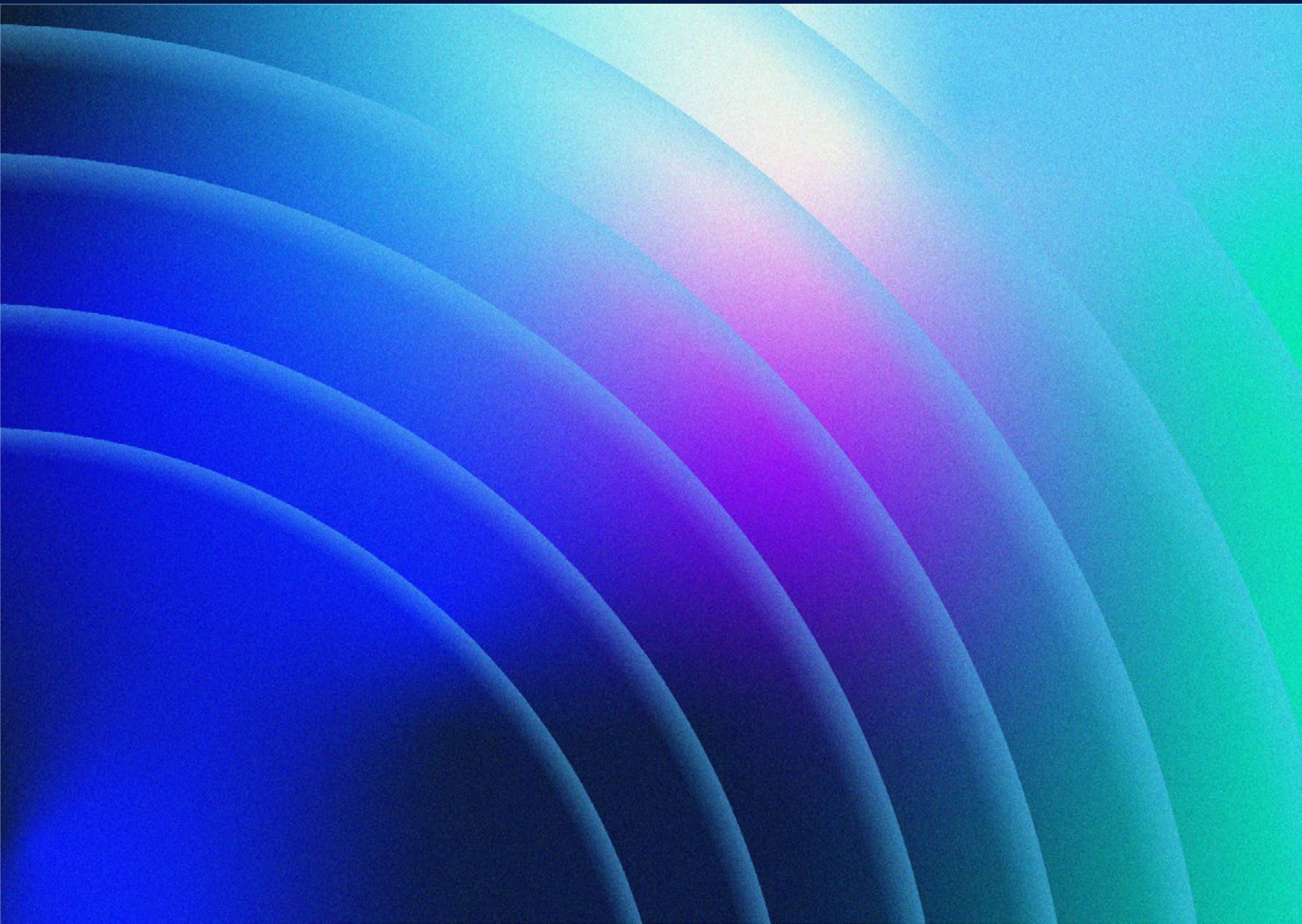
ISTARI Perspectives

# Can AI reverse the Defender's Dilemma in Cybersecurity?

Abel Archundia
Alexander Pabst

10 • 2024



# Can AI reverse the Defender's Dilemma in Cybersecurity?

On the traditional war battlefield, some military theorists say that attack is preferable over defence, as it serves a more "positive" purpose: conquest and control. Defence, on the other hand, is seen as passive, and at a disadvantage. How well does this analogy apply to a digital battlefield?

