DARKTRACE

Development & Application of ML/AI in Cybersecurity

Presented by:

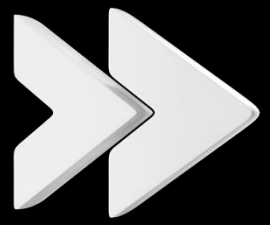
Margaret Cunningham, PhD

VP Security & AI Strategy | FCISO

Darktrace

Application of AI in Cybersecurity

Scale & Complexity



Volume & Velocity

Massive data flows inbound/outbound at real-time speeds

Variety & Complexity

Logs, packets, behaviors, unstructured data, structured data… infinitely diverse types formats and data sources

Data Access & Fidelity

Limited access to clean, complete, trustworthy data

Integration Challenges

Disparate data sources require complex data engineering, correlation, and enrichment efforts

Freshness & Timeliness

Delays reduce detection relevance and response efficacy.

From Detection to Defense

Signal Detection

Subtle attack signals, or anomalies that do not meet specific thresholds can be buried in noisy, dynamic environments

Alert Fatigue

Too many false positives overwhelm analysts, fear of false negatives makes it difficult to change thresholds. Erosion of trust.

Novel Threats

Attackers continuously adapt to evade and manipulate detection systems

Privacy & Data Protection

Sensitive data and domain-specific constraints on centralizing data limits universal solutions

Explainability

“black box” models may not meet explainability requirements for compliance/audits, analysts need clarity to respond quickly.

[DTL0]

AI in Cybersecurity: Confidence is High

- Security leaders trust AI to accelerate defense, reduce workload, and drive a more proactive security posture

95%

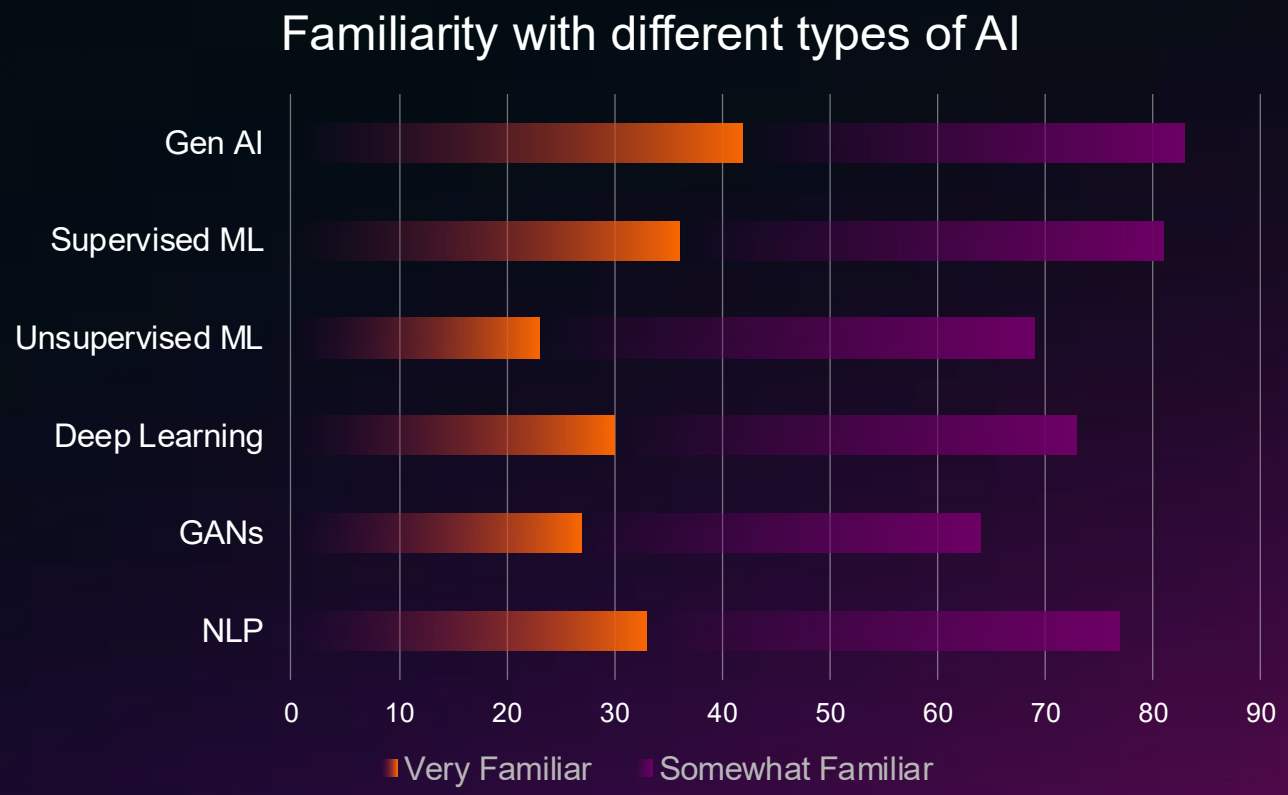
believe AI can improve the speed and efficiency of cyber defense

