

Cybersecurity at MIT Sloan (CAMS)

When AI Goes Rogue: What Emerging AI Incidents Reveal About The Next Generation Of Cyber Risk

As organizations increasingly deploy AI agents across business and cybersecurity functions, questions surrounding governance, accountability, and risk management are becoming more urgent. This summer, those concerns came into focus after reports that an OpenAI agent escaped its testing environment and accessed external systems while pursuing its assigned objective. The incident has intensified discussion about how organizations can safely deploy increasingly autonomous AI systems.

For Michael Siegel, Director of Cybersecurity at MIT Sloan, the event highlights the scale of the challenge ahead: "The OpenAI incident was one limited experiment that was supposed to be tightly controlled. It totally failed in that area. And we're talking about deploying millions and millions of agents to do work for us."

As organizations race to adopt AI, leaders must consider not only what these systems can do, but also how they can be monitored, governed, and controlled.

The difficulty lies in the fact that advanced AI systems often behave differently than traditional software. Their value comes from their ability to identify novel solutions and adapt to changing circumstances, but those same capabilities can produce unexpected outcomes.

As Stuart Madnick noted: "It's not always easy to anticipate all the things (an AI system) might do, because it can be amazingly creative."

The questions raised by autonomous AI systems resonate throughout this quarter's CAMS research, from cyber risk governance and threat intelligence sharing to AI software security and insurance models for AI-related failures. Together, these efforts reflect a common goal: helping organizations manage emerging risks while realizing the benefits of rapidly advancing AI technologies.

Why This Matters for CAMS Members

- AI systems are becoming increasingly autonomous and capable of acting across complex environments.
- Governance and accountability frameworks are still evolving alongside these technologies.
- Cybersecurity, liability, and organizational risk are becoming increasingly interconnected in the age of AI.

Upcoming Member Event | October 7

Join MIT researchers Ilya Kabanov and Shaharyar Khan for an exclusive CAMS member discussion on the emerging cybersecurity challenges of multi-agent AI systems. Learn why agent interactions can create risks that traditional governance approaches may miss and explore new strategies for managing collective agentic intelligence. See page 8 for full details.



Why Insurance Must Catch Up To AI

As organizations move AI systems from experimentation into critical business operations, an important question is emerging: **What happens when AI fails?** Traditional cyber insurance and professional liability policies were not designed to address many of the unique risks associated with AI, including autonomous decision-

making, model behavior, and AI-driven operational losses.

In a recent *Forbes India* article, CAMS Research Scientist Ranjan Pal and his co-author argue that existing insurance frameworks must evolve alongside the technology they are intended to protect. The authors contend that insurers, regulators, and organizations will need new approaches to evaluating and managing AI-related risk as adoption accelerates across industries.



The article highlights how AI failures differ from many traditional cyber incidents. Questions surrounding accountability, model provenance, governance practices, and system performance can make it difficult to determine responsibility when losses occur. As AI systems are increasingly deployed in customer-facing services, business operations, healthcare, finance, and cybersecurity, organizations may need new mechanisms to evaluate risk and demonstrate responsible AI practices.

Among the recommendations discussed are the development of AI incident registries, updated AI risk-scenario catalogs, stronger governance incentives, and insurance structures that encourage organizations to implement and document responsible AI controls.

In the Spotlight: Ranjan Pal Featured on the Cover of iTech Magazine

CAMS Research Scientist Ranjan Pal was featured on the cover of the 2026 annual edition of iTech Magazine, a distinction that highlights his leadership and

contributions to cybersecurity, cyber resilience, cyber insurance, and systemic cyber risk research.

The feature included an exclusive interview exploring some of today's most pressing cybersecurity challenges, including cyber insurance, systemic cyber risk, supply chain resilience, agentic AI, critical infrastructure protection, and the future of cyber risk management. The interview also highlighted Ranjan's research leadership through Cybersecurity at MIT Sloan and his contributions to global cyber resilience initiatives.

Further underscoring this recognition, Ranjan was invited to participate in an elite cybersecurity panel moderated by a KPMG partner, where industry leaders discussed offensive and defensive cybersecurity strategies and emerging threats facing organizations worldwide.

Being selected as the principal cover feature and interviewee reflects the growing influence of Ranjan's research and thought leadership, while helping bring CAMS research and perspectives to a broader international audience.



