WHITE PAPER

AI-Augmented Offensive Security

Improving Cybersecurity Through Human Expertise, Defense in Depth, and Artificial Intelligence

By Tyler Shields, Principal Analyst
Enterprise Strategy Group

May 2025

This White Paper from Enterprise Strategy Group was commissioned by HackerOne and is distributed under license from TechTarget, Inc.

Contents

Executive Summary

Emulating an attacker's thought process helps defenders gain a better understanding of potential threats. Just as defenders acquire a deeper knowledge of actual risks by concentrating on attacker actions, an offensive security approach examines applications and infrastructure holistically, promoting enhanced secure code development while observing and learning from production systems. Modern offensive security concepts combine manual efforts with automation and AI technologies to achieve defense-in-depth security. Today, offensive security techniques provide quicker discovery of exposures, improving remediation recommendations and helping organizations resolve issues quickly. This paper analyzes offensive security concepts, concentrating on the effects of AI, automation, and humans on defense in depth.

Offensive Security Builds a Strong Foundation

Offensive security is a collection of proactive security techniques designed to identify vulnerabilities and rapidly fix security issues. This broad view focuses on simulating attacks to test and discover weaknesses. As a subset of offensive security, adversarial security often focuses on a more targeted and in-depth analysis of specific attack scenarios within a larger ecosystem of issues identified by an offensive security model. Offensive security provides an overarching framework for proactive security testing in which adversarial models flourish. Together, they form a robust and comprehensive framework upon which enterprise security teams can build a defense-in-depth approach for their cybersecurity program.

Offensive Security at Each Defensive Layer

Defense in depth is a cybersecurity strategy that uses security controls at every level of a technology stack. Layering defenses helps detect and fend off attacks across the entire system, minimizing the risk of failure by putting backup security capabilities in place. This strategy isn't just about one security control. Instead, it's a holistic approach that improves the chances of spotting threats by integrating offensive thinking into each existing process.

While traditional defensive techniques focus on detection and protection, they often fall short in keeping up with fast-changing and adaptive attack techniques. Offensive and adversarial security techniques enhance ongoing threat management by combining human and automated threat scenarios. This method complements defensive tools, providing them with the latest threat context. When adversarial and defensive strategies work together, the defense-in-depth approach becomes highly effective.

Application and Cloud Modernization Demands Innovation and Adaptation

The rise of cloud-native applications and infrastructure requires changes to offensive security approaches. To remain essential in a defense-in-depth strategy, offensive techniques must be updated in tandem with application and software modernization. Security risks to organizations no longer focus solely on penetrating and exploiting single or multiple hosts. Application modernization has diminished the importance of traditional security perimeters and shifted many of an organization's critical assets ("crown jewels") into cloud infrastructure, which lies beyond direct organizational security control.

Now more than ever, a defense-in-depth approach to cybersecurity is essential for success. AI is driving significant innovation in these products, broadening the scope of analysis that can be conducted on modern infrastructure and applications through both human-augmented and purely AI-driven methods. When executed effectively, offensive security capabilities, AI-powered triage, and bespoke security advisory services enhance the efficiency, accuracy, and speed of an organization's entire cybersecurity program while simultaneously reducing tool sprawl and risk.

