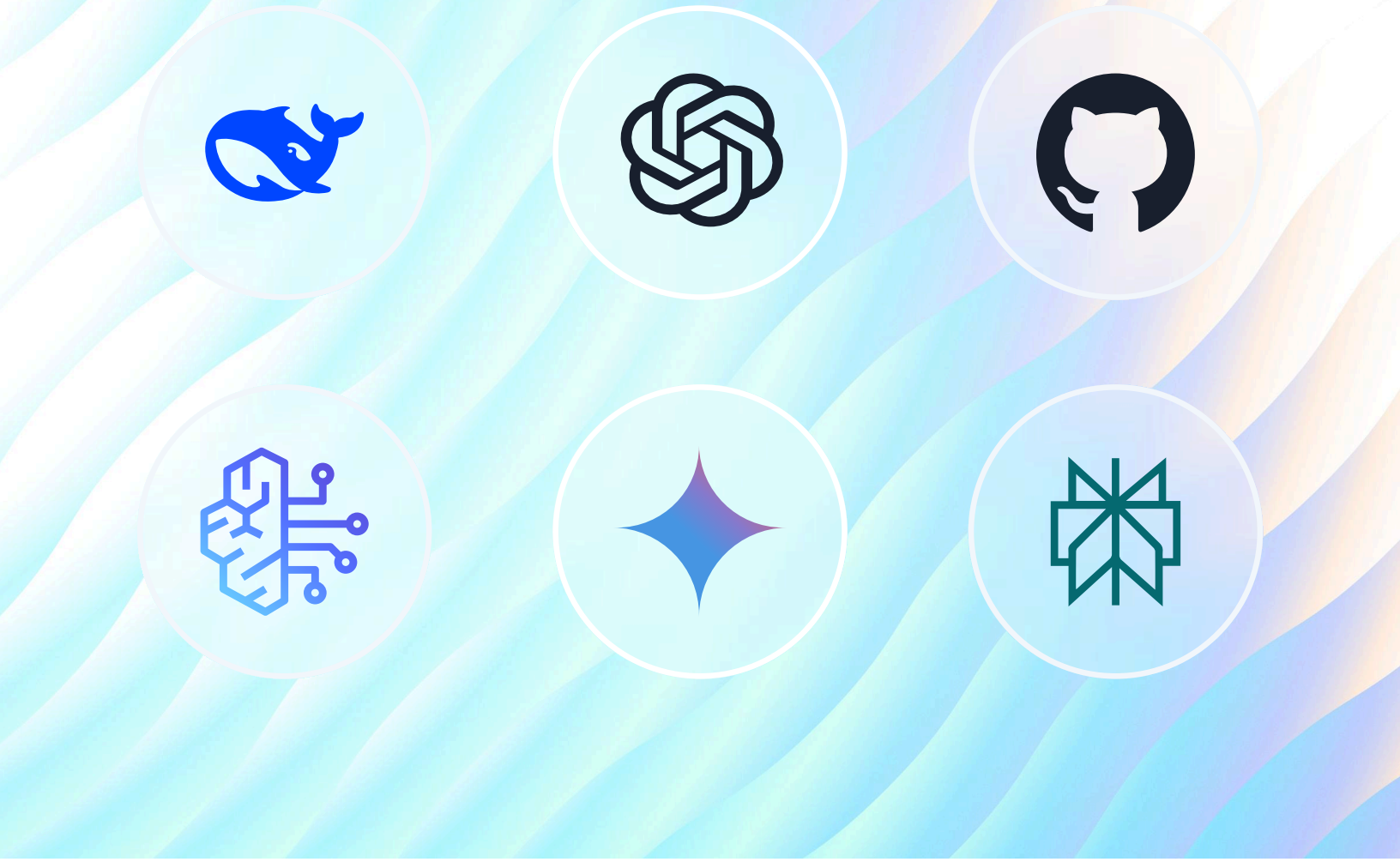