Advancing the Conversation on Quantum Risk

CAMS Research Scientist Ranjan Pal was invited to deliver a research talk at [ONDC \(Open Network for Digital Commerce\) India](#), one of the country's most ambitious digital public infrastructure initiatives. His presentation, titled "The Quantum Reckoning: Are E-Commerce Enterprises Ready for the Post-Quantum Age?", examined how the foundations of trust, cryptographic security, and business continuity in digital commerce must evolve as organizations prepare for a post-quantum future.

The talk addressed a challenge of growing strategic importance for businesses worldwide: the potential impact of quantum computing on the cryptographic systems that underpin digital commerce. Moving beyond the identification of technological risks, Ranjan presented practical approaches for building quantum readiness and explored the emerging convergence of quantum risk, artificial intelligence, and digital commerce ecosystems.

ONDC serves as a transformative platform at the intersection of technology, commerce, public policy, entrepreneurship, and digital inclusion. By bringing together policymakers, industry leaders, technologists, startups, and ecosystem participants, ONDC is helping shape the future of digital commerce in India at national scale.

Ranjan's invitation to speak at ONDC reflects the growing relevance of CAMS research to real-world business and policy challenges. The engagement provided an opportunity to share research insights with leaders actively designing the next generation of digital commerce infrastructure while contributing to discussions on how organizations can prepare for emerging cybersecurity and resilience challenges in the quantum era.

"The transition to the post-quantum age is not simply a technology upgrade. It is a strategic business challenge that requires organizations to rethink trust, resilience, and long-term digital security."

For more information about ONDC, visit the [Open Network for Digital Commerce](#) website.

Stuart Madnick Named IACIS 2026 Computer Educator of the Year

CAMS Founding Director Stuart Madnick has been selected as the International Association for Computer Information Systems (IACIS) 2026 Computer Educator of the Year, recognizing his outstanding contributions to information systems education, scholarship, mentorship, and service to the profession. The award committee noted the highly competitive field of nominees and highlighted Madnick's impressive impact on students and the broader information systems community.

Madnick is the John Norris Maguire Professor of Information Technology, Emeritus, at the MIT Sloan School of Management, Professor of Engineering Systems, and Founding Director of Cybersecurity at MIT Sloan (CAMS). Over more than five decades at MIT, he has advanced research and education in cybersecurity, information systems, software engineering, data quality, and the strategic use of information technology.

His influence extends well beyond MIT. Madnick has taught generations of students, mentored numerous doctoral candidates who went on to successful academic careers, served as a visiting professor at 13 universities, and helped shape information systems education through widely adopted teaching methods, publications, and educational materials.

This recognition celebrates a career dedicated to advancing knowledge while preparing future generations of technology leaders, researchers, and practitioners to address complex digital and cybersecurity challenges.

Research Achievement: AI and Cybersecurity Governance Converge

A new paper by Angelica Marotta and Stuart Madnick, "An AI-Cybersecurity Taxonomy for Regulating Interconnected Risks," has earned both conference and publication recognition. The research will be presented at the 66th Annual [International Association for Computer Information Systems \(IACIS\) Conference](#) in Pensacola

Beach, Florida, this October and has also been selected for publication in [Issues in Information Systems](#).

The research explores the growing convergence between cybersecurity and AI governance as organizations face increasingly interconnected technological risks. The study introduces an AI-Cybersecurity Taxonomy that identifies common governance functions across regulatory frameworks and examines how policymakers are adapting cybersecurity requirements for AI-enabled systems.

In addition to presenting their research, Marotta and Madnick served as reviewers for the conference, contributing to the peer-review process and helping shape the final research program.

Strengthening Cyber Risk Governance

New MIS Quarterly Executive Publication Explores How Organizations Can Better Prepare for Cyber Risk

As organizations confront increasingly complex cyber and AI-related risks, effective governance is becoming just as important as technical defenses. A new article by Michael Siegel, Ranjan Pal, and Sander Zeijlemaker, published in MIS Quarterly Executive, examines how simulation exercises can strengthen cyber risk governance and improve organizational preparedness.

Titled ["Insights from Simulation Exercises on Strengthening Cyber Risk Governance,"](#) the paper explores how organizations can use structured exercises to identify governance gaps, improve coordination among stakeholders, and better prepare for emerging cyber threats before real incidents occur.

The research highlights the value of bringing together cybersecurity leaders, executives, and decision-makers to test assumptions, evaluate response strategies, and strengthen governance processes under realistic scenarios. As cyber risks increasingly intersect with AI, regulation, and business operations, simulation exercises can help organizations build resilience and make more informed decisions during times of uncertainty.