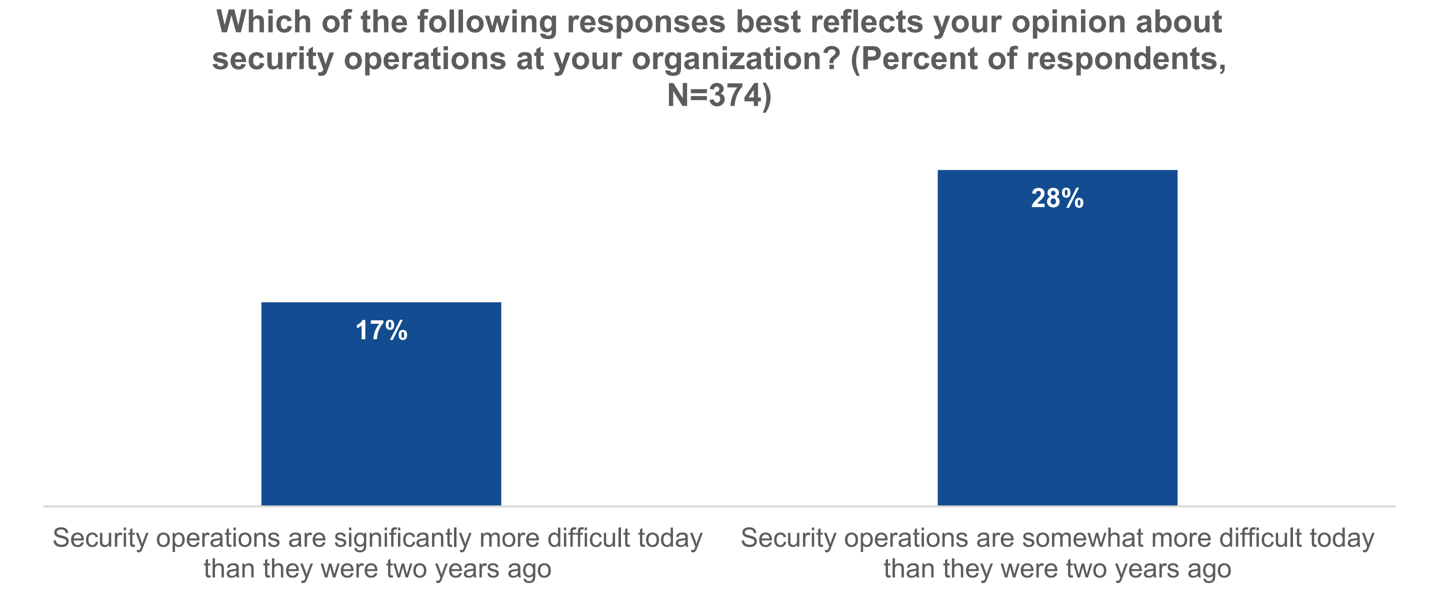
Security Complexity in an Evolving Threat Landscape

Over the last twenty years, digital transformation has turned businesses into software-focused organizations that rely heavily on applications and code. The quick rise of cloud-native applications has changed how technology is built and operated. Today, most business systems function beyond the limits of traditional data centers, taking advantage of cloud flexibility and the ability to connect via APIs. This new level of connectivity significantly increases the number of technology assets that need security checks and controls, resulting in a need for a complete defense-in-depth model based on offensive security capabilities. Research from Enterprise Strategy Group indicated that about 45% of organizations have seen a notable rise in the difficulty of their security operations compared to two years ago (see Figure 1).[1]

Security Operations Complexities

- Growing and evolving attack surface
- Rapidly changing threat landscape
- Volume and complexity of alerts
- Too much security data
- Too many tools and manual processes

