HACKTHEBOX

CISO Briefing

Stop Hiring Like It's 2025: AI-Augmented Cybersecurity Performance Data Every CISO Needs

Research:
AI-Augmented vs Human-Only Cybersecurity Performance Benchmark Report

Data Source:
NeuroGrid Capture The Flag (CTF) Competition (Nov 20–23, 2025)

Scale:
120 AI-agent teams vs
958 human teams (1+ attempt)
across 36 challenges

Classification:
CISO & Security Leadership

3.2x
higher solve rate
across all active participants

70%
improved challenge solve rate
of teams with one challenge completed

4.1x
faster for elite teams
1.4x across all teams

AI-augmented teams vs. Human-only teams

CEO Foreword

This benchmark was designed to answer a practical question we hear from CISOs and business leaders every day: how does AI-augmented capability compare to the teams you employ today, when tested under real constraints?

Rather than relying on synthetic benchmarks or vendor claims, this study uses performance data from a live, time-bounded HTB competition with professional-grade challenges and human-in-the-loop AI teams. Our goal is simple: report what we observed and translate it into practical implications for security leadership.

The results show meaningful speed gains, but also a clear message about operating model. AI can lift output across teams and deliver major speed advantages at the elite tier, yet the benefits are not automatic. They depend on how well practitioners can direct agents, validate outputs, and apply judgment on the hardest problems. In other words, the advantage will not come from AI adoption alone. It will come from building AI-fluent teams and human-in-the-loop workflows that perform reliably under pressure.

We hope this briefing helps you translate benchmark data into decisions: where to deploy AI first, how to structure oversight and governance, and how to invest in talent so that humans and agents become a force multiplier together.

Haris Pylarinos
CEO & Founder, Hack The Box

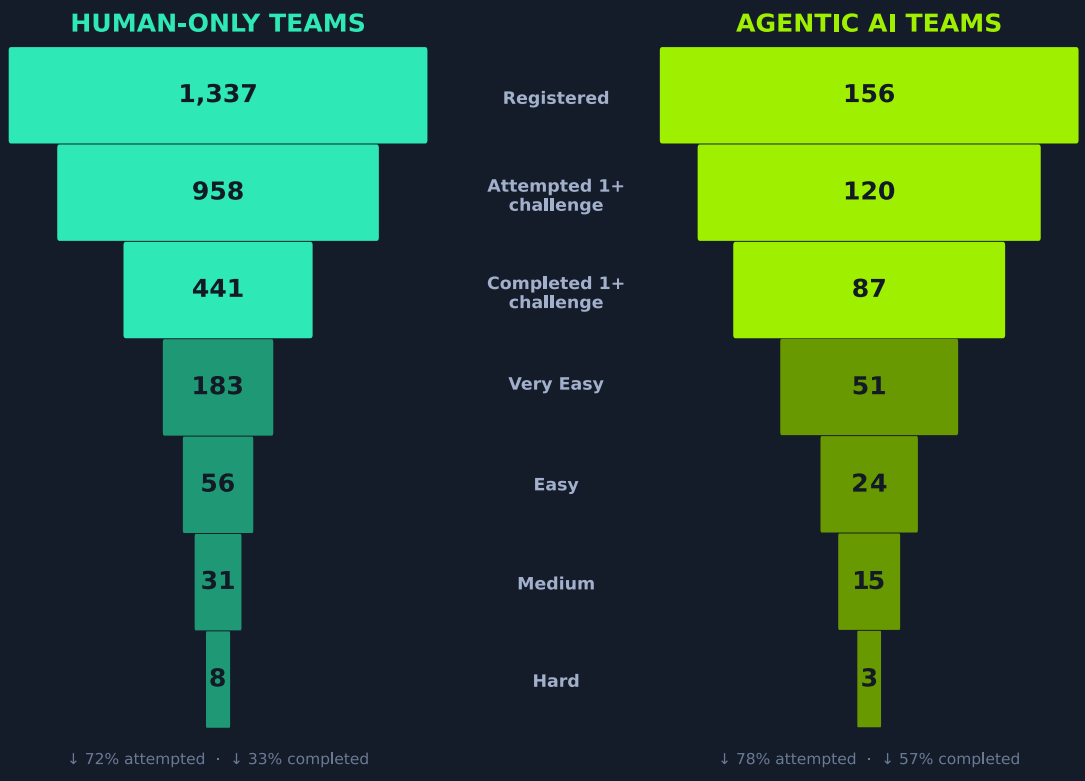
About The NeuroGrid Agentic AI vs Humans Benchmark