Attackers have several advantages: they can devise plans and tactics using initiative, and can take advantage of the element of surprise. Defenders must react.

The same logic can be applied to the domain of cybersecurity. Attackers are also perceived to have the advantage. The attacker only needs to find one vulnerability, whereas defending from attack must be done right each time, all of the time. This is known as the defender's dilemma.

Artificial intelligence (AI) represents an opportunity to gain an edge for both attack and defence. But is one side gaining a greater advantage than the other? Can AI reverse that dilemma by enhancing enterprise or even systemic cybersecurity capabilities?

Over the past year, the rapid adoption of large language models (LLMs) such as ChatGPT has sparked intense media speculation and hype. Entrepreneurs feel compelled to include ″AI″ in their business plans to signal value to investors, while enterprises are shifting from exploring AI to actively experimenting with it. Companies recognise the need for action— but many are unclear on the practicalities and how to best use AI. We believe a powerful use case is to apply it to enhance cyber defence.

ISTARI has completed a project for a large European multinational enterprise that assessed how AI is changing the threat landscape, and conversely how AI can enhance internal cybersecurity capabilities. Below is an excerpt of the results.

## *AI-powered cyberattacks*

Geopolitical macro trends are having a profound impact on the activities of threat actors. State-sponsored criminals continue to gain access to military-grade cyber weapons, which rapidly filter down to criminal groups. While these groups may appear fragmented, they are highly organised and adept at turning sophisticated tools into tangible, often profitable, outcomes.

Recent findings illustrate how criminal groups are leveraging both freely available open-source models and customised, subscription-based tools for malicious purposes. With AI-powered technology at their disposal, these actors are able to orchestrate faster, more complex attacks at an alarming frequency. The result is an increasingly dynamic threat landscape that demands both vigilance and adaptive defensive strategies. Here are three key impacts on attacks:

1. *Faster attacks*

AI accelerates the speed of attacks by quickly discovering optimal targets based on a set of criteria (such as industry, geography or revenue). Attackers can then use AI to identify and exploit vulnerabilities, and to quickly write and evolve exploitative code. Moreover, AI-driven, real-time decision making allows attackers to adjust their tactics dynamically during an ongoing attack, making them more effective and harder to defend against.

2. *Sophisticated attacks*

The easy access to code libraries (whether open source or by subscription) means even entry-level criminals can unlock sophisticated capabilities. They can use AI to orchestrate their attacks, coordinating multiple stages of an attack with precision. Perhaps most concerning are AI-generated adaptive malware, which automatically changes characteristics or behaviour to bypass defences by adjusting and defeating cybersecurity measures in real-time.

Furthermore, AI enables highly personalised, accurate and effective phishing campaigns, using detailed insights to target victims. The creation of deepfake voice, image, and video content that is indistinguishable from reality adds a new layer of deception, making it harder for traditional detection methods to identify and block these attacks.

3. *Frequency of attacks*

Intelligent and persistent AI-enabled reconnaissance allows cybercriminals to continuously monitor and assess potential targets. AI also helps automation and the continuous deployment of attacks, allowing attackers to strike efficiently, repeatedly and at scale. This scalability means that multiple systems can be attacked concurrently with minimal effort.

From a threat scenario perspective, our analysis of model adoption and the models themselves reveals that AI is likely to have the most significant impact on ransomware, phishing and social engineering, as well as technical vulnerability exploits. The risk posed by third-party vendors grows in parallel with their exposure to AI-driven attacks, especially when those relationships are poorly monitored.

