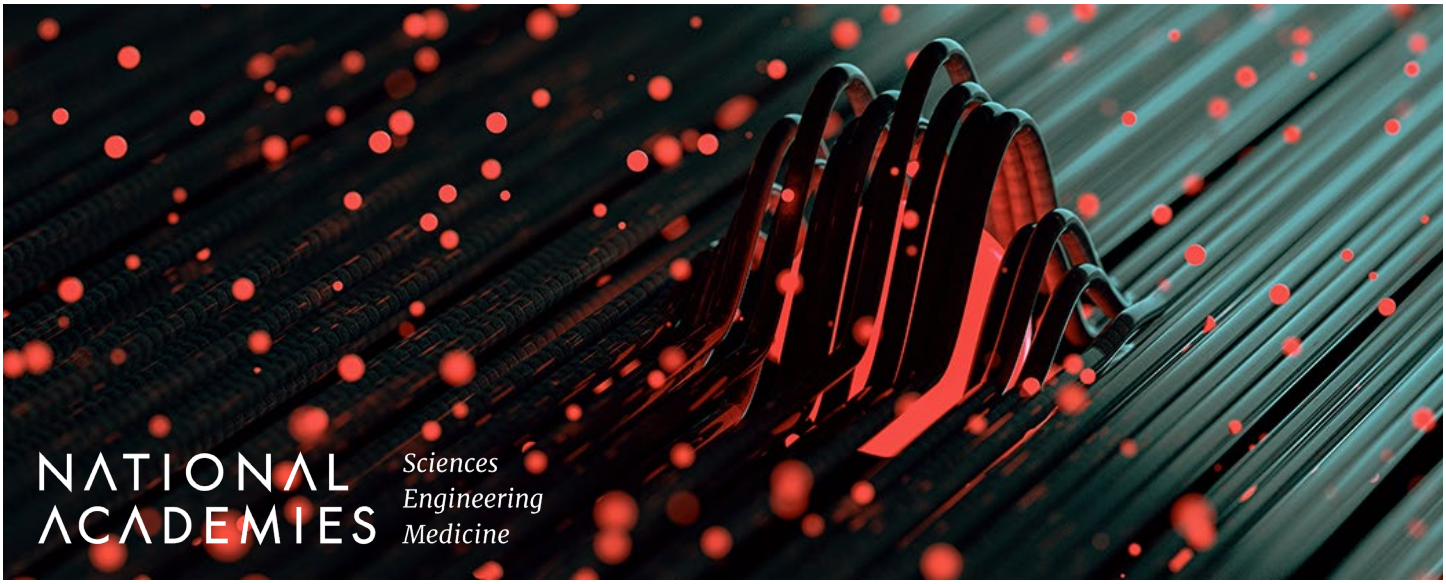

Securing AI Systems: New Challenges and Research Priorities

National Academy of Sciences, 2101 Constitution Ave NW, Washington, DC
APRIL 20 – 21, 2026

Purpose

At the request of the National Science Foundation, the National Academies' Forum on Cyber Resilience will hold a convening to identify research priorities for securing machine learning and AI-enabled systems. The meeting will examine frameworks and concepts for measuring AI security, assess how existing cybersecurity tools and practices can be adapted, identify novel and unique AI security challenges and explore emerging risks in high-impact applications such as scientific research, drug discovery, and financial services.

Bringing together experts in cybersecurity and AI, the meeting will frame research priorities and inform future research programs and strategies for academia, industry, and government aimed at advancing the security and resilience of AI systems.

About the Forum

The Forum on Cyber Resilience serves as an independent, trusted venue in which experts from industry, academia, and government can work collaboratively to explore emerging critical challenges related to the security, trustworthiness, and resilience of the nation's computing, communications systems, and critical infrastructures. Blending expertise in technology, policy, national security, and the law, the Forum convenes senior representatives and serves both as a readily available source of insight and expertise and as a body committed to anticipating and thinking about future trends.

Dear Colleagues,

On behalf of the *National Academies Forum on Cyber Resilience*, we are pleased to welcome you to the national convening, **AI Security: New Challenges and Research Priorities**. This hybrid event will take place Monday–Tuesday, April 20–21, 2026, in person at the National Academies and online.

The goal of this meeting is to identify key issues and inform a research agenda on the timely and important topic of securing AI systems. Accordingly, we intend this to be a working meeting, with discussions that are both in-depth and wide-ranging. Each session will include two components: (a) a panel discussion featuring focused, substantive dialogue among panelists, and (b) an open discussion with participants in the room and online—many of whom bring significant expertise—to surface additional perspectives and ideas.

We encourage your active participation throughout the event and invite you to share your insights, questions, and comments.