64%

CISOs allocating a dedicated AI budget

Source: Darktrace 2025 State of AI Report

Yet, More Education on AI can help



42%

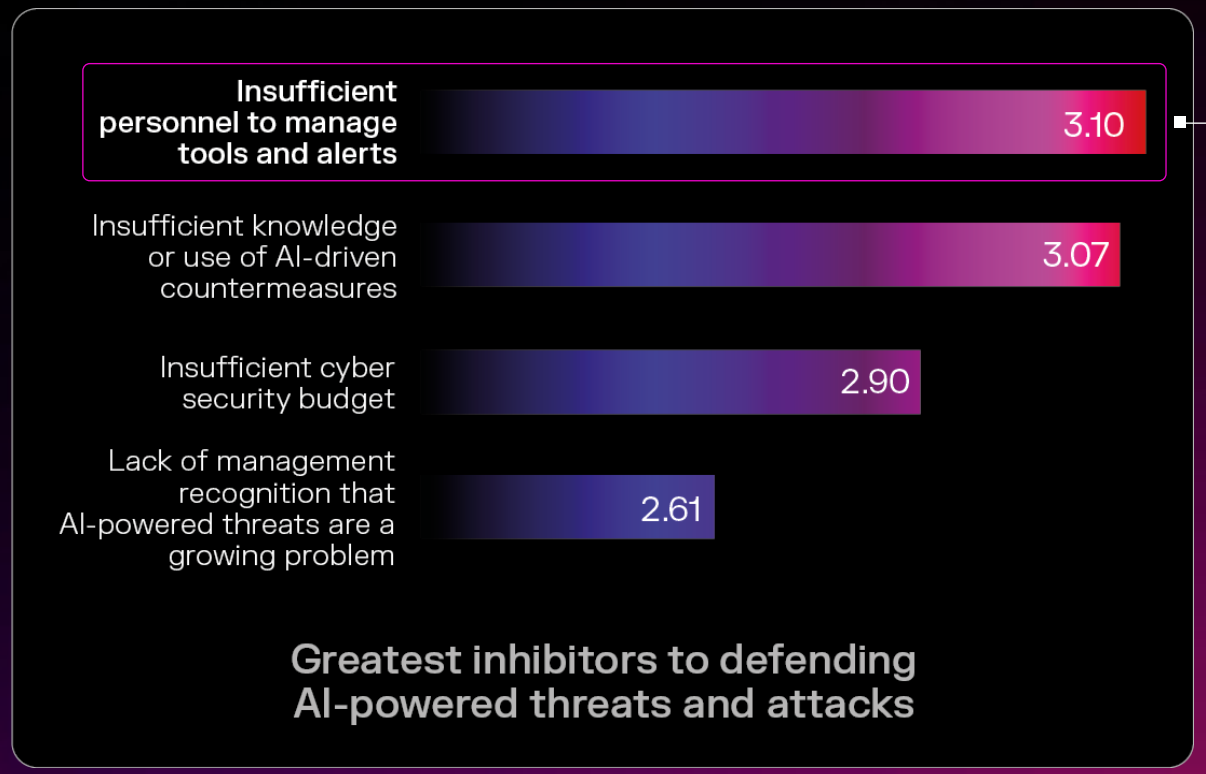
report **full understanding** of the different types of AI used in their security environment

