Artificial intelligence and cyber intrusion

An NCSC assessment that examines current AI-enabled threat trends and projections of the threat landscape to 2028.

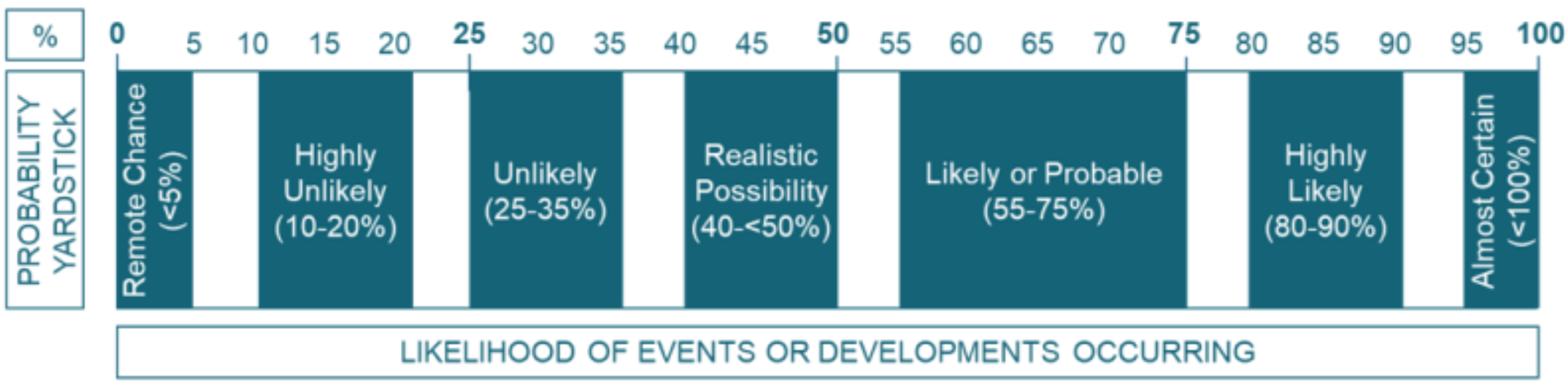
NCSC Assessment

NCSC Assessment (NCSC-A) is the authoritative voice on the cyber threat to the UK. We combine source information – classified intelligence, industry knowledge, academic material and open source – to provide independent key judgements that inform policy decision making, to improve UK cyber security. We work closely with government, industry and international partners for expert input into our assessments.

NCSC-A is part of the Professional Heads of Intelligence Assessment (PHIA). PHIA leads the development of the profession through analytical tradecraft, professional standards, to build and sustain a cross-government community.

Probabilistic language

NCSC Assessment uses the PHIA probability yardstick every time we make an assessment, judgement or prediction. The terms used correspond to these likelihood ranges:



Context

This assessment examines current AI-enabled threat trends and projections of the threat landscape to 2028 through three scenarios.

The scenarios are assessed to be plausible possibilities in the medium-term **but are not predictions**. Instead, they are designed to provide insights into the potential trajectories of AI-driven cyber threats and inform the development of proactive security strategies. Development in AI capability and its application to cyber operations is changing quickly. The prospect for surprise is high.

The three scenarios are:

1. **Baseline scenario:**
   Increased global threat from AI-cyber capability
2. **Pessimistic scenario:**
   AI-cyber capability proliferation far outstrips the UK's defensive ability
3. **Optimistic scenario:**
   AI-cyber development slows, allowing the UK to develop robust countermeasures

---

Key judgements

- By 2028, it is highly likely that Russia, China and other nation states will routinely use AI-cyber to achieve their geopolitical goals, with the UK facing an increase in scale and speed of sophisticated cyber attacks.
- By 2028, it is almost certain that many systems will remain vulnerable to attack from AI-cyber tools which identify and exploit system weaknesses, enabling attackers to display less caution over losing an access, as the cost to replace is minimal.
- By 2028, it is likely that attackers will apply AI-cyber capability against known vulnerabilities in critical national infrastructure (CNI) legacy technology, and the new opportunities in the attack surface which the deployment of AI tools within CNI allows.
- It is highly likely that between now and 2028, AI-powered malware will have been used in a ransomware attack on UK CNI.

