XBOW

Insights

The Next Six Months of Offensive Security

What CISOs Need to Change Now to Prepare for the Post-Mythos Era

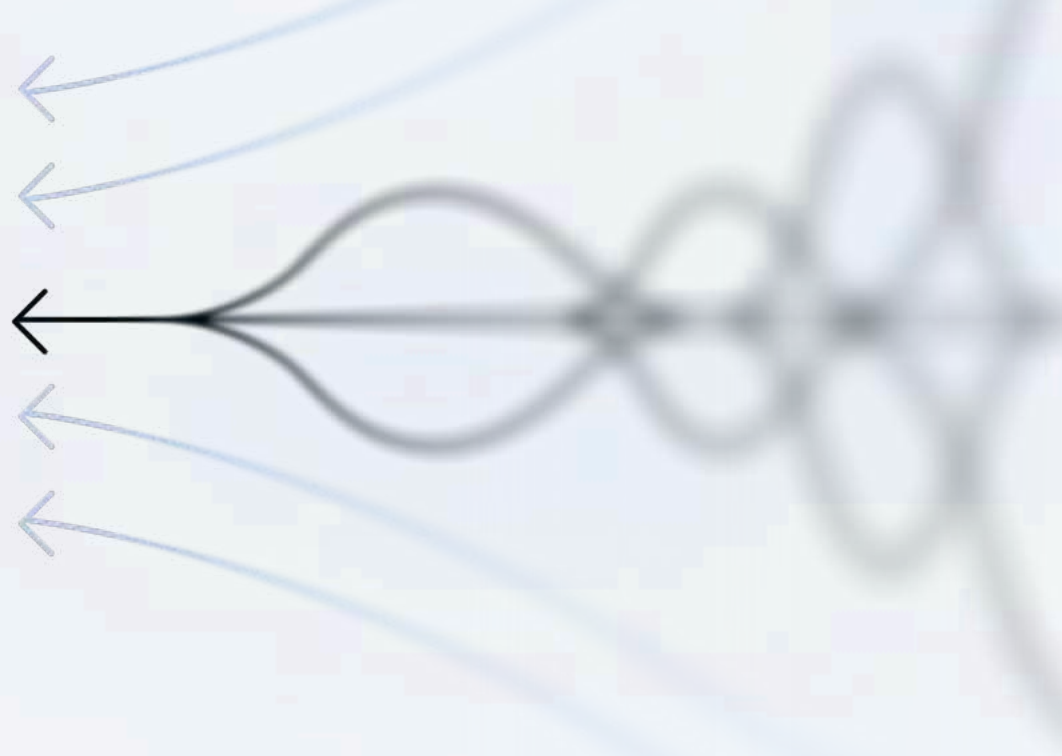
About the Panel:

This paper is based on a panel discussion at RSA Conference 2026 featuring XBOW CISO Nico Waisman; Hacker, CEO, and CISO Jason Haddix; and OpenAI Technical Staff Member Dave Aitel, focused on how AI is reshaping offensive security and what security leaders should do next.

AI does not need to invent a new class of attacks to change the threat landscape. It only needs to make existing offensive workflows faster, cheaper, and easier to scale.

AI is changing cybersecurity, but not in the way many headlines suggest.

Attackers will not suddenly have entirely new goals or novel classes of tradecraft. But the economics of offense will be dramatically altered. More code is being written, faster, by more people with less security expertise. At the same time, attackers can use AI to accelerate research, tooling, and exploit development. Mythos will intensify this activity. The result is a larger attack surface and a faster, more scalable offensive model.



That creates a temporary but meaningful imbalance.

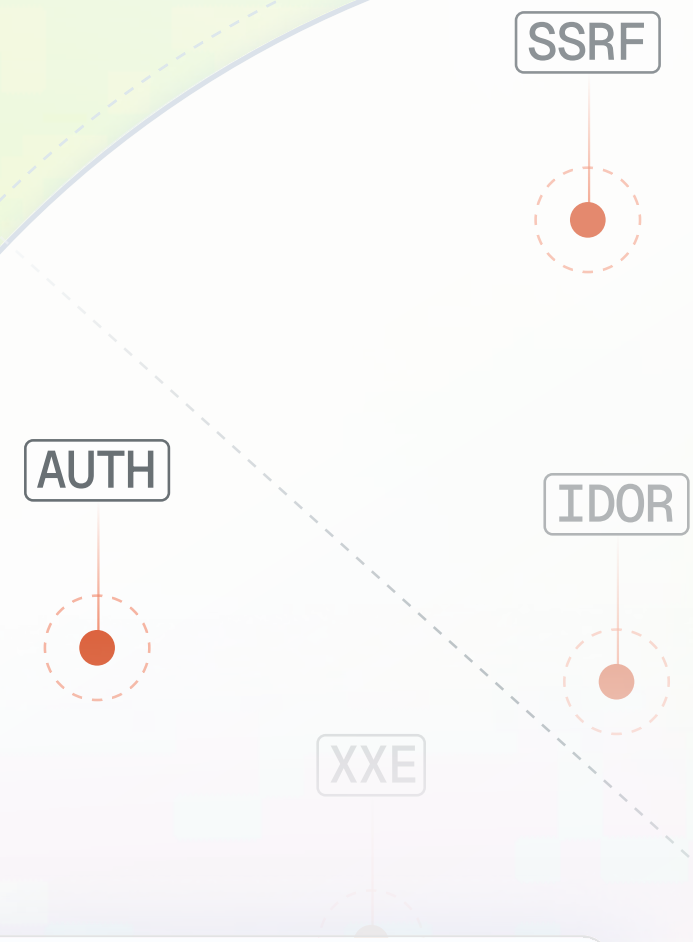
We recently came together to examine this shift at a panel discussion for RSAC 2026. This paper pulls from that discussion to highlight our thoughts about, questions on, and recommendations for this next phase of cybersecurity. We hope it offers a realistic (non-FUD) view of the current and near-future state, and some practical guidance for security teams.

