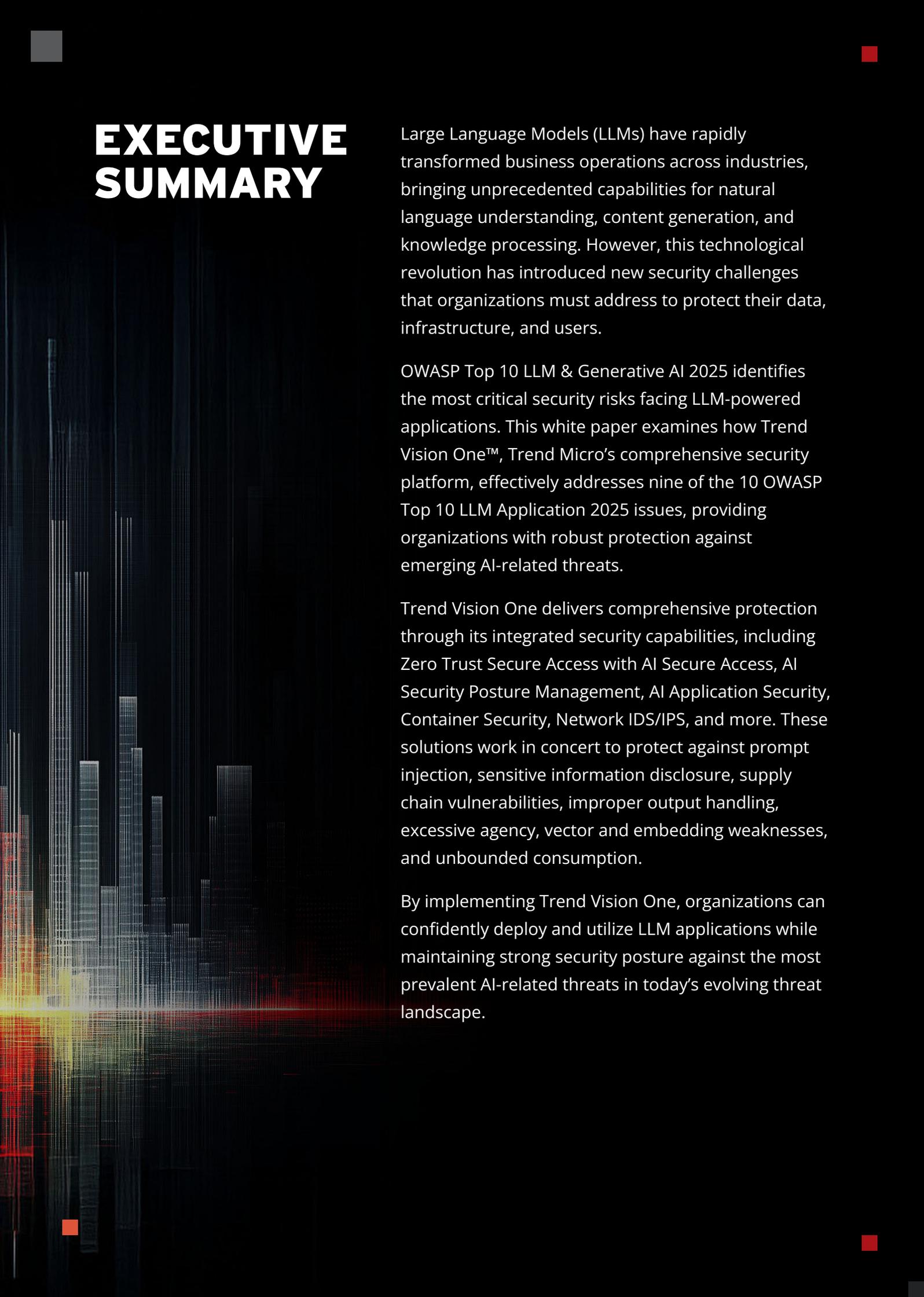


SECURING THE FUTURE

How Trend Vision One™ Addresses OWASP
Top 10 LLM & Generative AI Security Risks 2025

Fernando Cardoso, Dave McDuff, Fernando Tucci,
Kim Kinahan, and David Girard

UPDATED OCTOBER 14, 2025



EXECUTIVE SUMMARY

Large Language Models (LLMs) have rapidly transformed business operations across industries, bringing unprecedented capabilities for natural language understanding, content generation, and knowledge processing. However, this technological revolution has introduced new security challenges that organizations must address to protect their data, infrastructure, and users.