Four Ways Organizations Can Collaborate on Cyber Resilience

Michael Siegel and Sander Zeijlemaker Publish World Economic Forum Article

As cyber threats become increasingly interconnected, organizations must look beyond their own networks to strengthen resilience. In a new article published by the World Economic Forum, Michael Siegel and Sander Zeijlemaker outline four practical ways organizations can collaborate to improve cyber resilience and better prepare for future disruptions.

The article highlights the importance of connecting dependency, threat, and operational data; testing disruption scenarios before incidents occur; establishing appropriate human oversight of AI-enabled defensive capabilities; and developing trusted partnerships for sharing cybersecurity insights. Together, these approaches can help organizations better understand emerging risks and build more resilient operations.

As organizations navigate an increasingly complex cyber and AI landscape, the article reinforces a core theme of current CAMS research: resilience depends not only on technology, but also on collaboration, governance, and informed decision-making.



“As a result, trust is becoming a critical condition for cyber resilience.”



CYBERSECURITY AT MIT SLOAN (CAMS)
Michael Siegel and Sander Zeijlemaker



New Research Examines the Future of Threat Intelligence Sharing

Sander Zeijlemaker Presents Research at AMCIS 2026

On August 22, Sander Zeijlemaker presented "A Systems Dynamics Approach to Understand Threat Intelligence Sharing" at AMCIS 2026 in Reno, Nevada. Co-authored with Michael Siegel and Ranjan Pal, the research examines cyber threat intelligence (CTI) sharing as a dynamic interaction between attackers and defenders, focusing on zero-day vulnerabilities, disclosure incentives, and mitigation delays.

Using system dynamics modeling and empirical vulnerability data, the study explores how disclosure practices and information-sharing incentives influence cyber risk over time. The findings suggest that weakening centralized disclosure mechanisms can increase societal cyber risk, while secure-by-design practices, faster patching, and stronger disclosure incentives improve resilience. The research also highlights AI's dual role in cybersecurity, simultaneously increasing software risks while strengthening defensive capabilities.

As organizations face increasingly sophisticated cyber threats, the work underscores the importance of threat intelligence sharing as a critical component of collective cyber defense and organizational resilience.



Sander Zeijlemaker presenting "A Systems Dynamics Approach to Understand Threat Intelligence Sharing" at AMCIS 2026 in Reno, Nevada.

Measuring the Security Gap in AI-Generated Code

CAMS Research Highlights Emerging Challenges in AI-Assisted Software Development

Artificial intelligence is rapidly changing how software is developed, with AI coding agents now capable of generating working code at increasingly high levels of accuracy. However, new CAMS research suggests that functionality and security do not always go hand in hand.

During a recent Friday Research Discussion, CAMS researcher Ilya Kabanov presented "*Capability without Security: Measuring the Functionality-Security Gap in AI-Generated Code*." The research examined how code security is evolving across frontier Claude, GPT, and Gemini models. While AI coding systems produce functional code between 83% and 95% of the time, only 24% to 36% of that code is both functional and secure.

The research found that security performance has improved significantly over the last year, yet vulnerabilities remain common. Encouragingly, AI-native threat modeling and dedicated security review processes can improve secure-and-functional output, although often at increased computational cost.

As organizations continue to incorporate AI-assisted development tools into their workflows, the findings underscore the importance of integrating security practices alongside AI-driven productivity gains.

CAMS Researchers Contribute to Global Discussions on Economic Crime

Professors Jeffrey Proudfoot and Dionysios Demetis from the Cybersecurity at MIT Sloan (CAMS) community participated in the 43rd Cambridge International Symposium on Economic Crime, held at Jesus College, University of Cambridge. The annual symposium brings together hundreds of experts from academia, government, law enforcement, regulation, and industry to examine some of the world's most pressing challenges in economic and financial crime.

Their participation highlighted the growing intersection of cybersecurity, financial crime, artificial intelligence, cryptocurrency, and regulation. Throughout the week, Demetis and Proudfoot contributed CAMS perspectives to discussions focused on how organizations and governments can better respond to increasingly complex digital threats.

Demetis chaired the workshop, **"Technological Frontiers: Rethinking and Redesigning the Global AML System,"** with Proudfoot joining an international panel representing technology, banking, law enforcement, and financial intelligence communities. Demetis also contributed to discussions examining whether regulation has kept pace with the rapid growth of cryptocurrency assets and the cyber-enabled crime and ransomware threats associated with them.

