XBOW

SECURITY BRIEF

The Chaos Phase: Why AI Has Broken the Old Security Model

How CISOs Regain Confidence When AI Breaks Security Models

Executive Summary

Cybersecurity has entered a period of instability that many CISOs and security leaders recognize, even if they have not fully articulated why.

AI-driven attackers now operate at machine speed. They automate reconnaissance, exploitation, and lateral movement continuously and at scale. Security programs built around periodic testing, reactive detection, and manual validation were not designed for this reality.

The good news is that this shift makes clear what an effective response now requires: the problem is solvable, but in a fundamentally new way. Human-level reasoning, when paired with autonomous, machine-speed execution, is necessary to outpace security programs built around human-paced cycles.

This brief introduces the **Chaos Phase:** a transitional period where offense is evolving faster than defense. It explains what has structurally changed, why familiar security practices are breaking down, and why the next 12 months represent a critical window for security leadership to adapt.

What you should take away

- AI removes human pacing from attack execution, exposing security models built around manual validation and delayed response.
- The first thing to fail is confidence. Security leaders can no longer prove controls work at the speed attacks unfold.
- Periodic testing forces leaders to explain past posture while real risk evolves continuously.
- As offense accelerates, security programs must shift from snapshot-based assurance to models that can continuously validate exposure.

1) What is the Chaos Phase

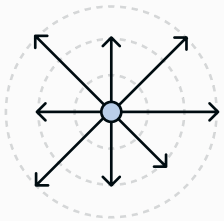
The Chaos Phase is the time between two realities:

1. **Offense has already become more autonomous, continuous, and scalable with AI.** Most defense programs are still structured around human-paced cycles and human-in-the-loop validation.

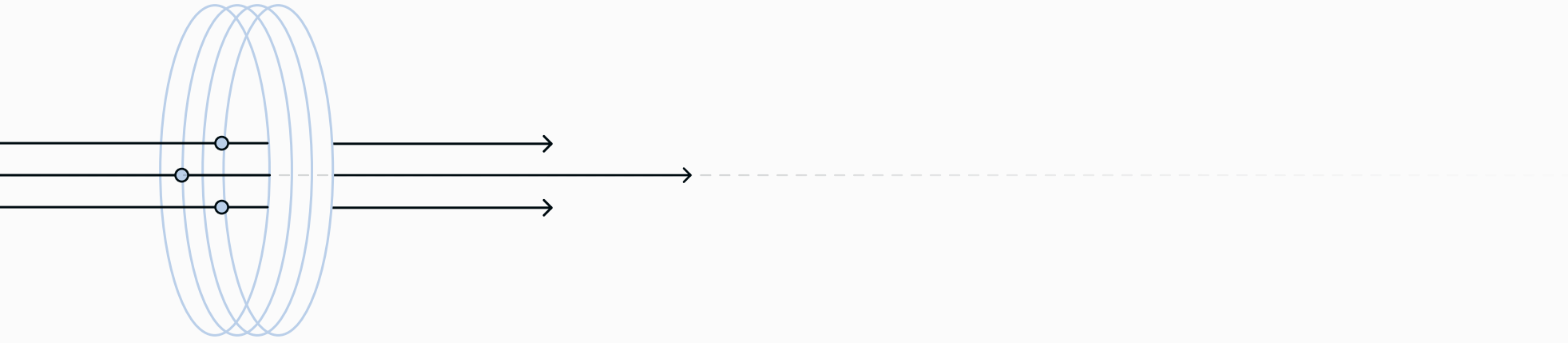
2. **This creates volatility.** Teams that have invested heavily in tools and telemetry increasingly feel less certain about their actual exposure.

If you are hearing variations of these statements inside your organization, you are already in it:

- “We have the data, but we cannot tell what is real fast enough.”
- “We test, but not at the pace we ship.”
- “We detect, but by the time we respond, the situation has changed.”



The Chaos Phase is not about a single technique or a single actor.
It is about a structural mismatch. That mismatch is created when offense removes human gating from execution, while defense still depends on humans to initiate, validate, and respond.