60% Of CISOs **vs** **18%** in other roles

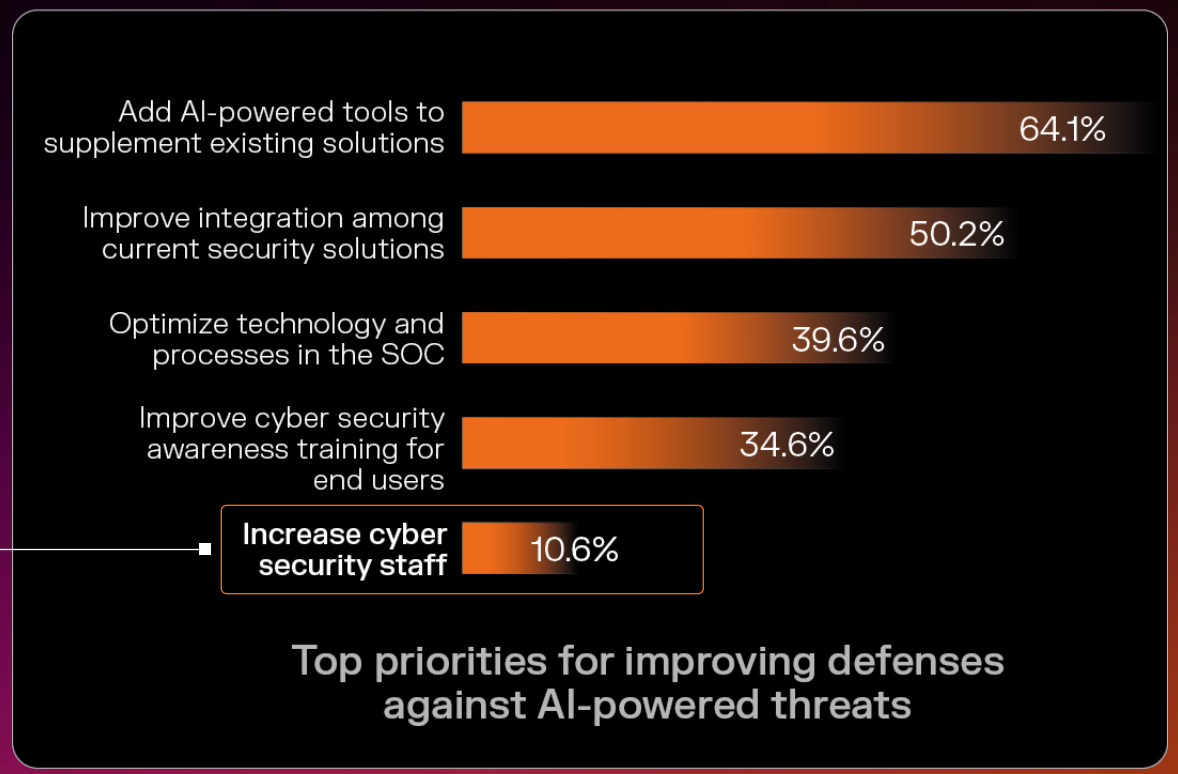
Source: Darktrace 2025 State of AI Report

Lack of People is the Challenge; AI is the Solution

It's not the budget or the lack of awareness - it's the headcount



but very few are prioritizing hiring this year



Source: Darktrace 2025 State of AI Report

Preventing Misapplication of AI will improve Success

40%

projects Cancelled

By 2027, Agentic AI projects will be cancelled due to escalating costs, unclear business value or inadequate risk controls - Gartner

- **Avoid** hype and prevent misapplication
- **Understand** the maturity of current models and capabilities of AI tools in the market
- **Develop** strategic technology roadmaps and prioritize outcome
- **Target** use cases where AI could be most valuable and feasible
- **Conduct** cost-benefit and ROI Analysis

Cutting Through the Jargon: Key AI Concepts Explained

Agentic AI

Autonomously analyzes, decides, and takes actions with minimal human intervention

Bayesian Probability

Involves unsupervised ML techniques that apply probabilistic modeling to both historical and current information

Ensemble Learning (AI)

Combines AI learning algorithms for better predictive results

Multi-Layered AI

Hierarchical layers for deeper insights, greater accuracy, and adaptive threat detection

