Coalition for Secure AI Marks First Anniversary with New Principles for Agentic Systems and Defender Frameworks

Global Participation Expands as the Coalition Releases Essential AI Security Guidance

BOSTON, MA, UNITED STATES, July 17, 2025 /EINPresswire.com/ -- The Coalition for Secure AI (CoSAI), an OASIS Open Project, celebrates its first anniversary since launching at the Aspen Security Forum in 2024. Over the past year, CoSAI has grown into the industry's leading collaborative ecosystem for AI security, expanding from its initial founding sponsors to more than 45 partner organizations worldwide. Its mission to enhance trust and security in AI development and deployment has resonated widely, attracting premier sponsors EY, Google, IBM, Microsoft, NVIDIA, Palo Alto Networks, PayPal, Protect AI, Snyk, Trend Micro, and Zscaler. Through multiple workstreams, the coalition has produced practical frameworks and research addressing real-world challenges in securing AI systems. Central to CoSAI's impact this year are the most recent releases of the “Principles for Secure-by-Design Agentic Systems," which establishes three core principles for autonomous AI, and the “Preparing Defenders of AI Systems” whitepaper.

Security Principles Help Safeguard Agentic AI Systems

CoSAI’s Technical Steering Committee (TSC) has released the “Principles for Secure-by-Design Agentic Systems,” a foundational document aimed at helping technical practitioners address the unique security challenges posed by autonomous AI.

The principles offer practical guidance on balancing operational agility with robust security

controls, establishing that secure agentic systems should be Human-governed and Accountable, architected for meaningful control with clear accountability, constrained by well-defined authority boundaries aligned with risk tolerance, and subject to risk-based controls ensuring alignment with expected business outcomes. They must be Bounded and Resilient, with strict purpose-specific entitlements, robust defensive measures including AI-specific protections, and continuous validation with predictable failure modes. Finally, they should be Transparent and Verifiable, supported by secure AI supply chain controls, comprehensive telemetry of all system activities, and real-time monitoring capabilities for oversight and incident response.

This blog post provides additional context on the principles and how they can be applied in real-world environments.

“As agentic AI systems become more embedded in organizations' operations, we need frameworks to secure them," said David LaBianca, Project Governing Board co-chair at CoSAI. "These principles provide a technical foundation for organizations to adopt AI responsibly and securely."

New Defender Frameworks Help Organizations Operationalize AI Security

CoSAI has published another landscape paper, “Preparing Defenders of AI Systems,” developed through our workstream on Preparing Defenders for a Changing Cybersecurity Landscape. The paper provides practical, defender-focused guidance on applying AI security frameworks, prioritizing investments, and enhancing protection strategies for AI systems in real-world environments.

A companion blog post offers additional insights on how this evolving resource bridges high-level frameworks with practical implementation and will continue adapting as AI threats and technologies advance.

“This paper provides defenders with specific guidance on how security frameworks must be adapted to mitigate risks in the AI transformation pinpointing gaps in current approaches and prioritizing critical investments,” said Josiah Hagen of Trend Micro and Vinay Bansal of Cisco, CoSAI’s Workstream 2 Leads. “As security practices are aligned with AI adoption realities, organizations are empowered to make informed decisions and protect their assets while ensuring innovation doesn't outpace defenders. This exemplifies CoSAI's commitment to connecting emerging threats to AI systems with practical security solutions."

These foundational outputs from CoSAI's first year set the stage for even greater impact ahead.

Looking Ahead: Building a Secure AI Future

As CoSAI enters its second year, the coalition is positioned to further accelerate AI security innovation through expanded research initiatives, practical tool development, and increased

global engagement. With active workstreams producing actionable guidance and a growing community of practitioners, CoSAI continues to drive the adoption of secure-by-design AI systems across industries. Its commitment to open source collaboration and standardization remains central to establishing trust in AI technologies. Realizing this vision requires continued collaboration across the AI security community.

Get Involved

Technical contributors, researchers, and organizations are invited to join CoSAI's open source community and help shape the future of secure AI. To learn more about how to get involved, contact join@oasis-open.org.

One year in: What CoSAI members are saying about our impact: [https://www.oasis-open.org/2025/07/17/coalition-for-secure-ai-marks-first-anniversary/].

About CoSAI

The Coalition for Secure AI (CoSAI) is a global, multi-stakeholder initiative dedicated to advancing the security of AI systems. CoSAI brings together experts from industry, government, and academia to develop practical guidance, promote secure-by-design practices, and close critical gaps in AI system defense. Through its workstreams and open collaboration model, CoSAI supports the responsible development and deployment of AI technologies worldwide.

CoSAI operates under OASIS Open, the international standards and open source consortium. www.coalitionforsecureai.org

About OASIS Open

One of the most respected, nonprofit open source and open standards bodies in the world, OASIS advances the fair, transparent development of open source software and standards through the power of global collaboration and community. OASIS is the home for worldwide standards in AI, emergency management, identity, IoT, cybersecurity, blockchain, privacy, cryptography, cloud computing, urban mobility, and other content technologies. Many OASIS standards go on to be ratified by de jure bodies and referenced in international policies and government procurement. www.oasis-open.org

Media Inquiries:
communications@oasis-open.org

Jane Harnad
OASIS Open
email us here