Speakers, Presenters, Panelists

Opening Remarks: Kevin Stine, Director, ITL, NIST

Session 0: Sketching the Problem Space, M S Raunak and Peter Cihon

Keynote: Neil Thompson, Director of MIT FutureTech (Virtual)

Title: **A Global Perspective on AI Incident Management: Risks, Responses, and Gaps**

Abstract: AI systems are creating new kinds of incidents, while also changing the frequency, scale, and character of familiar risks such as fraud, cyberattacks, privacy harm, discrimination, and system failures. This keynote will give a global perspective on how organizations identify, prioritizing, and responding to AI risks and incidents. Drawing on work from MIT FutureTech and the MIT AI Risk Initiative, Prof. Thompson will discuss which AI risks appear most severe, how organizations are currently responding, where existing incident response practices are useful, and where they may break down. The talk will also highlight gaps in today's guidance, frameworks, and best practices, and suggest areas where better standards and coordination may be needed.

Bio: Dr. Neil Thompson is Director of MIT FutureTech and a Principal Research Scientist at MIT's Computer Science and Artificial Intelligence Laboratory and the Initiative on the Digital Economy at the MIT Sloan School of Management. His research examines the foundations of progress in computing and AI, and how these technologies shape scientific discovery, economic prosperity, policy, and industry. Since founding MIT FutureTech in 2019, he has grown the group to more than 110 researchers and attracted over $25 million in funding, with research partnerships including Google, IBM, Amazon, Accenture, Microsoft, and Los Alamos National Laboratory. Dr. Thompson's work has been cited over 3,000 times and he has presented to congressional staff, the Federal Reserve, the Pentagon, the Department of Commerce, the Department of Energy, and other policy and industry leaders. He holds a PhD from UC Berkeley, master's degrees from UC Berkeley and the London School of Economics, and bachelor's degrees from Queen's University.

Session 1: AI Under Attack: Incident Response for AI Systems

Session Chair: Craig Schlenoff, Chief, AI Research, Measurement, and Standards Division

Speaker: Kyriakos Lambros, Director of AI Standards and Governance, Zenity

Title: AI Under Attack: Where your IR Lifecycle Cracks for Agentic Systems

Abstract: Most AI incident response advice falls into one of two camps. Either AI is just IT and your existing plan covers it, or AI requires a brand-new model, and you should start over. Both are wrong. The classical IR lifecycle still applies for AI systems. Each phase has one specific crack that agentic AI widens until current plans stop working. This session walks the lifecycle and names the four cracks: detection signals that move into the model's reasoning layer, containment that has to operate on authorization scope rather than network segments, recovery that must produce governance evidence as well as patched artifacts, and forensics that collapse without a reproducibility envelope captured before the incident. Each crack is anchored in a publicly documented incident. The audience leaves with three operational changes to make by next quarter.

Bio: Kyriakos "Rock" Lambros serves as Director of AI Standards and Governance at Zenity, where he leads AI security standards work, policy advocacy, and cybersecurity governance consulting for Fortune 500 clients. A recovering CISO with 30 years of experience, Rock has built and led security programs for multibillion dollar organizations across energy, eCommerce, government, banking, and manufacturing. He co-leads the OWASP Top 10 for LLMs, serves as a core team member of the OWASP GenAI Security Project and OWASP Agentic Security Initiative, and is a project author for the OWASP AI Exchange. He is a contributor to CoSAI and AIUC-1, holds a Distinguished Fellow position with the Enterprise Risk Quantification Institute (ERQI), and is a certified AI Governance Professional (AIGP) and Qualified Technology Expert (QTE). Rock co-authored “The CISO Evolution: Business Knowledge for Cybersecurity Executives and publishes “RockCyber Musings” a bi-weekly newsletter distilling AI security developments for executive leaders. He is a frequent keynote speaker, panelist, and industry commentator on agentic AI security, AI governance, and enterprise risk. Rock holds an MBA in Finance and Entrepreneurship from Arizona State University, a B.S. in Management Information Systems from the University of Nevada Las Vegas and is currently pursuing an M.S. in Applied Data Science and AI at the University of Denver.

Speaker: James Perry, Global Vice President and Head of Incident Response, CrowdStrike

Title: Investigating Artificial Intelligence Attacks: A Framework

Abstract: As artificial intelligence systems become deeply embedded in enterprise environments, incident responders face a fundamentally new challenge: traditional IR methodologies were not designed for the ephemeral, often unlogged, and architecturally complex nature of AI infrastructure. This presentation provides a brief overview of a draft technical framework for investigating attacks against AI systems, from hosted LLM platforms to self-hosted inference servers, agentic frameworks, and the connective tissue between them.