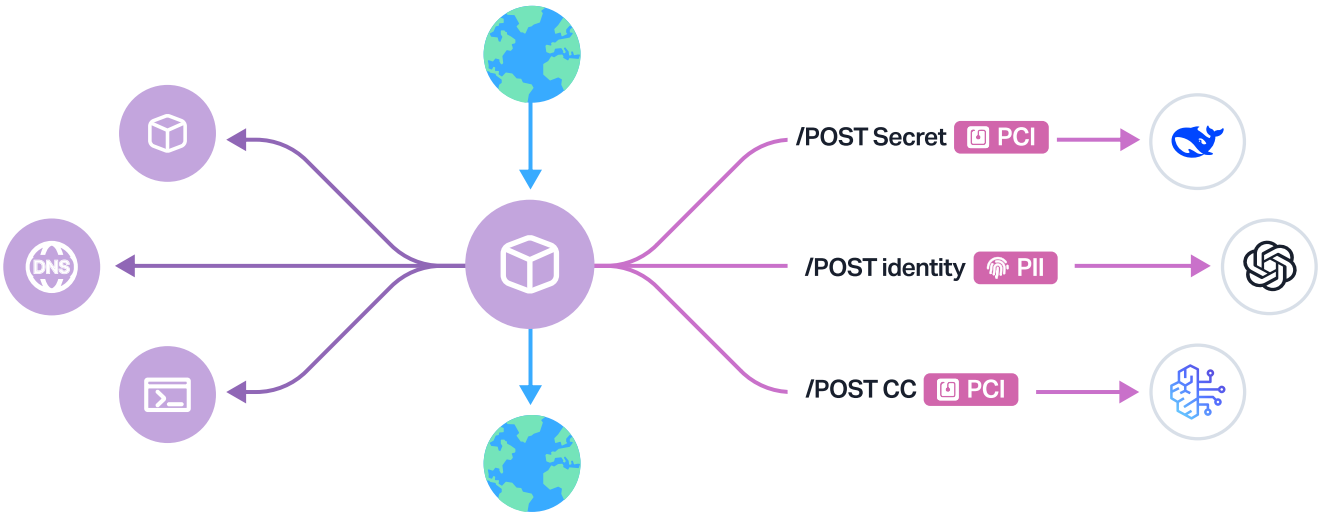
upwind