Thank you for your support of the National Academies and the Forum on Cyber Resilience. We look forward to a productive and engaging convening.

Sincerely,

John Manferdelli, NAE, Forum Chair

Tho Nguyen, Forum Staff Director

AGENDA

DAY 1 – MONDAY, APRIL 20, 2026

OPEN SESSION

9:00 AM	**Welcome and Opening Remarks** • Ellen Zegura, National Science Foundation • John Manferdelli (NAE), National Academies Forum on Cyber Resilience
9:15 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 AM	BREAK
10:20 AM	**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 to guide research and policy discussions.* 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 PM	LUNCH
1:00 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)

- Malachi 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 PM — BREAK

3:00 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
- Galen Hunt, Microsoft
- George Kesidis, Pennsylvania State University and Anomalee Inc.
- Moderator: Paul England (NAE), Microsoft (retired)

Moderated Audience Discussion (3:00 – 3:40 PM)

- Moderator: Dan Massey, National Science Foundation

5:00 PM — ADJOURN DAY 1

DAY 2 – TUESDAY, APRIL 20, 2026

OPEN SESSION

9:00 AM — **Day 1 Summary**

- John Manferdelli (NAE), National Academies Forum on Cyber Resilience

9:15 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 AM — BREAK

11:20 AM — **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
- Manish Parashar, University of Utah
- Erin Kenneally, Elchemy
- Moderator: Anita Nikolich

Moderated Audience Discussion (12:20 – 1:00 PM)

- Moderator: Athina Markopoulou, University of California, Irvine

1:00 PM — LUNCH

1:40 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.

- Moderator: Angelos Keromytis, Georgia Institute of Technology

2:30 PM — ADJOURN DAY 2

SPEAKER BIOGRAPHIES

Ellen Zegura is a Senior Science and Engineering Advisor in the Office of the Director at the National Science Foundation (NSF) with responsibility for the agency's AI funding strategy and portfolio. She previously served as Acting Assistant Director in the Computer and Information Science and Engineering (CISE) Directorate. She is on loan to NSF from her faculty position at Georgia Tech where she is Fleming Professor of Computer Science. She is a Fellow of the IEEE and the ACM.

John Manferdelli, (NAE), chair of the forum on cyber resilience is an independent consultant. Before that, he was, most recently, Confidential Computing, Incubation Project Leader in the Office of the CTO at VMware. Previously, he was Professor of the Practice and executive director of the Cybersecurity and Privacy Institute at Northeastern University. Immediately prior, Manferdelli was Engineering Director for Production Security Development at Google. Prior to Google, he was a senior principal engineer at Intel Corporation and co-PI (with David Wagner) for the Intel Science and Technology Center for Secure Computing at the University of California at Berkeley. He was also a member of the Information Science and Technology advisory group at DARPA and is a member of the Defense Science Board. Prior to Intel, J Manferdelli was a distinguished engineer at Microsoft and was an affiliate faculty member in computer science at the University of Washington. He was responsible for computer security, cryptography, and systems research, as well as research in quantum computing. At Microsoft, John also worked as a senior researcher, software architect, product unit manager, general manager at Microsoft and was responsible the development of the next-generation secure computing base technologies and the rights management capabilities currently integrated into Windows, for which he was the original architect. He joined Microsoft in February 1995 when it acquired his company, Natural Language Inc., based in Berkeley, California. At Natural Language, Manferdelli was the founder and, at various times, vice president of research and development and CEO. Other positions he has held include staff engineer at TRW Inc., computer scientist and mathematician at Lawrence Livermore National Laboratory, and principal investigator at Bell Labs. He was also an adjunct associate professor at Stevens Institute of Technology. Manferdelli's professional interests include cryptography and cryptographic mathematics, combinatorial mathematics, operating systems, and computer security. He is also a licensed Radio Amateur (AI6IT). Manferdelli has a bachelor's degree in physics from Cooper Union for the Advancement of Science and Art and a PhD in mathematics from the University of California, Berkeley.

Kathleen Fisher is the CEO of ARIA, she joins from RAND, where she established a new centre applying AI and formal methods to cybersecurity challenges. Previously, she served as Director of DARPA's Information Innovation Office (I2O) from May 2022 to September 2025, leading a portfolio of over 50 programmes and 20 programme managers. This was Kathleen's second tour at DARPA. As a programme manager, her High-Assurance Cyber Military Systems (HACMS) programme created provably secure software for military vehicles and was designated DARPA's 2023 "Game Changer" – the most influential

programme from a decade ago. Before joining DARPA, Fisher was a professor and chair of the Computer Science Department at Tufts University and a principal member of the technical staff at AT&T Labs. She holds a PhD in Computer Science from Stanford University and is an AAAS fellow and ACM fellow.

