The near-term impact of AI on the cyber threat

An NCSC assessment focusing on how AI will impact the efficacy of cyber operations and the implications for the cyber threat over the next two years.

UK Cyber Policy comment

During the Bletchley AI Safety Summit in November 2023, international leaders came together to discuss the vast potential of AI models in promoting economic growth, propelling scientific advances, and providing a wide range of public benefits. They also underscored the security risks that could arise from the irresponsible development and use of AI technologies. The UK government is evaluating and addressing the potential threats and risks associated with AI.

While it is essential to focus on the risks posed by AI, we must also seize the substantial opportunities it presents to cyber defenders. For example, AI can improve the detection and triage of cyber attacks and identify malicious emails and phishing campaigns, ultimately making them easier to counteract.

The Summit Declaration highlighted the importance of ensuring that AI is designed, developed, deployed, and used in a manner that is safe, human-centric, trustworthy, and responsible for the benefit of all. The NCSC continues to work with international partners and industry to provide guidance on the secure development and use of AI, so that we can realise the benefits that AI offers to society, publishing Guidelines for Secure AI System Development in November 2023.

NCSC Assessment

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

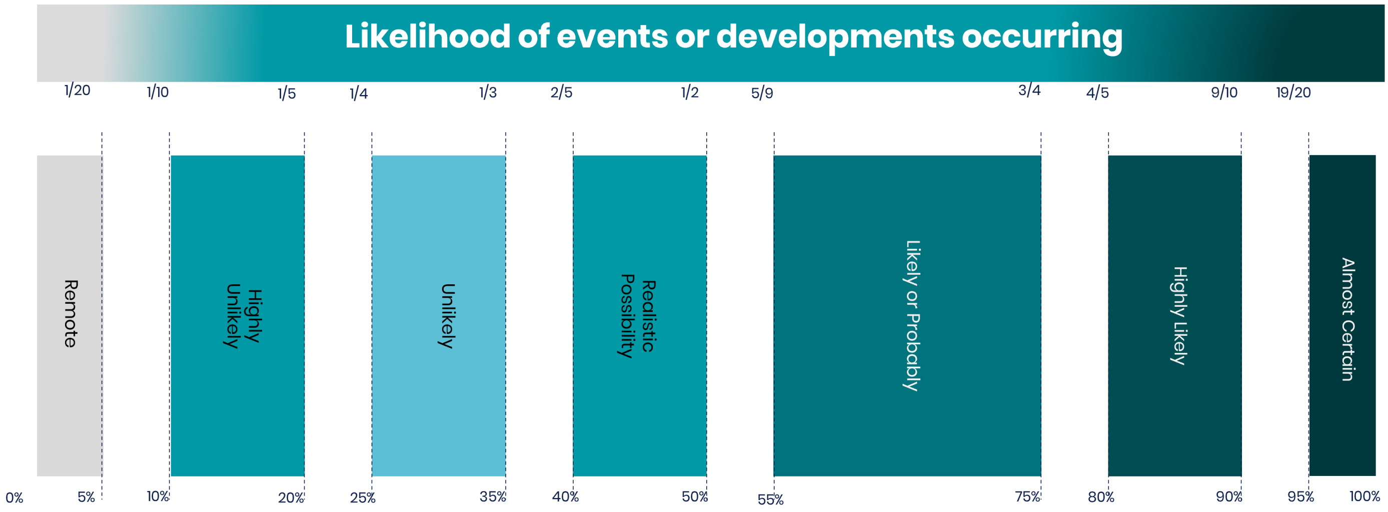
NCSC-A is part of the Professional Heads of Intelligence Assessment (PHIA). PHIA leads the development of the profession through analytical tradecraft, professional standards, and building and sustaining a cross-government community.

This report uses formal probabilistic language (see yardstick) from NCSC-A product to inform readers about the near-term impact on the cyber threat from AI. To learn more about NCSC-A, please contact enquiries@ncsc.gov.uk.

How likely is a 'realistic possibility'?

