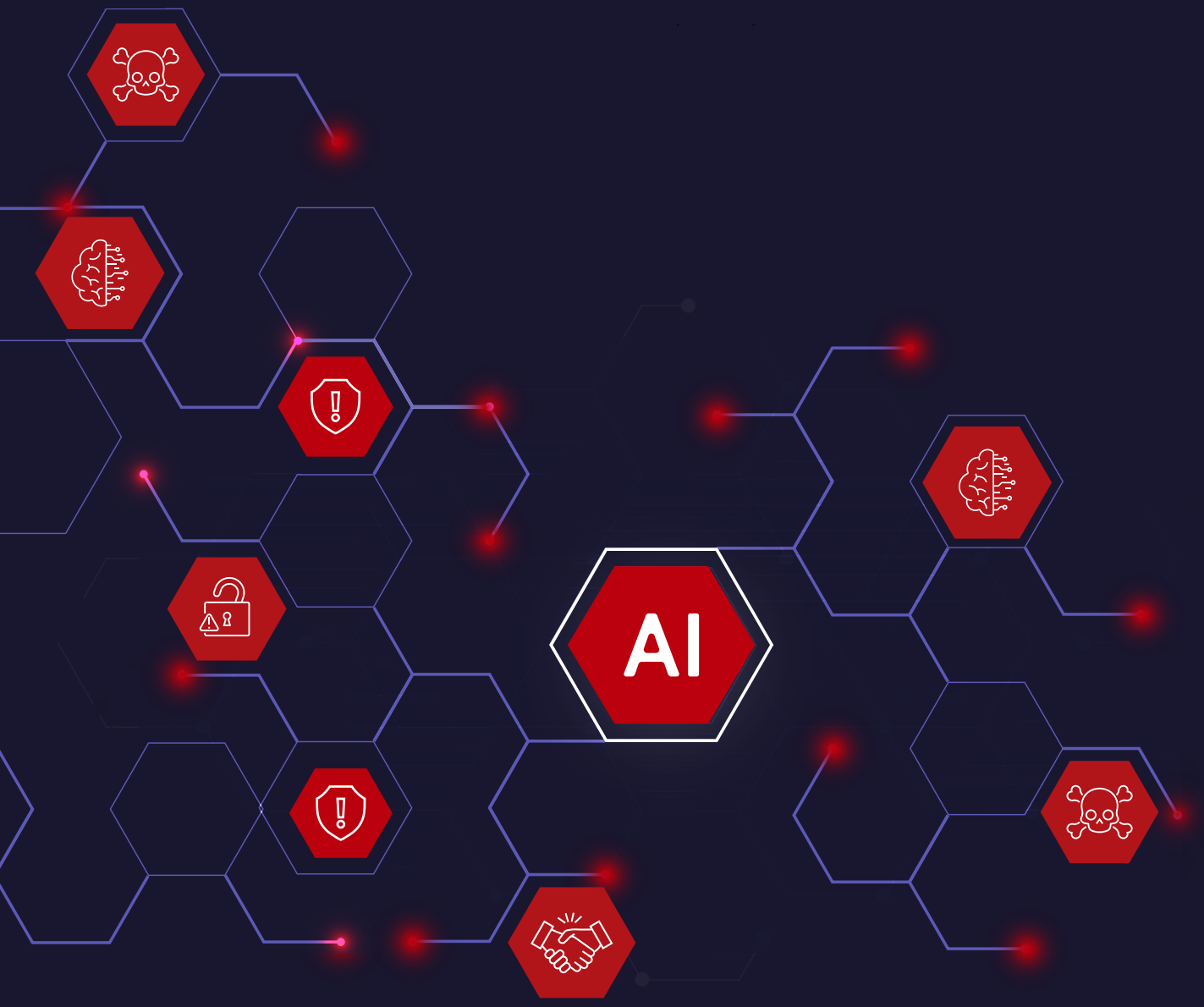
Deepkeep
AI-Native Security



AI-NATIVE SECURITY AND TRUSTWORTHINESS

SECURING GENAI WITH GENAI

EMPOWERING AI WITH SECURITY AND TRUST

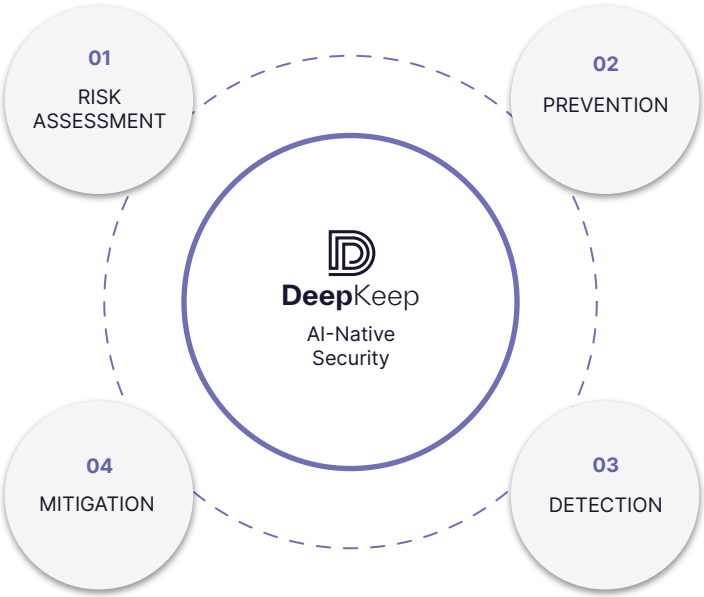
AI systems must be trustworthy, explainable, private, and secure for both the enterprise and its customers. These characteristics could be compromised by mistakes or malicious activity. Meanwhile, the EU AI Act, American AI Bill of Rights and other regulatory frameworks in local US state governments, Japan, Korea and India are already starting to define compliance controls, requiring the assessment, monitoring and control to AI.

PRE AND POST DEPLOYMENT

- Penetration testing
- Vulnerability assessment
- Poisoning and backdoor detection
- Protection recommendations