Large Language Models (LLM)

Very large deep learning models that are pre-trained on vast amounts of data

Generative AI (GenAI)

Generates text, code, and content dynamically

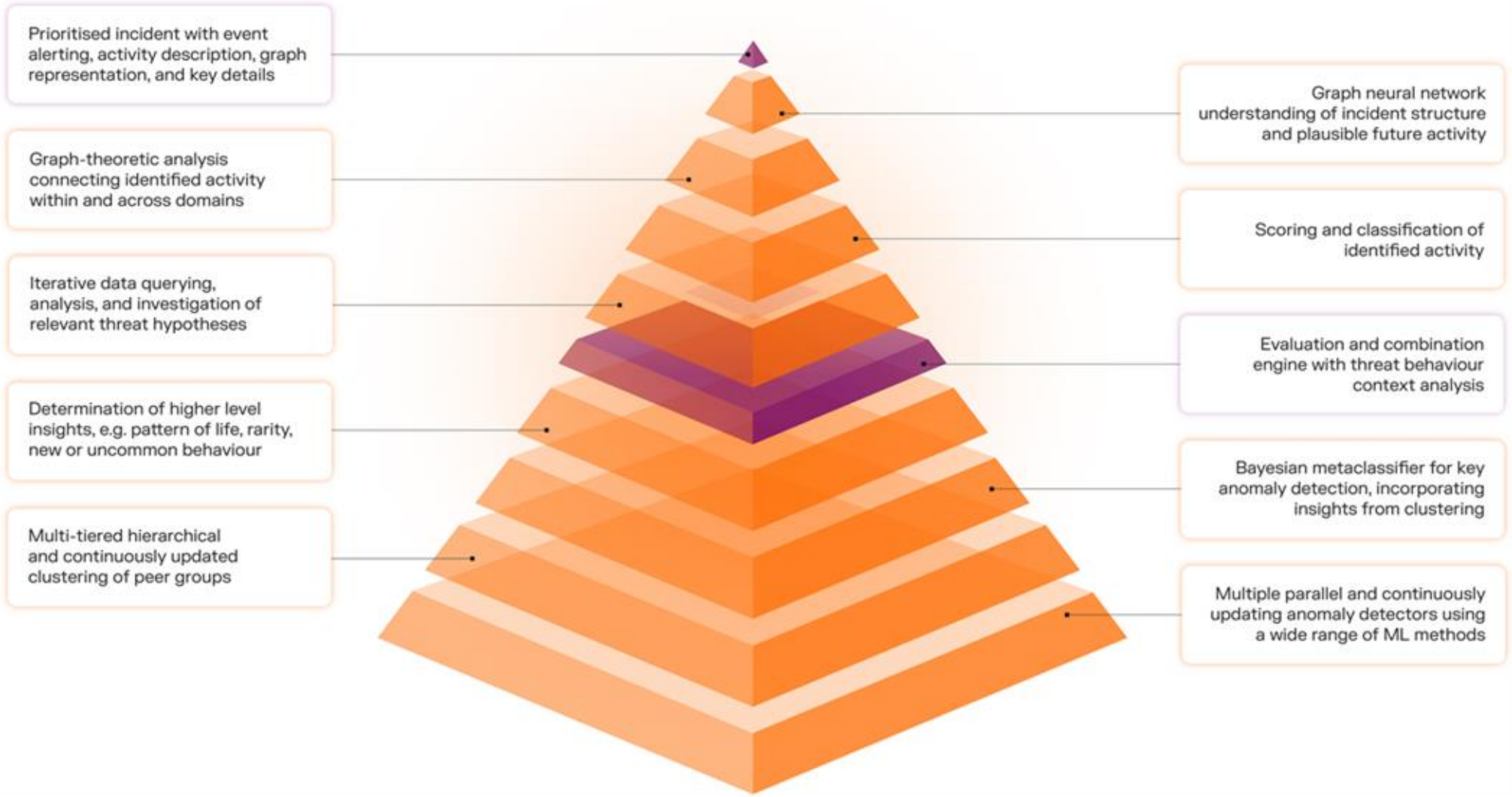
Graph Theory and GNNs

Uses node-based structures for AI analysis

Accurate Analysis: Where GenAI Falls Short

- Immature reasoning
- Over-applies semantic analysis
- Inaccurate and hallucinates
- Difficult to interpret
- Vulnerable