Professional Head of Intelligence Assessment (PHIA) probability yardstick

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



Key judgements

- Artificial intelligence (AI) will **almost certainly increase the volume and heighten the impact of cyber attacks** over the next two years. However, the **impact on the cyber threat will be uneven** (see table 1).

- The threat to 2025 comes from **evolution and enhancement of existing tactics, techniques and procedures** (TTPs).
- **All types of cyber threat actor** – state and non-state, skilled and less skilled – are already using AI, **to varying degrees.**
- AI provides **capability uplift in reconnaissance and social engineering**, almost certainly making both more effective, efficient, and harder to detect.
- More **sophisticated uses of AI in cyber operations** are highly likely to be restricted to threat actors with access to **quality training data, significant expertise (in both AI and cyber), and resources.** More advanced uses are unlikely to be realised before 2025.
- AI will almost certainly make cyber attacks against the UK more impactful because **threat actors will be able to analyse exfiltrated data faster and more effectively, and use it to train AI models.**
- AI lowers the barrier for novice cyber criminals, hackers-for-hire and hacktivists to carry out effective access and information gathering operations. This **enhanced access will likely contribute to the global ransomware threat** over the next two years.
- Moving towards 2025 and beyond, commoditisation of AI-enabled capability in criminal and commercial markets will almost certainly make **improved capability available to cyber crime and state actors.**

Context

This assessment focuses on how AI will impact the effectiveness of cyber operations and the implications for the cyber threat over the next two years. It does not address the cyber security threat to AI tools, nor the cyber security risks of incorporating them into system architecture.

The assessment assumes no significant breakthrough in transformative AI in this time period. This assumption should be kept under review, as any breakthrough could have significant implications for malware and zero-day exploit development and therefore the cyber threat.

The impact of AI on the cyber threat will be offset by the *use* of AI to enhance cyber security resilience through detection and improved security by design. More work is required to understand the extent to which AI developments in cyber security will limit the threat impact.

Assessment

1. The impact of AI on the cyber threat is uneven; both in terms of its use by cyber threat actors and in terms of uplift in capability.
2. Table 1: Extent of capability uplift caused by AI over next two years.

	Highly capable state threat actors	**Capable state actors, commercial companies selling to states, organised cyber crime groups**	**Less-skilled hackers-for-hire, opportunistic cyber criminals, hacktivists**
Intent	High	High	Opportunistic
Capability	Highly skilled in AI and cyber, well resourced	Skilled in cyber, some resource constraints	Novice cyber skills, limited resource
Reconnaissance	Moderate uplift	Moderate uplift	Uplift
Social engineering, phishing, passwords	Uplift	Uplift	Significant uplift (from low base)
Tools (malware, exploits)	Realistic possibility of uplift	Minimal uplift	Moderate uplift (from low base)
Lateral movement	Minimal uplift	Minimal uplift	No uplift
Exfiltration	Uplift	Uplift	Uplift
Implications	Best placed to harness AI's potential in advanced cyber operations against networks, for example use in advanced malware generation.	Most capability uplift in reconnaissance, social engineering and exfiltration. Will proliferate AI-enabled tools to novice cyber actors.	Lower barrier to entry to effective and scalable access operations - increasing volume of successful compromise of devices and accounts.

KEY: MINIMAL UPLIFT ▯ MODERATE UPLIFT ▯ UPLIFT ▯ SIGNIFICANT UPLIFT

3. AI will primarily offer threat actors capability uplift in social engineering. Generative AI (GenAI) can already be used to enable convincing interaction with victims, including the creation of lure documents, without the translation, spelling and grammatical mistakes that often reveal phishing. This will highly likely increase over the next two years as models evolve and uptake increases.

4. AI's ability to summarise data at pace will also highly likely enable threat actors to identify high-value assets for examination and exfiltration, enhancing the value and impact of cyber attacks over the next two years.