2) Offense has moved beyond human speed

The most important change in the threat landscape is not sophistication. **It is tempo.**

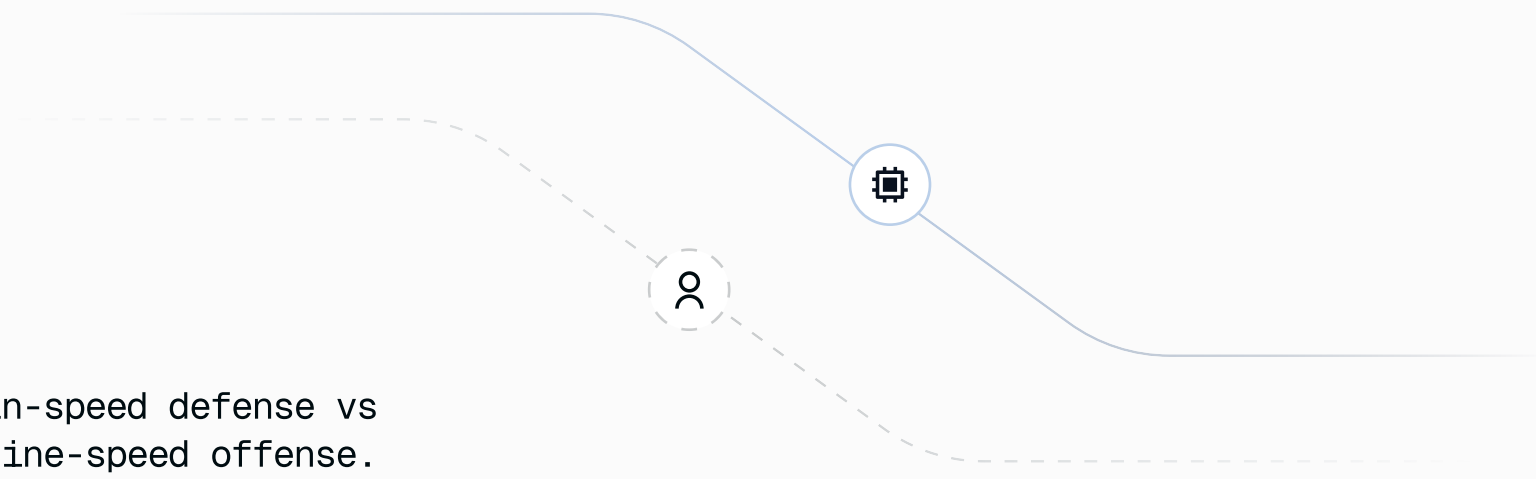
AI removes the pauses between discovery and exploitation, signal and response, and change and validation that defenders depended on.

Reconnaissance, exploitation, and lateral movement no longer occur in clean, sequential stages. These activities now run continuously, in parallel, and without rest. Tasks that once required weeks of expert human effort can be executed autonomously across many targets at once

What changes for defenders

- The window between discovery and exploitation shrinks.
- The number of parallel attempts rises.
- The cost of running sophisticated operations falls.

The Chaos Phase is not about a single technique or a single actor. It is about a structural mismatch. That mismatch is created when offense removes human gating from execution, while defense still depends on humans to initiate, validate, and respond.



Human-speed defense vs machine-speed offense.