Figure 1. Security Operations Have Increased in Complexity

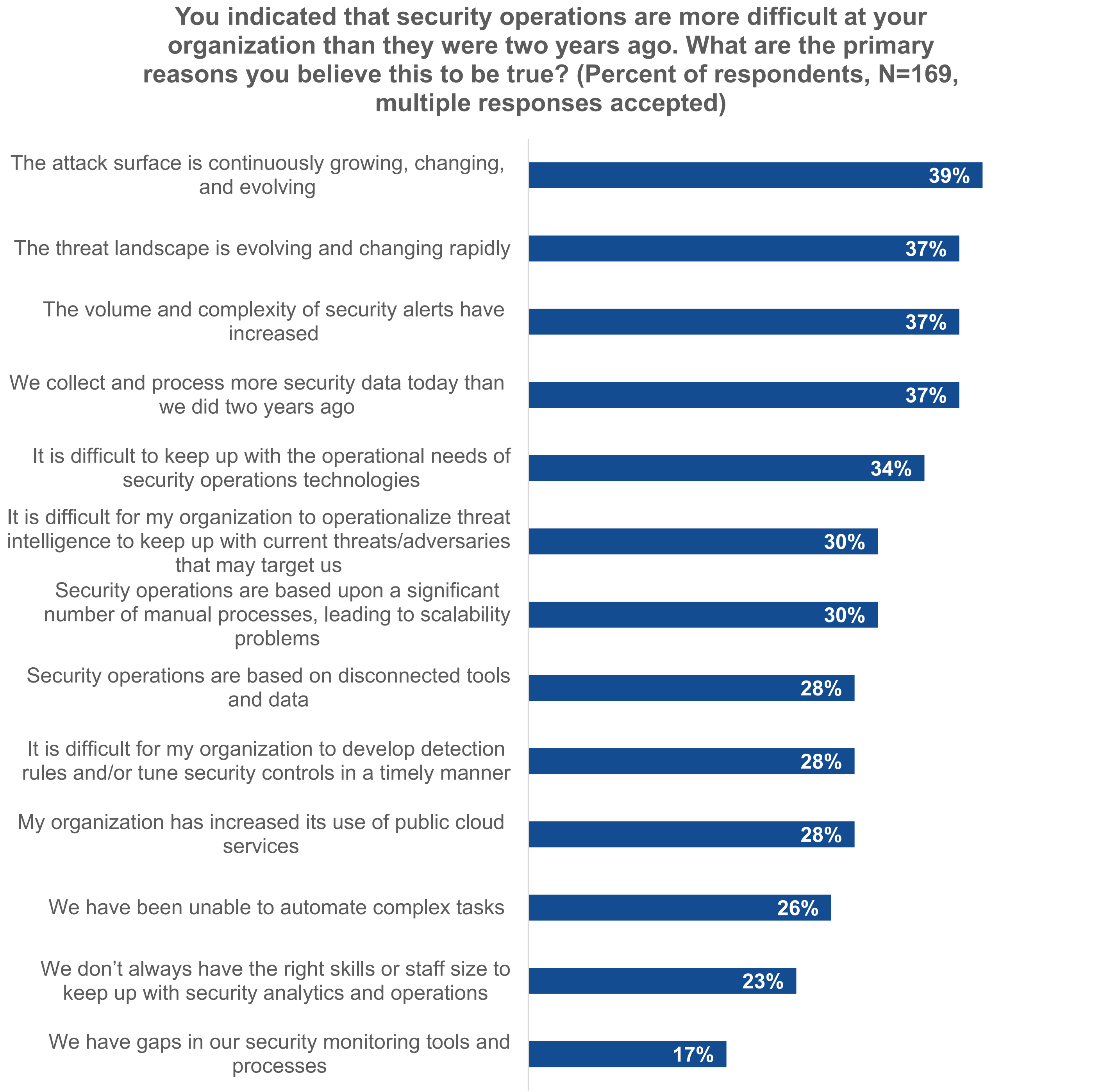


Source: Enterprise Strategy Group, now part of Omdia

Respondents identified the expanding attack surface (39%) and the rapidly evolving threat landscape (37%) as the primary reasons security operations are more difficult at their organization. Additional concerns included the rising volume and intricacy of security alerts, the surge in security data requiring analysis, and the challenges of tool sprawl and manual processes (see Figure 2).[2] These converging challenges create a rapidly changing technology stack that requires defense in depth to ensure consistent security. Offensive security enables continuous discovery and assessment of the entire attack surface in a thorough approach.

[1] Source: Enterprise Strategy Group Research Report, *The Triad of Security Operations Infrastructure: XDR, SIEM, and MDR*, June 2024.
[2] Ibid.

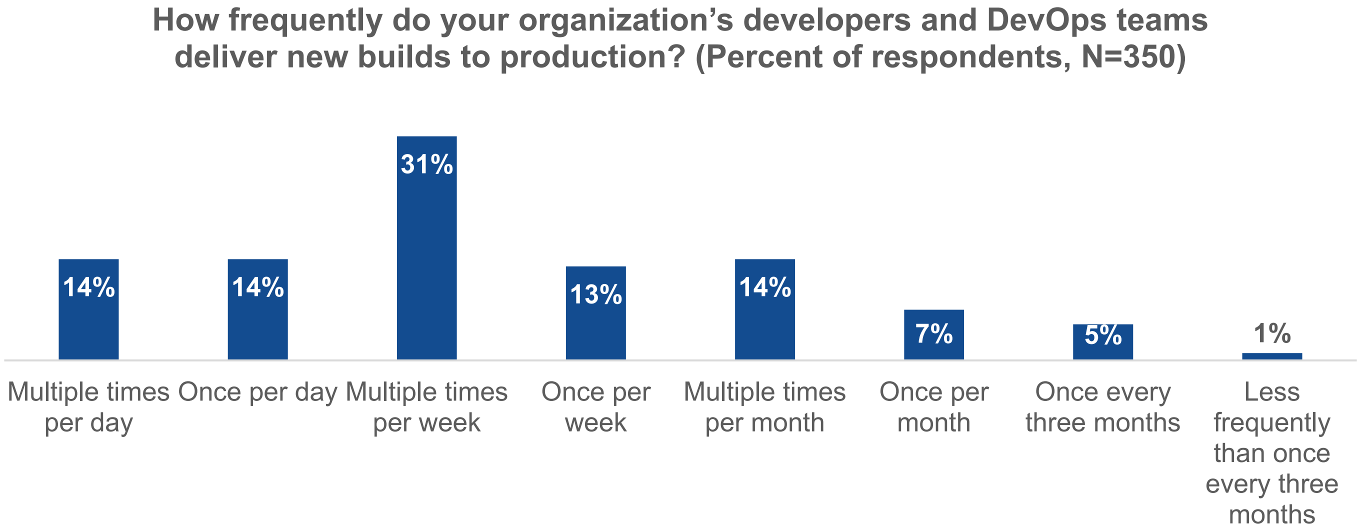
Figure 2. Causes of Security Operations Complexity