WHITE PAPER

Top 5 GenAI Risks & How to Protect Against Them

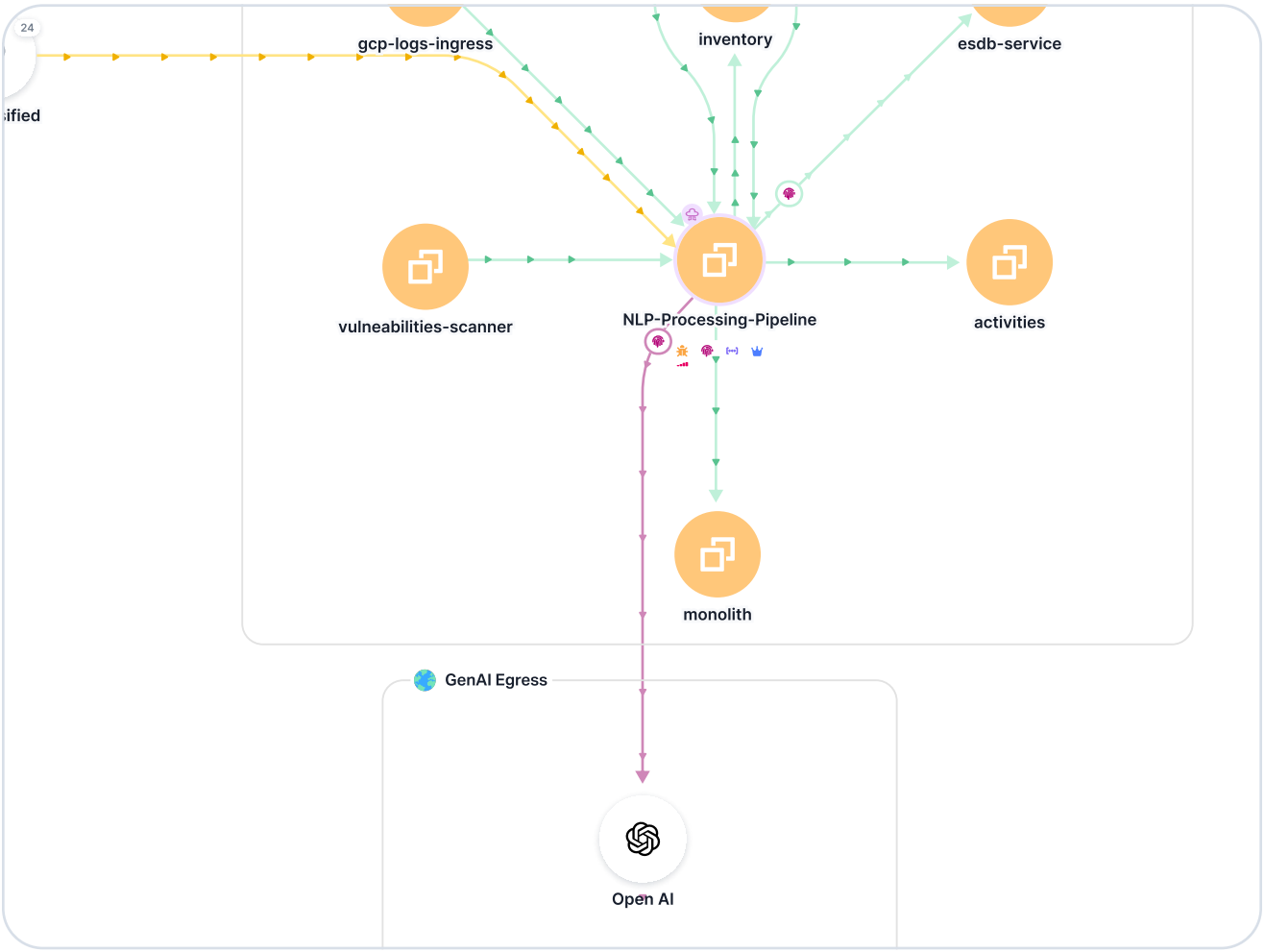
Generative AI (GenAI) encompasses technologies that create content—ranging from text to images and music—by **leveraging extensive datasets.** As Gartner has indicated, "generative AI is becoming a **general-purpose technology** with an impact similar to that of the steam engine, electricity and the internet." Its integration into our daily tasks promises to **revolutionize productivity and creativity** across various industries.