Bio: As Global Vice President and Head of Incident Response at CrowdStrike, James Perry leads the worldwide team responding to sophisticated cyber threats, including high-profile breaches that have shaped the cybersecurity landscape. Since joining in 2016, he has advanced through leadership positions including Head of Incident Response for America and Senior Director of Consulting.

His role provides unique visibility into emerging threat trends, enabling him to advise C-suite executives during active incidents, deliver comprehensive forensics, lead adversary hunting initiatives, and develop remediation strategies for clients across commercial and government sectors.

Previously, he served as Chief Technologist at a major consulting firm, co-leading Incident Response services, designing security operations centers, and developing enterprise cyber defense strategies.

Mr. Perry holds an M.S. in Information Systems and Technology from Johns Hopkins University and a B.S. in Systems Engineering from the University of Virginia. He is based in Washington, D.C.

Speaker: Robert Saul, General Manager of Customer Incident Response Team (CIRT), Amazon Web Services (AWS)

Title: The New IR Landscape: Defending AI Assets and Models

Abstract: Organizations are deploying AI systems faster than they are updating their incident response plans to account for them. Traditional IR frameworks, including NIST SP 800-61, provide a proven four-phase lifecycle (Preparation, Detection and Analysis, Containment/Eradication/Recovery, Post-Incident Activity), but each phase requires rethinking when the asset under attack is a model, a training pipeline, an inference endpoint, or an autonomous agent.

This presentation examines how the IR planning cycle changes when AI infrastructure enters scope. In the Preparation phase, responders face new asset classes to inventory (models, training data, vector stores, agent tool chains) and new telemetry to pre-position (prompt logs, model drift baselines, embedding access patterns). Detection and Analysis demand familiarity with failure modes that have no equivalent in traditional IT: data poisoning, model inversion, prompt injection, and silent drift in model behavior that may not trigger conventional alerting. Containment decisions become more complex when you must choose between killing an agent mid-workflow, rolling back a model version, or isolating a RAG knowledge base while downstream systems depend on it. Recovery requires coordination with ML engineers and data scientists who own model lineage but may never have participated in an incident response before.

The core message is that responding to incidents involving AI systems is not a new discipline. It is the same planning cycle applied to a new set of assets, stakeholders, and failure modes. The organizations that will respond well are the ones planning now: inventorying their AI assets, identifying the log sources they will need before an incident occurs, pre-building containment playbooks for their model serving infrastructure, and pulling ML engineers into tabletop exercises. Planning is the differentiator.

Bio: With nearly 30 years of experience in security and network engineering, Robert Saul currently serves as the General Manager of the Customer Incident Response Team (CIRT) at Amazon Web Services (AWS). In this role, he establishes the strategy and measures the operations of a global network of security incident responders dedicated to supporting the investigation of security events that occur on the customer side of the shared responsibility model. The CIRT's focus is on coordination, analysis, mitigation, and recovery from cyber incidents, ensuring AWS customers receive top-tier support to accomplish their business objectives.

Robert is incredibly proud to be a part of this team of talented professionals. Their collective experience, dedication, and expertise allow the team to provide invaluable guidance, often in high-pressure situations where time is critical. He is continually inspired by their commitment to excellence in incident response and their unwavering customer obsession.

Prior to AWS, he engineered and secured tactical communications platforms for the defense and intelligence sectors. This background provides him with a unique perspective on the evolving landscape of cyber threats and the importance of robust incident response strategies.

Session 2: Misuse and Malfunction: Understanding Emerging AI-Induced Incidents

Session Chair: Jon Boyens, Chief, Computer Security Division, ITL, NIST

Speaker: Sean McGregor, Executive Director, AI Incident Database (Virtual)

Title: The Next 6,173 AI Incident Reports

Abstract: AI Incidents in isolation offer little insight into the probability of incident recurrence. Advances in data collection and modeling are required to advance industrial safety and security cultures to achieve reliable measurements of "risk." In this presentation, Dr. McGregor will share notes derived from cataloging thousands of AI incident reports with incidentdatabase.ai with the goal of moving beyond spectacle to provide measured insight.

Bio: Sean McGregor launched the AI Incident Database after earning his 2017 PhD in computer science from Oregon State University. He since went on to start the neural accelerator company Syntiant as a founding engineer, lead the assessment of teams competing for the AI XPRIZE, start the Digital Safety Research Institute at the UL Research Institutes, and most recently cofound the AI Verification and Evaluation Research Institute (AVERI).

Sean's open source development work has earned media attention in Time, the Atlantic, Der Spiegel, and Wired, among others, while his technical publications have appeared in a variety of machine learning, human-computer interaction, ethics, and application-centered proceedings.

