MirrorMire
PROACTIVE CYBER RESILIENCE

Becoming Mythos-Ready through Proactive Cyber Resilience

Table of Contents

Executive Summary

Artificial intelligence is becoming a defining force in offensive cybersecurity. Frontier models can now read, reason about, and rewrite code at a level that meaningfully compresses the time required to find and weaponize software flaws. Purpose-built research systems are pushing this further: Anthropic's Project Glasswing reported that Claude Mythos Preview identified thousands of estimated high and critical severity vulnerabilities across more than 1,000 open-source projects.

This creates a widening speed gap for defenders. CrowdStrike reported that average eCrime breakout time fell to 29 minutes in 2025, while IBM reported that organizations still took a mean of 241 days to identify and contain a breach. In practical terms, attackers are moving from initial access to lateral movement far faster than most organizations can detect, investigate, and respond.

This paper builds on the emerging "Mythos-ready" security framing and argues one central point: as AI accelerates vulnerability discovery and zero-day exposure, controls that depend on knowing the flaws in advance lose ground. Deception holds its value because it detects intruders by behavior, not by prior knowledge of the exploit. The closing sections show how MirrorMire AMaze applies that principle through Synthetic Cognitive Agents (Advanced Decoys), Neural Echoes (Advanced Lures), attacker engagement, and high confidence detection.

Attack Capabilities Are Evolving Faster Than Defenses

For most of the last decade, finding a serious vulnerability took skill, time, and money. AI has removed all three constraints at once. The change is visible across three independent strands of evidence: how many flaws AI now finds, how broadly the capability has spread, and how quickly disclosed flaws are exploited.

23,019
Issues flagged by one AI model across 1,000+ open-source projects

6,202
rated high or critical severity

>90%
of independently reviewed findings confirmed as real bugs

Project Glasswing, the output of a single program (Anthropic, 2026).

The Barrier to Offensive Capability Is Shrinking

The clearest example is Anthropic's Project Glasswing. In its initial update, Anthropic reported that Claude Mythos Preview scanned more than 1,000 open-source projects and surfaced an estimated 23,019 issues, including 6,202 high or critical severity findings. Six independent security firms reviewed a sample of 1,752 findings and confirmed more than 90% as valid.

The results show that AI-assisted vulnerability discovery is moving from theory to operational reality. Cloudflare reported that Mythos found roughly 2,000 bugs in its critical path systems with a false-positive rate better than human testers. In another case, Mythos produced a working exploit for a wolfSSL vulnerability, demonstrating that these systems can move beyond discovery into exploit construction when used in controlled research settings.

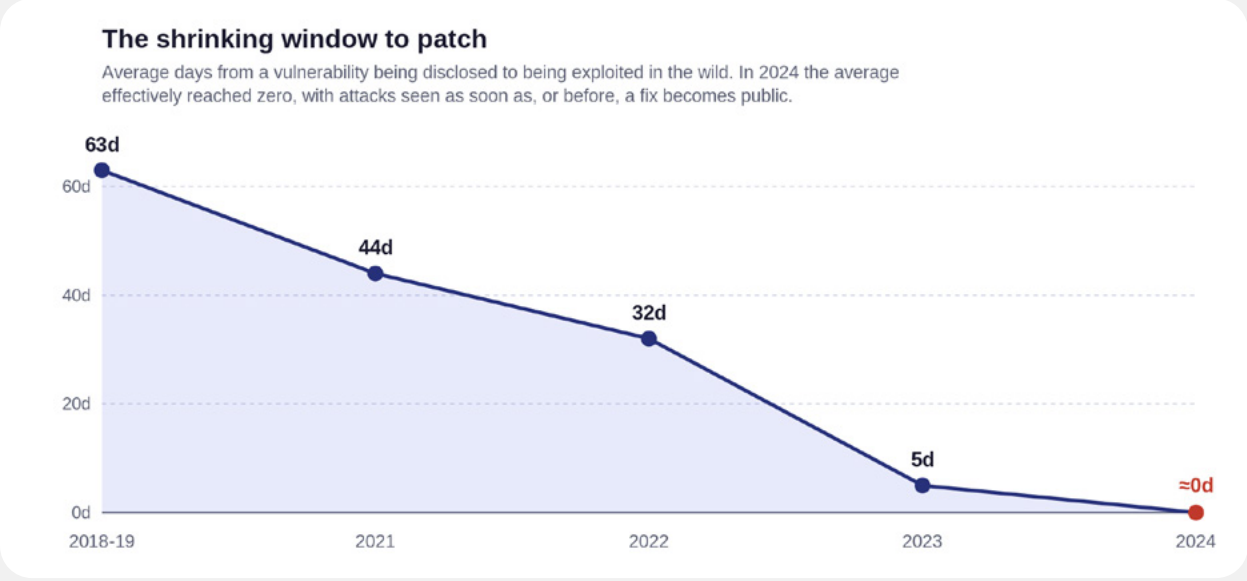
Anthropic is not the only example. Google's Big Sleep agent discovered CVE-2025-6965, a SQLite vulnerability that Google said was known only to threat actors and at risk of exploitation. At the DARPA AI Cyber Challenge final, autonomous systems analyzed 54 million lines of code and uncovered 18 previously unknown real-world flaws.