As enterprises rush to integrate generative AI (GenAI) into core products, workflows, and services, they're also facing a wave of emerging security risks. GenAI introduces complex and novel attack surfaces—ranging from dynamic API interactions to model-level threats—that traditional security stacks were never designed to handle.

The following pages contain the top 5 GenAI-specific risks every technical team should be aware of—and best practices to mitigate them.

1. Sensitive Data Leakage via AI Interactions



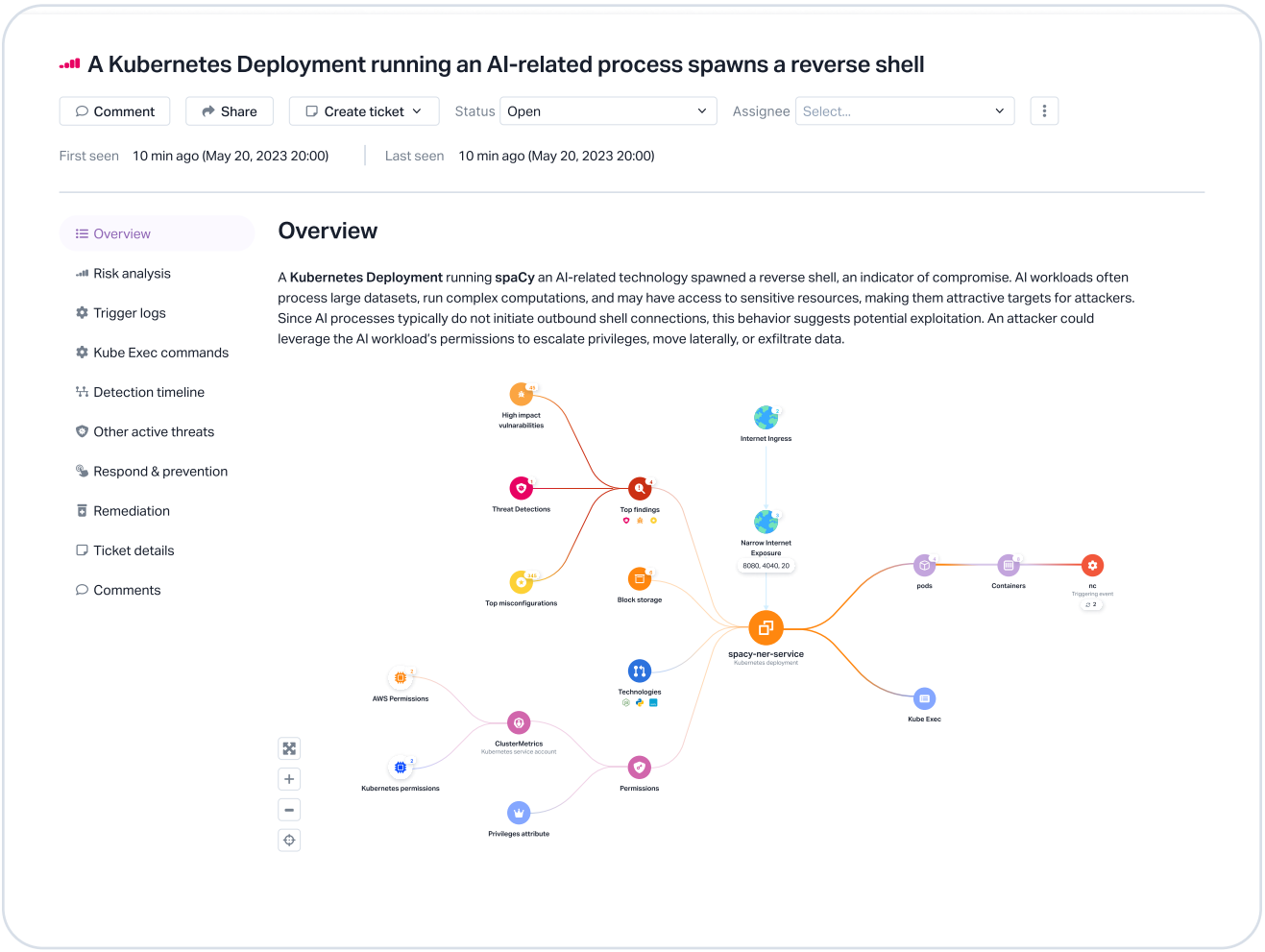
The Risk:

Prompts, completions, and API calls to GenAI services often involve transmitting internal or regulated data. Without visibility into these flows, teams risk exposing PII, secrets, proprietary code, or other sensitive data to third-party AI systems.

Best Practice:

Deploy payload inspection and AI-aware data loss prevention (DLP) controls that analyze traffic in real-time. Use pattern-matching (e.g., regex) alongside machine learning to detect PII, credentials, keys, and other sensitive fields leaving your environment.