Dr. McGregor currently serves as executive director for the AI Incident Database, lead of the ML Commons Agentic Workstream, and co-founder of AVERI. These efforts thematically align with his fellowship at the Berkman Klein Center of Harvard University.

Speaker: Chris Meserole, Director of the Frontier Model Forum

Title: Information-sharing, Incident Reporting, and Incident Response for Frontier AI

Abstract: Information-sharing, incident reporting, and incident response are distinct, complementary tools for establishing a trusted AI ecosystem. Yet they are often treated or referred to interchangeably, in ways that risk producing mechanisms that are poorly suited to their intended function. This talk will walk through the different purposes of information-sharing, incident reporting, and incident response and the role the FMF plays in information-sharing specifically.

Bio: Chris Meserole is the executive director of the Frontier Model Forum, an industry-supported non-profit dedicated to advancing frontier AI safety and security. He works closely with experts at leading AI developers to advance the safe development and deployment of the most advanced general-purpose AI systems. Under his tenure, the FMF has established a first-of-its-kind information-sharing mechanism for frontier AI threats, allocated more than $10 million in funding for leading-edge AI safety research and evaluations, and outlined early and emerging best practices for frontier AI risk management. He also chairs the expert group tasked with drafting a US standard for frontier AI frameworks within INCITS AI. Meserole previously served as director of the Brookings AI and Emerging Technology Initiative.

Speaker: Jolanda Kumakaw, Crisis Response, Google (Virtual)

Bio (Incomplete)**:** Over a decade in big tech leading crisis response and major launch moments at Apple, Meta, & Google. Helped streamlining Apple App Store operations to reduce app reviews from weeks to hours. Mobilized Meta company-wide response to Cambridge Analytica, FTC's multi-billion dollar fine, COVID (business continuity planning), and year 1 of the Privacy program and FTC audit. Currently working on crisis and escalation management for Google's GenAI products, major world events impacting the platform, and ensuring safety of users.

Mini-talks: Landscape of Existing Guidance

Session Chair: Harold Booth, CSD, ITL, NIST

Mini-talk Presenter: Alex Nelson, NIST

Alex Nelson is a Computer Scientist at NIST in the Computer Security Division of the Information Technology Laboratory. His research is in digital forensics and ontology-driven interoperability. Dr. Nelson is one of the authors of 800-61r3.

Mini-talk Presenter: Katerina Megas, NIST

Katerina Megas is the Program Manager for the NIST Cybersecurity for Internet of Things (IoT) and the AI and Cybersecurity programs. With a Masters in Information Systems, PMP and ScrumMaster certifications, she has over 25 years of experience developing and leading technology and corporate strategies for organizations in both the private and public sectors. She has over 25 years of experience working in a wide range of technology areas ranging from organizations' development and execution of technology strategies to achieving their CMMI certification. She loves traveling and appreciates her wonderful

colleagues who cover for her at work while she piles her family into a minivan taking road trips across Europe and the U.S. in search of the non-touristy experience.

Mini-talk Presenter: Andrea Brennen, IQT (Virtual)
Andrea Brennen is a designer, technologist & recovering architect who lives in Boston with her husband. She helps start-ups tell better stories, builds tools to help people make sense of data, and is generally interested in how we interact with machines that engineers tell us are "intelligent."

Andrea is VP of Design & Data Visualization at IQT Labs, where she works on visualizing uncertainty, explaining AI, and communicating technical concepts in clear and intuitive ways. Previously, at MIT Lincoln Laboratory, she managed the development of new software tools that helped researchers analyze data about communication networks.

Andrea has an M.Arch in Architectural Design from MIT and bachelor's degrees in Mathematics and Studio Art from Grinnell College. Her design work has been exhibited at the Venice Biennale, the Rotterdam Biennale, the Sao Paolo Biennale, the Canadian Center for Architecture and published internationally. She spends nearly all of her free time rock climbing and hopes to climb V10 someday.

Mini-talk Presenter: Michael Garris, MITRE

Mike Garris serves as the Chief Artificial Intelligence Officer (CAIO) at GBS Solutions, where he leads AI initiatives, research, governance, and practical adoption to help federal clients meet their mission objectives. In this role, he focuses on strategic planning and fostering ethical, secure approaches to emerging technologies.

Previously, he was the Senior Principal Technical Advisor to MITRE Labs' AI and Autonomy Innovation Center, where he guided research roadmaps, external partnerships, and innovation strategies. Before MITRE, Mike dedicated over 34 years as a federal civil servant at the National Institute of Standards and Technology (NIST). During his tenure, he advanced to the highest levels of U.S. government science and technology policy, serving as an advisor within the Office of Science and Technology Policy and the National Security Commission on Artificial Intelligence.