The direction is clear: AI-assisted vulnerability discovery is becoming faster, broader, and more repeatable.

For defenders, this changes the economics of detection. As vulnerability discovery accelerates, controls that depend on knowing the flaw in advance face a widening gap. Deception remains valuable because it detects behavior after adversary access begins: scanning, probing, credential misuse, lateral movement, and interaction with assets legitimate users have no reason to touch.

Patches now arrive too late

As discovery scales up, the time defenders have to react has collapsed. Mandiant's analysis of time to exploit, the gap between a vulnerability becoming known and being used in an attack, shows a steady fall to near zero, and in 2024 the average went negative because attacks were observed before fixes were public.



Real-world breach data reflects the shift. In Verizon's 2025 Data Breach Investigations Report, vulnerability exploitation rose 34% year over year as an initial-access vector and now accounts for about one in five breaches, while only 54% of known vulnerabilities were fully remediated, at a median of 32 days (Verizon, 2025). The math no longer works: exploitation is measured in days or hours, and remediation is measured in weeks.

The Deception Gap

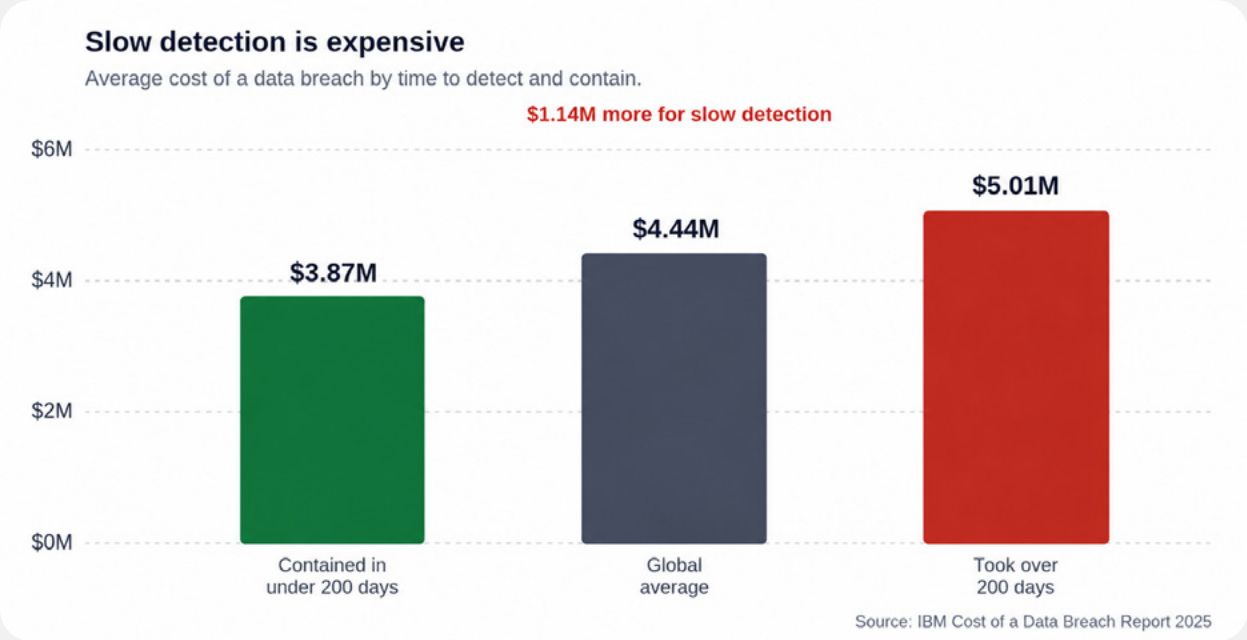
If patching can no longer keep attackers out, the question shifts from prevention to detection: once an intruder is inside, how quickly are they found and stopped? This is a different clock from the one in the previous section, and it is where most programs are weakest.

Detection, not prevention, decides the outcome

Once an attacker has a foothold, detection speed is what decides the outcome, and it is badly outmatched. CrowdStrike puts average breakout time, the interval from initial access to lateral movement, at 29 minutes, with the fastest observed case at 27 seconds, and reports that 82% of detections are malware-free, meaning intruders increasingly use legitimate tools and stolen credentials rather than files that an antivirus engine would catch (CrowdStrike, 2026). Set against that, the average breach still takes 241 days to identify and contain, split between 181 days to identify the breach and 60 days to contain it (IBM, 2025).

