Trust3 AI Launches Trust3 AI Labs to Secure the Entire Agent Lifecycle

SAN FRANCISCO, CA, UNITED STATES, August 6, 2026 /EINPresswire.com/ -- Trust3 AI today announced the launch of Trust3 AI Labs, a dedicated technical research hub built to address the full spectrum of security challenges across the agentic enterprise. The Labs will focus on five core pillars: agent discovery, full-fidelity observability, runtime enforcement, MCP and A2A security, and unified governance.

As enterprises accelerate the deployment of AI agents across production systems, the attack surface has expanded significantly. Trust3 AI Labs will research emerging threats targeting AI agents, Model Context Protocol (MCP), agent-to-agent protocols, models, tools, identities, and enterprise data systems, publishing findings across four areas: Threat Research, Tools, Engineering, and Talks.

"AI agents collapse reasoning, access, and action into a single runtime," said Neeraj, co-founder of Trust3 AI. "That changes the security model. Enterprises need to discover every agent, trace every decision, evaluate every action, and stop unsafe behavior before it reaches production systems. Trust3 AI Labs will research the attacks, architectures, and controls required to make that possible."

The five research pillars address distinct and compounding risks in agentic systems. Agent discovery targets the identification of both registered and shadow agents deployed across Amazon Bedrock, Microsoft Copilot Studio, Databricks, LangChain, SaaS platforms, internal APIs, and custom SDKs. Full-fidelity observability focuses on capturing and replaying complete agent traces, spanning Prompt, Retrieval, Reasoning, Tool Selection, Tool Call, Response, and Action, to give security teams the visibility required to audit and investigate agent behavior at every step.

Runtime enforcement introduces controls that operate directly inside the agent's decision loop, kill switches, credential revocation, and purpose drift detection. MCP and A2A security research addresses MCP server discovery, tool provenance, authentication and authorization, tool description poisoning, cross-server trust, and agent-to-agent identity propagation, an area of increasing concern as multi-agent architectures become standard in enterprise deployments. Unified governance ties these pillars together, enabling continuous policy enforcement across frameworks, platforms, and clouds.

Trust3 AI Labs will make its research accessible to the engineers, architects, and security

practitioners building and defending agentic systems, publishing threat research, technical experiments, security tools, reference architectures, and engineering insights.

About Trust3 AI

Trust3 AI provides one control plane to secure enterprise data and AI agents. The Trust3 AI platform helps organizations discover every agent, observe every decision, secure every action, and govern every connected data source across frameworks, platforms, and clouds. Built on Privacera's production-grade policy enforcement heritage, Trust3 AI combines agent discovery, full-fidelity observability, runtime security, policy intelligence, native enforcement, and continuous governance for the agentic enterprise.

For more information, visit labs.trust3.ai

Angela Chen
Trust3 AI
+1 510-413-7300
email us here
Visit us on social media:
LinkedIn