Source: Enterprise Strategy Group, now part of Omdia

Modern businesses are dealing with an ever-growing attack surface, with quickly evolving threats. This increasing complexity calls for better innovation in security operations. Today's applications rely on short-lived workloads and components, with a focus on continuous development and integration. Many top organizations have sped up their processes, pushing new code into production several times daily. The most forward-thinking companies go even further, offering constant micro-updates. The constant-update approach thrives when secured with a continuous offensive security approach. As the threat landscape transforms, offensive capabilities are the most agile and can adapt and innovate quickly. Research from Enterprise Strategy Group showed that 72% of organizations roll out new builds at least once a week (see Figure 3).[3]

Figure 3. Enterprises Are Deploying to Production More Frequently Than Ever Before



Source: Enterprise Strategy Group, now part of Omdia

Attackers are increasingly automating their processes to exploit evolving infrastructure and application strategies. Unlike past threats driven by manual human action, modern attacks leverage scalable, adaptive, and fully automated enterprise-grade systems.

Public disclosures, such as the "MOVEit attack," highlight the rapid weaponization and exploitation of vulnerabilities. In this case, attackers leveraged a widespread vulnerability in a popular file transfer software to automate simultaneous attacks across multiple targets.[4] This demonstrated significant weaponization, enabling post-exploitation and system manipulation. The attackers' flexibility and adaptability across compromised environments maximized their financial gain.

Effective security programs must understand and counter these automated attack processes. This requires a combination of human expertise, automation, and AI augmentation to match the speed, scale, and adaptive nature of modern threats.

Defense in Depth in Modern Applications

Defense in depth provides a methodology that strengthens security capabilities for applications and systems. Instead of relying on a single security control, it builds multiple layers of protection to minimize risks at each stage of the application development and deployment lifecycle. This means embedding security into the application's design from the outset, beginning with secure-by-design principles woven into the application's foundation. This is followed by continuous application testing, encompassing both automated and human-augmented methods to identify vulnerabilities as early as possible. With adversarial testing and analysis, conducted both in time-bound

[3] Source: Enterprise Strategy Group Complete Survey Results, *Modernizing Application Security to Scale for Cloud-native Development*, October 2024.
[4] Source: Caitlin Condon, "Rapid7 Observed Exploitation of Critical MOVEit Transfer Vulnerability," Rapid7.com, June 1, 2023.

engagements and continuous cycles, organizations learn how attackers attempt to breach their defenses, measure the potential impact of security issues, and accurately assess overall risk.

The defense-in-depth process doesn't stop at validation; organizations must follow through with mitigation, workflow enhancements, and continuous reinforcement to ensure sustained risk reduction. This layered, proactive approach helps organizations actively reduce risks and strengthen their security over time, ensuring a robust defense against evolving threats.

Humans and AI: Augmenting Each Layer of Defense

The future of security is a mix of AI automation and human expertise, layered within a defense-in-depth strategy. Automation helps operations scale with predictable processes, while AI provides flexibility and quick problem-solving, mimicking human abilities. This blend enables both attackers and defenders to adapt and enhance their operations without losing accuracy. In offensive security, this hybrid model strikes a good balance between efficiency and effectiveness. The best technologies use AI to boost human efforts in finding and verifying vulnerabilities, ensuring precise results at scale. Ultimately, this AI-human combination leads to actionable outcomes, including more accurate vulnerability validation, better prioritization of remediation efforts, and faster risk mitigation. By combining AI's speed with human insight, organizations can proactively strengthen their defense posture and minimize the impact of security incidents.

The following sections will cover how humans and AI complement each other at each layer of the security stack.

Secure By Design

Fixing security issues after production is much more expensive than catching them before release.[5] By integrating security into the design phase, whether it's for software development or infrastructure deployment, organizations use resources more efficiently and justify the time spent improving security.

When companies follow secure-by-design principles, security becomes a foundation of the entire development lifecycle rather than a last-minute addition. This forward-thinking approach helps spot potential threats early, making the system stronger from the start. Key secure-by-design practices include threat modeling, input validation, default secure settings, least privilege access, and ongoing security testing throughout the software development and deployment lifecycle.

Using AI in the design phase can improve threat modeling decision-making, helping to identify vulnerabilities before production, where fixes can get expensive. When done right, early code assessments cut down risk effectively without needing many resources. When applied incorrectly, traditional noisy application security assessment technologies can cause difficulties for developer productivity. AI-driven secure-by-design methods enable quick detection by analyzing code changes, maximizing risk reduction.

Defense-in-Depth Modern Security

- Implement secure-by-design principles to eliminate risks early in development.
- Conduct automated and human-augmented testing during coding and post-development stages.
- Use automation and AI at scale while human expertise improves accuracy and remediation outcomes.
- Achieve broad contextual awareness.