OWASP Top 10 LLM & Generative AI 2025 identifies the most critical security risks facing LLM-powered applications. This white paper examines how Trend Vision One™, Trend Micro's comprehensive security platform, effectively addresses nine of the 10 OWASP Top 10 LLM Application 2025 issues, providing organizations with robust protection against emerging AI-related threats.

Trend Vision One delivers comprehensive protection through its integrated security capabilities, including Zero Trust Secure Access with AI Secure Access, AI Security Posture Management, AI Application Security, Container Security, Network IDS/IPS, and more. These solutions work in concert to protect against prompt injection, sensitive information disclosure, supply chain vulnerabilities, improper output handling, excessive agency, vector and embedding weaknesses, and unbounded consumption.

By implementing Trend Vision One, organizations can confidently deploy and utilize LLM applications while maintaining strong security posture against the most prevalent AI-related threats in today's evolving threat landscape.

TABLE OF CONTENTS

4

Introduction

5

Understanding the 2025
OWASP Top 10 Risks and
Mitigations for LLMs
and Gen AI Applications

7

Trend Vision One™
Platform Overview

9

How Trend Vision One
Addresses OWASP Top 10
LLM Application Issues

18

Implementation
Considerations

20

Conclusions

Introduction

The emergence of Large Language Models (LLMs) and Generative AI has changed how organizations process, analyze, and generate information. These powerful AI systems have been integrated into numerous applications, across all industries, from customer service chatbots to content generation tools and decision support systems. However, this technological advancement comes with a new set of security challenges that organizations must address.



The OWASP LLM and Gen AI Security Project, a global initiative dedicated to identifying and mitigating security risks associated with large language models, has recognized these emerging threats and developed the OWASP Top 10 for LLM Applications 2025. This list, which was created in May 2023 and updated annually, identifies the most critical security risks facing LLM-powered applications, reflecting the evolving landscape of AI threats and vulnerabilities. These risks range from prompt injection attacks to sensitive information disclosure, supply chain vulnerabilities, and more.

As organizations increasingly adopt LLM technologies, they need robust security solutions to protect against these new threats. Trend Vision One, Trend Micro's comprehensive security platform, offers a range of capabilities designed to address the unique security challenges posed by LLM applications.

This white paper explores how Trend Vision One effectively addresses nine of the 2025 OWASP Top 10 LLM and Gen AI Application issues, providing organizations with the tools they need to secure their LLM implementations. By understanding these threats and implementing appropriate security measures, organizations can confidently leverage the power of LLMs while maintaining a strong security posture.

Understanding the 2025 OWASP Top 10 Risks and Mitigations for LLMs and Gen AI Applications

The OWASP Top 10 for LLM Applications 2025 represents the most critical security risks facing LLM-powered applications. Understanding these risks is essential for organizations looking to secure their LLM implementations. The ten risks identified by OWASP are:

1. **LLM01:2025 Prompt Injection:** Occurs when user prompts alter the LLM's behavior or output in unintended ways, potentially causing them to violate guidelines, generate harmful content, enable unauthorized access, or influence critical decisions.
2. **LLM02:2025 Sensitive Information Disclosure:** Involves the risk of LLMs exposing sensitive data, proprietary algorithms, or confidential details through their output, resulting in unauthorized data access, privacy violations, and intellectual property breaches.
3. **LLM03:2025 Supply Chain:** Addresses vulnerabilities in LLM supply chains that can affect the integrity of training data, models, and deployment platforms, potentially resulting in biased outputs, security breaches, or system failures.



