Securing AI Systems: New Challenges and Research Priorities

A convening of the National Academies Forum on Cyber Resilience

April 20–21, 2026
National Academy of Sciences Building - 2101 Constitution Avenue, NW Washington, DC

Day 1 — Monday, April 20, 2026

All times are EDT.

9:00 – 9:15 AM
Welcome and Opening Remarks
Ellen Zegura, National Science Foundation
John Manferdelli (NAE), National Academies Forum on Cyber Resilience

---

9:15 – 10:00 AM
Session 1: Why Are We Here?
Perspectives on the importance and urgency of securing AI systems.

Kathleen Fisher, ARIA, "The convergence of AI, cyber, and formal methods"
Hyrum Anderson, SSIL, "AI's role in cyber"
Anita Nikolich, UIUC, "why we need to shorten research cycles"

10:00 – 10:20 AM
Break

10:20 AM – 12:00 PM
Session 2: Defining AI Security
Define what "AI security" encompasses, distinguishing it from traditional cybersecurity and AI safety. It will identify key concepts, stakeholders, system boundaries, and shared vocabulary.

Panel Discussion (10:20 – 11:20 AM)
Jonathan Petit, Qualcomm
Siwei Lyu, University at Buffalo
Giovanni Vigna, University of California, Santa Barbara
Moderator: Greg Shannon, Idaho National Laboratory

Moderated Audience Discussion (11:20 AM – 12:00 PM)
Moderator: Howie Shrobe, DARPA

12:00 – 1:00 PM
Lunch

1:00 – 2:40 PM
Session 3: Threat Modeling for AI Security
Deep dive into adversaries, emerging attack surfaces, failure modes, and trust boundaries; consider how threat modeling must evolve to account for advancing AI capabilities, with examples from key application domains.

Panel Discussion (1:00 – 2:00 PM)
Malichai Jones, Microsoft
Ads Dawson, Dreadnode
Bo Li, University of Illinois Urbana-Champaign
Barton Miller, University of Wisconsin–Madison
Moderator: Anita Nikolich, University of Illinois Urbana-Champaign

Moderated Audience Discussion (2:00 – 2:40 PM)
Moderator: Window Snyder, Thistle Technologies

2:40 – 3:00 PM
Break

3:00 – 4:40 PM
Session 4: Adapting Classical Security to AI Security
Explore how established cybersecurity principles—such as prevention, audit, defense-in-depth, identity, least privilege, and secure development lifecycles—can be translated to AI systems, and where adaptation is needed.

Panel Discussion (3:00 – 4:00 PM)
Hyrum Anderson, SSIL
Fred Schneider (NAE), Cornell University
Peter Weinberger, Google
George Kesidis, Pennsylvania State University and Anomalee Inc.
Moderator: Paul England (NAE), Microsoft (retired)

Moderated Audience Discussion (4:00 – 4:40 PM)
Moderator: Dan Massey, National Science Foundation

5:00 PM
Reception

Day 2 — Tuesday, April 21, 2026

9:00 – 9:15 AM
Day 1 Summary – John Manferdelli (NAE), National Academies Forum on Cyber Resilience

---

9:15 – 11:00 AM
Session 5: Securing Agentic Systems
How do we characterize and secure AI systems that act autonomously, use tools, and interact dynamically with other systems. Topics include orchestration, identity, access control, containment, integration, and resiliency challenges.

Panel Discussion (9:15 – 10:15 AM)
Apostol Vassilev, National Institute of Standards and Technology
George Fletcher, Practical Identity
Ken Huang, Cloud Security Alliance
Pete Bryan, Microsoft
Moderator: Rich Harang, NVIDIA

Moderated Audience Discussion (10:15 – 11:00 AM)
Moderator: Alex Gantman, Qualcomm

11:00 – 11:20 AM
Break

11:20 AM – 1:00 PM
Session 6: Evaluation Frameworks and Infrastructure
How do we assess and benchmark AI security, including metrics, red-teaming methodologies, leaderboards, and evaluation infrastructure?

Panel Discussion (11:20 AM – 12:20 PM)
Lauren Deason, Meta
Maia Hamin, National Institute of Standards and Technology
Manish Parashar, University of Utah
Erin Kenneally, Elchemy
Moderator: Anita Nikolich

Moderated Audience Discussion (12:00 – 1:00 PM)
Moderator: To Be Determined

1:00 PM – 1:40 PM
Lunch

1:40 – 2:30 PM
Session 7: Research Priorities and Actionable Steps

Session moderators will share key takeaways, followed by a community discussion to identify research priorities, collaboration gaps, and strategic investments to strengthen AI security.

Moderated by Angelos Keromytis, Georgia Institute of Technology

---

2:30 PM
Adjourn