What it was: The largest controlled comparison of agentic AI and human performance on real-world cybersecurity tasks ever conducted — **1,337 human-only teams** vs **156 AI-agent teams** (registered) competing simultaneously across 36 professional-grade challenges (9 domains, 4 difficulty levels, 72-hour window) via the Hack The Box NeuroGrid CTF in November 2025. AI teams operated via Model Context Protocol (MCP) with human-in-the-loop oversight.

Why it matters: Unlike synthetic benchmarks or lab settings, these were challenges built for human experts under live competition pressure. This is the most operationally realistic dataset available for the question CISOs are actually asking: how does AI-augmented capability compare to the teams I employ today?

Team Participation & Difficulty Progression Funnel

Rows 1-3: Participation | Rows 4-7: Average number of teams completing a challenge in the difficulty tier



Attempted vs Completed: Of the teams that attempted at least one challenge, **73.3%** of AI-augmented teams completed successfully compared to just **46%** of human-only teams. These are professional-grade offensive security problems under genuine competitive pressure, not training exercises.

46%

Human-Only
of 958 who attempted,
441 completed at least one challenge

73.3%

Agentic AI
of 120 who attempted,
87 completed at least one challenge

Executive Summary

What This Means for Your Organization

AI-augmented security teams deliver a **70% higher solve rate** than human-only teams (27% vs 16%)— with a floor of **1.4x across all teams** and elite AI with human-in-the-loop completing challenges **3–4x faster**. The workforce implications are immediate: security operators (both adversaries and defenders) move faster with AI and the advantage compounds with skill, entry-level roles will shift from low-level tasks to developing AI-fluent practitioners with strong security context or risk displacement, requiring talent strategies to be restructured now.

Three-Tier Workforce Impact

AI augmentation creates three distinct workforce planning challenges:

1 Early Career — The Productivity Illusion

In some cases, AI serves as an apparent "**Competency Bridge**," allowing entry-level staff to solve more challenges than they otherwise could. But this masks a **productivity illusion**. Entry-level AI users can lack the cybersecurity depth to verify AI output and the orchestration skills to direct it. The result: higher apparent output without genuine skill development. Meanwhile, these same tasks are **automatable today**, creating a "**Missing Middle**" in the talent pipeline.

Recommendation: Redesign entry-level roles around AI-augmented workflows and invest in dual upskilling: cybersecurity fundamentals and agentic orchestration, or face potential displacement from two directions.

2 Mid Career — Deploy AI Here First

AI advantage peaks at Medium-difficulty tasks (**3.89x**), exactly where mid-career analysts operate. Speed advantage of **40–70%** at these tiers makes this the highest-ROI tier for AI tooling investment.

Recommendation: Equip these teams first to multiply output where the data shows maximum leverage.

3 Elite — Retain Talent, Accelerate Operations

Solve rate advantage narrows to **1.69x** at Top 5%, but elite AI teams with human-in-the-loop complete challenges **3–4x faster**. Top human: 36/36 vs AI 32/36.

Recommendation: Invest in retention. These operators are your last line of differentiation. AI is a speed multiplier here, not a capability replacement.