Hyrum Anderson is cofounder and CEO of Security Superintelligence Labs. Previously, he was Senior Director of AI & Security at Cisco, via an acquisition of Robust Intelligence where he served as Chief Technology Officer. Much of his career has been focused on defense and security, having directed research projects at MIT Lincoln Laboratory, Sandia National Laboratories, Mandiant, as Chief Scientist at Endgame (acquired by Elastic), and Principal Architect of Trustworthy Machine Learning at Microsoft. While at Microsoft, he was the architect for Microsoft's AI Red Team inaugural work on production AI systems as chair of the AI Red Team governing board. Hyrum co-founded the Conference on Applied Machine Learning in Information Security (CAMLIS), was appointed to serve on the 2024 National Academies study on Cyber Hard Problems, and has been an advisor to the US and UK governments on AI Safety and Security. He has spoken at numerous academic and industry conferences at the intersection of security and machine learning, including RSA, BlackHat and DEFCON. He has authored over 60 peer-reviewed academic publications, and co-authored the book Not With a Bug, But With a Sticker: Attacks on Machine Learning Systems and What To Do About Them. He received his PhD in Electrical Engineering from University of Washington, with an emphasis on signal processing and machine learning, and BS and MS degrees in Electrical Engineering from Brigham Young University.

Anita Nikolich is a Research Scientist and Director of Research and Technology Innovation at the University of Illinois-Urbana Champaign. She began her career as a military cryptologist and has worked in multiple technical leadership positions in government, industry, and academia.

Jonathan Petit is Director of Engineering at Qualcomm, where he leads the Artificial Intelligence Security research team. He is widely recognized as a pioneer in automotive and AI cybersecurity, with early, influential work exposing vulnerabilities in automated driving systems and AI based perception. At Qualcomm, Dr. Petit's research spans across adversarial machine learning, robustness of perception systems, and security of on device AI. He also plays a central role in shaping global AI security practices through active leadership and coordination in international standardization bodies, including CEN/CENELEC, ISO, SAE, ETSI, NIST, and the FCC. Prior to joining Qualcomm in 2019, he was Senior Director at OnBoard Security, where he worked closely with government agencies, industry partners, standardization organizations, and academic institutions to secure automotive systems. From 2011 to 2014, Dr. Petit served as technical coordinator of the EU funded PRESERVE project, which developed Hardware Security Modules for Vehicle to Everything (V2X) communications. Dr. Petit holds a PhD in Computer Science from Paul Sabatier University (Toulouse III), France.

Securing AI Systems: New Challenges and Research Priorities

Siwei Lyu is a SUNY Distinguished Professor and a SUNY Empire Innovation Professor in the Department of Computer Science and Engineering at the University at Buffalo, State University of New York, USA. He is currently the Director of the Institute for Artificial Intelligence and Data Science (IAD) and the founding Director of the UB Media Forensic Lab (UB MDFL). Before joining UB, Dr. Lyu served in the Department of Computer Science at the University at Albany, State University of New York. He is the founding Director of UAlbany's Computer Vision and Machine Learning Lab (CVML). Dr. Lyu earned his Ph.D. in Computer Science from Dartmouth College in 2005, and his M.S. (2000) and B.S. (1997) degrees in Computer Science and Information Science, respectively, from Peking University, China.
Dr. Lyu's research interests include media forensics, computer vision, and machine learning. He has published over 240 refereed journal and conference papers. His research has been funded by the U.S. National Science Foundation, DARPA, and the U.S. Department of Homeland Security. Dr. Lyu has received numerous awards, including the IEEE Signal Processing Society Best Paper Award (2011), the National Science Foundation CAREER Award (2010), the University at Albany Presidential Award for Excellence in Research and Creative Activities (2017), the SUNY Chancellor's Award for Excellence in Research and Creative Activities (2018), the Google Faculty Research Award (2019), and the IEEE Region 1 Technological Innovation (Academic) Award (2021).
Dr. Lyu has served on the IEEE Signal Processing Society's Information Forensics and Security Technical Committee and held editorial positions with several prestigious journals. He is a Fellow of the IEEE, the IAPR, and the AAIA. He is also a Distinguished Member of the ACM, a Senior Member of the Sigma Xi Society, and a Member of the Omicron Delta Kappa Society.