4. **LLM04:2025 Data and Model Poisoning:** Occurs when pre-training, fine-tuning, or embedding data is manipulated to introduce vulnerabilities, backdoors, or biases, compromising model security, performance, or ethical behavior.
5. **LLM05:2025 Improper Output Handling:** Refers to insufficient validation, sanitization, and handling of LLM-generated outputs before they are passed to other components and systems, potentially leading to security vulnerabilities.
6. **LLM06:2025 Excessive Agency:** Involves the risks associated with granting LLMs too much autonomy or access to functions and systems, which can lead to unintended consequences if the LLM malfunctions or is manipulated.
7. **LLM07:2025 System Prompt Leakage:** Concerns the risk that system prompts or instructions used to guide the LLM's behavior might contain sensitive information that could be exposed to unauthorized users.
8. **LLM08:2025 Vector and Embedding Weaknesses:** Addresses security risks in systems using Retrieval Augmented Generation (RAG) with LLMs, where weaknesses in vector and embedding handling can lead to data leakage or manipulation.
9. **LLM09:2025 Misinformation:** Focuses on the risk of LLMs producing false or misleading information that appears credible, potentially leading to security breaches, reputational damage, and legal liability.
10. **LLM10:2025 Unbounded Consumption:** Involves the risk of LLM applications allowing excessive and uncontrolled inferences, leading to denial of service, economic losses, model theft, and service degradation.

Trend Vision One currently addresses nine of these 10 risks, with capabilities to address LLM09 in development. The following sections will detail how Trend Vision One's existing capabilities help organizations mitigate these critical security risks.

Trend Vision One™

Platform Overview

Trend Vision One is Trend Micro's enterprise cybersecurity platform that provides integrated protection across AI, endpoints, networks, cloud environments, email systems and many more. The platform combines exposure and risk management, advanced threat detection, response capabilities, and runtime protection into a unified solution, enabling organizations to effectively defend against a wide range of cyber threats, including those targeting LLM applications.

Key Components of Trend Vision One

ZTSA AI Secure Access

AI Secure Access is an advanced capability within Zero Trust Secure Access that provides content inspection and access control for public and private Generative AI services. It helps prevent prompt injection, malicious large language model attacks, and data leakage through advanced content inspection and filtering.

Key features include:

- Monitoring and blocking based on data filtering rules, content violations, or potential malicious content.
- Support for popular public Generative AI services like ChatGPT, Gemini, Microsoft Copilot and Claude and DeepSeek.
- Supports content inspection on major GenAI API providers such as Anthropic API and Amazon Bedrock.
- Bidirectional traffic inspection of both user prompts and AI-generated responses.

AI Security Posture Management (AI-SPM)

AI Security Posture Management provides visibility into cloud assets used to build AI services, including threats, misconfigurations, and potential attack paths. It displays interactive widgets summarizing key information about AI-related assets in connected cloud accounts across categories like cloud services, workloads, models, data storage, and entitlements.

AI Application Security

AI Application Security protects AI models and applications from vulnerabilities, malicious prompts, and data leaks using two capabilities. AI Scanner identifies and helps fix risks before deployment, while AI Guard blocks real-time threats such as prompt injection and sensitive data leakage. Together, they help ensure compliance without slowing performance.

Trend Micro TippingPoint™

Trend Micro TippingPoint delivers real-time, in-line threat protection for AI infrastructure by preventing the exploitation of vulnerabilities through network-based attacks without performance degradation. It enables the detection and blocking of zero-day vulnerabilities and leverages threat intelligence from the Zero Day Initiative to protect critical network assets before patches are available.

Endpoint Security: Server & Workload Protection (SWP) + Intrusion Prevention System (IPS)

Within Trend Vision One our SWP IPS rules safeguard AI servers and workloads against known and zero-day vulnerabilities through automated virtual patching. This host-based protection is especially valuable for continuous protection of critical AI infrastructure by automatically applying recommended intrusion prevention rules based on security priorities and vulnerability scan results.

Container Security

Container Security provides flexible control over what is deployed into your containerized environments, ensuring that only trusted containers get deployed and keeping your pipeline monitored for potential vulnerabilities, malware, secrets, and or compliance violations within the images even before they are deployed to production. Once a container is running, it is protected by our runtime scanning engine that provide visibility into your container activity, detecting events mapped to the MITRE ATT&CK framework and notifying you of container drift.