Possible drivers

To develop the three scenarios in this paper, we have identified five key drivers likely to influence the AI-cyber threat to the UK in 2028:

- **Geopolitical tensions**
- **UK CNI cyber defence**
- **The pace of development for AI technologies**
- **Availability of AI-cyber capability**
- **AI and a cyber-literate workforce**
- **AI-cyber threat in 2028**

Baseline scenario: Increased global threat from AI-cyber capability

In 2028, the AI-cyber threat landscape facing the UK has almost certainly increased. Geopolitical tensions highly likely remain high, and nation states are almost certainly leveraging AI-powered cyber capabilities to project power and achieve strategic objectives.

State-sponsored actors, such as China and Russia, highly likely remain the greatest threat, through their developing AI capabilities. China, despite some quality lag in its chip development, has made significant strides in AI capabilities, with a focus on developing its own indigenous technologies. However, the increasing efficiency of AI models has reduced its dependence on cutting-edge western chip technology.

Meanwhile, it is highly likely that Russian cyber actors have leveraged open-source AI models, adapting and retraining them to suit their intelligence requirement needs, rather than developing in-house capabilities from scratch. This approach has allowed Russia to maintain a significant cyber capability without investing heavily in AI research and development. This has highly likely led to an increase in pre-positioning, espionage and intellectual property theft against UK interests.

The widespread availability of AI-powered cyber tools has almost certainly led to a significant increase in the discovery of zero-day vulnerabilities. Defensive AI-cyber tools have reduced the window of a vulnerability moving from zero-day to an n-day vulnerability, and has made patching a quicker process. However, without correct and widespread configuring of this technology, many systems are still vulnerable to attackers using AI-cyber tools to identify and exploit system weaknesses. Additionally, an organisation's internal patching and testing processes will need to evolve to meet the pace of accelerating vulnerability discovery. Attackers have subsequently displayed less caution over losing an access to achieve an objective, as the cost to replace is minimal. Cyber capabilities are now used for a broader range of activity due to the lower risk involved, making it easier for adversaries to access information and use it operationally.

In 2028, it is likely that attackers will apply AI-cyber capability against known vulnerabilities in CNI legacy tech and the new opportunities in the attack surface provided by the deployment of AI tools within CNI. The use of AI within CNI has almost certainly presented an increased attack surface for adversaries to exploit. UK CNI has likely seen an increased volume of sophisticated attacks as the result of AI-cyber tools, some combatted by defensive AI tools, but implementation of defensive capabilities in UK CNI lags behind offensive ones.

Pessimistic scenario: AI-cyber capability proliferation far outstrips the UK's defensive ability

In 2028, the widespread availability of frontier-AI capability to all actors, from state-level threat actors to independent neophyte hackers, has led to a significant increase in the frequency and sophistication of cyber attacks worldwide. The UK faces a significant increase in cyber attacks targeting its CNI, industry, government and civil society.

The proliferation of advanced AI-cyber capability has made it increasingly difficult to attribute attacks to specific threat actors. China, as a leading developer of AI-cyber capability, has likely played a role in disseminating advanced capability to other actors, enabling them to hide their activities within the noise. The use of proxies has become more common, further complicating attribution efforts.

A new threat has emerged in the form of an automated cyber actor, using frontier AI, which is no longer under its operator's control. It is a realistic possibility that such an entity could self-replicate and conduct cyber attacks at will, with no apparent motivation or restraint. The lack of intentionality of such an actor, and the multiple parties affected by such activity, would likely complicate geopolitical issues. This would also likely make more deniable intentional harm caused by adversaries.

The situation is further complicated by the increased commercial availability of AI tools in the criminal marketplace, including jailbroken models. This has widened the availability of advanced AI-cyber capability beyond sophisticated actors to include inexperienced cyber actors, such as novice or amateur non-professionals, who may not fully understand the risks and consequences of their actions.

