AI Institute for Agent-based Cyber Threat Intelligence and Operation

https://action.ucsb.edu

ACTION Year 2 Annual Report: Summary

Reporting period: June 2024 to May 2025 (Published November 2026)

The NSF AI Institute for Agent-based Cyber Threat Intelligence and Operation (ACTION) develops AI-enabled security agents that work with human defenders across the full cybersecurity life cycle: assessing vulnerabilities, detecting attacks, attributing them, and responding and recovering. ACTION is led by UC Santa Barbara with 11 partner universities and is supported by the U.S. National Science Foundation, with additional support from the Department of Homeland Security and IBM.

Research. In its second year, ACTION researchers published more than 60 papers, bringing the Institute's total to 100. The work spanned four foundational AI thrusts, four cybersecurity thrusts, and the Institute's shared testbed, GATE. Highlights include:

- A framework that helps networks of AI agents reach the correct conclusion even when each agent has limited or unreliable information.
- A learning method that keeps improving as conditions change, demonstrated on malware detection, where attackers constantly adapt.
- Graphene, developed with Cisco, which uses AI-generated attack graphs to assess an organization's overall security posture.
- ArtPrompt, which showed that ASCII art can bypass the safety measures of leading large language models, and DeepCapa, which identifies malicious capabilities in Windows malware even when the code is disguised.

Open tools and datasets. ACTION released MABEL, a malware benchmark of more than 82,000 samples across 470 malware families, on GitHub and through the National AI Research Resource (NAIRR) pilot. Researchers also released MAGPIE, a tool for generating instruction data for large language models, along with early versions of ACTION's AI stack and GATE code. GATE added realistic background traffic, a new command line interface, and a starter kit, and began a partnership with USC's SPHERE testbed.

Competitions. Teams advised by ACTION faculty, Shellphish and Team Atlanta, were two of the seven finalists selected for DARPA's AI Cyber Challenge. A UIUC team advised by Dr. Gang Wang was one of 10 teams selected for Amazon's Trusted AI Challenge.

Education and workforce. ACTION expanded its outreach from high school through graduate school:

- iCTF 2024 drew 16 high school teams, a 60% increase over 2023, and 27 undergraduate teams. Six ACTION institutions built 53 new challenges, and 12 walkthrough videos on YouTube help newcomers get started.
- A student-led hacking club at Santa Barbara High School produced a curriculum that Norfolk State University and UC Santa Cruz are adapting for their own outreach.
- ACTION's first summer research experience brought 23 undergraduate and graduate students to UC Santa Barbara, Rutgers, and Norfolk State, with student work presented at UCSB's Undergraduate Research Symposium and CSCI'24.
- Dr. Giovanni Vigna introduced new graduate and undergraduate courses on AI and security at UC Santa Barbara.

Impact and engagement. ACTION Director Dr. Giovanni Vigna received the 2024 ACM SIGSAC Outstanding Innovation Award. ACTION researchers joined panels at ACSAC 2024 and the Billington State and Local Cybersecurity Summit, presented to the University of California Regents and Georgetown's Center for Security and Emerging Technology, and took part in the Summit for Digital Trust in Switzerland, which opened new connections in Europe and Africa.

Looking ahead. In February 2025, ACTION revised its Strategic and Implementation Plan for Years 3 to 5. The new plan replaces fixed research phases with focused cross-campus projects selected through an internal call for proposals, broadens SEC-3 from attribution to cyber threat intelligence, and refocuses SEC-4 on multi-step attacks.

The full report is available at https://action.ucsb.edu/knowledge-transfer. ACTION's current Strategic & Implementation Plans are available by request.
Please direct any questions to action@cs.ucsb.edu.

This material is based upon work supported by the U. S. National Science Foundation under grant no. 2229876 and is supported in part by funds provided by the U. S. National Science Foundation, by the Department of Homeland Security, and by IBM.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U. S. National Science Foundation or its federal agency and industry partners.