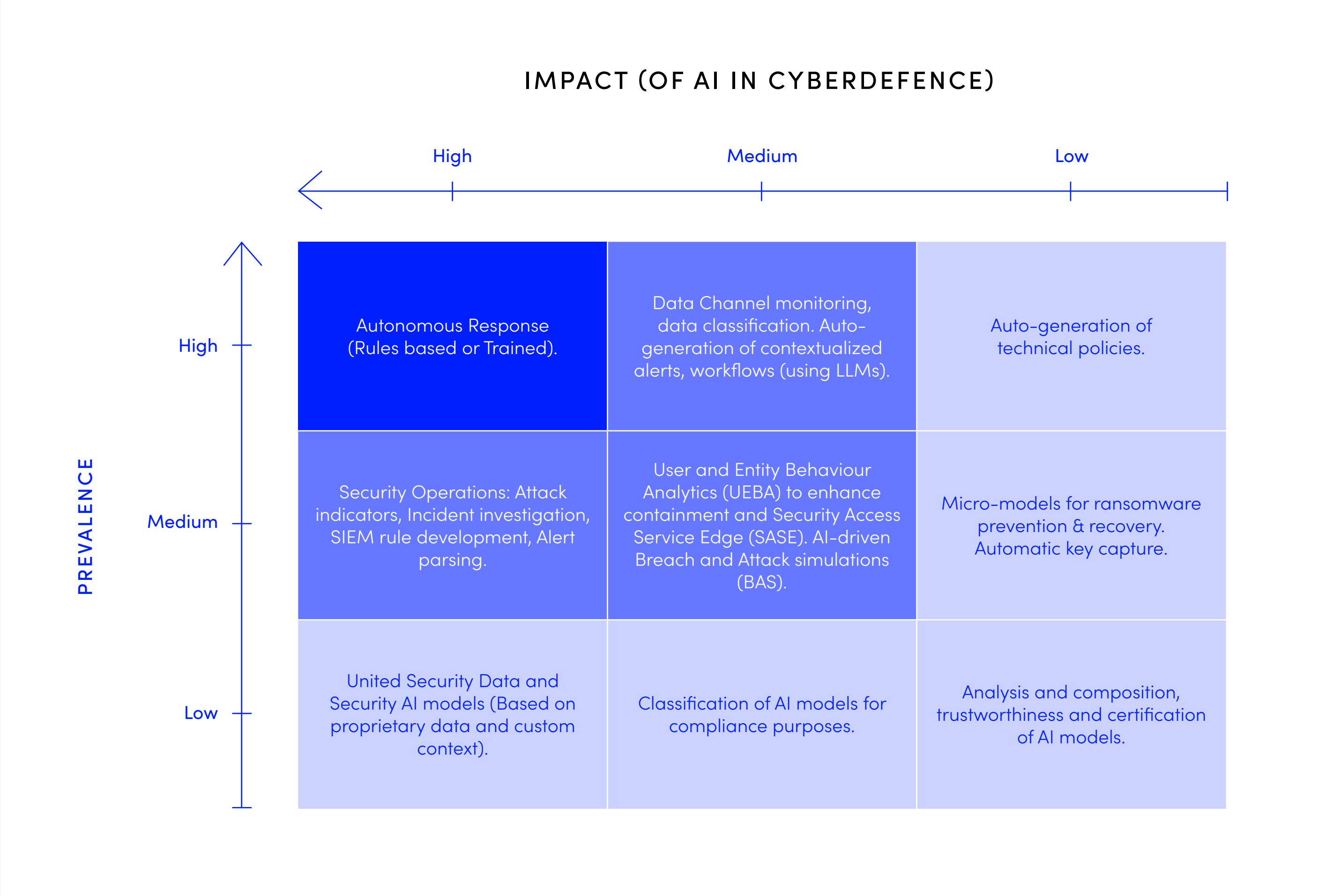
## *AI-powered cybersecurity: Strengthening defence mechanisms*

Cybersecurity is often framed as an arms race between attackers and defenders. At ISTARI, we believe defenders can reverse the dilemma and establish an advantage by strategically adopting advanced AI capabilities to protect their assets and information, and by leveraging their partner ecosystem to access the power of the collective.