3) This shift is not hypothetical

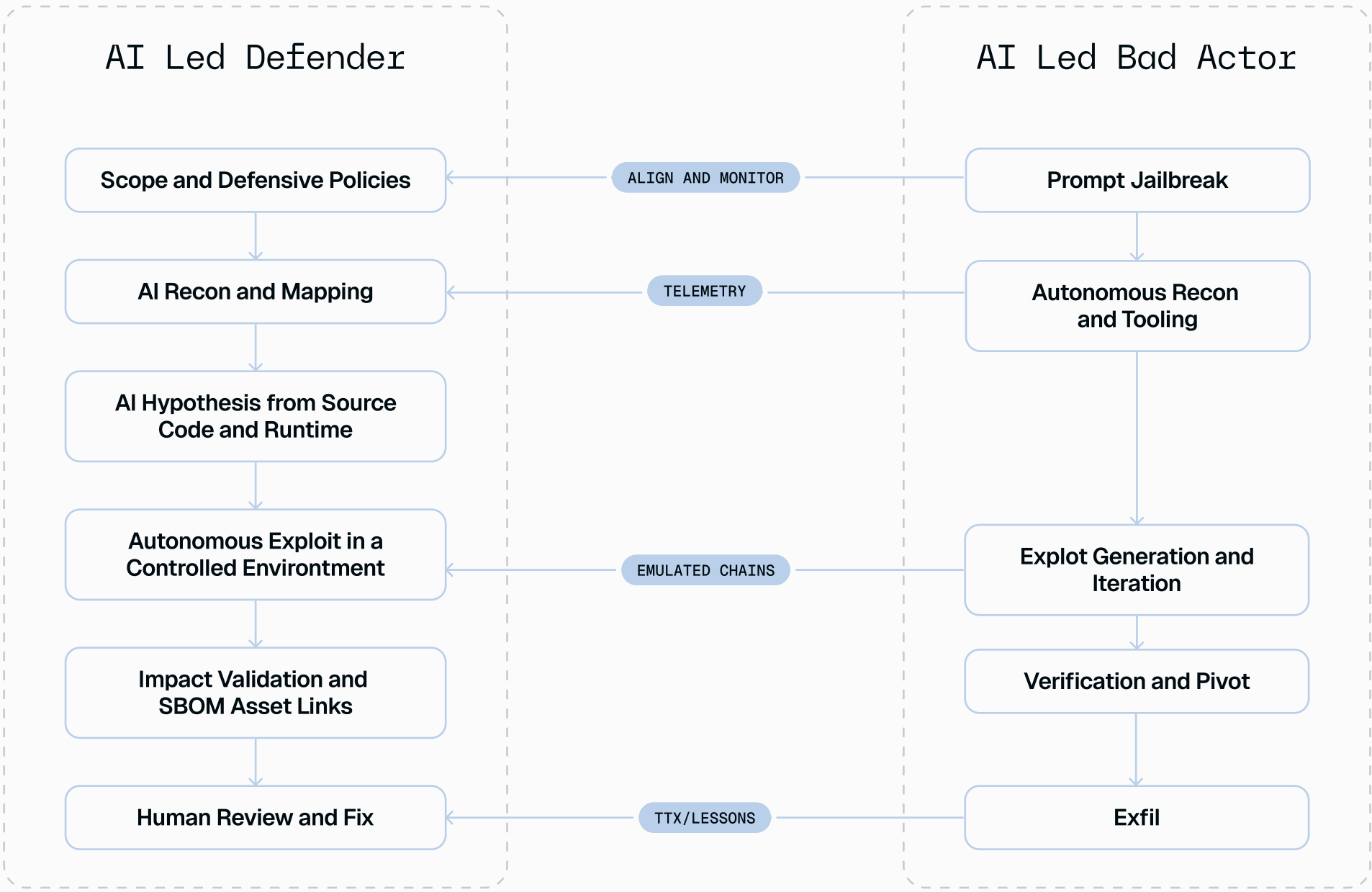
Major platform and intelligence teams are now publishing concrete observations of adversaries using generative AI to move faster and at higher volume.

Microsoft has documented how threat actors use LLMs as productivity tools across offensive workflows, accelerating reconnaissance, targeting, and exploitation. **Google Threat Intelligence** describes threat actors misusing AI tools to support stages across the attack lifecycle, including reconnaissance, phishing lure creation, lateral movement assistance, and data exfiltration support.

These are broad signals. They point to a consistent direction. AI is being used to reduce time, increase parallelism, and lower the cost of running sophisticated operations.

In late 2025, **Anthropic** disclosed GTG-1002, a cyber espionage campaign it described as the first reported AI-orchestrated campaign of its kind. Anthropic reports that an AI system executed the majority of tactical activity, with humans stepping in only at key decision points.

Autonomy here does not mean the absence of human strategy. It means humans no longer sit in the execution path, allowing operations to run continuously without bottlenecks.
GTG-1002 provided clear evidence of a shift already well underway.



4) What breaks first under AI-driven attacks

When attackers don't wait, security practices that assume time exists between discovery and impact fail by default.