PRE-DEPLOYMENT HARDENING

- Preprocessing
- Post-processing
- Model reparation



AI FIREWALL

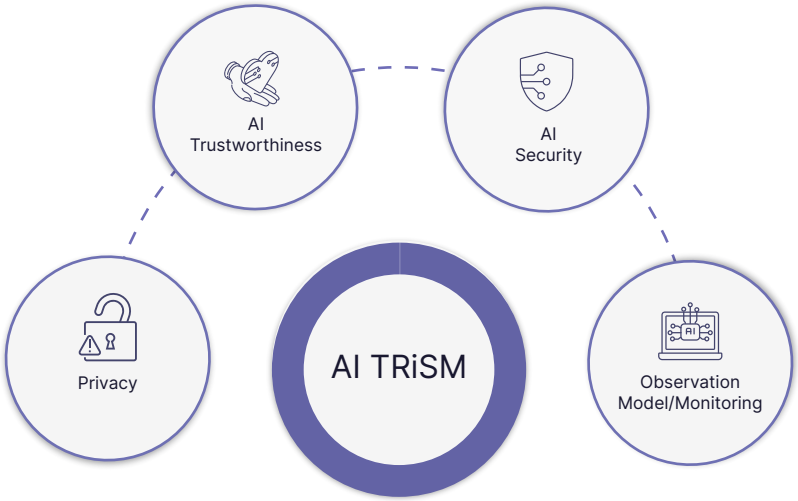
- Access restriction
- Real-time alert triggering
- Dynamic protection
- Operation center and response

DETECTION

- Online/offline
- Stateful/stateless
- Anomaly detection
- Context based

DEEPKEEP'S AI-NATIVE TRiSM SOLUTION

AI TRiSM - Trust, Risk and Security Management – is Gartner's top strategic technology trend for 2024. TRiSM is a set of solutions that proactively identify and mitigate the risks arising from AI models and applications, as well as risks related to reliability, trustworthyness, fairness and security.



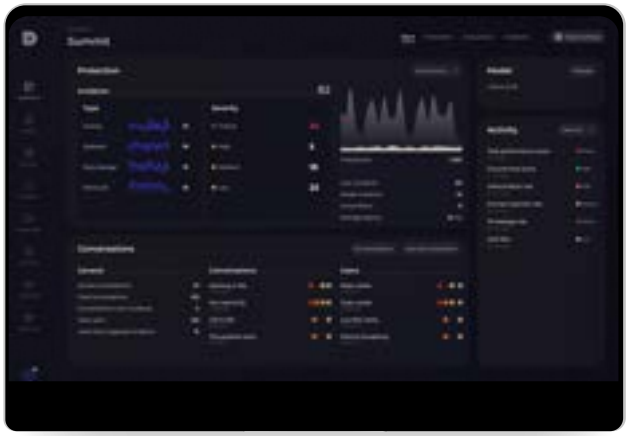
Trustworthiness safeguards from mistakes, ethical considerations and fairness in decision-making.