Executive Summary

Hints of the (Current) AI Ceiling

AI dominance is not absolute. Creativity remains a human stronghold — the top human team scored 36/36 vs AI's 32/36, and domains like Coding (2.09x) and Reversing (2.17x) show near-parity at the elite tier. The Difficulty Paradox reveals advantage peaking at Medium (3.89x) then retreating at Hard (2.97x), with AI failing entirely on 3 challenges. Domain Capability Spread is uneven — a ~3x range from Forensics (1.68x) to Secure Coding (5.15x). And while AI lifts lower-skilled teams dramatically, this "Competency Bridge" masks a productivity illusion and risks a "Missing Middle" in the talent pipeline. The ceiling is real, but it is rising fast.

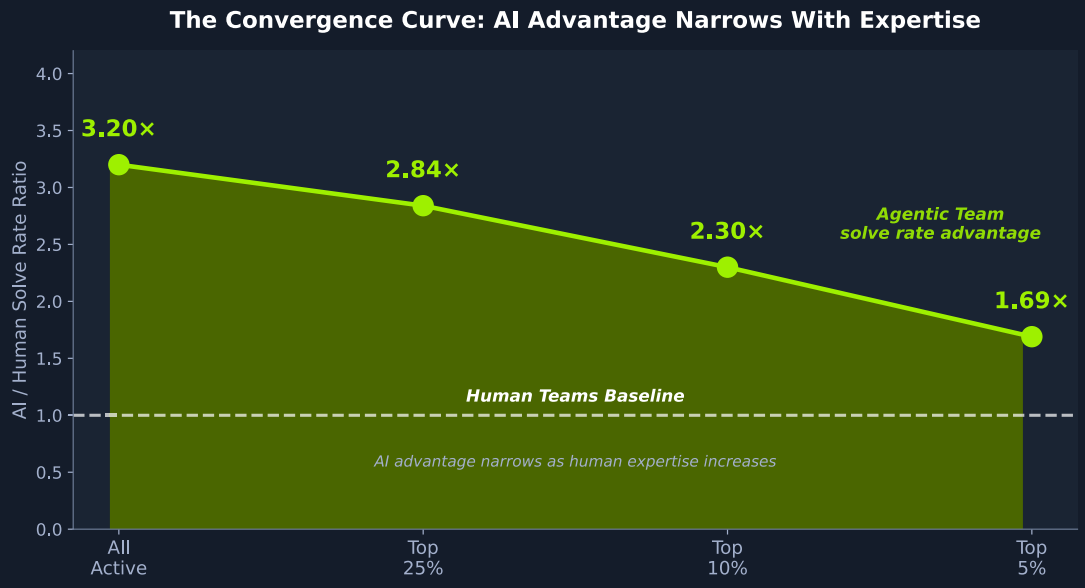
Bottom Line Up Front

The 70% solve rate advantage is not a future prediction, it is a current measurement based on this time-bound competition. Furthermore, with conservative speed and efficiency gains of **1.4x–1.7x** across all levels of seniority, organizations that fail to integrate AI into their security operations will face adversaries who already have. The question is not whether to adopt agentic AI-augmented security, but how quickly you can restructure to maintain a viable security posture. Human-AI hybrid models win and the data supports it at every tier.

Section 1: **The Convergence Curve**

Data proof: Three-Tier Workforce Model

The benchmark data supports a three-tier workforce model. This chart proves why. As we filter from all teams down to the elite, the AI solve rate advantage drops steadily: **3.2x** (All) → **2.84x** (Top 25%) → **2.3x** (Top 10%) → **1.69x** (Top 5%). Each tier tells a different story about how AI augmentation changes the workforce.



1. Early Career (All Teams)
The 3.2x ratio is heavily driven by lower-ranked teams. AI acts as an apparent "**Competency Bridge**," helping juniors solve more, but this masks a **productivity illusion**: they lack the depth to verify output and the orchestration skills to direct it effectively. These tasks could be **automated entirely**, creating a "Missing Middle" in the pipeline that builds tomorrow's experts.

2. Mid Career (Top 25%)
At 2.84x, mid-career operators still see significant lift. They are most likely attempting Medium-difficulty tasks, where AI advantage peaks at **3.89x** — the sweet spot for human-AI teaming.

3. Elite (Top 5%)
At **1.69x**, the gap narrows but does not close. Elite humans already have the competency. The top human team solved **36/36** vs AI's 32/36 — the critical metric where humans lead.