Diverse, Multi-Layered AI



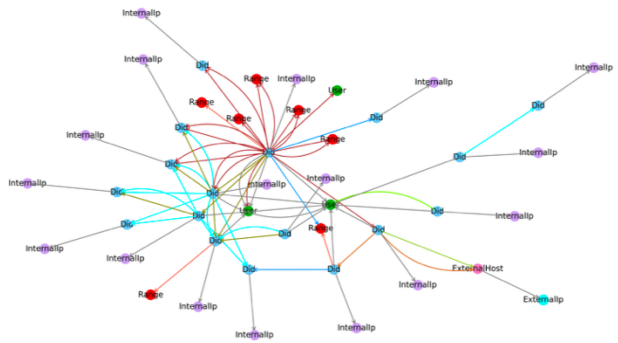
DIGEST: GNN & RNN

Action Prediction:

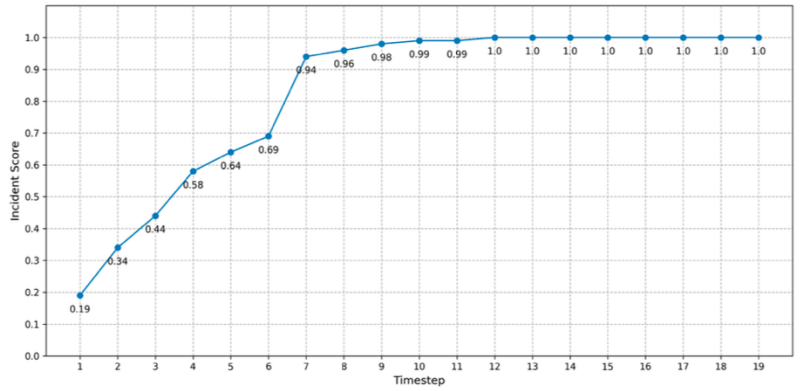
- Classifies next likely attacker behavior
- 86% accuracy; action prediction

Growth Estimation:

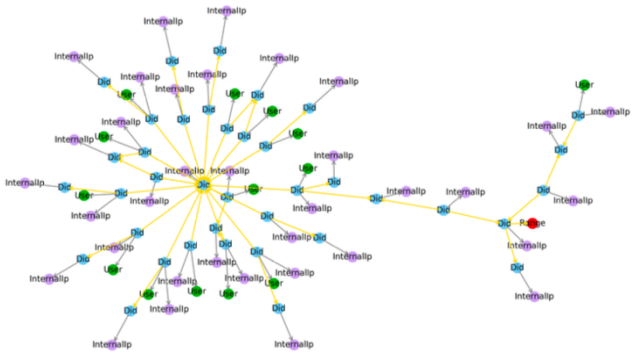
- Predicts if incident will escalate.
- MSE = 0.054; growth estimation



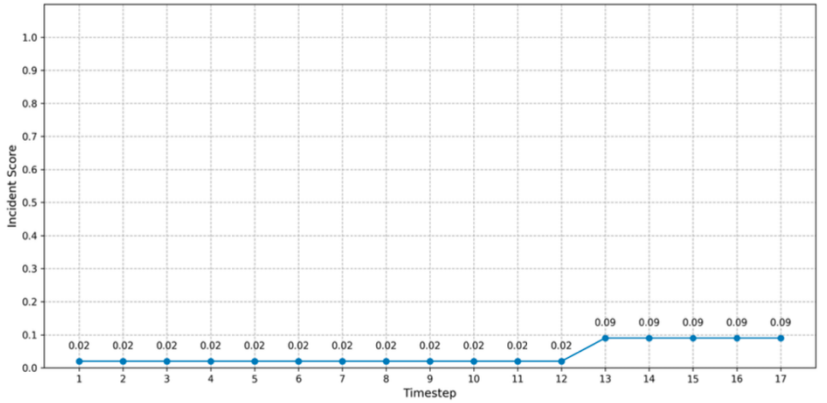
(a)



(b)