Application Security Testing

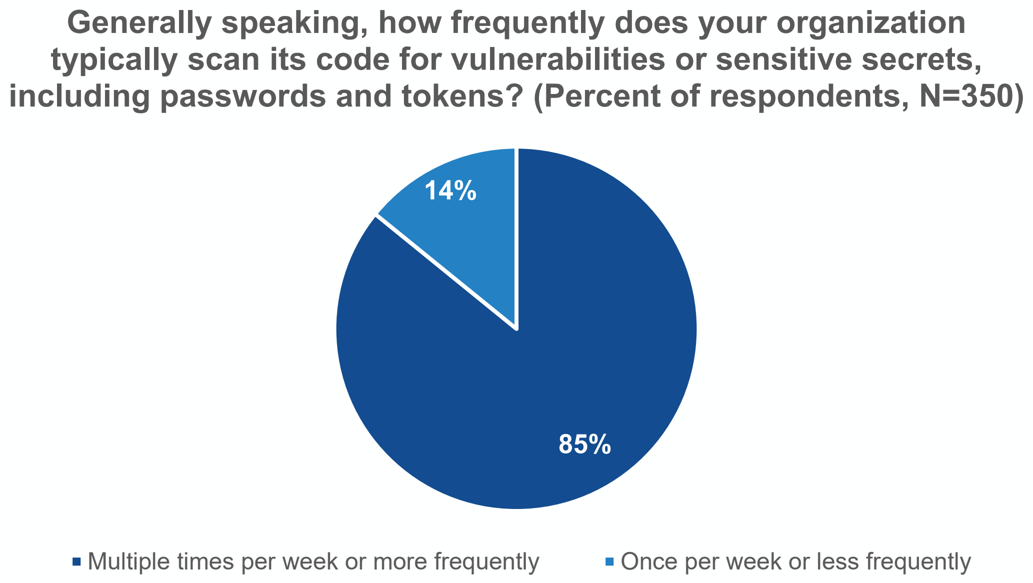
Effective security programs depend on secure design processes that include code reviews during and after development. Application security testing tools assess code

[5] Source: Dan Mateer, "The Cost Savings of Fixing Security Flaws in Development," HackerOne.com, February 25, 2025.

dependencies, run static analysis on the application source code, and review infrastructure as code to find and fix vulnerabilities before code goes live.

After development, a secondary layer of application security testing takes place, often at build time, to catch any problems that earlier steps might have overlooked, focusing on code composition and the use of libraries and open source components. Research from Enterprise Strategy Group showed that 85% of organizations scan their code multiple times per week or more frequently, underlining the necessity for security teams to automate scanning to increase frequency without slowing down development (see Figure 4).[6]

Figure 4. Automated Code Scanning Should Be Incorporated into the Development Process



Source: Enterprise Strategy Group, now part of Omdia

Automation is key in application security testing, functioning both before and sometimes during production. Automated tools, combined with human validation, enhance testing accuracy and deliver valuable insights for development and security teams. At this layer, AI enables quick detection by analyzing code changes, maximizing risk reduction. Combining human insight with AI analysis decreases false positives and helps development teams focus on what really matters.

While AI searches for vulnerabilities, human judgment is vital for validating findings and filtering out irrelevant concerns, ensuring that organizations prioritize the biggest risks. Furthermore, human validation of AI suggestions improves the analysis process and builds trust within the team, empowering developers to tackle real threats confidently and efficiently.

Penetration Testing

Adding AI and human integration into the penetration testing process can relieve testers of repetitive tasks while providing a baseline of consistent analysis. AI-augmented penetration testing processes help find vulnerabilities and issues quickly and accurately, enabling the human tester to focus on the more esoteric and difficult-to-discover issues that automation might struggle to find.

[6] Source: Enterprise Strategy Group Complete Survey Results, *Modernizing Application Security to Scale for Cloud-native Development*, October 2024.

Connecting AI with the penetration testing process also provides contextualization and personalization of the assessment by adding knowledge of the specific characteristics of the target to the approach. While AI brings a holistic analysis engine to the penetration testing team, the human capacity for creativity and innovation often outshines current AI engines. The optimal level of integration is to augment automated tools with both AI and human logic and intuition capabilities.

Finally, AI can streamline reporting tasks at the end of the penetration test. This saves testers valuable time, enabling them to spend more time on testing the actual asset. Additionally, AI can help ensure that reporting and communications are clear and consistent.