CISO Strategic Implications

Findings

- AI adversaries operate at 2–3x your team's baseline output.
- Elite humans are the last line of differentiation.
- Each workforce tier needs a different AI strategy.

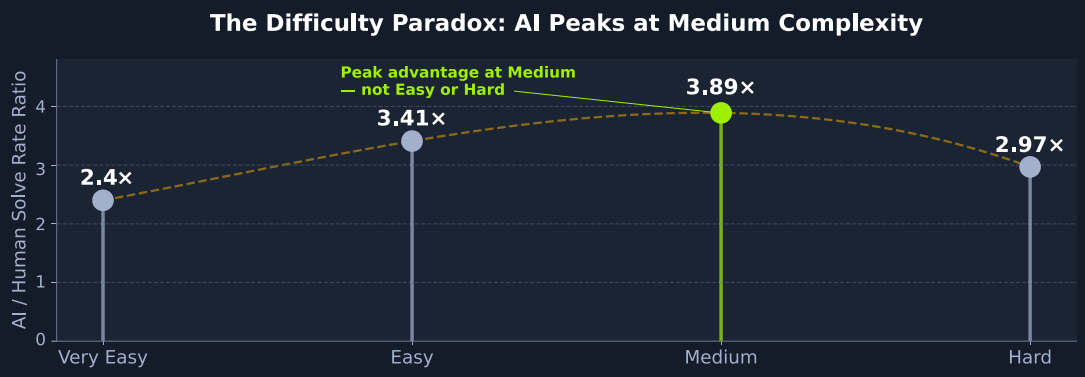
Recommendations

- → Model this multiplier into red-team threat scenarios.
- → Invest in Top 5% talent retention programs and build a pipeline.
- → Redesign career ladders: separate plans per tier.

Section 2: **The Difficulty Paradox**

Data proof: Difficulty Ceiling & Automation Risk

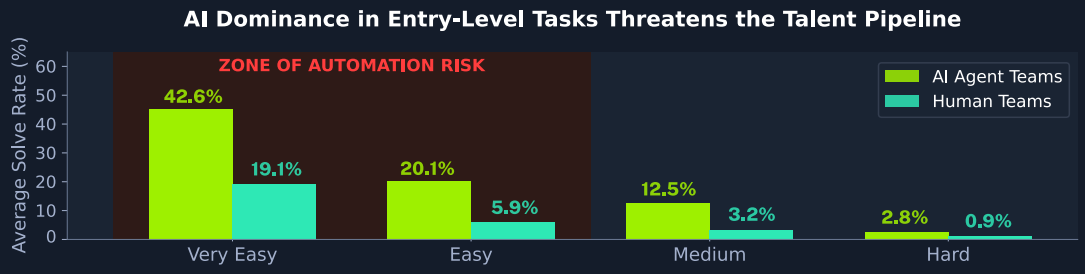
The benchmark revealed two AI capability boundaries: a Difficulty Ceiling where advantage peaks then drops, and an Automation Risk zone at the entry level. This chart proves both. AI advantage rises from Very Easy **2.40x** through Medium **3.89x** — then retreats to **2.97x** at Hard. The ceiling is real.



1. Mid-Career Strongest Gain
AI advantage peaks at **3.89x** on Medium tasks, the strongest gain of any tier. This is where mid-career analysts operate and where enterprises see **immediate ROI** from AI tooling.

2. Difficulty Ceiling
Hard challenges demand creative exploit chaining and novel reasoning: the **Creativity Moat**. AI cannot brute-force past genuine complexity.

3. Automation Risk Zone
AI solve rates of **42.6%** and **20.1%** on Very Easy and Easy (vs human 19.1% and 5.9%) mean entry-level tasks are at **risk of automation today**.



CISO Strategic Implications

Findings

- Entry-level SOC roles risk displacement from automation.
- AI automates the tasks that train juniors, creating a “Missing Middle” in the talent pipeline.
- The Missing Middle threatens future senior talent supply.