A responsible CISO will develop a programmatic strategy which should include a roadmap to adopt AI capabilities to gain and sustain this advantage. However, enterprises face resource constraints, and CISOs must constantly balance competing demands to meet ambitious security objectives. From our research, it's clear that AI adoption in cyber defence must focus on outcomes that deliver resilience. The priorities will naturally differ by industry – what a manufacturing firm emphasises will not be the same as a financial institution.

## *Defending and sustaining advantage: Integrating key recommendations*

AI adds a crucial layer to a multi-layered defence strategy, particularly in high-risk scenarios such as ransomware, phishing, and technical vulnerabilities. This can be achieved by deploying AI across each layer, enhancing filtering capabilities, revising per-layer security controls, and implementing



| | High | Medium | Low |
|---|---|---|---|
| High | Autonomous Response (Rules based or Trained). | Data Channel monitoring, data classification. Auto-generation of contextualized alerts, workflows (using LLMs). | Auto-generation of technical policies. |
| Medium | Security Operations: Attack indicators, Incident investigation, SIEM rule development, Alert parsing. | User and Entity Behaviour Analytics (UEBA) to enhance containment and Security Access Service Edge (SASE). AI-driven Breach and Attack simulations (BAS). | Micro-models for ransomware prevention & recovery. Automatic key capture. |
| Low | United Security Data and Security AI models (Based on proprietary data and custom context). | Classification of AI models for compliance purposes. | Analysis and composition, trustworthiness and certification of AI models. |

threat-specific solutions. For example, using AI-enabled anti-ransomware tools alongside endpoint detection and response (EDR), complemented by AI-powered third-party monitoring and remediation.

The effectiveness of AI capabilities largely depends on the quality of the data it is trained on. CISOs must establish robust data management mechanisms and governance for their cybersecurity data. We strongly recommend centralising and aggregating security data from diverse channels in a way that reflects their priority and context, and to create a comprehensive repository to serve enterprise needs in the foreseeable future. Practical examples include expansive log feeds into Security Operations, and building dynamic playbooks supported by AI —adhering to enterprise rules rather than vendor-specific guidelines.

The third pillar of AI-powered defence is building strong cybersecurity foundations, such as User and entity behaviour analytics (UEBA), breach and attack simulations, and threat detection/hunting that adapt to AI-driven threat vectors, such as deepfakes—an emerging and highly effective threat. This requires deploying AI defences that evolve in real-time, learning from new threats to predict, prevent, and respond faster than AI-enabled adversaries. CISOs must also address the widening gap between risk controls and the frequency of oversight and reviews. AI-powered solutions like continuous automated red teaming, breach and attack simulations, and AI-informed continuous threat exposure management can help mitigate these risks.

## *The path forward*

In brief, enabling AI requires strong cybersecurity foundations from data management to the continuous evolution of multi-layered defences, to stay ahead of adversaries. As Sun Tzu articulated in the principles of warfare, adaptability and proactivity are key to gaining an advantage—and reversing the defender's dilemma in favour of your enterprise. Now is the time to strengthen your capabilities and not yield ground to attackers. Continuous learning and collaboration will be more essential than ever in building strategic resilience and staying ahead of emerging threats.

> “The art of war teaches us to rely not on the likelihood of the enemy not coming, but on our readiness to receive him.”
>
> SUN-TZU, THE ART OF WAR

ABOUT THE AUTHORS:

### Abel Archundia

Abel Archundia is the CTO of ISTARI. Previously he was Global CIO at Bayer Pharma, and at Novartis' Sandoz Division. He started his career as a consultant at BCG and was later responsible for Dell's business unit in Mexico.

### Alexander Pabst

Alex is the global Deputy CISO of Allianz SE, the world's biggest insurance company and the largest financial services company in Europe.

He brings both the top-management view, from previously working at McKinsey, as well as a technical background from Cisco and Intel. Alex holds an MBA from INSEAD.