers, this means:

- automation of workflows to allow rapid national-scale mitigation, such as automated blocking of known malicious domains and networks

---

Why working in partnership is vital

The Cyber Shield vision is ambitious and wide-reaching, and faces significant delivery challenges. It cannot be developed and operated by the NCSC or government alone. In addition, there is a clear potential benefit to UK economic growth from nurturing innovation in this domain.

We need to work in partnership:

- across **government**, and with **industry** and **academia** to develop innovative approaches and evidenced solutions from national- to enterprise-scale
- with **organisational and sector-based network defenders** to deploy and integrate agentic capabilities under their control and authority into their activities

To this end, the NCSC and DSIT are working together to establish effective pathways for partners to come together and contribute.

Get involved

We invite all organisations who are interested in partnering to develop the Cyber Shield to contact the NCSC at cybershield@ncsc.gov.uk.

Harry G, Deputy Director Capability, NCSC

Peter Haigh, Deputy Chief Technology Officer, NCSC

WRITTEN BY

Peter Haigh
Deputy Chief Technology Officer, NCSC

Harry G
Deputy Director Capability, NCSC

PUBLISHED

7 July 2026

WRITTEN FOR

Cyber security professionals Large organisations Public sector

PART OF BLOG

Events and initiatives