2. Model Manipulation and Prompt Injection



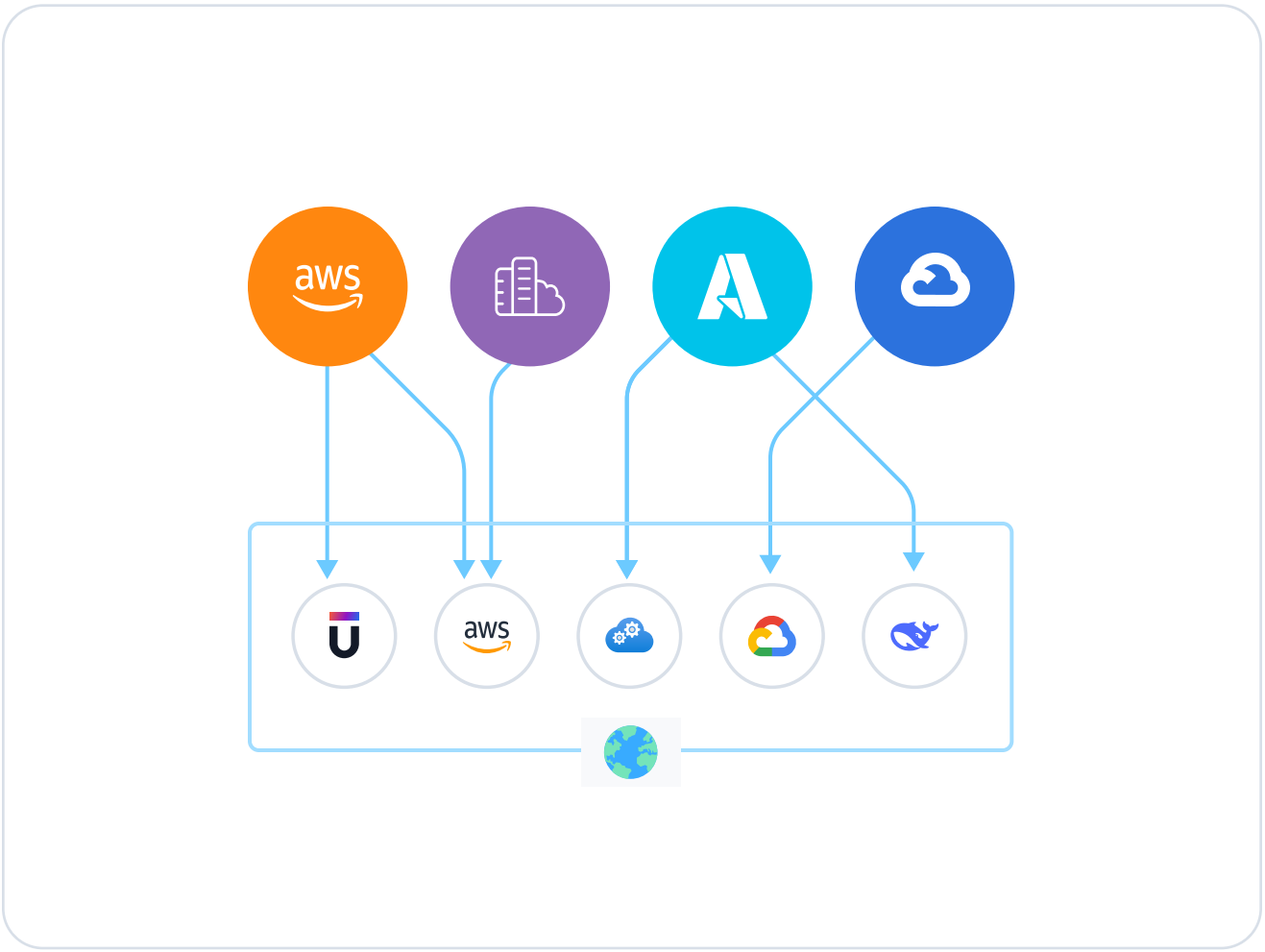
The Risk:

GenAI models are vulnerable to adversarial prompts, malicious context injections, and unauthorized fine-tuning. These attacks can result in corrupted model behavior, hallucinated outputs, or malicious actions triggered by crafted inputs.

Best Practice:

Implement dynamic behavioral baselining and AI-specific threat detection rules. Monitor model interactions for anomalies in usage patterns or unexpected input/output sequences. Alerts should trigger for abnormal prompt chains, excessive token usage, or unauthorized fine-tuning attempts.