File Security

Trend Vision One File Security delivers multi-vector, automated scanning to detect and stop file-based threats across cloud storage, endpoints, servers, and hybrid environments. For AI workloads, it safeguards training data, model artifacts, and production pipelines from malware, data poisoning, and supply chain threats—ensuring secure AI development and deployment.

By combining these components, Trend Vision One offers a comprehensive approach to securing LLM applications against the various threats identified in the OWASP Top 10 for LLM Applications 2025.

How Trend Vision One Addresses OWASP Top 10 LLM Application Issues

LLM01:2025 Prompt Injection

Prompt Injection occurs when user prompts alter the LLM's behavior or output in unintended ways. These inputs can affect the model even if they are imperceptible to humans — such as through adversarial character strings, encoded instructions, hidden directives in images accompanying text, or split payloads across multiple inputs. Such attacks can cause models to violate guidelines, generate harmful content, reveal sensitive system information, execute unauthorized commands in connected systems, or manipulate critical decision-making processes while evading human detection.



How Trend Vision One Addresses Prompt Injection

1. **Define and validate expected output formats:** ZTSA AI Secure Access implements strict validation of output formats to ensure that LLM responses conform to expected patterns and structures, making it harder for malicious prompts to generate harmful outputs.

2. **Implement input and output filtering:** ZTSA AI Secure Access provides advanced content inspection and filtering capabilities that analyze both user prompts and AI responses. This bidirectional inspection helps identify and block potential prompt injection attempts before they can affect the LLM's behavior.

3. **Enforce privilege control and least privilege access:** ZTSA AI Secure Access implements strict role-based access controls that limit what users can do with AI services, reducing the potential impact of successful prompt injection attacks.
4. **Conduct adversarial testing and attack simulations:** Trend Vision One AI Application Security, provides capabilities for adversarial testing and attack simulations to identify and address prompt injection vulnerabilities before they can be exploited.

LLM02:2025 Sensitive Information Disclosure

Sensitive Information Disclosure involves the risk of LLMs exposing sensitive data, proprietary algorithms, or confidential details through their output. This can result in unauthorized data access, privacy violations, and intellectual property breaches.

How Trend Vision One Addresses Sensitive Information Disclosure

Trend Vision One provides multiple layers of protection against sensitive information disclosure:

1. **Apply mitigations from OWASP Top Ten's "A06:2021 – Vulnerable and Outdated Components":** Trend Vision One Cyber Risk and Exposure Management (CREM) capabilities help detect and mitigate vulnerabilities in components that could lead to sensitive information disclosure. The platform provides both agentless vulnerability scanning for cloud resources and agent-based scanning for endpoints, along with virtual patching through TippingPoint and Endpoint Security SWP IPS.
2. **Conduct comprehensive AI Red Teaming and Evaluations:** In future development, Trend Vision One will provide capabilities for comprehensive AI red teaming and evaluations to identify sensitive information disclosure vulnerabilities introduced from training data, RAG databases, or agent tooling.
3. **Maintain an up-to-date inventory of components:** AI Security Posture Management (AI-SPM) maintains an up-to-date inventory of AI-related cloud assets, including software components, helping organizations identify and address potential vulnerabilities that could lead to information disclosure.

3. **Implement strict monitoring and auditing practices:** ZTSA AI Secure Access provides robust monitoring and auditing capabilities for collaborative model development environments, helping organizations detect and prevent supply chain attacks that could compromise their LLM applications.
4. **Secure container environments for LLM deployments:** Trend Vision One Container Security scans container images for vulnerabilities, malware, and compliance violations throughout the development pipeline, ensuring that containerized LLM environments remain secure from supply chain threats before deployment to production.
5. **Enforce security policies for container deployments:** Trend Vision One Container Security enables organizations to implement admission policies that ensure only containers meeting specific security requirements can run in production environments, reducing the risk of compromised components in the LLM supply chain. By implementing these controls, Trend Vision One helps organizations protect their LLM applications from supply chain vulnerabilities that could otherwise lead to security breaches or system failures.

