Deepkeep

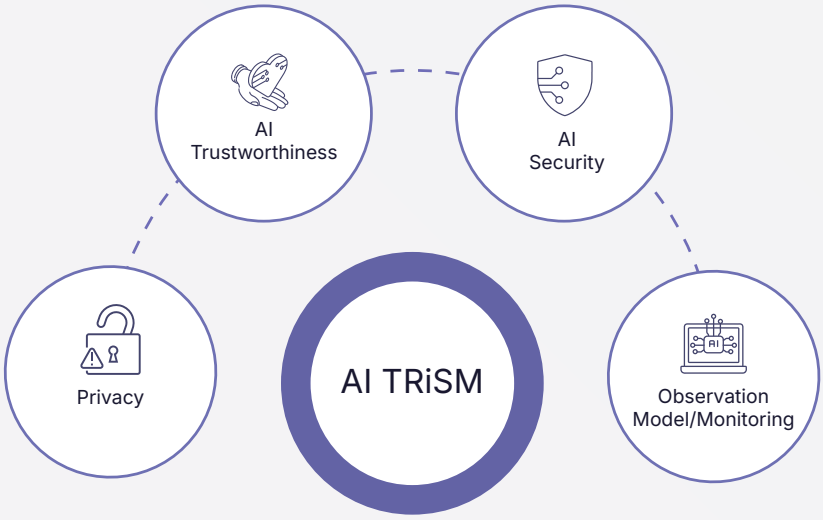
END-TO-END AI SECURITY & TRUSTWORTHINESS

Empowering Secure AI Adoption

AI adoption and usage must be done in trustworthy, explainable, private, and secure manners for both the enterprise and its customers. These characteristics could be compromised by mistakes or malicious activity.

AI TRISM

AI models and applications can pose significant risks if left unchecked. Gartner's AI Trust, Risk and Security Management (AI TRiSM) framework provides proactive solutions to identify and mitigate these risks, ensuring reliability, trustworthiness and security.



Trust safeguards from mistakes, ethical considerations and fairness in decision-making.

Risk is about identifying potential vulnerability and threats to an AI system's security, trustworthiness and privacy.

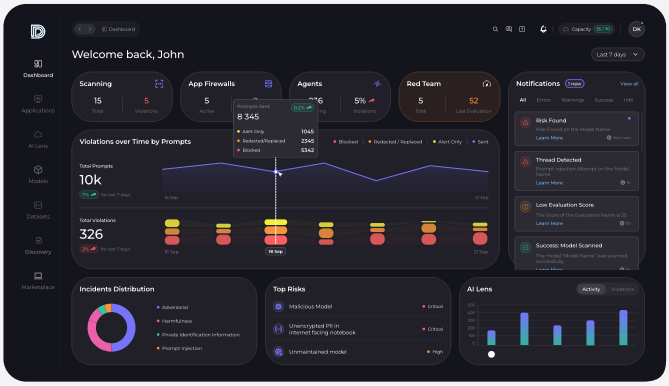
Security Management safeguards models and datasets from attacks, unauthorized access and manipulation.

DEEPKEEP'S AI TRISM SOLUTION

DeepKeep delivers end-to-end AI security and trustworthiness, to mitigate risk across the full AI lifecycle. Built with GenAI at its core, the platform protects both large language model and vision language model systems against evolving threats, ensuring resilience, compliance and confidence in every AI interaction.

By continuously adapting to the rapid pace of AI innovation, DeepKeep helps organizations maintain control, visibility, and assurance across all their AI assets.

Cybersecurity teams worldwide use DeepKeep's multilingual solution to secure AI agents, employee AI use, and homegrown AI applications.



FROM RISK TO RESILIENCE

AI introduces new forms of exposure that go beyond typical cybersecurity risks. Models can produce biased or inaccurate results, unintentionally share confidential data, or be influenced by external manipulation. DeepKeep provides continuous oversight to identify and reduce these risks before they impact operations or reputation.

46%

of AI PoCs get cut before deployment to production, due to risks in data privacy and security, and high costs.

The platform establishes a clear picture of AI trustworthiness, monitoring for weaknesses, anomalies, and signs of misuse. Its proactive approach turns AI risk management from a reactive process into a predictable, auditable practice.

Through comprehensive reporting and clear risk indicators, leadership teams gain the insight needed to make informed decisions about where AI can be safely deployed and how to meet emerging regulatory obligations.

COMPLIANCE, GOVERNANCE AND TRANSPARENCY

As regulations evolve, organizations must demonstrate responsible oversight of their AI systems. DeepKeep helps achieve this by providing a clear audit trail and continuous documentation of AI security.

This transparency simplifies compliance reviews and supports internal governance programs, ensuring that accountability is built in rather than added later. DeepKeep's approach aligns with recognized frameworks for data protection, fairness, and reliability, allowing enterprises to stay ahead of new requirements without constant manual effort.

Organizations that have implmemented real-time AI monitoring and oversight committees are **34% more likely** to see improvements in revenue growth and **65% more likely** to realize cost savings.

TURNING TRUST INTO A COMPETITIVE ADVANTAGE

Trust is now a business asset. Customers, regulators, and investors are increasingly asking how AI systems make decisions, protect data, and avoid unintended harm.

By combining continuous risk oversight with strong governance capabilities, DeepKeep transforms AI trust from an aspiration into a measurable reality. Organizations gain not only protection but also credibility - the ability to demonstrate that their AI systems are safe, transparent, and accountable.

MODEL SCANNING

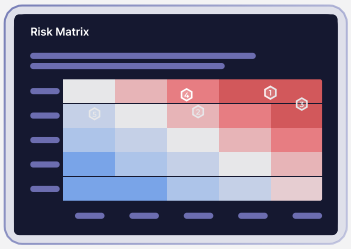
Ensure your AI model is secure and safe to use.

Static (SAST) and dynamic (DAST) scanning of first-party and open-source models, to detect malware and vulnerabilities.

AUTOMATED AI RED TEAMING

Context-based evaluations of your AI model.

Model testing, tailored to your specific use case, identifies flaws and vulnerabilities, suggesting mitigations to secure and ensure your apps and agents act according to policy.



AI FIREWALL

Monitor & stop threats originating from AI interactions.

Apply a runtime AI firewall across every app, user or agent interaction. Inspect prompts and evaluate model responses before they're seen or acted on.

From prompt injection and jailbreaking to data leaks and toxic outputs, apply context-aware guardrails that reflect your compliance standards and data handling policies.

AI AGENTS

Establish AI agent visibility and tackle risks.

Identify and mitigate risks in AI agent behavior before they escalate. Monitor activity, enforce MCP server usage, trace data flows, and assess how agents interact with systems and make decisions - keeping you ahead of security incidents.

Want to learn more?

www.deepkeep.ai

AICPA SOC2 · ISO 27001 Certified · United Registrar of Systems ISO 9001