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.

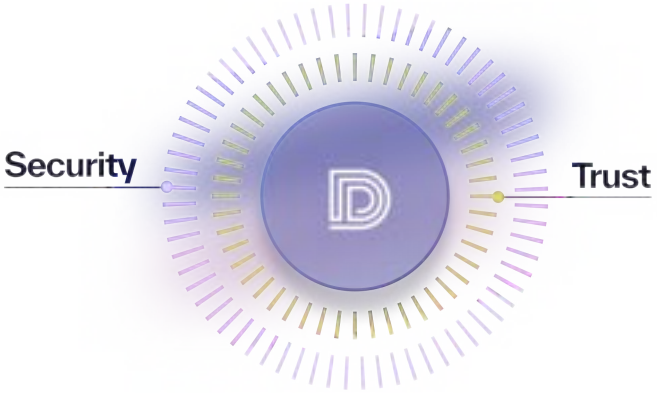
DeepKeep ensures the health and robustness of ML models to safeguard AI from errors and threats.

DeepKeep offers a complete security and trust solution for the entire ML model lifecycle, covering all stages from data curation, model training, risk assessments, prevention, detection, monitoring and deployment to mitigation.



INDISPENSIBLE PILLARS – SECURITY AND TRUSTWORTHINESS

AI trustworthiness and robustness are complementary and inseparable – an AI model cannot be trusted if it is not robust and vice versa.



When data used to train an AI model is biased or incomplete, that model may learn to replicate and amplify mistakes, leading to dangerous outcomes.

When algorithms used to develop an AI model are flawed or incomplete, the model may produce inaccurate results.

DEEPKEEP'S SOLUTION FOR LLM

1. Protects against LLM attacks, including prompt injection, adversarial manipulation and semantic attacks.
2. Identifies and alerts against hallucination using a hierarchical system of data sources, including both internal and trusted external references.
3. Safeguards against data leakage, protecting sensitive data and personally identifiable information (PII).
4. Detects and removes toxic, offensive, harmful, unfair, unethical, or discriminatory language.

DEEPKEEP'S SOLUTION FOR COMPUTER VISION

Images encapsulate a wealth of visual cues, encompassing textures, colors, shapes, and, most importantly, contextual elements which serve as crucial indicators for object detection models, influencing their ability to accurately detect and classify objects.

Evaluates the integrity of datasets used for model training as well as the performance, reliability and robustness of models during the entire AI journey, data curation and model building to deployment in production environments.

Detects and mitigates malicious incidents and trustworthiness challenges in real-time, including fairness, biases, weak spots, data drift, out-of-distribution (OOD), poisoning, evasion and denial-of-service.

Protects and monitors image classification and object detection models from physical and digital attacks, enabling secure and trustworthy deployments.

DEEPKEEP'S SOLUTION FOR MULTIMODAL

AI usage goes beyond any single mode of interaction, whether it's language through LLMs or perception through computer vision. To truly imitate human intelligence, AI Agent systems process multiple modals at once: seeing, reading, interpreting, and responding like we do. This introduces new risks across every input and output.

DeepKeep's unified multimodal solution combines the best of both worlds and ensures your AI remains safe, reliable, and aligned - no matter how it sees or speaks.

Unlock the power of trust and security in AI with DeepKeep.

ABOUT DEEPKEEP

DeepKeep safeguards machine learning pipelines from biases, errors and cybersecurity risks, ensuring trustworthy and reliable AI.

DeepKeep's AI security safeguards machine learning pipelines, promoting secure, unbiased, error-free, explainable and trustworthy AI solutions. This includes vision data models, LLM and tabular models in risk assessments, prevention, detection, monitoring, and mitigation. Only AI-Native security - built itself with generative AI - can protect limitless borders and endless content generation across diverse source domains, models and datasets. DeepKeep's enterprise software platform provides security and trustworthiness from the research and development phase of machine learning models through deployment and the entire product lifecycle.

CONTACT

sales@deepkeep.ai

www.deepkeep.ai

Deepkeep
AI-Native Security