LLM04:2025 Data and Model Poisoning

Data and model poisoning occurs when pre-training, fine-tuning, or embedding data—or the model artifact itself—is manipulated to introduce biases, backdoors, or degraded behavior. Attacks may remain dormant until a specific trigger phrase or context activates them, causing harmful outputs, integrity loss, or downstream exploitation in production.

How Trend Vision One Addresses Data and Model Poisoning

1. **Detect hidden triggers and block poisoned content at runtime:** Using the capabilities within AI Application Security, AI Scanner can perform automated adversarial evaluations against LLM endpoints (e.g., backdoor triggers, prompt driven integrity drift, leakage). Then convert findings into AI Guard policies that enforce input/output guardrails, including RAG context screening and harmful content filtering. Poisoned prompts or retrieved corpus fragments are blocked before they reach the model or users.

2. **Control and sanitize datasets and model artifacts:** Use AI Security Posture Management (AI-SPM) to gain visibility into AI-related cloud assets, detect misconfigurations in data storage and model infrastructure, identify vulnerabilities, and assess potential attack paths that could enable data or model poisoning attacks. Use Trend Vision One File Security to scan the training data, files used for fine-tuning, embedded data, and model files for malware before they are used or deployed. This helps ensure that any files or data given to your AI system are safe and free from malicious software.
3. **Harden the build and deployment pipeline to prevent poisoned dependencies:** Use Container Security (image scanning, admission control, runtime protection) and Code Security to block risky libraries and flag unsafe model-loading patterns (e.g., pickle-based deserialization or remote-code loaders) from entering CI/CD, and to detect anomalous behavior at runtime
4. **Use File Integrity Monitoring** on the training data & LLM model files to detect poisoning attempts. This is available in Workload Security agents.

LLM05:2025 Improper Output Handling

Improper Output Handling refers to insufficient validation, sanitization, and handling of LLM-generated outputs before they are passed to other components and systems. This can lead to security vulnerabilities such as XSS, CSRF, SSRF, privilege escalation, or remote code execution.

Trend Vision One Addresses Improper Output Handling

ZTSA AI Secure Access

ZTSA AI Secure Access protects **user access to commercial AI services** (e.g., ChatGPT, Gemini) and mitigates improper output handling through:

1. **Output validation and sanitization**
Ensures LLM responses are cleansed of potentially malicious content before reaching downstream systems.
2. **Rule-based content filtering**
Detects and blocks or logs harmful output patterns using customizable filtering rules.
3. **Rate limiting and throttling**
Controls the volume and frequency of LLM outputs to prevent abuse or system overload.

AI Application Security: AI Scanner and AI Guard

AI Application Security is designed to protect **an organization's own AI applications**—such as those embedded in websites or mobile apps—from improper output handling risks.

1. **AI Scanner – Pre-deployment Output Risk Detection**

Simulates adversarial prompts and output flows to identify vulnerabilities like prompt injection, unsafe output formatting, and data leakage before deployment. It helps teams proactively fix output handling flaws before they reach production.

2. **AI Guard – Runtime Output Enforcement**

Monitors and enforces output safety policies in real time. It blocks malicious or non-compliant LLM outputs—such as those containing executable code, sensitive data, or unsafe formatting—before they are delivered to users or integrated systems.

LLM06:2025 Excessive Agency

Excessive Agency involves the risks associated with granting LLMs too much autonomy or access to functions and systems. This can lead to unintended consequences if the LLM malfunctions or is manipulated, particularly through extensions or plugins that allow the LLM to interact with other systems.

How Trend Vision One Addresses Excessive Agency

Trend Vision One prevents excessive agency through several key capabilities:

1. **Implement strict role-based access controls:** ZTSA AI Secure Access implements strict role-based access controls that limit what LLMs can do and what systems they can interact with, reducing the risk of excessive agency leading to security breaches.
2. **Implement robust logging and auditing mechanisms:** AI Security Posture Management (AI-SPM) provides robust logging and auditing capabilities for LLM operations, helping organizations detect and respond to instances of excessive agency before they can cause significant harm.
3. **Regularly review and update LLM permissions and capabilities:** AI Security Posture Management (AI-SPM) helps organizations detect misconfigurations between cloud resources and LLMs, ensuring that LLMs only have the permissions and capabilities they need to perform their intended functions.

