

CrowdStrike and Meta Deliver New Benchmarks for the Evaluation of AI Performance in Cybersecurity

September 15, 2025

New benchmarks define how LLMs should be tested in the SOC – measuring real threats, workflows, and outcomes to help defenders

AUSTIN--(BUSINESS WIRE)--Sep. 15, 2025-- Fal.Con 2025, Las Vegas -- [CrowdStrike](#) (NASDAQ: CRWD) today, in partnership with Meta, introduced a new suite of benchmarks – [CyberSOCEval](#) – for evaluating how AI systems perform in real-world security operations. Built on Meta’s [CyberSecEval framework](#) and CrowdStrike’s leading threat intelligence and cybersecurity AI data expertise, this suite of open source benchmarks helps establish a new framework for testing, selecting, and leveraging large language models (LLMs) in the security operations center (SOC).

Cyber defenders face an overwhelming challenge from the influx of security alerts and evolving threats. To outpace adversaries, organizations must embrace the latest AI technologies. Many security teams are still early in their AI journeys, particularly in using LLMs to automate tasks and drive efficiency in security operations. Without clear benchmarks, it’s difficult to know which systems, use cases, and performance standards deliver a true AI advantage against real-world attacks.

Meta and CrowdStrike are addressing this challenge by introducing CyberSOCEval, a suite of benchmarks that help define what effective AI looks like for cyber defense. Built on Meta’s open source CyberSecEval framework and CrowdStrike’s frontline threat intelligence, CyberSOCEval evaluates LLMs across critical security workflows such as incident response, malware analysis, and threat analysis comprehension. By testing AI systems’ ability against a combination of real-world adversary tradecraft and expert-designed security reasoning scenarios based on observed adversarial tactics, organizations can validate performance under pressure and prove operational readiness. With these benchmarks, security teams can pinpoint where AI delivers maximum value, while model developers gain a North Star for improving capabilities that enhance ROI and SOC effectiveness.

“At Meta, we’re committed to advancing and maximizing the benefits of open source AI – especially as large language models become powerful tools for organizations of all sizes,” said Vincent Gonguet, Director of Product, GenAI at Superintelligence Labs at Meta. “Our collaboration with CrowdStrike introduces a new open source benchmark suite to evaluate the capabilities of LLMs in real world security scenarios. With these benchmarks in place, and open for the security and AI community to further improve, we can more quickly work as an industry to unlock the potential of AI in protecting against advanced attacks, including AI-based threats.”

“When two leaders like CrowdStrike and Meta come together, it’s larger than collaboration, it’s about setting the direction of cybersecurity for the AI era,” said Daniel Bernard, chief business officer at CrowdStrike. “By combining CrowdStrike’s adversary intelligence and leadership in AI-native cybersecurity, with Meta’s AI research expertise and vast dataset, we’re helping customers – and cybersecurity as a sector – adopt AI systems with confidence. This partnership sets a new bar for how AI in the SOC should be built and deployed, empowering defenders to stay ahead of the adversary.”

The CyberSOCEval open source benchmark suite is now available for the AI and security community to use to evaluate model capabilities. To access the benchmarks, please visit [Meta’s CyberSecEval framework](#). For more information about the benchmarks, [visit here](#).

About CrowdStrike

[CrowdStrike](#) (NASDAQ: CRWD), a global cybersecurity leader, has redefined modern security with the world’s most advanced cloud-native platform for protecting critical areas of enterprise risk – endpoints and cloud workloads, identity and data.

Powered by the CrowdStrike Security Cloud and world-class AI, the CrowdStrike Falcon® platform leverages real-time indicators of attack, threat intelligence, evolving adversary tradecraft and enriched telemetry from across the enterprise to deliver hyper-accurate detections, automated protection and remediation, elite threat hunting and prioritized observability of vulnerabilities.

Purpose-built in the cloud with a single lightweight-agent architecture, the Falcon platform delivers rapid and scalable deployment, superior protection and performance, reduced complexity and immediate time-to-value.

CrowdStrike: We stop breaches.

Learn more: <https://www.crowdstrike.com/>

Follow us: [Blog](#) | [X](#) | [LinkedIn](#) | [Facebook](#) | [Instagram](#)

Start a free trial today: <https://www.crowdstrike.com/free-trial-guide/>

© 2025 CrowdStrike, Inc. All rights reserved. CrowdStrike and CrowdStrike Falcon are marks owned by CrowdStrike, Inc. and are registered in the United States and other countries. CrowdStrike owns other trademarks and service marks and may use the

brands of third parties to identify their products and services.

View source version on [businesswire.com](https://www.businesswire.com/news/home/20250914633899/en/): <https://www.businesswire.com/news/home/20250914633899/en/>

Media Contact

Jake Schuster

CrowdStrike Corporate Communications

press@crowdstrike.com

Source: CrowdStrike