The core message is straightforward: offense is likely to benefit first, but not forever. The organizations that respond best will be the ones that strengthen fundamentals, increase remediation throughput, and adopt governed AI in their security operations before the gap widens further.

– **Nico Waisman**, XBOW CISO, **Jason Haddix, Arcanum** CEO, **Dave Aitel**, Technical Staff, OpenAI

The threat landscape is changing faster than most security programs

Security leaders already know the attack surface is expanding. What AI changes is the speed at which attackers can operate against that surface.



AI-powered attacks will not look unfamiliar. There will simply be more of them because they are faster, cheaper, and easier to scale.

In the near term, most AI-enabled attacks will not look radically different from the attacks defenders already know. What changes is the operational tempo behind them. Threat actors can move faster from idea to proof of concept, from proof of concept to weaponization, and from one successful technique to many variants.

In many cases, defenders will not be able to tell whether a human, an agent, or a human using an agent is behind an intrusion attempt. But that distinction matters less than the measurable increase in speed, volume, and persistence.

Will the AI models know their platforms are being used for attacks and stop them?

Also unlikely. Although there have been reports released from the frontier model companies detailing how attackers have used their models to carry out attacks, and how they have stopped potential attackers, the bottom line is that the majority of attackers (good ones anyway) will use private models.

For defenders, that means the more relevant question is not whether an attack was “AI-generated,” but whether their controls can withstand adversaries whose research and execution loops have become materially faster.

The biggest near-term change is operational leverage for attackers

The most immediate impact of AI is **improved attacker operations.** Their teams are typically smaller, less encumbered, and more experimental than enterprise security organizations. That makes them well positioned to adopt AI quickly.

Simply put, **AI lowers the cost of capability.** Sophisticated techniques that once required deep expertise, time, and custom tooling can now be adapted, packaged, and reused more quickly by less sophisticated groups. That does not erase the gap between elite and lower-tier adversaries, but it narrows it enough to matter.

For example, a recent iOS zero-day for jailbreaking phones emerged from a nation-state group. It was immediately cloned by two smaller, less sophisticated groups. After vibecoding the framework to weaponize the exploit and deliver it to phones, they sold it for a fraction of the price.

The practical consequence for CISOs is that more attackers will be able to operate with higher-quality tooling and shorter iteration cycles. Defenders should expect more rapid exploit adaptation, more efficient reconnaissance, and faster follow-on activity after an initial breakthrough.

Offense has the advantage, for now

In the short term, offense has the AI advantage.

That is not because defenders lack talent. It is because enterprise security programs operate inside constraints that attackers do not: budget cycles, fragmented tooling, compliance obligations, change-management processes, and organizational politics. All of those are real, and all of them slow adoption.

XBOW CEO Oege de Moor has described **this period as a "chaos phase"**: a temporary window in which offensive capabilities accelerate faster than most defensive programs can absorb. That framing is useful, as long as it is understood correctly. This is not a permanent new normal.

It is a transition period in which the gap between offensive and defensive adoption creates higher risk.

There is precedent for this kind of cycle. Fuzzing followed a similar pattern. Early on, attackers gained disproportionate value from it. Over time, defenders integrated fuzzing into development and testing workflows, and entire categories of bugs became harder to find. AI is likely to follow a similar path, but on a much shorter timeline.

That matters because it points to a realistic conclusion: **the current imbalance is serious,** but temporary. Defenders will catch up. The timeline will vary by sector, budget, and technical maturity, but the broader pattern is clear.

What AI defense actually requires

The answer is not simply to "use AI on defense." The more useful question is where AI works today, where it still struggles, and what kind of human oversight remains necessary.

AI is strong at pattern recognition. It is weaker at governed, end-to-end security judgment.

Today's models can already be effective at identifying likely application weaknesses, especially when they have access to source code. In environments that have not received sustained security attention, that kind of pattern matching can produce meaningful findings quickly.