Slow detection has a direct, measurable cost

The delay is not only an exposure problem; it is a financial one. IBM puts the global average cost of a breach at $4.44M. Breaches identified and contained in under 200 days averaged $3.87M, while those that took longer than 200 days averaged $5.01M, a difference of $1.14M tied directly to detection speed (IBM, 2025).

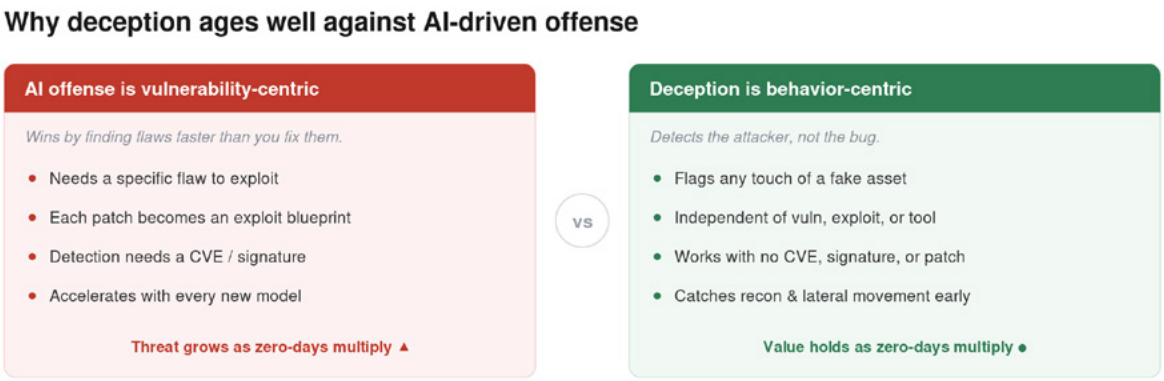


The same report found that organizations making extensive use of AI and security automation shortened breach identification and containment timelines by an average of 80 days compared with organizations that did not use these capabilities (IBM, 2025).

Why Deception Changes the Math

Most detection methods ask a version of the same question: does this activity match something we already know is bad? Signatures, CVE feeds, and threat intelligence all depend on prior knowledge of the specific flaw or tool. That model is exactly what AI-driven offense undermines, because it produces a constant supply of flaws that have no signature, no CVE, and no patch yet.

Deception asks a different question entirely. It places assets that no legitimate user or process has any reason to touch, so any interaction is suspicious by definition. It does not need to recognize the exploit; it only needs to notice that something touched a resource that should never be touched. The Cloud Security Alliance (CSA), SANS, and OWASP briefing makes the same point directly, describing deception as attack-tool and vulnerability independent, identifying attackers by their techniques rather than by the specific bug they use (CSA et al., 2026).

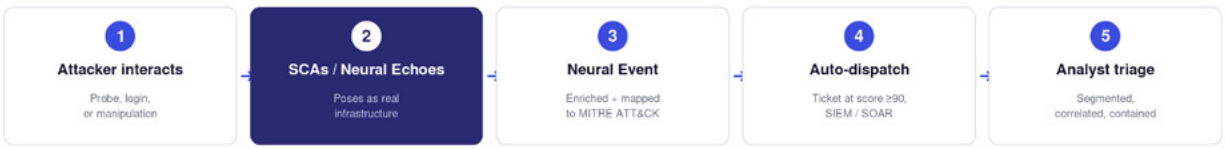


The evidence for deception is strong

Because every interaction with a decoy is unauthorized by definition, deception produces high-confidence alerts with very few false positives, and it can cut attacker dwell time from months to hours by surfacing intrusions while they are still in progress. In a world where 82% of attacks are malware-free and breakout takes 29 minutes, a control that flags an intruder the moment they probe a fake asset is aimed precisely at the gap the data exposes.

MirrorMire AMaze in Practice

AMaze deploys Synthetic Cognitive Agents, or SCAs, and Neural Echoes across IT and OT environments to create a believable deception layer inside the network. SCAs appear as realistic enterprise and industrial assets, communicating through real protocols and responding to attacker activity like legitimate systems. Neural Echoes extend that deception surface with lightweight traces, lures, and signals that guide unauthorized users toward controlled SCA interactions. Together, they give MirrorMire AI's Proactive Cyber Resilience platform a high-confidence way to expose reconnaissance, credential misuse, lateral movement, and exploitation attempts before attackers reach critical systems or data.



High-severity events auto-create tickets at a threat score of 90 or above (configurable) and can dispatch to SIEM or SOAR. Each event carries its mapped MITRE ATT&CK technique, so analysts get context rather than a raw alert, and the very low false-positive rate of SCA interactions keeps that signal clean.