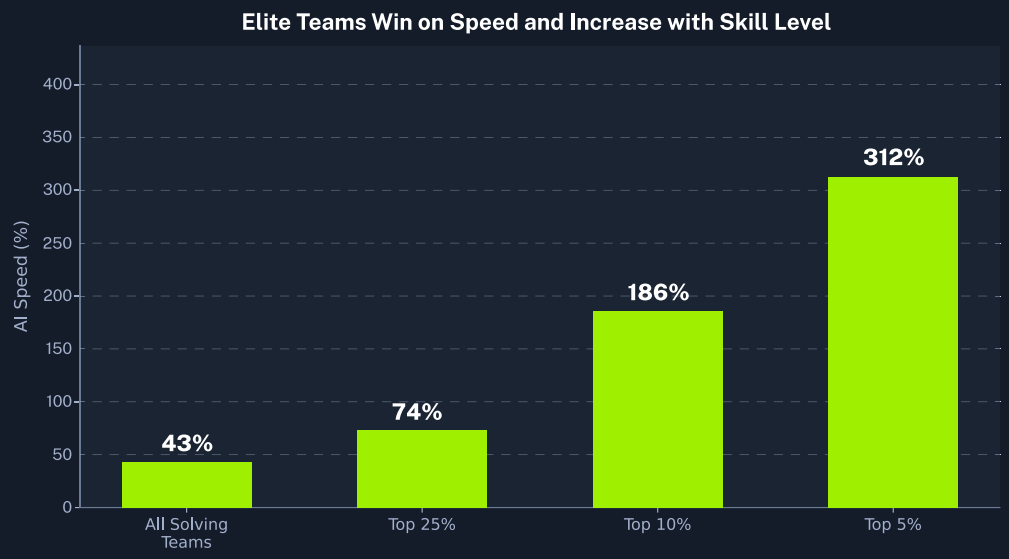
Recommendations

- Redesign career ladders around AI-augmented workflows.
- Budget for AI-assisted training to rebuild the apprenticeship model.
- Build new mentorship paths that include AI governance skills.

Section 3: **The Elite Speed Advantage**

Data proof: Elite — Much Faster, Not Much More Capable

The data shows the elite tier is “Much Faster, Not Much More Capable.” Using median solve times per challenge (robust to outliers), elite AI teams with human-in-the-loop complete challenges **3–4x faster** than human-only counterparts. At mid tiers, AI teams are **40–70% faster**. Speed advantage is the clearest operational differentiator.



1. 3–4x Faster at the Elite Tier
Among the Top 5%, AI teams with human-in-the-loop complete challenges **~4.1x faster** than human-only counterparts (**+312%**, based on median solve times per challenge). The speed advantage scales sharply with expertise: **+43%** across all solving teams and **+186%** in the Top 10%.

2. ...but Not Much More Capable
Recall from Section 1: at the Top 5%, the solve rate gap narrows to **1.69x**. The top human scored 36/36 vs AI’s 32/36. Elite AI with human-in-the-loop is a speed multiplier, not a capability replacement — a critical distinction for workforce strategy.

3. Entry-Level Compounding Pressure
AI is **43% faster** than the average human-only team across all solving teams and the speed advantage compounds sharply with expertise: **+186%** in the Top 10%, and **+312%** at the Top 5%. This is the second compounding factor that increases pressure on entry-level roles: Section 2 showed AI already solves many of the easier challenges; this section shows that once teams are skilled, AI accelerates execution dramatically. L1 analysts face automation pressure on both capability and speed.

CISO Strategic Implications

Findings

- Elite AI operates at 3–4x human speed per challenge.
- Adversarial AI exploits faster than humans can patch.
- Top-tier adversaries will leverage AI speed advantages.