Where things get harder is in dynamic analysis, exploit validation, and contextual judgment. It is one thing to identify a buffer overflow or authentication flaw. It is another to determine exploitability, business impact, and the best path to remediation in a live environment with incomplete information. However, this challenge will likely be short-lived; the models are getting better (and quickly) at understanding business logic and applying that knowledge to findings.

Beyond the business-logic context issue, the bigger concern, and reason humans will play a role for the foreseeable future, is that the LLMs need a fair amount of structure and scaffolding. Effective AI security systems need orchestration, validation layers, and clear testing boundaries. Without that scaffolding, models can confidently pursue the wrong path, compound small misunderstandings, or operate outside intended constraints.

That is why the future is unlikely to be "AI replaces security professionals." A more accurate view is that strong security outcomes will come from a combination of:

- **Model capability**
- **Expert methodology**
- **Governed orchestration**
- **Human oversight where judgment matters most**

Albert Ziegler, XBOW head of AI, sums it up perfectly in a recent blog post when he says,

> “In pentesting, it’s easy to make plausible mistakes based on small misinterpretations, both for humans and AI. But while a human will move on from the mistake, an AI agent will build upon them. This is why you can’t just wrap an LLM and call it a pentest - effective AI pentesting needs a structured, governed system.”

The implication for teams: security talent must become more AI-native

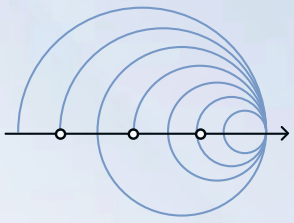
Over the next six months, one of the most valuable combinations in security will be domain expertise plus AI engineering literacy. Teams will need people who understand offensive methodology, but also how to encode workflows into tools, agents, prompts, evaluation loops, and guardrails.

That shift has hiring implications. It also has buying implications. CISOs should look closely at whether a vendor’s AI approach is grounded in real security methodology and strong governance, or whether it is simply wrapping a general-purpose model with thin automation.

What CISOs should do now

The right response to AI-enabled attacks is not panic. It is disciplined modernization.

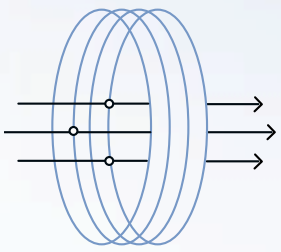
- **Recommit to fundamentals:** Asset inventory, identity hygiene, incident response, disaster recovery, credential rotation, secure-by-default
- **Increase remediation throughput:** Use AI to reduce the operational friction around remediation
- **Adopt governed AI:** Choose the right model for the right task, and add orchestration and validation



Recommit to the fundamentals

When attacker speed increases, weak fundamentals become more expensive. Asset inventory, identity hygiene, incident response readiness, disaster recovery, credential rotation, secure-by-default configuration, and basic control discipline all matter more in a faster threat environment, not less.

As offensive activity accelerates, organizations will need stronger operating discipline around response, recovery, and accountability.

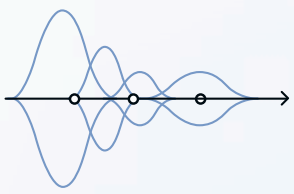


Increase remediation throughput, not just detection coverage

Many security programs are already good at finding issues. Far fewer are good at moving them efficiently through validation, prioritization, ownership, and remediation.

That has to change. Vulnerability management is full of repetitive workflow: triage, ticketing, routing, follow-up, status tracking, and retesting. Those processes are highly amenable to automation. AI can help security teams increase throughput here far sooner than it can fully replace expert judgment in complex testing.

For CISOs, this is one of the clearest near-term opportunities: use AI to reduce the operational friction around remediation, so the organization can keep pace with a faster offensive cycle.



Adopt governed AI, not uncontrolled experimentation

Security teams should absolutely use AI as a force multiplier. But the goal is not to spray tokens at every problem. The goal is to build repeatable, governed systems.

In practice, that means choosing the right model for the right task, adding orchestration and validation, and codifying successful workflows into tools rather than repeating expensive ad hoc prompting. Over time, the winning approach will be systematic: models where they are useful, tooling where tasks can be codified, and humans where judgment and accountability remain critical.

How to keep up without infinite token spend

Most security teams do not have unlimited budget for inference. That makes operating discipline important.

Two practical principles matter:

Diversify model usage. You'll need scaffolding, built or bought, that can determine which models are good for what tasks, and switch between them as needed.

Codify repeatable work. When a model successfully solves a recurring problem, turn that workflow into a tool or structured process. That reduces repeated inference costs and creates more consistent outcomes over time.

The broader point is that token efficiency is not just a cost issue. It is an operating model issue. Teams that codify, route, and govern AI usage will outperform teams that rely on one-off prompting.

AI is compressing the time, cost, and expertise required for offensive work faster than most enterprises are currently increasing defensive capacity.

Over the next six months, CISOs should focus on strengthening fundamentals, increasing remediation throughput, and adopting AI in ways that are structured, governed, and operationally useful. The organizations that do this well will be in a much better position not only to absorb the current disruption, but to emerge from it with a stronger and more scalable security program.