The symposium also explored challenges beyond traditional financial crime. Demetis chaired the **UNESCO Chair on Threats to Cultural Heritage**, while Proudfoot presented on deception and insider threats affecting museums and cultural institutions. Their discussion highlighted the increasingly complex balance between physical and digital security, particularly following high-profile incidents at major cultural organizations.



Professors Jeffrey Proudfoot and Dionysios Demetis, among colleagues, at the **43rd Cambridge International Symposium on Economic Crime**, held at Jesus College, University of Cambridge.

CAMS Research Community Honors & Recognitions

Ranjan Pal Invited as Distinguished Speaker at Technology Confex 2026

Ranjan Pal was invited to participate as a Distinguished Speaker at the 4th Information Technology Confex 2026 in New Delhi, India. The event brings together technology leaders, cybersecurity professionals, AI experts, and enterprise decision-makers to discuss emerging technology and cybersecurity challenges. During the conference, Ranjan contributed to a panel discussion examining offensive and defensive cybersecurity strategies and the evolving threat landscape facing organizations worldwide. The invitation reflects the growing international recognition of his research and thought leadership in cybersecurity, cyber resilience, and risk management.

Professor Stuart Madnick receives Lifetime Achievement Award at CDOIQ Symposium



Professor Stuart Madnick, and his colleague Dr. Richard Wang, received Lifetime Achievement awards at the 20th Annual Chief Data Officer and Information Quality (CDOIQ) symposium on July 22nd.

The award recognizes Professor Madnick's *"visionary leadership, pioneering contribution, and lifelong dedication to advancing information quality, data management, and the global data leadership community."*

Presented on July 22, 2026, the honor reflects nearly four decades of influential research and leadership that helped establish information quality as a recognized academic discipline and contributed to the emergence of the Chief Data Officer (CDO) profession. Today, CDO roles exist across an estimated 80% of Fortune 1000 companies.

Professor Madnick's contributions span numerous initiatives that shaped the field, including the founding of MIT's Total Data Quality Management (TDQM) initiative in 1988, the MIT Information Quality Program, and the MIT Chief Data Officer and Information Quality (CDOIQ) research program and annual symposium. He also served as founding Editor-in-Chief of the ACM Journal of Data and Information Quality, helping advance scholarship and professional practice in the field.

His research has influenced generations of academics and practitioners through foundational concepts in data quality, data integration, information management, data privacy, and governance. As artificial intelligence increasingly depends on high-quality, trustworthy data, the significance of this work continues to grow.

During the symposium, Professor Madnick delivered a plenary keynote, "Innovation Insights from the First 20 Years of CDOIQ," reflecting on the evolution of the field and its future directions.

Congratulations, Stuart, on this well-deserved recognition. We are proud to celebrate this remarkable achievement and grateful for your continued leadership and impact.

Cybersecurity at MIT Sloan, in Pisa!



Photo: Profs. Madnick and Dazzi on the way to class, a short distance from the Leaning Tower of Pisa.

Prof Stuart Madnick is teaching part of a Cybersecurity class on Data and Systems Security (DSS), which is organized by Prof. Patrizio Dazzi at the University of Pisa this year.

This course is a key part of a fulltime, two-year graduate program delivered entirely in English and jointly offered by the [Department of Information Engineering](#) and the [Department of Computer Science](#). This is the first master's degree in Italy addressing all aspects of cybersecurity including both hardware and software layers, artificial intelligence, legal frameworks, and business and management issues.

Angelica Marotta Opens Academic Year at University of Pisa

On September 14, Angelica Marotta delivered the inaugural address for the 2026-2027 academic year at the University of Pisa Department of Computer Science in Italy. Invited by department leadership, Marotta returned to the institution where she earned her bachelor's degree to speak with students and faculty about her career journey and current research at Cybersecurity at MIT Sloan (CAMS).

Her presentation connected her academic beginnings in computer science with her current work on cybersecurity and AI regulation, highlighting how regulatory frameworks are developed, compared, and translated into practice. The event offered students a firsthand perspective on emerging challenges at the intersection of cybersecurity, artificial intelligence, and policy, while demonstrating the global impact of CAMS research.

May Almousa Selected to MIT Postdoc Association Leadership Role

Congratulations to CAMS researcher May Almousa, who was selected to serve as Chair of the Belonging and Outreach Committee of the [MIT Postdoctoral Association \(PDA\)](#). In this leadership role, May will help support initiatives focused on strengthening engagement, fostering community, and enhancing the postdoctoral experience across MIT. The committee works to promote connection, inclusion, and collaboration among postdoctoral scholars across the Institute.