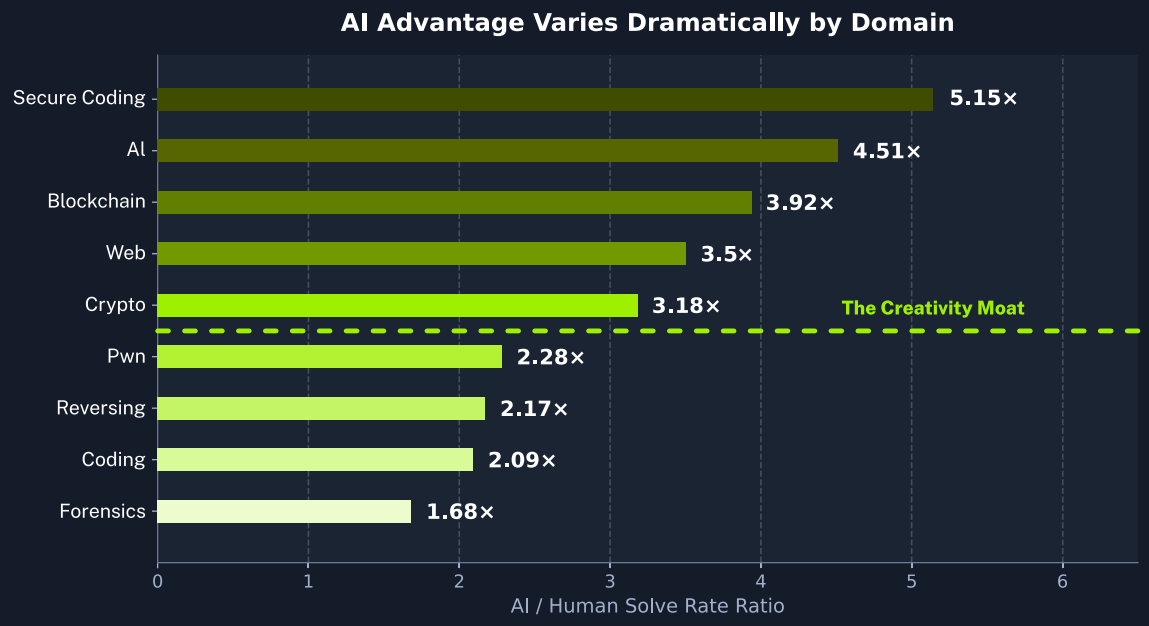
Recommendations

- Don’t dismiss AI speed based on aggregate totals alone.
- Compress incident response window assumptions.
- Redefine SLAs assuming AI-speed adversaries.

Section 4: **The Domain Spectrum**

Data proof: Domain Capability Spread & Creativity Moat

The research identified two further capability boundaries. First, when looking at all skill levels, AI's advantage varies dramatically by domain — from **5.15x** in Secure Coding to just **1.68x** in Forensics. Second, that advantage narrows sharply among elite performers, particularly in creative domains like Coding and Reversing. The chart below quantifies both patterns across all nine domains.



1. Domain Capability Spread
Across all skill levels, AI outperformed humans in every domain, but by vastly different margins — from **5x in structured tasks to just 2x** where improvisation matters most.

2. Creativity Moat
Among elite performers, **AI's edge nearly vanishes** in creative domains. The elite analysts still match or beat AI where deep reasoning is required.

CISO Strategic Implications

Findings

- Creative domains (Coding, Reversing) show near-parity at elite tier.
- Systematic domains (Secure Coding, AI, and Blockchain) show 4–5x AI advantage.
- AI shows a 3 to 1 capability spread across tested domains.

Recommendations

- Prioritise human investment in novel exploit development.
- Deploy AI tooling here first for maximum ROI.
- Build hybrid teams: AI breadth and human depth.

Operationalising the Three-Tier Model

Each data section maps to a workforce action

Tier 1: Early Career — Retrain or Risk Headcount Loss

Sections 1 & 2 | 3.2x solve rate · Automation Risk Zone · Productivity Illusion

Findings

- Entry-level tasks are automatable today (42.6% Very Easy, 20.1% Easy) and the "Competency Bridge" masks a productivity illusion.
- Automating training tasks creates a "Missing Middle" in the pipeline that produces future senior talent.
- L1 analysts face displacement from two directions — lacking both cybersecurity depth and agentic orchestration skills.