Giovanni Vigna is a Professor in the Department of Computer Science at the University of California in Santa Barbara, and the director of the NSF AI Institute for Agent-based Cyber Threat Intelligence and Operation (ACTION) at UCSB. He was the CTO and co-founder of Lastline, Inc., a company that provides anti-malware solutions. Lastline was acquired by VMware, Inc., in June 2020, which, in turn, was acquired by Broadcom, Inc., in November 2023. Since then, Dr. Vigna leads the Advance Threat Prevention group in the ANS business unit at Broadcom. His research interests include vulnerability assessment, malware analysis, the underground economy, the security of social networks, voting security and misinformation detection, and the applications of machine learning and artificial intelligence to security problems. He has been the Program Chair of the International Symposium on Recent Advances in Intrusion Detection (RAID 2003), of the ISOC Symposium on Network and Distributed Systems Security (NDSS 2009), of the IEEE Symposium on Security and Privacy (Oakland 2010-2011), and of the ACM Conference on Computer and Communications Security (CCS 2020-2021). He is known for organizing and running, since 2003, a yearly educational Capture The Flag hacking contest, called iCTF, that every year involves dozens of teams around the world. Giovanni Vigna is also the founder of the Shellphish hacking group, who has participated in more DEF CON CTF competitions than any other group in history. Giovanni Vigna received his Ph.D. from Politecnico di Milano, Italy. He is an IEEE Fellow and an ACM Fellow.

Greg Shannon works at the Idaho National Laboratory as a Laboratory Fellow and the Chief Cybersecurity Scientist, leading strategic approaches to protecting critical infrastructure from cyber-physical threats. Previously, he was Chief Scientist for Carnegie Mellon's CERT Division and served at the White House Office of Science and Technology Policy. He currently serves as Chief Science Officer for the Department of Energy's Cybersecurity Manufacturing Innovation Institute. Greg holds a B.S. in Computer Science from Iowa State University and a Ph.D. in Computer Sciences from Purdue University.

Howard Shrobe joined DARPA in April 2022 to develop, execute, and transition programs in computing systems, cyber security, and artificial intelligence (AI). Shrobe joins DARPA from the Massachusetts Institute of Technology's (MIT) Computer Science and Artificial Intelligence Lab (CSAIL), where he leads research creating high-performance, reliable, and secure computing systems and technology. He first joined MIT's research staff in 1978, working initially on computing systems and AI, and later adding cyber security as a focus area. Shrobe has previously worked at DARPA with two tours as a program manager. He has also worked in industry as the technical director and vice president of technology for Symbolics Inc., from 1982 to 1992. He is a Fellow of both the Association for the Advancement of Artificial Intelligence and the American Association for the Advancement of Science. His work has been included in more than 100 publications.

Malachi Jones is a Principal Cybersecurity AI/LLM Researcher and Manager at Microsoft, where he leads efforts to develop autonomous red team agents within Microsoft Security AI. With over 15 years of experience across academia and industry, his work focuses on applying AI/ML and LLMs to cybersecurity and automated reverse engineering. Previously, he conducted advanced research at MITRE and Booz Allen Dark Labs, specializing in machine learning–driven reverse engineering and embedded security, and co-authored U.S. Patent 10,133,871. He also served as an Adjunct Professor at the University of Maryland, College Park, and has taught advanced courses at leading conferences including Black Hat USA and RECON Montreal. Dr. Jones is the founder of Jones Cyber-AI, an initiative dedicated to independent research and education. He holds a Ph.D. from Georgia Tech and continues to drive innovation at the intersection of AI and cybersecurity.

Adam (Ads) Dawson is a Staff AI Security Researcher at Dreadnode, where he red teams frontier AI models for government, military, and tier-1 model providers and builds automated offensive tooling to find vulnerabilities in AI-enabled systems at scale. He is lead author of AIRTBench (arXiv:2506.14682), the first comprehensive benchmark for autonomous AI red teaming, and of "The Automation Advantage in AI Red Teaming" (arXiv:2504.19855). As a contractor to the European AI Office, he authored adversarial risk scenarios and threat actor rubrics for the EU AI Act's Code of Practice. He is the Founding Technical Lead of the OWASP Top 10 for LLM Applications, a technical expert for the MITRE AI Working Group, and author of AI Native LLM Security (Packt Publishing, 2025).

Bo Li is a Research Associate Professor in the Computer Science Department and Data Science Institute at UChicago. Bo's research addresses trustworthy machine learning from

both theoretical and practical aspects and aims to enable reliable machine learning algorithms and systems in the real world, such as safe autonomous vehicles and federated (distributed) learning. She focuses on three interconnected aspects: robustness, privacy, generalization, and their underlying connections. Bo received her Ph.D. in Computer Science from Vanderbilt University in 2016. She was a Postdoctoral Researcher at UC Berkeley 2017-2018 (working with Prof. Dawn Song) and joined the faculty at UIUC in 2018. She been recognized by a long list of notable awards and fellowships for young faculty. She is a Sloan Fellow, MIT Technology Review TR-35 innovator, and recipient of the IJCAI Computers and Thought Award, NSF CAREER, Intel Rising Star Faculty award, Symantec Research Labs Fellowship, Rising Stars in EECS, Research Awards from Amazon, Facebook, and Google, and best paper awards at multiple top machine learning and security conferences. Her research has been featured by major publications and media outlets such as Nature, Wired, New York Times, Fortune, and is on display at the Science Museum in London.