Adversarial Testing

Adversarial testing unifies all previous layers of defense. It checks for vulnerabilities and exposures while recommending remediation through a combination of automated, human, and AI techniques. Both human and AI expert systems act as adversaries, utilizing extensive knowledge of the target environment to enhance vulnerability discovery and detection capabilities that adapt to an organization's resource needs. This model strengthens the ability of organizations' in-house security teams to identify exploits that modern tools might overlook.

Adversarial testing can function as either a time-bound or a continuous model. Time-bound testing is well-suited for targeted validation during key development milestones, while continuous testing enables persistent visibility into emerging threats across the software lifecycle. Both approaches surface exposures that require triage and validation. The results of AI detection are best validated by a human, as humans possess the creative ability to understand and detect novel and elusive vulnerabilities that automation often misses.

Adversarial testing uniquely identifies real-time exposures, particularly vulnerabilities that have bypassed earlier controls, providing a vital safety net to catch potential "escapes" in production environments. Together, human ingenuity and AI acceleration create a dynamic and adaptive defense, solidifying adversarial testing as an essential final layer in any comprehensive defense-in-depth strategy.

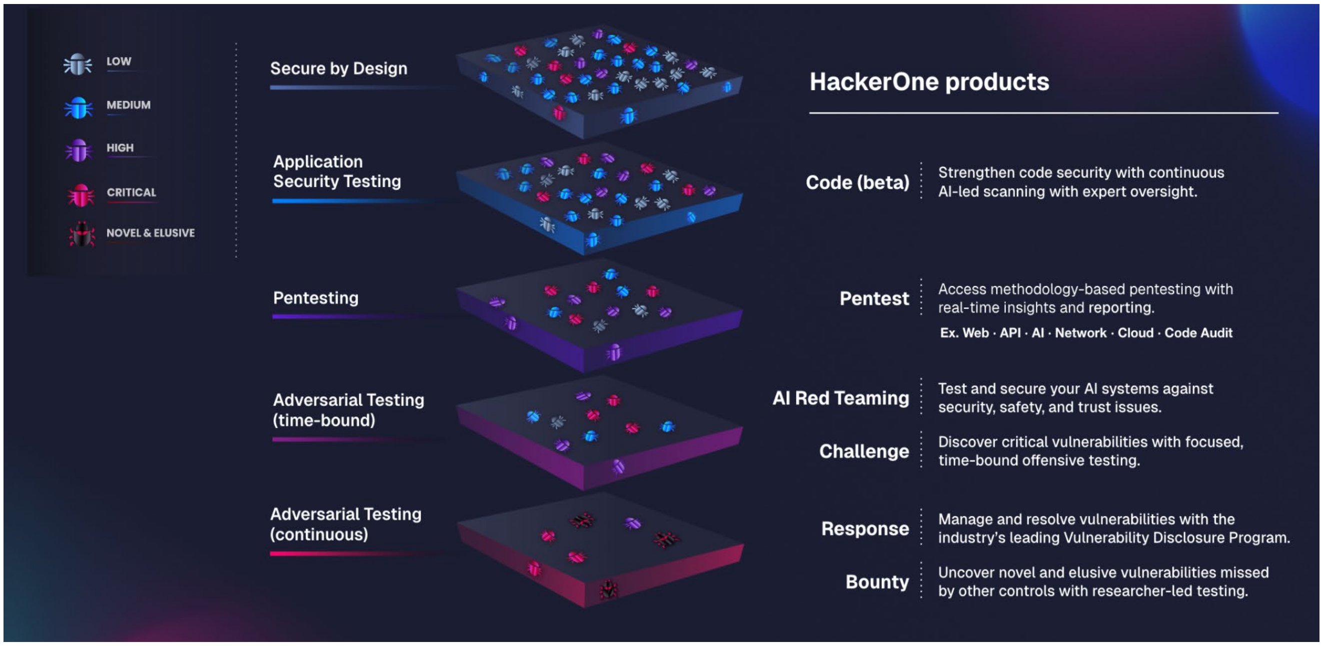
HackerOne Addresses Human and AI-Augmented Defense in Depth

HackerOne provides a cutting-edge offensive security testing platform that proactively identifies and mitigates vulnerabilities before adversaries can exploit them (see Figure 5). By leveraging the scalability of Hai, HackerOne's AI security agent, with the deep knowledge and experience of its security researchers, HackerOne offers unmatched protection. Supported by years of expertise, access to the world's largest security researcher community, and a proven track record across diverse industries, HackerOne delivers comprehensive and effective vulnerability detection and remediation guidance at every stage of the software development lifecycle.

While Hai's AI-powered analysis rapidly identifies potential vulnerabilities and exposures, HackerOne's human validation ensures accuracy, reduces false positives, and delivers clear, actionable recommendations, enabling security and development teams to confidently accelerate secure development.