Reactive detection struggles next

Detection systems rely on enough time existing between intrusion and outcome for alerts, investigation, and response. When attacks adapt and complete rapidly, detection becomes forensic rather than preventative. By the time an alert fires, it’s already too late.

Manual validation becomes the bottleneck

As AI accelerates discovery and attack iteration, human teams cannot triage and confirm findings fast enough to matter. The result is a growing confidence gap.

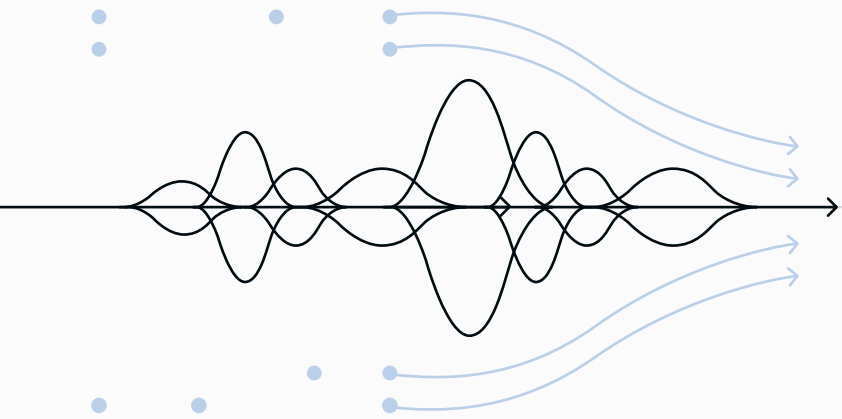
- SANS survey findings show false positives are a major issue for threat detection and can overwhelm teams, leading to alert fatigue and missed threats.
- Forrester has been blunt about the operational impact. In its guidance for network analysis and visibility, Forrester warns that excessive alerts cripple security operations.

These failures are not the result of poor execution. They reflect a structural mismatch between machine-speed offense and human-speed defense.

Periodic assurance breaks first

Annual penetration tests, quarterly risk reviews, and compliance-driven snapshots were designed for an era when attackers were constrained by human labor. That assumption no longer holds.

When attacks unfold in hours or days, a security assessment that happens once a quarter becomes obsolete almost immediately. Risk does not accumulate quietly between tests. It compounds continuously.



7) What this means for CISOs and Security leaders today

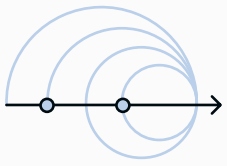
The Chaos Phase is a transitional moment. Offense currently holds a temporary advantage by removing human constraints. That advantage will not last forever. Defense will catch up by adopting similar principles of automation, scale, and speed.

What cannot continue is reliance on security models that assume time, pause, and predictability.

The next 12 months will determine which organizations adapt and which remain exposed. Leaders who recognize this shift early can modernize deliberately rather than reactively.

Questions to take into your next staff meeting

- Where do we still rely on periodic assurance to represent real risk?
- Where does manual validation slow our ability to act?
- What would it take to prove, not assume, that our controls work?



The Chaos Phase is not an endpoint. It is a forcing function.
The programs that survive it will be designed to operate continuously, prove effectiveness quickly, and scale without humans becoming the bottleneck.

What to do next

The Chaos Phase is already reshaping security programs. Learn how autonomous attacks operate today, and how leading teams are adapting validation models.

Watch a breakdown of AUTONOMOUS INTRUSIONS IN ACTION

Have a focused conversation about what this shift means for your security program TALK TO AN EXPERT

Sources

- Microsoft Security Blog: Staying ahead of threat actors in the age of AI.
- Google Cloud Blog: Advances in threat actor usage of AI tools.
- Anthropic: Disrupting the first reported AI-orchestrated cyber espionage campaign (GTG-1002) and accompanying report PDF.
- Gartner Newsroom: Gartner identifies the top cybersecurity trends for 2024 (CTEM).
- Gartner: How to manage cybersecurity threats, not episodes (CTEM steps and validation).
- Forrester: Predictions 2025, cybersecurity, risk, and privacy.
- Forrester: The Forrester Wave, Network Analysis and Visibility Solutions, Q4 2025 (alert fatigue callout).
- SANS Institute: 2024 Detection and Response Survey announcement (false positives and alert fatigue).