Barton Miller is the Vilas Distinguished Achievement Professor at the University of Wisconsin-Madison. He leads the UW-Madison effort on Trusted CI, the National Science Foundation Cybersecurity Center of Excellence, where he helps direct the software assurance effort. His research interests include software security, in-depth vulnerability assessment, and binary code analysis. From 2012 to 2020 he was chief scientist for the DHS-funded Software Assurance Marketplace (SWAMP). In 1988, Miller founded the field of Fuzz random software testing, which is the foundation of many security and software engineering disciplines. In 1992, Miller (working with his then-student Prof. Jeffrey Hollingsworth) founded the field of dynamic binary code instrumentation and coined the term “dynamic instrumentation”. Miller is a Fellow of the ACM and recipient of the IFIP WG 10.4 Jean-Claude Laprie Award in Dependable Computing and an R&D 100 Award. Miller has been a leader in the software security education. His free and open software security textbook materials (written with Prof. Elisa Heymann) has over 100,000 chapter downloads per year. Miller and Heymann have taught in-depth software security tutorials at conferences, labs, companies on six continents. Miller has been chair of the Institute for Defense Analysis Center for Computing Sciences Program Review Committee; member of the U.S. Federal Aviation Administration (FAA) VECTOR Task Force; member of the U.S. Department of Energy National Nuclear Security Agency Los Alamos and Lawrence Livermore National Labs Security Review Committee; member of the Los Alamos National Laboratory Computing, Communications and Networking Div. Review Committee; member of the U.S. Secret Service Electronic Crimes Task Force (Chicago area); and is currently an advisor to the Wisconsin National Guard and advisor to the Wisconsin Security Research Consortium. Miller is an active participant in NATO’s maritime cybersecurity program.

Window Snyder is a security industry pioneer and CEO and Founder of Thistle Technologies. Ms. Snyder is the former Chief Security Officer at Square and Fastly. She previously spent five years at Apple responsible for security and privacy strategy and features for OS X and iOS. Other roles include Chief Software Security Officer at Intel, Chief Security Something-or-Other at Mozilla and a founder at Matasano, a security services and

product company based in New York. Ms. Snyder is co-author of Threat Modeling, a manual for security architecture analysis in software.

Fred B. Schneider, (NAE), is the Samuel B. Eckert Professor of Computer Science at Cornell University. He joined Cornell's faculty in Fall 1978, having completed a Ph.D. at Stony Brook University and a B.S. in engineering at Cornell in 1975. Dr. Schneider's research has always concerned various aspects of trustworthy systems —systems that will perform as expected, despite failures and attacks. Most recently, his interests have focused on system security. His work characterizing what policies can be enforced with various classes of defenses is widely cited, and it is seen as advancing the nascent science base for security. He is also engaged in research concerning legal and economic measures for improving system trustworthiness. Dr. Schneider was elected a fellow of the American Association for the Advancement of Science (AAAS;1992), the Association of Computing Machinery (1995), and the Institute of Electrical and Electronics Engineers (IEEE; 2008). He was named a professor-at-large at the University of Tromso (Norway) in 1996 and was awarded a doctor of science (honoris causa) by the University of Newcastle-upon-Tyne in 2003 for his work in computer dependability and security. He received the 2012 IEEE Emanuel R. Piore Award for contributions to trustworthy computing through novel approaches to security, fault-tolerance and formal methods for concurrent and distributed systems. The U.S. National Academy of Engineering (NAE) elected Dr. Schneider to membership in 2011, and the Norges Tekniske Vitenskapsakademi (Norwegian Academy of Technological Sciences) named him a foreign member in 2010. He is founding chair of the Academies' Forum on Cyber Resilience.

Galen Hunt leads research in the use of AI and Large Language Models to address the challenges of engineering and maintaining mission critical software systems at scale. Dr. Hunt has worked at Microsoft and Microsoft Research for over 26 years. Most recently, Dr. Hunt founded and led engineering for the Microsoft team responsible for Azure Sphere—with the mission of ensuring that every IoT device on the planet is secure and trustworthy. Dr. Hunt has pioneering work in cloud computing, confidential computing, light-weight virtualization and cross-OS containerization, in porting the CLR to ARM, memory safe OSes and OS kernels. Dr. Hunt holds over 100 US Patents and is a global expert in Operating Systems and technologies—with Best Paper and Test-of-Time Awards from the top OS-related research conferences including ACM SIGOPS, EuroSys, and USENIX OSDI. Dr. Hunt holds a Ph.D. and M.S. in Computer Science from the University of Rochester, a B.S. in Physics from the University of Utah, and A.S. from Utah Tech University. Dr. Hunt is a previous member of the National Academies Forum on Cyber Resilience.