The HackerOne platform provides continuous and scalable security testing through six interconnected solutions.

Figure 5. HackerOne Products


LOW
MEDIUM
HIGH
CRITICAL
NOVEL & ELUSIVE
Secure by Design
Application Security Testing
Pentesting
Adversarial Testing (time-bound)
Adversarial Testing (continuous)

HackerOne products

Code (beta) Strengthen code security with continuous AI-led scanning with expert oversight.

Pentest Access methodology-based pentesting with real-time insights and reporting.
Ex. Web · API · AI · Network · Cloud · Code Audit

AI Red Teaming Test and secure your AI systems against security, safety, and trust issues.

Challenge Discover critical vulnerabilities with focused, time-bound offensive testing.

Response Manage and resolve vulnerabilities with the industry's leading Vulnerability Disclosure Program.

Bounty Uncover novel and elusive vulnerabilities missed by other controls with researcher-led testing.

Source: HackerOne

HackerOne Code: AI + Human Code Security

HackerOne Code helps developers ship more secure software by embedding remediation guidance directly into the development process. Designed as an intelligent evolution of traditional static application security testing, HackerOne Code leverages specialized AI security agents informed by HackerOne security intelligence to analyze pull requests in real time, flagging security flaws before the code reaches production. Complex and high-risk issues are further reviewed by HackerOne's network of expert security engineers, ensuring only accurate, relevant vulnerabilities are passed to development teams. This not only validates risks, virtually eliminating false positives and noise, but also delivers hands-on remediation advice within familiar developer tools like GitHub, Bitbucket, and Azure DevOps. With human-in-the-loop validation layered over AI scanning and seamless integration into existing CI/CD pipelines, HackerOne Code helps teams move fast and stay secure, prioritizing prevention over detection without interrupting development speed.

HackerOne Pentest: Programmatic, On-Demand Pentests by Experts

HackerOne Pentest delivers comprehensive security evaluations using proven testing methodologies and specialized human expertise. Organizations engage a carefully selected group of security professionals to assess web apps, APIs, networks, and cloud environments against frameworks like OWASP, NIST, and PTES. Findings are detailed, prioritized, and paired with practical guidance to accelerate remediation. A centralized interface simplifies everything from test scoping to retesting, offering clear oversight and faster turnaround than legacy consultancies. Designed to align with compliance and risk management objectives, Pentest uncovers complex vulnerabilities that scanners might miss, while fitting the pace and demands of modern software delivery cycles.

HackerOne AI Red Teaming: Testing AI for Security, Safety, and Trust

HackerOne AI Red Teaming is purpose-built to expose weaknesses in AI and machine learning models before they're deployed in production. Expert security researchers simulate targeted attacks to identify issues like model bias, prompt injection, logic manipulation, and data leakage—threats unique to AI environments. Each engagement is customized to an organization's use case, whether it's a generative large language model, recommendation engine, or autonomous agent. Findings are supported by HackerOne's security advisors to help teams secure models against both technical exploits and trust-related failures. This offering supports growing regulatory and ethical expectations for responsible AI deployment and is essential for organizations operating at the intersection of innovation and risk.

HackerOne Challenge: Time-Bound, Adversarial Testing

When speed matters, HackerOne Challenge delivers focused, competitive testing that rapidly surfaces serious security gaps. These time-bound engagements activate top-ranked security researchers in a race to uncover impactful vulnerabilities across a defined scope. Commonly used before major product launches, compliance deadlines, or infrastructure changes, Challenges provide actionable results in days, instead of weeks. The gamified format encourages deep exploration, while the time-boxed nature ensures high signal-to-noise and a strong return on investment. Challenge serves as a fast-response layer in an organization's security strategy, surfacing critical exposures quickly and enabling rapid fixes when timing is critical.