How it closes the gap

- **Flaw-independent by design.** Detection is behavioral, so AMaze needs no CVE, KEV (Known Exploited Vulnerabilities) entry, or signature for the zero-day in play. It catches the action, not the bug.
- **Detection inside the breakout window.** SCA touch fires instantly, closing the 29-minute breakout window instead of the 241-day average. (CrowdStrike, 2026; IBM, 2025)
- **Built for malware-free attacks.** With 82% of intrusions using legitimate tools and stolen credentials, antivirus signals can disappear, but unauthorized interaction with a SCA still creates a high-confidence alert. (CrowdStrike, 2026)
- **A tripwire for lateral movement.** Neural Echoes and SCAs seeded across segments turn a flat or weakly segmented network into a detection surface, alerting the moment an attacker pivots toward real systems.
- **Wired into response.** Events flow through webhook, Slack, or Teams into Splunk, Wazuh, Elastic, XSOAR, Jira, and more, with any outbound action gated by an analyst.

IT and OT Coverage

AMaze deploys two families of decoys. IT SCAs emulate the office and data-center world, including SSH, FTP, RDP, HTTP, MySQL, Redis and Windows domain controllers and workstations. OT SCAs mimic real industrial devices using native OT, MQTT and IoT protocols.

Protocol	Port	Emulates	Industry
Modbus TCP	502	Process gas analyzer (H2S/CH4), power monitor	All OT (universal)
EtherNet/IP (CIP)	44818	Allen-Bradley CompactLogix / ControlLogix PLCs	Manufacturing (N. America)
S7comm	102	Siemens S7-1500 PLC	Manufacturing (global)
DNP3	20000	RTU / SCADA outstation	Electric, water
BACnet	47808	Building automation controller Facilities	Facilities
IEC-104	N/A	Grid telecontrol (via VAPT probe)	Electric / substations
OPC-UA	N/A	Modern OT data gateway (via VAPT probe)	IT/OT bridge

Decoys on both the IT and OT sides give AMaze a view of the crossover an attacker would use to pivot from a compromised workstation toward control systems, the boundary where pure IT or pure OT tools tend to have blind spots. When an attacker reads decoy Modbus registers, AMaze labels it reconnaissance (T1046). Writing surge values maps to parameter falsification (T0827), and suppressing an active alarm maps to T0838, each in MITRE ATT&CK for ICS terms an analyst can act on.

Recommendations

No single control answers AI-driven offense. The evidence points to a layered response that keeps the fundamentals while adding speed where prevention falls short.

- Turn AI on your own code first. Run LLM-based review across your code and dependencies before attackers do. The same capability behind Glasswing is available to defenders.
- Keep hardening the basics. Segmentation, egress filtering, phishing-resistant MFA, and least privilege raise attacker cost regardless of which flaw is used, and they limit blast radius when one lands.
- Measure resilience, not just prevention. Track early detection, containment time, and recovery time. IBM's data shows that breaches identified and contained in under 200 days cost $1.14M less on average than those that take longer (IBM, 2025).
- Deploy deception across IT and OT. A decoy layer gives high-confidence detection in the breakout window, independent of any specific vulnerability, and stays effective against malware-free attacks.
- Automate response where you safely can. Pre-authorized, analyst-gated containment and machine-speed dispatch close the distance between a 29-minute attacker and a human-paced SOC.

The threat is real and well documented, but so is the response. The industry has met systemic, deadline-driven problems before. The difference now is that defenders have powerful tools of their own, and that detecting an attacker by behavior, rather than by a flaw you may not yet know about, is the part of the strategy that holds up best as the next wave arrives.

References

Anthropic (2026). Project Glasswing: An initial update. https://www.anthropic.com/research/glasswing-initial-update

CSA CISO Community, SANS, OWASP GenAI Security Project (2026). The “AI Vulnerability Storm”: Building a “Mythos-ready” Security Program (v0.4). https://cloudsecurityalliance.org

Fortinet. What Is Deception Technology? https://www.fortinet.com/resources/cyberglossary/what-is-deception-technology

Google (2025). Cloud CISO Perspectives: Our Big Sleep agent makes a big leap. https://cloud.google.com/blog/products/identity-security/cloud-ciso-perspectives-our-big-sleep-agent-makes-big-leap

CrowdStrike (2026). 2026 Global Threat Report. https://www.crowdstrike.com/en-us/global-threat-report/

IBM (2025). Cost of a Data Breach Report 2025. https://www.ibm.com/reports/data-breach

DARPA (2025). AI Cyber Challenge marks pivotal inflection point for cyber defense. https://www.darpa.mil/news/2025/aixcc-results

Cloudflare (2026). Project Glasswing: what Mythos showed us. https://blog.cloudflare.com/cyber-frontier-models

Verizon (2025). 2025 Data Breach Investigations Report. https://www.verizon.com/business/resources/reports/dbir/