George Kesidis received his MS (1990, neural networks and stochastic optimization) and PhD (1992, performance evaluation and networking) in EECS from UC Berkeley. Following eight years as a professor of ECE at the University of Waterloo, he has been a professor of EE and CSE at the Pennsylvania State University since 2000. His research interests include problems in networking, games, network security, cloud computing, performance evaluation, and AI/ML. His research has previously been supported by over 20 NSF grants,

several DoD grants, and several Cisco gifts. He has been working on problems of secure and trustworthy AI for 10 years under the support of AFOSR, ONR and Cisco (through PSU) and NSF SBIR (through his start-up Anomalee Inc.). His book with D.J. Miller and X. Ziang on Adversarial Learning and Secure AI was published by Cambridge University Press in Fall 2023. He has published in "major" AI and security venues like ICLR, AAAI, NeurIPS and IEEE S&P, and has patents in this field.

Paul England, (NAE), is an entrepreneur and consultant in the areas of confidential and trusted computing, as well as hardware-based security. He recently retired from Microsoft Research, where he was a distinguished engineer and manager of a team of researchers and engineers. Paul led or contributed to many of the computer industry's hardware-based security innovations of the last 20 years. Most notable is the field of Trusted and Confidential Computing: a combination of novel cryptographic operations together with hardware/software environments for secure computation. Trusted Computing primitives are now a feature of most mobile, client, server and cloud computer systems, and the field remains an area of active research. Paul also contributed to the design of the first Trusted Platform Module and led the team that developed the current version. Paul became interested in cyber-resilient systems though his work with NIST in developing NIST SP 800-193 – Platform Firmware Resiliency Guidelines. Based on this, he subsequently worked with hardware partners and standards groups to develop architectures and hardware/software building blocks to enable secure and high assurance recovery of devices that have been compromised by malware or misconfiguration. Dr. England received his Ph.D. in condensed matter physics from Imperial College, London.

Dan Massey is a Program Director at the National Science Foundation's Office of Advanced Cyberinfrastructure where he leads efforts on cybersecurity and networking. Prior to joining NSF, he also served as a Program Director at the Office of the Under Secretary of Defense - Research and Engineering where he led the DoD Operate Through 5G Initiative aims to ensure DoD can securely operate through commercial 5G networks across the globe. As a Program Manager in the Cyber Security Division, Science and Technology Directorate, U.S. Department of Homeland Security (DHS), he developed and managed programs that focused on cyber security for automobiles and other systems that combine the cyber and physical worlds. He participates on several cybersecurity educational advisory boards. Dr. Massey has more than 25 years of research and management experience and is the author over 100 peer reviewed publications on networking and cyber security including co-editor of the DNS Security Standard (RFCs 4033, 4034, and 4035) and early work on Named Data Networking. He has served as the Principal Investigator on research funded by the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation (NSF), the Department of Homeland Security (DHS), and industry. He earned his doctorate in computer science from the University of California, Los Angeles.

Apostol Vassilev is a leading expert in Trustworthy and Responsible AI and Cybersecurity at the National Institute of Standards and Technology (NIST) and the National

Cybersecurity Center of Excellence (NCCoE). His work is characterized by a rare fusion of deep theoretical research and practical advances, driving the development of national and international standards that secure the next generation of AI technologies. Recently, Apostol made a significant contribution to the fundamental understanding of AI safety by extending Gödel's incompleteness theorem to the domain of artificial intelligence. He successfully proved that no finite set of guardrails is universally robust against adaptive adversarial prompts. This landmark result offers a formal mathematical boundary for AI Security and Alignment, suggesting that safety in current and future AI systems cannot be a static achievement but must be a dynamic, evolving process. Beyond his theoretical breakthroughs, Apostol is a practical force in the AI security community. He serves on the Distinguished Expert Review Board of the OWASP GenAI Security Project and is a founding member of the OWASP AI Vulnerability Scoring System project. His research also focuses on Adversarial Machine Learning (AML) and Robust Physical AI for autonomous vehicles. With a Ph.D. in Mathematics from Texas A&M University, Apostol has authored over 60 scientific papers and holds five U.S. patents. His leadership and dedication to public service have earned him numerous accolades, including a medal from the U.S. Department of Commerce. A respected authority and frequent conference speaker, his insights are regularly featured in prominent publications such as the Wall Street Journal, Politico, and Forbes.