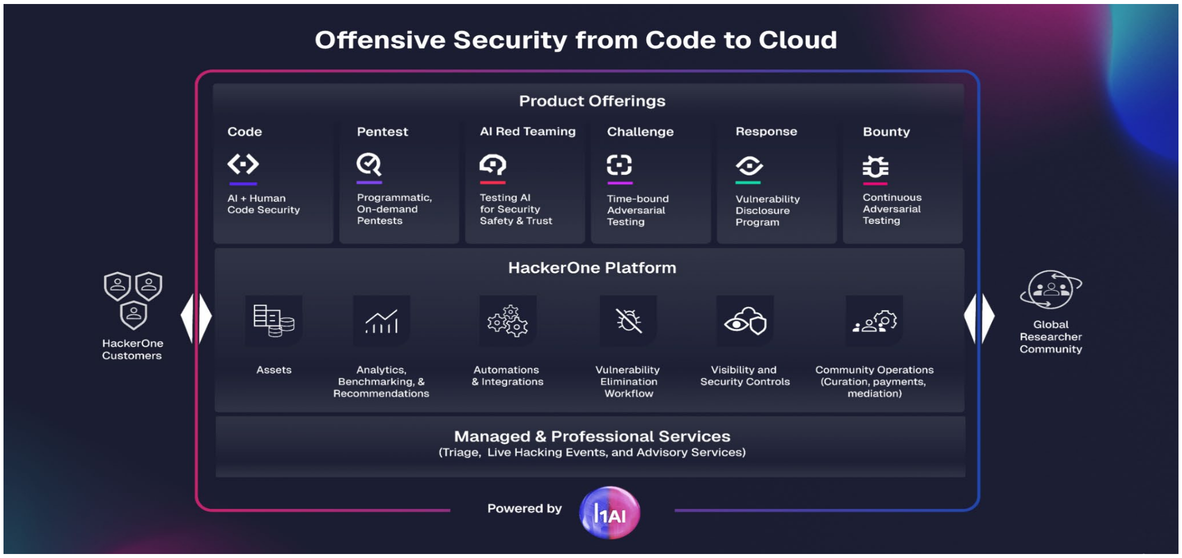
HackerOne Response: Always-on Vulnerability Disclosure Program

HackerOne Response enables organizations to receive, manage, and resolve externally reported vulnerabilities through a coordinated, always-on vulnerability disclosure program (VDP). Whether driven by compliance, public policy, or proactive security strategy, a VDP provides a secure, accessible channel for ethical hackers, researchers, and users to disclose issues responsibly. HackerOne Response includes the workflows, infrastructure, and operational best practices needed to triage submissions efficiently and collaborate securely with reporters. Backed by program management expertise and built on lessons from hundreds of VDPs, it offers security teams a unified view of vulnerability trends, helping them reduce risk, improve coverage, and foster a transparent security culture across the organization.

HackerOne Bounty: Continuous Adversarial Testing

HackerOne Bounty harnesses the power of collective intelligence to scale adversarial testing alongside an organization's growth. As a fully managed bug bounty program, HackerOne Bounty connects organizations with a global network of security experts to uncover vulnerabilities and exposures in real-world conditions, before adversaries can exploit them. Unlike one-time assessments, Bounty runs continuously, adapting to changes in infrastructure, applications, and threat landscape. Security teams receive validated, high-impact reports prioritized by severity and business risk, enabling rapid and efficient remediation. As the final layer of defense, Bounty integrates learning loops with Hai to deliver intelligent recommendations that span across every layer of defense, driving proactive vulnerability elimination at scale.

Figure 6. HackerOne's Offensive Security From Code to Cloud



Offensive Security from Code to Cloud

Product Offerings

Code: AI + Human Code Security

Pentest: Programmatic, On-demand Pentests

AI Red Teaming: Testing AI for Security Safety & Trust

Challenge: Time-bound Adversarial Testing

Response: Vulnerability Disclosure Program

Bounty: Continuous Adversarial Testing

HackerOne Platform

Assets

Analytics, Benchmarking, & Recommendations

Automations & Integrations

Vulnerability Elimination Workflow

Visibility and Security Controls

Community Operations (Curation, payments, mediation)

Managed & Professional Services
(Triage, Live Hacking Events, and Advisory Services)

HackerOne Customers

Global Researcher Community

Powered by

Source: HackerOne

A Layered, Human-Integrated Approach Ensures Security for Today and Tomorrow

In today's complex threat landscape, offensive security is indispensable for building a robust defense-in-depth security program. By integrating offensive techniques into a comprehensive defense-in-depth strategy, organizations can proactively improve their security posture by identifying weaknesses before malicious actors exploit them. This approach, which leverages human expertise, automated tools, and AI-driven technologies, enhances traditional security processes at every stage of the secure software development lifecycle. The strength comes from layering multiple testing and security controls: Offensive testing complements secure-by-design principles, application testing, and penetration testing, creating a multi-faceted defense that minimizes attack surfaces and maximizes an organization's security outcomes.

As attackers increasingly utilize automation and AI, adopting advanced offensive security capabilities becomes a necessity. This layered approach ensures continuous validation, ultimately strengthening an organization's ability to handle sophisticated threats.

HackerOne integrates several layers to review code, identify vulnerabilities, conduct penetration tests, and utilize security experts—all within a single platform that fortifies organizations against today's security threats. Explore the HackerOne Platform in depth.

About Enterprise Strategy Group
Enterprise Strategy Group, now part of Omdia, provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

contact@esg-global.com
www.esg-global.com