His extensive expertise spans AI, biometrics, machine learning, testing and evaluation, and international standards development. Mike holds a B.S. in Computer Science from Clarion

Mini-talk Presenter: Nimura Kazuaki (Pre-recorded)

Kazuaki Nimura is Chief Researcher and Technology Lead at the Japan AI Safety Institute (J-AISI), where he focuses on AI safety and AI security. He has contributed to drafting official public documents for J-AISI, as well as to the crosswalk between the NIST AI RMF and the Japan AI Guidelines for Business.

Mini-talk Presenter: Violet Turri, CMU (Virtual)

Title: Exploring the State of AI Incident Documentation Practices

Abstract: This talk discusses the current state of AI incident documentation practices and identifies key gaps that limit the ability of AI engineers to learn from and prevent system failures. The presentation walks through a landscape analysis of existing AI incident documentation methods and compares them with incident documentation approaches in more established fields, including aviation and cybersecurity. Our work highlights several challenges in current practices, including a lack of phenotypical information about systems and a reliance on second-hand accounts of system failures. The talk concludes with a set of policy recommendations, including establishing a federal database with mandatory disclosure requirements, developing confidential submission systems, and implementing more proactive documentation approaches.

Bio: Violet Turri is an Engineering Team Lead within the AI Division at the Carnegie Mellon University Software Engineering Institute. She has more than five years of experience in public-sector innovation, R&D software development, and human-centered design. Her recent work explores how agentic AI systems can drive innovation and advance mission outcomes across government and nonprofit organizations.

Panel: Roles and Responsibilities in the Era of AI Agents

Moderator: Sanjay Rekhi, CSD, ITL, NIST

Panelist: Ian Reynolds, Hugging Face

Ian Reynolds works on AI public policy at Hugging Face. He previously served as the Post-Doctoral Futures Fellow at the Center for Strategic and International Studies (CSIS) Futures Lab. He earned his PhD in International Relations from American University's School of International Service in 2024. Dr. Reynolds has held several prestigious research positions, including serving as a Research Fellow at Stanford University's Center for International Security and Cooperation (CISAC) and the Institute for Human-Centered AI (HAI) from 2022 to 2023.

His research focuses on Artificial intelligence and national security decision-making, the intersection of science and politics, and the relationship between digital technologies and international security.

His doctoral dissertation examined the history and cultural politics of AI, specifically focusing on its relationship with military command and control practices in the United States.

Panelist: Apostol Vassilev, NIST

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*.

Panelist: Mikel Rodriguez, MITRE Fellow (Virtual)

Mikel Rodriguez, Ph.D. is a recognized leader in the field of artificial intelligence. He is a MITRE Fellow. Dr. Rodriguez has over two decades of experience bringing transformative, AI-enabled solutions to complex challenges in both the private and public sectors. At Google DeepMind, he co-founded and led a team focused on research to enable the secure deployment of frontier AI systems, particularly for high-stakes applications. As a core team member of Google's Project Gemini, he helped develop and deploy foundational models that impacted products used by

billions. He has also served as a special government employee for the Defense Innovation Unit, where he helped the U.S. government identify and adopt emerging technologies, and previously led MITRE's AI and Autonomy Innovation Center. Dr. Rodriguez works with hundreds of data scientists and AI engineers who provide deep expertise across the executive branch, utilizing resources like MITRE's Federal AI Sandbox, a supercomputer built in partnership with NVIDIA. The Sandbox can train large language models and tackle government-specific AI tasks, such as enhancing military command and control, securing infrastructure, and reducing fraud.

Panelist: Jim Reavis

Co-founder and Chief Executive Officer, CSA

For over 30 years, Jim Reavis has worked in cybersecurity industry as an entrepreneur, writer, speaker, technologist and business strategist. Jim's innovative thinking about emerging trends have been published and presented widely throughout the industry and have influenced many. Jim launched Cloud Security Alliance (CSA) in 2009 and has led its global growth and position as among the most vital cybersecurity communities worldwide. Under Jim's stewardship, CSA has made tremendous strides in security best practices for cloud, quantum computing, IoT, blockchain, zero trust and now generative AI with CSA's AI Safety Initiative.
Jim received a B.A. in Business Administration / Computer Science from Western Washington University in 1987 and formerly served on WWU's alumni board. Jim was recognized as a WWU Distinguished Alumnus in 2015. In 2016, Jim was inducted into the Information Systems Security Association (ISSA) Hall of Fame