LLM07:2025 System Prompt Leakage

System prompt leakage occurs when hidden instructions that steer model behavior are exposed through direct or indirect interaction with an LLM. While the disclosure can reveal internal rules, roles, or engineering choices, the real risk arises when secrets (e.g., credentials) or authorization logic are in the prompt, or when guardrails rely on the model itself rather than deterministic controls. Attackers commonly combine indirect prompt injection with output reflection to extract these details and then pivot to privilege bypass or data exfiltration.

How Trend Vision One Addresses System Prompt Leakage

1. **Shift left to find leakage paths pre-deployment:** Trend Vision One™ AI Scanner performs red teaming of LLM applications and APIs to detect system prompt reflection, indirect prompt injection, verbose error disclosures, and related behaviors that enable system prompt leakage. The scanner simulates attacker techniques and produces remediation guidance so teams remove leakage triggers before release.
2. **Enforce guardrails outside the model at runtime:** Trend Vision One™ AI Guard mediates inputs and outputs at the access layer (API/gateway), blocking prompts that coerce disclosure of hidden instructions and preventing responses that echo system prompts or leak sensitive data.
3. **Prevent sensitive data in prompts and outputs (data loss controls):** Use AI Guard's pattern / PII / secret detection with its ability to block found patterns in prompts and outputs during application runtime of your AI Applications. Additionally, also use Trend Vision One Code Security & Container Security secrets scanning to ensure credentials, keys, and connection strings are not placed in system prompts and cannot be exfiltrated if echoed.
4. **Keep authorization and privilege checks out of the LLM (posture & API hardening):** Use AI Security Posture Management (AI-SPM) to inventory AI assets, tighten identities/permissions, and enforce least privilege on tools the model can reach.
5. **Control employee use of external LLM services (SaaS mediation):** Use ZTSA AI Secure Access to inspect both prompts and responses of public AI Services like ChatGPT to prevent prompt-injection-driven leakage, or sensitive data disclosure.

6. **Continuously validate and monitor (scan → protect → validate → improve):** Operate a continuous loop in which AI Scanner re-tests production endpoints and AI Guard telemetry informs policy tuning. This ensures new jailbreaks and injection patterns that could re-expose system prompts are discovered and mitigated without code change.

LLM08:2025 Vector and Embedding Weaknesses

Vector and Embedding Weaknesses present significant security risks in systems utilizing Retrieval Augmented Generation (RAG) with LLMs. Weaknesses in how vectors and embeddings are generated, stored, or retrieved can be exploited to inject harmful content, manipulate model outputs, or access sensitive information.

How Trend Vision One Addresses Vector and Embedding Weaknesses

Trend Vision One provides protection against vector and embedding weaknesses through several key capabilities:

1. **Regularly update and patch vector database systems:** Trend Vision One's vulnerability patch management capabilities help organizations keep their vector database systems up-to-date with the latest security patches, reducing the risk of exploitation through known vulnerabilities. Additionally, if the vector database is containerized, Trend Vision One Container Security can detect exploitation attempts.
2. **Monitor for unusual patterns in vector queries or embeddings:** Trend Vision One AI Application Security provides visibility into vector queries and responses by intercepting them and inspecting responses for sensitive content. This helps organizations detect and respond to potential attacks targeting vector and embedding weaknesses.

