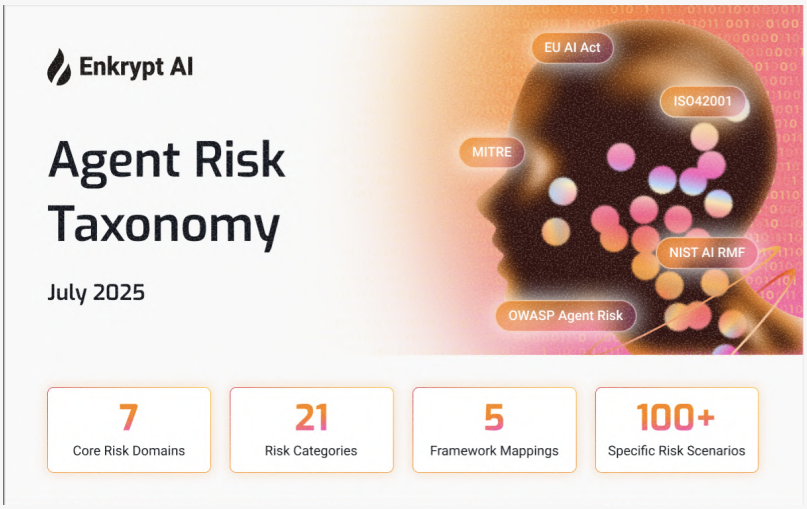
Enkrypt AI Releases Agent Risk Taxonomy to Secure Autonomous AI Systems

Agent Risk Taxonomy delivers actionable guidance for managing autonomous AI risks in real-world enterprise environments.



Enkrypt AI's Agent Risk Taxonomy introduces 7 core domains, 21 risk categories, and over 100 specific risk scenarios, mapped to frameworks like OWASP, NIST AI RMF, and MITRE. Released July 2025

BOSTON, MA, UNITED STATES, July 25, 2025 /EINPresswire.com/ -- Enkrypt AI today announced the release of its **Agent Risk Taxonomy**, a hands-on framework designed to help enterprise security, compliance, and engineering teams manage the fast-emerging risks introduced by autonomous and generative AI systems.

As organizations across finance, healthcare, and technology adopt AI agents that make decisions, call APIs, and access live data with minimal human oversight, traditional security frameworks fall short. Enkrypt AI's Agent Risk Taxonomy fills this critical gap by helping teams assess, monitor, and mitigate the real-world risks that arise when AI systems take autonomous action.

> We built this for CISOs, red teamers, and AI leads who need to move fast and secure even faster. It is practical, not theoretical and ready to plug into enterprise workflows."
>
> *Sahil Agarwal, CEO of Enkrypt AI.*

Purpose-built for AI security and risk practitioners:

While traditional AI frameworks address model development and fairness, they often overlook the unique behaviors of autonomous agents. Enkrypt AI's Agent Risk Taxonomy fills this gap by mapping **seven critical agent-specific risk domains** to industry-standard frameworks, including OWASP, MITRE ATLAS, and NIST AI RMF:

- Governance Failures: When agents ignore or circumvent instructions

- Output Quality Issues: Including hallucinations, bias, or misleading outputs
- Tool Misuse: Unauthorized use of APIs or systems
- Privacy Breaches: Exposure of sensitive or protected data
- Reliability Problems: Drift, inconsistency, and lack of explainability
- Behavioral Risks: Manipulative or deceptive agent behavior
- Access Control Failures: Credential compromise or privilege escalation

The Agent Risk Taxonomy framework by Enkrypt AI identifies seven core domains of risk in autonomous systems, including governance, output quality, privacy, and tool misuse.

What Security Teams Can Expect:

The Agent Risk Taxonomy is a framework built for real-world use. It equips security and engineering teams with:

- Detailed risk scenarios that reflect issues already emerging in production environments
- Monitoring patterns and technical controls aligned with enterprise deployment workflows
- Compliance mappings that integrate with existing security and audit frameworks
- A foundation for red teaming, secure-by-design development, incident response, and risk assessments**

Enkrypt AI's framework enables teams to proactively evaluate and secure AI systems at scale, supporting safe, compliant deployment in a rapidly evolving threat landscape.

Download the Framework: https://cdn.prod.website-files.com/6690a78074d86ca0ad978007/687f7fac66e8127aa565341d_Agent%20Risk%20taxonomy_enkryptai.pdf

Schedule the Demo: https://www.enkryptai.com/request-a-demo

Learn more at: https://www.enkryptai.com/agent-risk-taxonomy

About Enkrypt AI

Enkrypt AI is an AI security and compliance platform that safeguards enterprises against generative AI risks by automatically detecting, removing, and monitoring threats. The company's unified platform combines red teaming, security guardrails, and compliance automation to help enterprises move faster without sacrificing control. Fortune 500 companies are using Enkrypt AI to safely productionize their agents and chatbots. As adoption of generative AI accelerates, organizations face critical risks such as data leakage, jailbreaks, hallucinations, and compliance gaps. Enkrypt AI addresses these risks through end-to-end protection across the entire AI lifecycle.

The Agent Risk Taxonomy from Enkrypt AI visualizes 15 mapped risks across 7 core domains, categorized by agent failure, misuse, and tool-related vulnerabilities based on OWASP Agentic Risk.

The company has tested a wide range of language models, launched the first public AI Safety Leaderboard, and developed defenses against real-world threats, including prompt injection, bias, and misuse. Its solutions are gaining traction across finance, healthcare, and insurance industries, where security and compliance are non-negotiable. Founded by Yale PhD experts in 2022, Enkrypt AI is backed by Boldcap, Berkeley SkyDeck, ARKA, Kubera, and other investors. Enkrypt AI is committed to making the world a safer place by promoting the responsible and secure use of AI technology, ensuring that its benefits can be harnessed for the greater good.

Sheetal Janala
Enkrypt AI
sheetal@enkryptai.com
Visit us on social media:
LinkedIn
YouTube
X

in today's world. Please see our Editorial Guidelines for more information.