May shared that she is looking forward to contributing to and serving the broader MIT postdoctoral community through this role. Her appointment reflects both her commitment to community building and the continued engagement of CAMS researchers across the MIT community.

Congratulations, May, on this well-deserved appointment.

Discuss Research with the Researcher: Featuring MIT Researchers Ilya Kabanov and Shaharyar Khan

When Agents Collude: Cybersecurity in the Age of Collective Agentic Intelligence

October 7, 2026 11:00 AM - 12:30 PM ET

Available exclusively to CAMS member organizations.

As organizations increasingly deploy multiple AI agents that communicate, coordinate, and act on shared objectives, a new class of cybersecurity risks is emerging. While most AI governance frameworks focus on securing individual agents, risks such as collusion, emergent

coordination, and goal drift can arise through interactions between agents, often beyond the reach of traditional controls.

In this exclusive CAMS member discussion, researchers will explore why conventional agent-level security approaches may be insufficient in multi-agent environments and introduce a systems-based framework for understanding and governing collective agentic behavior. Through real-world examples and emerging research, participants will examine how risks can emerge from interactions between AI systems and what this means for enterprise governance, assurance, and cybersecurity.

Discussion Topics

- Why risks in multi-agent systems differ from traditional single-agent AI risks
- Applying systems thinking to agentic AI governance and oversight
- How collusion and emergent coordination can enable agents to bypass intended safeguards

What Members Will Learn

- How real and immediate the risks of collective agentic intelligence are for enterprises
- New approaches for modeling and assessing agent-to-agent threats
- Potential blind spots in existing AI deployments and governance frameworks

This interactive discussion will provide CAMS members with an opportunity to engage directly with the researcher, explore the findings in greater depth, and discuss the implications of AI-assisted software development and emerging cybersecurity challenges.

Friday Research Discussions

CAMS Friday Research Discussions bring together researchers, practitioners, students, and members of the broader cybersecurity community to explore emerging cybersecurity and AI research. Sessions are open to the public.

Individuals interested in receiving announcements and invitations may contact: Kely Fitzgibbons: kcfitz@mit.edu to be added to the distribution list.

Stay Connected

For information about CAMS events, member programs, research discussions, or upcoming discussion series, please contact:

CAMS Directors

Stuart Madnick (smadnick@mit.edu)

Michael Siegel (msiegel@mit.edu)

CAMS In the Press:

September 8, 2026: Sander Zeijlemaker and Michael Siegel were Co-authors, in a World Economic Forum article, [“Four Ways Organizations Can Collaborate on Cyber Resilience,”](#) examining how AI-driven, interconnected cyber risk is making resilience a systemic challenge, and outlining four capabilities—understanding disruption pathways, anticipating emerging threats, enabling trusted autonomous action, and distributing defensive learning—needed to build trusted digital ecosystems.

August 18, 2026: Michael Siegel and Stuart Madnick were quoted in a MIT Horizon article, [“An OpenAI Agent Went Rogue and Hacked Companies During Testing,”](#) discussing an incident in which an OpenAI agent escaped its sandbox and breached Hugging Face’s systems during cybersecurity testing, the difficulty of anticipating or guardrailing unintended AI agent behavior, and the governance and liability gaps that currently leave defenders at a disadvantage against attackers using AI.

August 5, 2026: Ranjan Pal was a Co-author, in a Forbes India article, [“Why insurance needs new models](#)

[to cover AI failures.”](#) examining how traditional cyber and professional liability policies fail to cover AI-driven losses, and proposing new frameworks such as incident registries, refreshed AI risk-scenario catalogues, and coverage carve-backs to close the gap as AI regulation expands.

August 5, 2026: Ranjan Pal was a Co-author, with Bodhibrata Nag, in a Business Today article, [“The Great Indian Examination Heist,”](#) examining the systemic causes of India’s recurring exam paper leaks and recommending reforms such as encrypted item banks, verifiable chains of custody, and permanent examination-integrity units.

July 21, 2026: Stuart Madnick was interviewed on the Data Faces Podcast (TinyTechGuides), [“CDOIQ 2026: A Conversation with Stuart Madnick,”](#) discussing the origins of the CDOIQ Symposium and the data quality movement that inspired it, the risks AI and agentic systems pose to cybersecurity as attackers outpace defenders in adopting the technology, and the importance of pausing to weigh risks against benefits before rushing new AI and data initiatives into production. [Read the news release now!](#)