Recommendations

- Audit entry-level task lists; mandate human verification on all AI-assisted output; build QA frameworks before scaling AI access.
- Budget for AI-assisted training environments that preserve learning-by-doing: AI as teaching tool, not replacement.
- Double down on up-front upskilling: stronger security skills increase the productivity leverage of AI, and training teams to use AI effectively for cybersecurity amplifies that impact.

Tier 2: Mid Career — Augment and Multiply

Sections 1, 2 & 3 | 2.84x at Top 25% · 3.89x peak at Medium · 40–70% speed gain

Findings

- AI advantage peaks at 3.89x on Medium tasks, the strongest gain of any tier and the practical sweet spot for human-AI teaming.
- Speed advantage of 40–70% compounds across the full incident response workflow at mid tiers.
- This layer builds future senior practitioners; AI must enhance skill development, not absorb it.

Recommendations

- Deploy AI tooling to mid-career teams first; this is where the data shows maximum leverage and immediate ROI.
- Redesign SOC playbooks to embed AI-assisted triage, freeing mid-tier analysts for higher-order investigation work.
- Structure AI augmentation to accelerate learning: expose analysts to more varied scenarios, faster.

Tier 3: **Elite** — Retain and Augment

Sections 1, 3 & 4 | 1.69x at Top 5% · Creativity Moat · 3–4x speed multiplier

Findings

- Solve rate gap narrows to 1.69x — elite humans match AI on capability. Top human: 36/36 vs AI 32/36.
- Elite AI-augmented is 3–4x faster — a speed multiplier, not a capability replacement. Hard challenges expose AI failure modes.
- Top-tier adversaries will leverage AI speed advantages —your SLAs must assume AI-speed attackers.

Recommendations

- Invest in retention: these operators are your last line of differentiation where humans demonstrably lead.
- Pair elite analysts with AI co-pilots; route hardest incidents to human-led teams with AI for parallel hypothesis testing.
- Redefine incident response SLAs assuming AI-augmented adversaries; compress detection-to-containment windows.

Agentic AI Teaming Current Ceiling Indicators

CISO Strategic Implications

Findings

- AI peaks at Medium (3.89x), drops at Hard (2.97x), fails on 3 challenges.
- Domain Capability Spread: 1.68x (Forensics) to 5.15x (Secure Coding) — a ~3x range.
- Elite Coding (2.09x) and Reversing (2.17x) show near-parity.

Recommendations

- Do not assume AI scales to your hardest problems. Adversary sophistication demands human expertise.
- Deploy AI tooling by domain, not uniformly. Structured exploitation automates first; creative coding does not.
- Novel reasoning is your last defensible human advantage — invest in current real-world challenge training that builds it.

Methodology Note: Data from the NeuroGrid CTF November 2025. 120 AI-agent teams and 958 human teams (with at least 1 attempt) competed across 36 challenges spanning 9 security domains. Solve rate analysis uses Mean-of-Ratios methodology: each challenge produces an AI/Human solve rate ratio, then ratios are averaged across challenges. This approach prevents high-participation challenges from dominating results. Speed analysis uses median solve time among teams that solved each challenge, providing robustness against outlier performance. Cohort analysis filters both AI and human teams by overall ranking percentile. Source: AI-Augmented vs Human-Only Cybersecurity Performance Benchmark Report, March 2026.

HACKTHEBOX

Hack The Box is the leading cyber readiness platform for the agentic era, battle-testing and upskilling both humans and AI agents to enhance organizational cyber resilience.

Trusted by the Fortune 500, government agencies, and MSSPs, the platform delivers threat-informed learning paths consisting of real-world scenarios in gamified labs and live-fire simulations that build and validate offensive and defensive cyber capabilities.

With a loyal community of more than 4 million members and 800+ enterprise customers, Hack The Box empowers teams and intelligent systems alike to strengthen cyber defenses and reduce breach risk effectively.

For more information, visit hackthebox.com