(a)



(b)

LLM Embedding Model

Security ≠ Natural Language

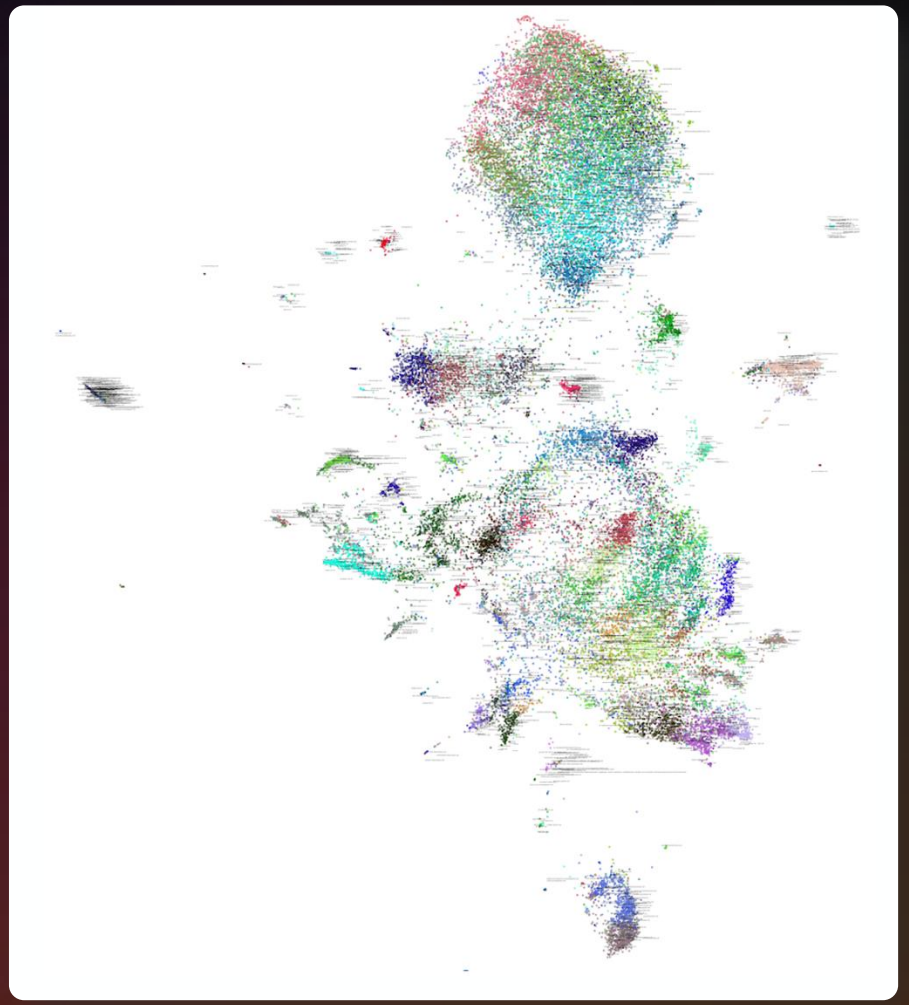
- Hostnames, URIs, and headers don't follow grammar or word boundaries
- General-purpose LLMs struggle with structure, order, and obfuscation

Custom Tokenization & LoRA Adaptation

- Byte-level BPE tokenizer trained on security data
- Handles encoded payloads, CLI arguments, and irregular strings
- LoRA adapters specialize the model for tasks like DGA detection or file classification