3. Unauthorized or Unmonitored AI API Usage



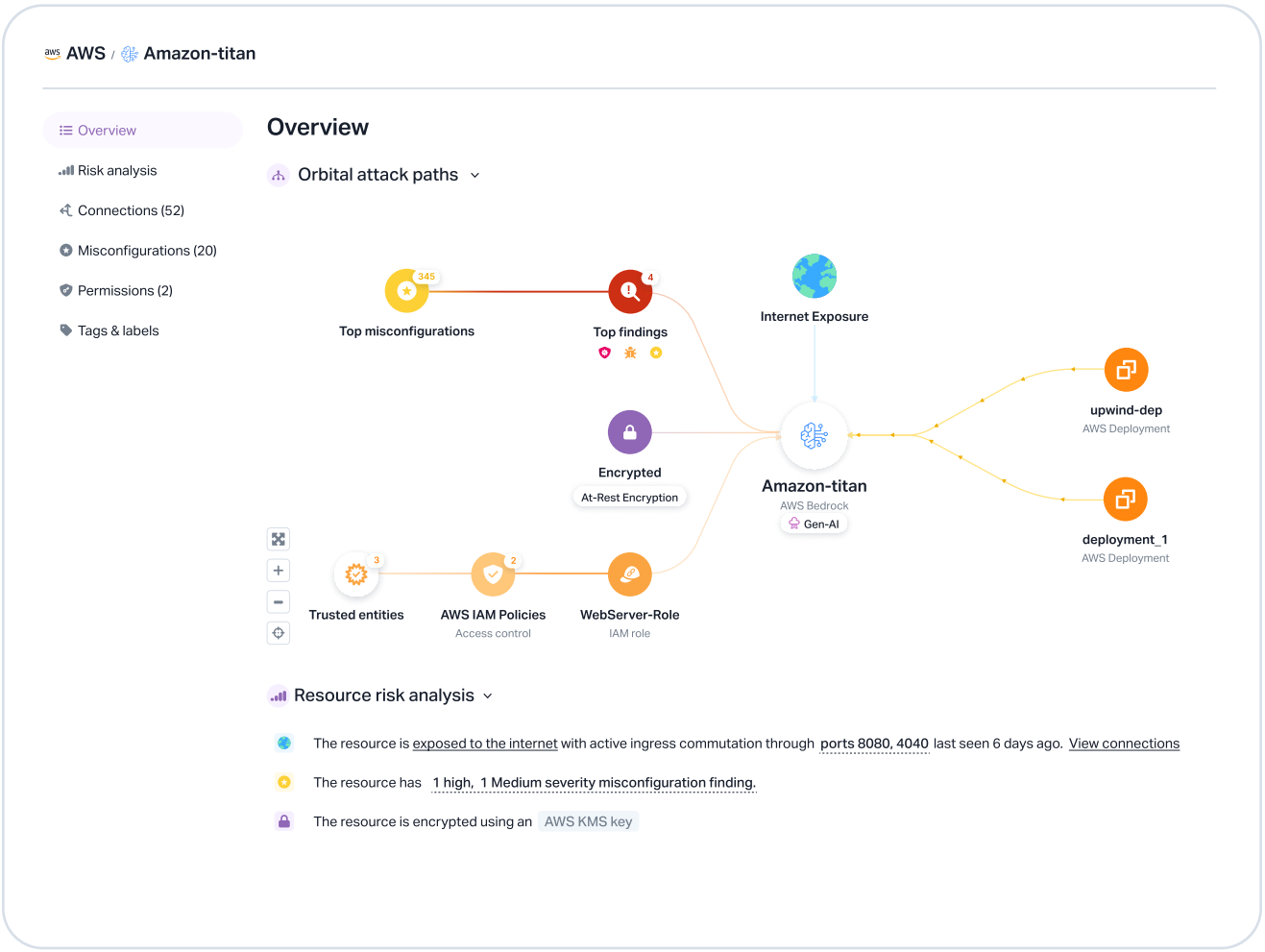
The Risk:

Developers and applications can invoke GenAI APIs (e.g., OpenAI, DeepSeek, etc) directly from cloud environments. These endpoints can easily bypass perimeter defenses, leading to shadow AI usage and compliance gaps.

Best Practice:

Continuously monitor Layer 3, 4, and 7 telemetry to detect and log outbound API traffic to GenAI services. Establish allow/block policies, enforce authentication, and classify traffic by model and provider to retain control over AI consumption.