In 2028, there is more potential for AI to coordinate cyber compromise impacting the physical world, because of the increased cyber-physical interface with more AI-controlled robots, cloud-controlled CNI and greater integration of cyber into physical (such as biological) labs.

The increased activity from non-state actors and immature threat actors has likely intensified geopolitical tensions, particularly where they are perceived to be emanating from a specific country or region. Espionage activity has also likely increased, with AI-enabled espionage capabilities making it easier for threat actors to remain undetected.

The UK's use of AI-cyber security tools for defensive purposes has helped to mitigate some of the threats, but it is unclear whether this will be enough to counter the rapidly evolving threat landscape. The situation is likely to be worse than is currently known, with many variables and unknowns still to be discovered.

Optimistic scenario: AI-cyber development slows, allowing the UK to develop robust countermeasures

In 2028, UK CNI has an incredibly robust defence against any, and all, cyber attacks from a range of actors using AI-cyber tools. UK CNI sectors are replacing ageing legacy systems with modern top-of-the-line systems that are less vulnerable to AI-cyber attack. The UK's cyber defence has also been enabled through widespread workforce education on AI and cyber security.

A series of private and public initiatives have been successful in educating the UK population on AI and cyber security. This has produced a workforce capable of deploying AI securely, and aware of the importance of following standard cyber security measures promptly and consistently. Organisations with well-implemented and strong cyber security measures have demonstrated their ability to deal with AI-empowered threats.

The UK's proactive approach to AI development and regulation has also paid off. It has given the defender an advantage, allowing the UK to catch up with its cyber literacy. The UK workforce, thoroughly educated in AI and cyber, has been instrumental in deploying AI within organisations effectively and promoting a culture of cyber security awareness.

The UK's robust regulatory framework has led to the development and release of fewer AI models, with a greater emphasis on responsible AI use. This has resulted in a more transparent AI ecosystem, with a better understanding of the capabilities and risks associated with AI.

It is likely that the AI-cyber threat to the UK in 2028 is significantly mitigated. UK CNI is well protected against cyber attacks from a range of actors, including state-sponsored threats and cyber criminals. Furthermore, existing prepositioning efforts from cyber threat actors have been identified and neutralised. Organisations have well-implemented cyber security measures and have demonstrated their ability to effectively deal with AI-empowered threats.

It is likely that the UK's robust defence measures and responsible use of AI will minimise the risk of a significant cyber attack, despite China and other nation states' continued development of AI capabilities. The UK's proactive approach has encouraged other countries to follow suit, leading to a more stable and secure global AI ecosystem.

Glossary

Artificial intelligence (AI)
Artificial intelligence (AI) refers to computer systems which can perform tasks usually requiring human intelligence. Generative AI (GenAI) is a form of AI which is capable of generating new content, such as text, images or video.

AI-cyber
AI-cyber refers to the application of AI developments in cyber intrusion capability.

Frontier AI
Frontier AI refers to AI systems that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced systems.

Large language model (LLM)
A large language model (LLM) is a machine-learning model that is equipped to summarise, translate, predict and generate human sounding text to convey ideas and concepts.

AI system
An AI system comprises the host infrastructure, management systems, access control systems and programming interface and can comprise multiple AI designs and models.

AI design
AI design refers to the mathematical and algorithmic processes and constraints that are used to convert inputs into outputs. They can vary widely, from narrow linear functions to execute single tasks, through to interconnected functionality that executes complex, decision-orientated tasks. An AI model describes the core AI algorithm.

Vulnerability
A vulnerability is a weakness or flaw in a computer system or process. An attacker may seek to exploit a vulnerability to gain access to a system. The code developed to do this is known as an exploit. A zero-day exploit exploits a vulnerability where there are no security fixes available. A zero-day becomes a 'known' or n-day vulnerability once a security fix has been issued by the vendor. Exploitation of known vulnerabilities rely on finding systems that have not been updated.

Also see



Frontier AI: what you need to know

Organisations need to be ready to counter the enhanced capabilities of AI-powered attacks.

PUBLISHED

3 November 2026

WRITTEN FOR

Cyber security professionals

Large organisations

Public sector