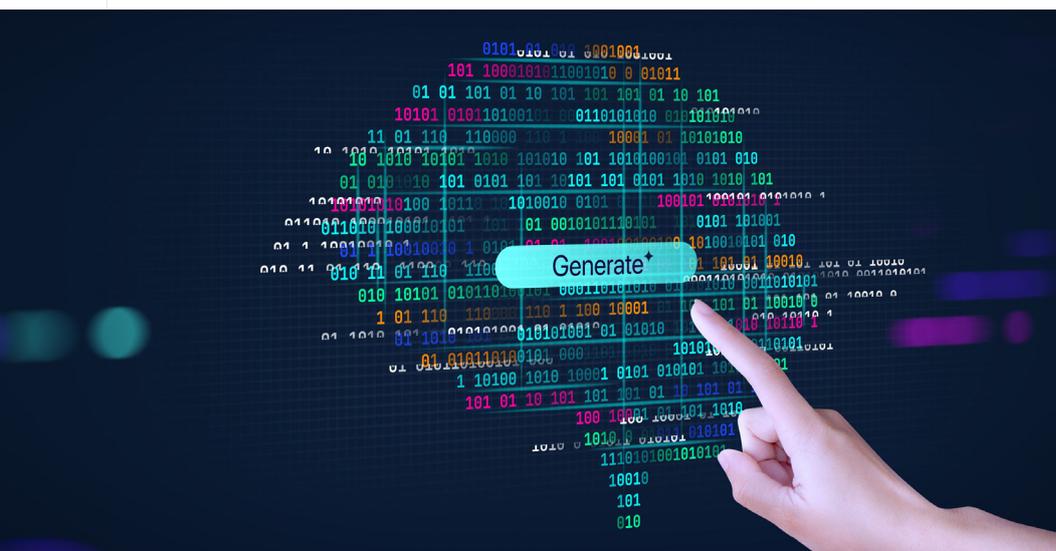
LLM10:2025 Unbounded Consumption

Unbounded Consumption involves the risk of LLM applications allowing excessive and uncontrolled inferences, leading to denial of service, economic losses, model theft, and service degradation. The high computational demands of LLMs, especially in cloud environments, make them vulnerable to resource exploitation and unauthorized usage.

How Trend Vision One Addresses Unbounded Consumption

Trend Vision One's ZTSA AI Secure Access provides protection against unbounded consumption through its rate limiting capabilities:

1. **Use rate limiting and throttling mechanisms:** ZTSA rate limiting helps control the number of requests that can be made to LLM services within a given time-period, preventing excessive consumption that could lead to denial of service or economic losses.



Implementation Considerations

When implementing Trend Vision One to address OWASP Top 10 LLM Application issues, organizations should consider the following:

Supported AI Services

ZTSA AI Secure Access currently supports most Generative AI services available to the public. Such as ChatGPT and Google Gemini. Check our documentation for up-to-date information [here](#).

Deployment Options

Organizations have multiple options for deploying ZTSA AI Secure Access:

1. Secure Access Module: Installed on end-user devices and integrated with IAM solutions
2. Traffic forwarding: Using proxy-based solutions like PAC files, proxy chaining, or port forwarding

The appropriate deployment option will depend on the organization's existing infrastructure and security requirements. Container Security can protect containers in multiple deployment environments:

1. Kubernetes: Both Cloud (examples like Amazon EKS, Microsoft Azure AKS, and Google GKE) and On-Premise through standard Kubernetes clusters and OpenShift
2. AWS Container Services including: Amazon ECS and AWS Fargate

Integration With Existing Security Controls

Trend Vision One works best when integrated with existing security controls. Organizations should consider how Trend Vision One will complement their current security architecture, including:

- Identity and Access Management (IAM) systems
- Endpoint security solutions
- Network security controls
- Cloud security posture management

Monitoring and Response Capabilities

To maximize the effectiveness of Trend Vision One in addressing LLM application security issues, organizations should establish robust monitoring and response capabilities:

- Regularly review AI Secure Access logs and alerts
- Establish incident response procedures for LLM-related security incidents
- Conduct regular security assessments of LLM applications
- Update security policies and controls based on emerging threats

By considering these implementation factors, organizations can effectively deploy Trend Vision One to address OWASP Top 10 LLM Application issues and maintain a strong security posture for their LLM applications.

Conclusion

As Large Language Models continue to transform business operations across industries, organizations must address the unique security challenges these powerful AI systems introduce. The OWASP Top 10 for LLM Applications 2025 provides a valuable framework for understanding these challenges, and Trend Vision One offers comprehensive protection against nine of these 10 critical security risks.

Through its integrated security capabilities — including Zero Trust Secure Access with AI Secure Access, AI Security Posture Management, AI Application Security, File Security, Container Security, TippingPoint and Endpoint Security, Trend Vision One helps organizations protect their LLM applications from prompt injection, sensitive information disclosure, supply chain vulnerabilities, improper output handling, excessive agency, vector and embedding weaknesses, and unbounded consumption.