4. Cloud-Native Attack Surface Expansion



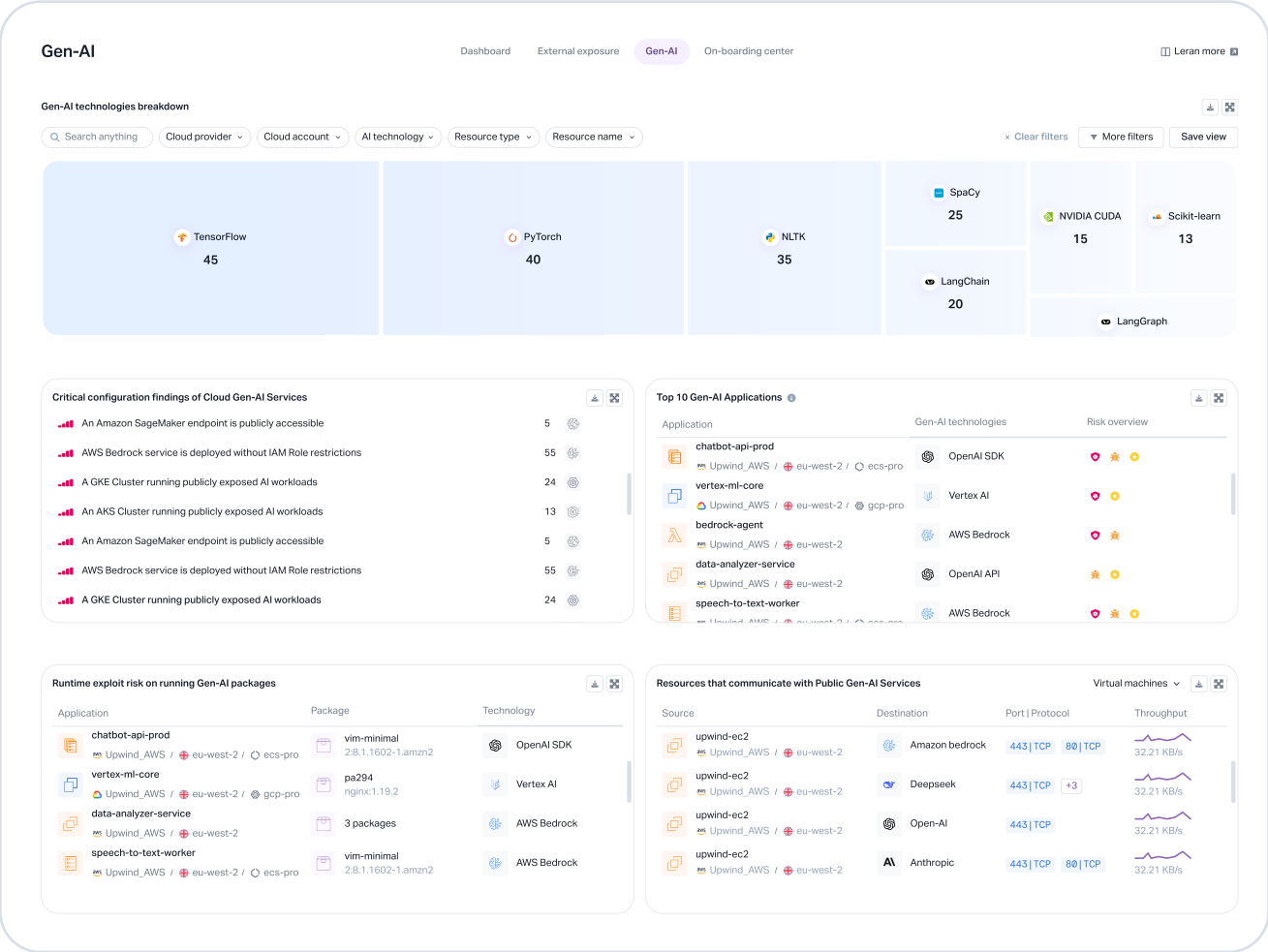
The Risk:

GenAI workloads often span multi-cloud environments, Kubernetes clusters, and ephemeral compute. Misconfigurations and poor cloud posture can expose critical paths where AI models interact with data or downstream systems.

Best Practice:

Use context-aware Cloud Security Posture Management (CSPM) focused on GenAI workloads. Customize posture rules to audit configurations of containers, storage, IAM roles, and network flows related to AI services. Enforce least-privilege principles and isolate AI compute pipelines.

5. Lack of GenAI-Specific Observability



The Risk:

Traditional observability tools don't provide visibility into AI-specific behaviors—such as data flowing through embeddings, token generation anomalies, or AI-assisted lateral movement across systems.

Best Practice:

Enable GenAI-aware observability with telemetry pipelines that correlate infrastructure data with AI-specific communication paths. Map how data flows between services, track interactions with model APIs, and surface visibility into encrypted or obfuscated payloads.

About Upwind

Upwind defends against advanced threats with GenAI-specific threat policies, real-time GenAI communication path mapping, and sensitive data discovery - offering deep visibility into data egress patterns to minimize exposure risks.

Secure Your GenAI Workloads with Upwind

Upwind delivers full-stack GenAI security by operationalizing these best practices:

- Real-time visualization of AI communication paths across layers 3, 4, and 7
- Sensitive data discovery using AI and regex-based payload inspection
- Threat detection with dynamic baselines and GenAI-focused policies
- Customizable CSPM rules for AI-centric cloud posture hardening

GenAI Inventory | GenAI Runtime activity | GenAI Threat detection | GenAI Security Posture

Want to know more about Upwind's GenAI security solution? Visit www.upwind.io or send us a note at hello@upwind.io to schedule a brief demo and see real-time security in action.