George Fletcher is an Identity Standards Architect at Practical Identity, LLC, specializing in Customer Identity and Access Management (CIAM) and B2B identity systems. With over 35 years in software architecture and development, he has spent the last two decades focused on identity standards and protocols, helping shape interoperable approaches to authentication, authorization, and trust across large-scale internet systems. He is an active contributor to global standards efforts, including work within the OpenID Foundation, IETF, and W3C communities, and serves as an editor for key OAuth specifications such as Transaction Tokens and OAuth for First-Party Applications. A former long-time community-elected board member and executive committee participant in the OpenID Foundation, his work spans foundational contributions to OpenID, OAuth, and earlier identity frameworks, reflecting a sustained commitment to building open, scalable identity ecosystems.

Ken Huang is a leading author and expert in AI applications and Agentic AI Security, serving as CEO and Chief AI Officer at DistributedApps.ai. He is Co-Chair of AI Safety groups at the Cloud Security Alliance and the OWASP AIVSS project, and Co-Chair of the AI STR Working Group at the World Digital Technology Academy. He is an EC Council instructor and Adjunct Professor at the University of San Francisco, teaching GenAI Security and Agentic AI security for data scientists respectively. He co authored OWASP's Top 10 for LLM Applications and contributes to the NIST Generative AI Public Working Group. His books are published by Springer, Cambridge, Wiley, Packt, and China Machine Press, including Securing AI Agents, LLM Design Patterns, Generative AI Security, Agentic AI Theories and Practices, Beyond AIand The Handbook for Chief AI Officers, A frequent global speaker, he engages at major technology and policy forums.

Pete Bryan is a Principal AI Security Researcher at Microsoft, where he works on the AI Red Team to proactively identify and mitigate security and safety risks in advanced AI systems. With over a decade of experience in security research, Pete focuses on red teaming large language models and agentic AI systems, exploring failure modes such as misuse, loss of control, and emergent security risks across real-world deployments. His background spans threat intelligence, cloud security, and offensive security, including prior leadership of the Sentinel Threat Research team and work within the Microsoft Threat Intelligence Center. Pete is also an active contributor to the security community, developing open-source tools and sharing practical approaches to AI red teaming and evaluation.

Rich Harang is a Distinguished Security Architect at NVIDIA, specializing in ML/AI systems, with over a decade of experience at the intersection of computer security, machine learning, and privacy. He received his PhD in Statistics from the University of California Santa Barbara in 2010. Prior to joining NVIDIA, he led the Algorithms Research team at Duo, led research on using machine learning models to detect malicious software, scripts, and web content at Sophos AI, and worked as a Team Lead at the US Army Research Laboratory. His research interests include adversarial machine learning, addressing bias and uncertainty in machine learning, and ways to use machine learning to support human analysis. Richard's work has been presented at USENIX, BlackHat, IEEE S&P workshops, and DEF CON AI Village, among others, and has also been featured in The Register and KrebsOnSecurity.

Alex Gantman is a product security executive with over 20 years of experience leading global organizations to deliver secure and reliable products at scale. As Vice President of Engineering for Qualcomm Technologies Inc., Gantman is responsible for making billions of Qualcomm-powered connected products secure and reliable against attacks. He leads a global team designing, implementing, and commercializing security capabilities across dozens of product lines spanning multiple industry verticals, including Mobile, Compute, Automotive, and IoT. Gantman has led the establishment and evolution of a broad-scale product security practice at Qualcomm, covering thousands of products, tens of millions of lines of code, and tens of thousands of engineers across the globe. He is a founding organizer of the Qualcomm Product Security Initiative (2006) and the Qualcomm Product Security Summit -- a premier industry conference focused on security of connected devices. He holds over 50 patents and is a recognized subject matter expert in hardware, software, and systems security across a wide range of domains. Gantman received Bachelor's (1998) and Master's (2001) degrees in Computer Science from the University of California, San Diego.

Lauren Deason is a software engineer at Meta focused on evaluating the cybersecurity risks and opportunities introduced by generative AI systems. She has previously worked on developing machine learning applications for security use cases including malware and network intrusion detection. Lauren serves on the governing board of CAMLIS and holds a PhD in Economics from UMD College Park, a Masters degree in Mathematics from UC Berkeley, and a BS in Applied Math from UVA.

Manish Parashar is the inaugural Chief AI Officer at the University of Utah. He is also Executive Director of the Scientific Computing and Imaging (SCI) Institute, and Presidential Professor in the Kalhert School of Computing. He leads the University's One-U Responsible AI Initiative. Manish's expertise is in high-performance parallel and distributed computing, large-scale data management, and cyberinfrastructure. His research has enabled new insights across multiple science domains. Manish has received several awards for his research and leadership, including the 2023 IEEE Computer Society Sidney Fernbach Award, the 2024 CRA Distinguished Service Award, and the 2025 ACM Distinguished Service Award. Manish is a Fellow of AAAS, ACM, and IEEE.