While capabilities for addressing misinformation (LLM09) are still in development, Trend Vision One's existing protections provide a solid foundation for securing LLM applications against the most prevalent threats.

By implementing Trend Vision One and following the recommended implementation considerations, organizations can confidently deploy and utilize LLM applications while maintaining a strong security posture. As the threat landscape continues to evolve, Trend Micro remains committed to enhancing Trend Vision One's capabilities to address emerging AI-related security challenges, ensuring that organizations can safely harness the power of LLMs to drive innovation and growth.

Sources and Reference Materials

1. [OWASP Top 10 for LLM and GenAI Applications 2025](#)
2. [Trend Vision One ZTSA AI Secure Access](#)
3. [Trend Vision One Cyber Risk Exposure Management – AI-SPM](#)
4. [Trend Vision One - AI Application Security](#)
5. [Trend Vision One Endpoint Security - Deepfake Protection](#)
6. [Trend Vision One Network Security - Network Detection and Response \(NDR\)](#)
7. [Trend Vision One Network Security - TippingPoint](#)
8. [OWASP Application Security Verification Standard \(ASVS\)](#)
9. [MITRE ATLAS \(Adversarial Threat Landscape for Artificial-Intelligence Systems\)](#)
 - a. [AI Model Tampering via Supply Chain Attack](#)

Glossary of Terms

AI Security Posture Management (AI-SPM): A capability within Trend Vision One that provides visibility into cloud assets used to build AI services, including threats, misconfigurations, and potential attack paths.

AI Secure Access: A capability within Zero Trust Secure Access that provides content inspection and access control for public and private Generative AI services.

Embedding: The process of converting text or other data into numerical vectors that capture semantic meaning, used by LLMs to process and understand information.

Excessive Agency: The vulnerability that occurs when an LLM is granted too much autonomy or access to functions and systems, potentially leading to unintended consequences.

Hallucination: When an LLM generates content that seems accurate but is fabricated, often occurring when the model fills gaps in its training data using statistical patterns.

Large Language Model (LLM): An advanced AI system trained on vast amounts of text data to understand and generate human-like language.

Prompt Injection: A vulnerability where user inputs manipulate an LLM's behavior or output in unintended ways, potentially causing it to violate guidelines or perform unauthorized actions.

RAG (Retrieval Augmented Generation): A technique that enhances LLM outputs by retrieving relevant information from external sources before generating responses.

Sensitive Information Disclosure: The risk of LLMs exposing sensitive data, proprietary algorithms, or confidential details through their output.

Supply Chain Vulnerability: Security risks in the components, data, or models used to create and deploy LLM applications.

System Prompt: Instructions provided to an LLM to guide its behavior and responses, which may contain sensitive information.

Unbounded Consumption: The vulnerability that occurs when an LLM application allows excessive and uncontrolled inferences, potentially leading to denial of service or economic losses.

Vector Database: A specialized database designed to store and query vector embeddings efficiently, often used in RAG systems.

Virtual Patching: A security technique that applies protection against known vulnerabilities before official patches are available. Trend Micro TippingPoint and Trend Vision One Endpoint Security (Server & Workload Protection with Intrusion Prevention System) offer these capabilities.

Zero Trust Secure Access: A security model that eliminates implicit trust and requires continuous verification of users and devices before granting access to resources.

Want more insights like this?

[TrendMicro.com/ai](https://www.trendmicro.com/ai)

SECURING THE FUTURE



Trend Micro, a global cybersecurity leader, helps make the world safe for exchanging digital information. Fueled by decades of security expertise, global threat research, and continuous innovation, Trend Micro's AI-powered cybersecurity platform protects hundreds of thousands of organizations and millions of individuals across clouds, networks, devices, and endpoints. As a leader in cloud and enterprise cybersecurity, Trend's platform delivers a powerful range of advanced threat defense techniques optimized for environments like AWS, Microsoft, and Google, and central visibility for better, faster detection and response. With 7,000 employees across 70 countries, Trend Micro enables organizations to simplify and secure their connected world.

For more information visit [www.TrendMicro.com](https://www.trendmicro.com).