5. Threat actors, including ransomware actors, are already using AI to increase the efficiency and effectiveness of aspects of cyber operations, such as reconnaissance, phishing and coding. This trend will almost certainly continue to 2025 and beyond. Phishing, typically aimed either at delivering malware or stealing password information, plays an important role in providing the initial network accesses that cyber criminals need to carry out ransomware attacks or other cyber crime. It is therefore likely that cyber criminal use of available AI models to improve access will contribute to the global ransomware threat in the near term.

6. AI is likely to assist with malware and exploit development, vulnerability research and lateral movement by making existing techniques more efficient. However, in the near term, these areas will continue to rely on human expertise, meaning that any limited uplift will highly likely be restricted to existing threat actors that are already capable. AI has the potential to generate malware that could evade detection by current security filters, but only if it is trained on quality exploit data. There is a realistic possibility that highly capable states have repositories of malware that are large enough to effectively train an AI model for this purpose.

7. Cyber resilience challenges will become more acute as the technology develops. To 2025, GenAI and large language models (LLMs) will make it difficult for everyone, regardless of their level of cyber security understanding, to assess whether an email or password reset request is genuine, or to identify phishing, spoofing or social engineering attempts. The time between release of security updates to fix newly identified vulnerabilities and threat actors exploiting unpatched software is already reducing. This has exacerbated the challenge for network managers to patch known vulnerabilities before they can be exploited. AI is highly likely to accelerate this challenge as reconnaissance to identify vulnerable devices becomes quicker and more precise.

8. Expertise, equipment, time and financial resourcing are currently crucial to harness more advanced uses of AI in cyber operations. Only those who

invest in AI, have the resources and expertise, and have access to quality data will benefit from its use in sophisticated cyber attacks to 2025. Highly capable state actors are almost certainly best placed amongst cyber threat actors to harness the potential of AI in advanced cyber operations. Other state actors and most commercial companies that offer capability to states worldwide will gain moderate capability uplift over the next eighteen months in social engineering, reconnaissance and exfiltration. Capable and established criminal groups are also likely to have enough training data and resource to gain some uplift.

9. However, it is a realistic possibility that these factors may become less important over time, as more sophisticated AI models proliferate and uptake increases. Publicly available AI models already largely remove the need for actors to create their own replica technologies, especially in low-sophistication operations such as spear-phishing. Less-skilled cyber actors will almost certainly benefit from significant capability uplifts in this type of operation to 2025. Commoditisation of cyber crime capability, for example 'as-a-service' business models, makes it almost certain that capable groups will monetise AI-enabled cyber tools, making improved capability available to anyone willing to pay.

10. To 2025, training AI on quality data will remain crucial for its effective use in cyber operations. The scaling barriers for automated reconnaissance of targets, social engineering and malware are all primarily related to data. But to 2025 and beyond, as successful exfiltrations occur, the data feeding AI will almost certainly improve, enabling faster, more precise cyber operations.

11. Increases in the volume and heightened complexity and impact of cyber operations will indicate that threat actors have been able to effectively harness AI. This will highly likely intensify UK cyber resilience challenges in the near term for UK government and the private sector.

Glossary

Artificial intelligence

Computer systems which can perform tasks usually requiring human intelligence. This could include visual perception, speech recognition or translation between languages. Modern AI is usually built using **machine learning** algorithms. The algorithms find complex patterns in data which can be used to form rules.

Generative AI (GenAI)
AI capable of generating new content, such as text, images or video. **Large language models (LLMs)** are an example of generative AI.

Transformative AI
An advanced AI system with transformative impact on society. One example is **artificial general intelligence**, the hypothetical concept of autonomous systems that learn to surpass human capabilities in most intellectual tasks.

Reconnaissance
First stage of the cyber attack chain which involves researching a target to identify potential access vectors for a future attack.

Spear-phishing
The practice of sending targeted emails, text messages, social media, calls or other messages to individuals to lure them into an interaction, such as clicking a link, to obtain access to their system, device or account or personal information.

Social engineering
The practice of manipulating people into carrying out specific actions, or divulging information, that is of use to an attacker.

PUBLISHED

24 January 2024

WRITTEN FOR

Public sector

Small & medium sized organisations

Self employed & sole traders

Large organisations

Cyber security professionals