Erin Kenneally is a scientist and licensed attorney by training, and technology risk innovation leader by trade. Erin is the Founder and CEO of Elchemy, building solutions in AI Risk Insurance. She previously directed programs in the Cyber Security Division for the U.S. Dept of Homeland Security, Science & Technology Directorate addressing Cybersecurity Data Infrastructure, Cyber Risk Economics, and Technology Ethics & Data Privacy. Kenneally was also previously the Global Director for Cyber Insurance at SentinelOne, and Director of Cyber Risk Strategy at Guidewire-Cyence Risk Analytics. She served as Technology Law Specialist at the Center for Internet Data Analysis (CAIDA) and Center for Evidence-based Security Research (CESR) at the UC San Diego. Kenneally has published, presented, advised, and built translational solutions at the crossroads of technology risk, law, policy, and ethics. She holds Juris Doctorate and Masters of Forensic Sciences degrees from Syracuse University, the George Washington University, and Cleveland-Marshall College of Law.

Athina Markopoulou is a Professor of Electrical Engineering and Computer Science (EECS) at UC Irvine. She currently serves as the director of the California Institute for Telecommunications and Information Technology (Calit2), as well as the director of ProperData (an NSF SaTC Frontiers project on "Protecting Personal Data on the Internet") and of ProperAI (a new multidisciplinary "Engineering+Society Institute on AI") in the Samueli School of Engineering at UC Irvine. Her research interests include privacy-enhancing technologies, AI safety and tech policy, in a range of application domains including social networks, web, mobile, virtual and augmented reality, voice assistants and LLM-powered services.

Angelos Keromytis is the John Weitnauer Technology Transfer Endowed Chair Professor and Georgia Research Alliance Eminent Scholar with the School of Electrical and Computer Engineering at the Georgia Institute of Technology (2018-present). Prior to Georgia Tech, he was with Columbia University (2001-2017). His research interests include systems and network security, and applied cryptography. He is an ACM Fellow, with the citation "for contributions to the theory and practice of systems and network security", and an IEEE Fellow in the Class of 2018, with the citation "for contributions to network security systems". Prior to Georgia Tech, he served as Program Manager with DARPA/I2O (2014-2018, when he received the Superior Public Service Medal) and as Program Director with

NSF/CISE/CNS (2013-2014). He received his Ph.D. (2001) and M.Sc. (1997) from the University of Pennsylvania, and his B.Sc. (1996) from the University of Crete, all in Computer Science. He has co-founded 4 technology startups, currently serving as President and Chairman of the Board for 2 of them.

FORUM ON CYBER RESILIENCE

The **Forum on Cyber Resilience** facilitates and enhances the exchange of ideas among scientists, practitioners, and policy makers concerned with urgent and important issues related to the resilience of the nation's computing and communications systems. Resilience is meant to encompass not only security in the face of attacks, resistance to degradation, and the ability to recover from adverse events but also the capacity for innovation and adaptation and the ability to absorb rapid technological disruption in a way that reflects the values--such as privacy--and needs of the infrastructure's many stakeholders. The focus of the Forum is accordingly expected to be two-pronged: traditional notions of cybersecurity, trustworthiness, and reliability of large-scale systems blended with considerations of the need to afford opportunities for innovation and respect for stakeholder values, needs, and priorities.

Members

John Manferdelli, NAE
Chair
Independent Consultant

Heather Lynn Adkins
VP, Security Engineering
Google

Hyrum Anderson
Director of AI & Security
Cisco

Steven Bellovin, NAE
Percy K. and Vidal L. W. Hudson Professor of Computer Science
Columbia University

Tom Berson, NAE
Chief Security Advisor
Salesforce

Nadya T. Bliss
Executive Director, Global Security Initiative
Arizona State University

Tim Booher
Senior Vice President of Special Projects
Leidos

Byron Cook
Professor of Computer Science
University College London
Vice President and Distinguished Scientist
Amazon

Srini Devadas
Webster Professor of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Paul England, NAE
Entrepreneur and Consultant

Jeremy Epstein
Co-director
Georgia Tech Research Institute / PNNL ICARIS Research Center

Kathleen Fisher
CEO
Advanced Research and Innovation Agency (ARIA)

Alex Gantman
Vice President of Engineering
Qualcomm Technologies, Inc.

James Gosler, NAE
Senior Fellow
Johns Hopkins Applied Physics Laboratory

NOTES