The Great Race: AI-Driven Cyber Offence vs AI-Driven Cyber Defence

By Andrew Horton

A new strategic arms race is underway. It is not being fought with aircraft carriers, missiles or nuclear weapons, but with algorithms. Its outcome may shape national power, economic competitiveness and institutional resilience more profoundly than any competition since the Cold War. Humanity has entered an era in which machines are competing directly against machines at machine speed. Artificial intelligence can now discover software vulnerabilities, analyse attack pathways, generate malicious code, identify targets and automate defensive responses at a pace beyond meaningful human comprehension. The result is a high-stakes contest with profound implications for governments, critical infrastructure operators, militaries, cybersecurity providers and corporations alike.

AI is simultaneously the most powerful shield ever engineered and the most sophisticated sword ever forged. Defensively, it can analyse vast quantities of data, identify subtle anomalies, automate threat hunting and assist with near real-time remediation. Offensively, those same capabilities enable automated reconnaissance, hyper-targeted phishing, rapid exploit development and the continuous mutation of malware designed to evade detection.The decisive question is whether AI-driven defence can outpace

AI-driven offence. On this point, the strategic balance remains deeply unsettled, and that uncertainty demands urgent attention in corporate boardrooms and national cabinets alike.

History records only a small number of strategic races that have reordered the international system: the Anglo-German naval race before 1914, the Cold War nuclear standoff and the space race. Artificial intelligence represents a similarly structural shift. Unlike earlier competitions, however, this one is unfolding largely out of public view, driven not only by nation-states but by a handful of commercial technology firms whose platforms increasingly underpin global digital infrastructure.

The AI-cyber race is asymmetric by design. A small number of technology providers now control the world's most advanced frontier models. As these systems become embedded within civilian infrastructure, intelligence analysis and defence platforms, states are accepting unprecedented forms of technical dependency. For nations outside the superpower duopoly, the challenge is not achieving dominance but preserving sovereign autonomy in a domain increasingly shaped by external forces.

Cybersecurity has always been a contest between attackers and defenders, but AI has fundamentally disrupted the element of time. For decades, organisations operated under an implicit assumption that a meaningful buffer existed between the discovery of a vulnerability and its exploitation. This interval allowed security teams to assess risk, allocate resources and deploy patches. That assumption is now obsolete.

As Kara Sprague, Chief Executive Officer of HackerOne, recently observed: *"Many are still operating as if they have time between when a vulnerability is discovered and when it is exploited—but with AI, that time is gone."*

AI compresses the discovery-to-exploitation window to seconds. Frontier AI systems can identify, prioritise and chain vulnerabilities with remarkable sophistication, moving from reconnaissance to enterprise compromise in a fraction of the time previously required. This speed overwhelms traditional governance structures, procurement lifecycles and manual approval processes.

Compounding this challenge is the sheer velocity of software development. Organisations are deploying AI-generated code at unprecedented scale to drive productivity and innovation. Yet every rapid deployment introduces additional dependencies, hidden integrations and unmapped attack surfaces. The paradox is stark: the same technology driving economic progress is also compounding systemic risk at an exponential rate.

The core strategic problem is that offence inherently benefits first from compressed timelines. While defenders must secure an entire, constantly expanding digital perimeter, an AI-driven attacker needs only a single vulnerability to achieve compromise.

This reality also forces a reassessment of the cybersecurity industry itself. For decades, cyber defence has been a human-intensive profession built around analysts, threat hunters and security operations centres. AI is

beginning to dismantle that model. Activities that once consumed thousands of analyst hours are increasingly being automated by machines.

The cybersecurity sector now finds itself competing in the very race it was created to manage.

This structural shift elevates cybersecurity from a technical challenge to a core institutional governance imperative. AI-era resilience cannot be measured through compliance checklists, retrospective audits or policy maturity frameworks. It must increasingly be measured through organisational velocity: the speed at which an institution can identify anomalies, make decisions and execute defensive actions.

Survival depends on answering three critical questions:

Identification: How rapidly can vulnerabilities be validated?

Decision: How quickly can executives make risk-bearing decisions?

Execution: How fast can distributed systems be patched and adapted?

Linear decision-making hierarchies designed to ensure stability now impede survival. In a machine-speed environment, bureaucracy itself becomes a security vulnerability.

The most profound implication of AI-driven cyber operations may have little to do with cybersecurity itself.

For centuries, governments, corporations and militaries have been built around a simple assumption: that human decision-making sits at the centre of important events. Humans observe, humans decide and humans act.

Artificial intelligence challenges that assumption.

As cyber operations accelerate towards machine speed, organisations may increasingly find that the pace of conflict exceeds the pace of governance. Machines identify vulnerabilities, analyse options and execute responses in milliseconds, while institutions continue to deliberate through committees, reporting lines and approval chains measured in days, weeks and months.

The strategic question is no longer whether humans remain in control. The strategic question is whether human institutions can remain relevant.

Cybersecurity is simply where this reality becomes visible first.

Anthropic's development of Mythos—a specialised AI capability designed to automate vulnerability discovery and analysis—offers an early glimpse of this trajectory. The significance of Mythos lies not in the act of discovery, but in the unprecedented speed and scale with which advanced AI performs the task. When the organisations building frontier models publicly acknowledge their inherent cyber risks and restrict deployment, leaders should pay attention.

For decades, nations recognised the risks associated with dependence on foreign energy, telecommunications and defence supply chains. The same logic must now be applied to artificial intelligence. As AI becomes embedded within threat analysis and decision-support systems, states risk becoming dependent not merely on foreign technology, but on foreign intelligence infrastructure.

For corporate leaders, the lesson is equally stark. The rush to adopt AI in pursuit of productivity and market advantage must be balanced by an equally rigorous focus on governance. Executives routinely assess what AI can optimise; far fewer ask what AI can expose, manipulate or compromise.

Every model deployed is both a capability multiplier and a potential strategic liability. Every autonomous workflow creates a new pathway for exploitation. In this environment, trust can no longer be assumed; it must be continuously and algorithmically verified.

The most resilient organisations will not be those deploying the greatest number of AI tools. They will be those embedding AI within disciplined governance frameworks, rigorous assurance testing and active human oversight.

Ultimately, technology alone will not determine the outcome of this race; institutions will.

The coming decade will not be defined by a contest between human attackers and human defenders. It will be defined by competing systems of machine intelligence operating at unprecedented speed and scale. One side is compressing the time required to exploit. The other must compress the time required to detect, respond and recover.

For global powers, this race offers a path to technological primacy. For everyone else, it presents a narrowing corridor of autonomy. Those who fail to adapt are unlikely to experience decline through a single catastrophic event, but through the gradual erosion of leverage, resilience and control.

The winners of this race will not be defined by the raw power of their algorithms, but by the agility and speed of the institutions that govern them.

In the intelligence age, the most dangerous vulnerability is not a flaw in software. It is the assumption that institutions operating at human speed can successfully govern threats evolving at machine speed.