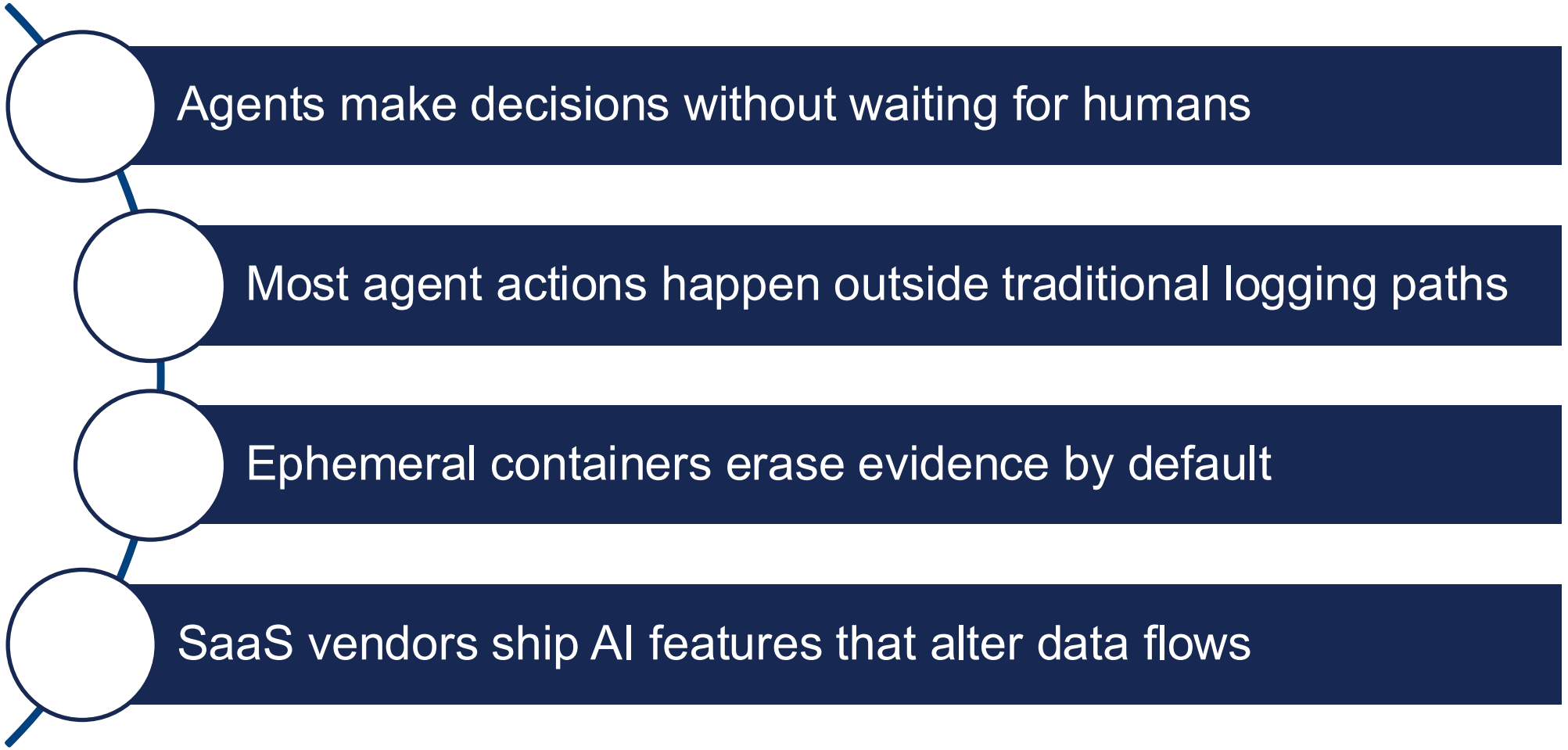
Privacy, Performance, and Portability

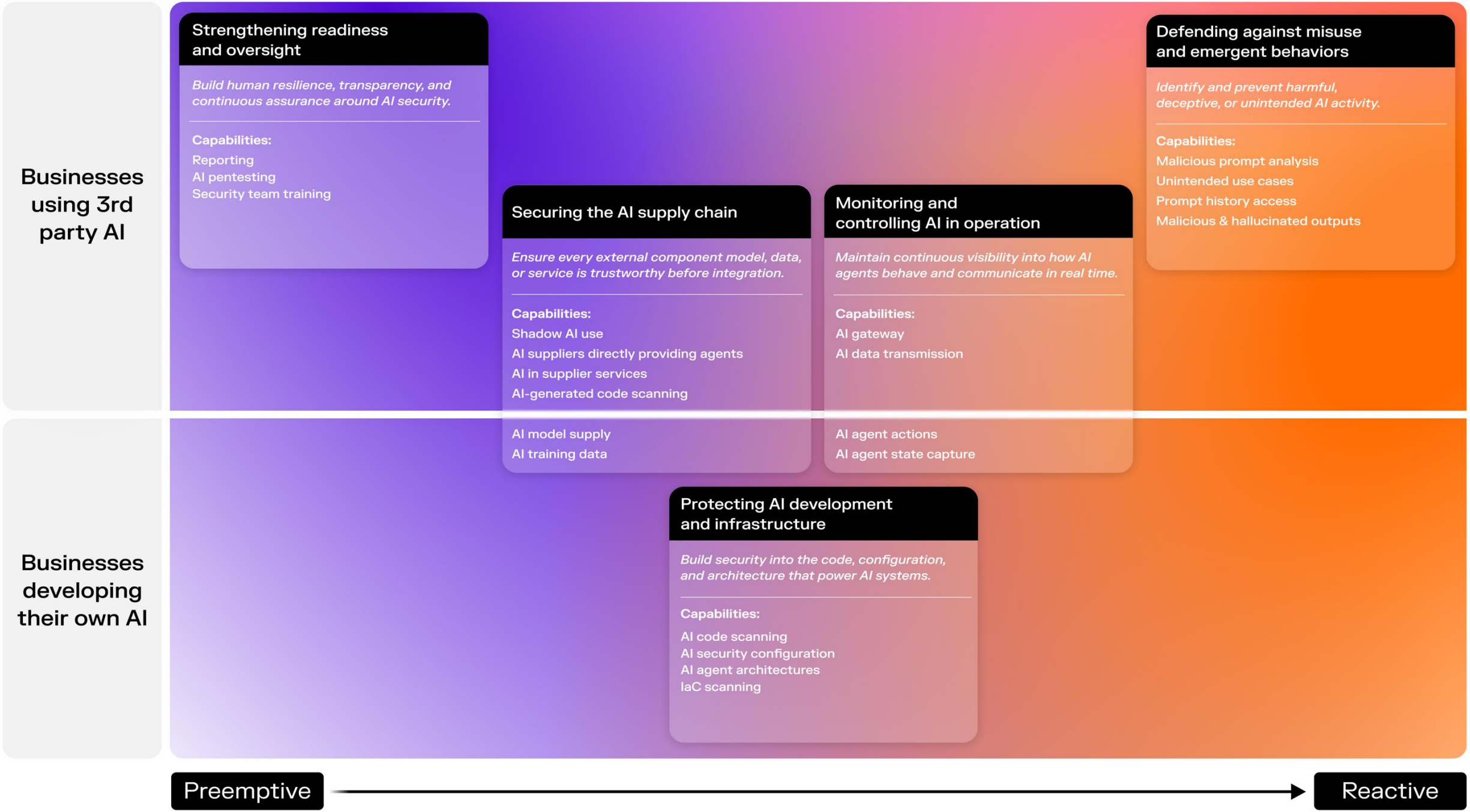
- Trained on anonymized, domain-specific data
- 95M parameters optimized for CPU (not GPU)
- Supports local, privacy-preserving inference in constrained environments



Managing AI Risk

The Operational Reality of Agentic AI




Businesses using 3rd party AI
Businesses developing their own AI
Strengthening readiness and oversight
Build human resilience, transparency, and continuous assurance around AI security.
Capabilities:
Reporting
AI pentesting
Security team training
Securing the AI supply chain
Ensure every external component model, data, or service is trustworthy before integration.
Capabilities:
Shadow AI use
AI suppliers directly providing agents
AI in supplier services
AI-generated code scanning
AI model supply
AI training data
Monitoring and controlling AI in operation
Maintain continuous visibility into how AI agents behave and communicate in real time.
Capabilities:
AI gateway
AI data transmission
AI agent actions
AI agent state capture
Defending against misuse and emergent behaviors
Identify and prevent harmful, deceptive, or unintended AI activity.
Capabilities:
Malicious prompt analysis
Unintended use cases
Prompt history access
Malicious & hallucinated outputs
Protecting AI development and infrastructure
Build security into the code, configuration, and architecture that power AI systems.
Capabilities:
AI code scanning
AI security configuration
AI agent architectures
IaC scanning
Preemptive
Reactive

THANK YOU

Margaret Cunningham, Ph.D.
VP Security & AI Strategy | Field CISO