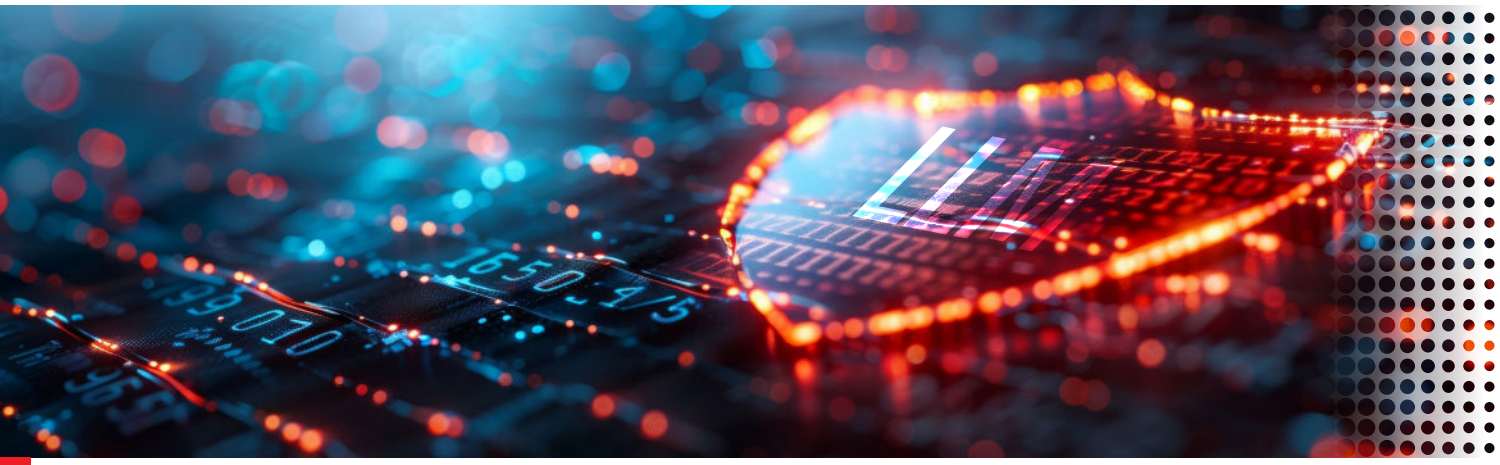


Radware LLM Firewall

Ensure safe, responsible AI for your modern organization.

Background

The use of AI to augment one's business and serve customers more effectively and efficiently continues to rise exponentially. This trend spans all industries ranging from education to healthcare, finance, commerce, government and more. Radware's 2025 cybersecurity survey, focusing on the adoption of AI and the security concerns that come with it, shows that 54% of organizations across the world are already adopting AI in their businesses. An additional 36% of these organizations have concrete plans to deploy AI within the next twelve months. When asked about AI-related security concerns, 65% to 70% of the organizations replied that they are concerned about a range of AI security issues including the disclosure of sensitive corporate and personal information, brand and reputational damage, and misinformation—all causing financial losses and business disruption.

Large language models (LLMs) are reshaping industries, unlocking unparalleled innovation and efficiency. But with this progress comes a serious concern: new cybersecurity risks that organizations must confront. LLM integration exposes businesses to vulnerabilities that can no longer be ignored. To stay ahead, it's crucial to understand these emerging threats. Identifying and fighting these risks now will better prepare you for the challenges ahead, allowing you to navigate the future of AI with confidence.

New Vulnerabilities Introduced by LLM

Integrating LLMs into applications introduces new vulnerabilities on top of existing ones.

- **Data Extraction from LLMs – Risk of Data Breaches and Competitive Loss**

  Attackers can extract sensitive data embedded within the LLM, potentially exposing private user information or business confidential data.

- **Adversarial Manipulation of Outputs – Leading to Misinformation**

  Attackers can manipulate LLMs to generate false or harmful content, resulting in the spread of misinformation, reputational damage, or influence over public opinion.

- **Model Inversion Attacks – Risk of Privacy Violations and Intellectual Property Theft**

  Attackers may reverse-engineer the LLM to reveal training data, exposing private and sensitive information such as personal details, trade secrets, or other confidential data.

- **Prompt Injection and System Control Hijacking**

  Malicious prompt injections can alter the behavior of the LLM, potentially leading to the exposure of sensitive information or bypassing security measures.

Radware's LLM Firewall Solution

Radware introduces its new LLM Firewall solution, which secures generative AI use with real-time AI-based protection at the prompt level. It stops threats before they even reach the organization's origin servers. The solution enforces enterprise-grade security and compliance by detecting risks like prompt injection, data leaks, harmful content, brand safety, and usage policies in real time. The Radware solution is fully model-agnostic, easy to onboard and secures AI use across platforms without disrupting workflows or innovation.

Radware LLM Firewall is part of Radware's Cloud Application Protection Service, allowing all Radware cloud customers to join and add the new LLM firewall protection immediately and seamlessly. It aligns with and references the OWASP Top 10 for LLM Applications (2025), which provides a structured view of these new threats, offering a starting point for secure LLM integration across teams.

Key Customer Benefits

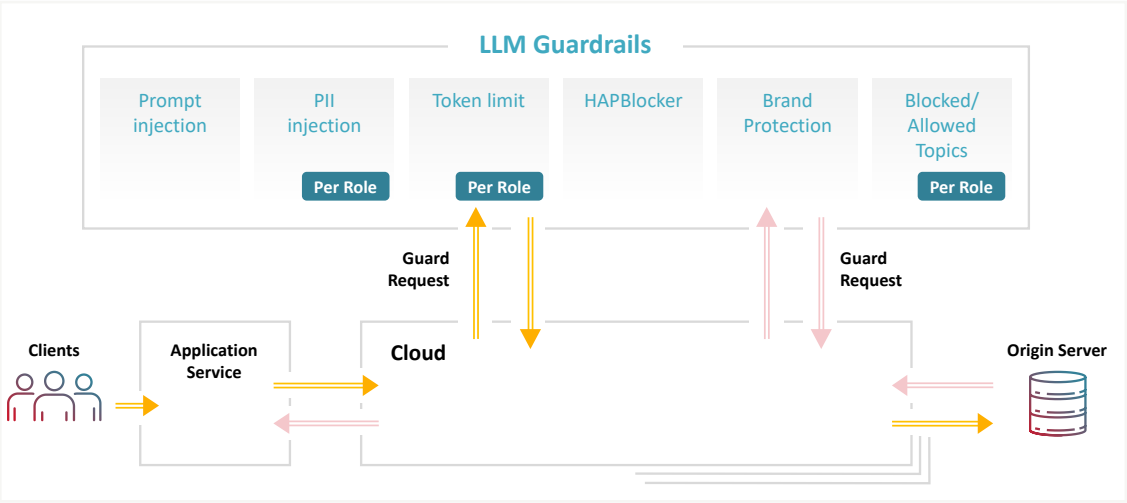
- **Protection Against OWASP Top 10 For LLM –** Radware LLM Firewall covers key threats from the OWASP Top 10 for LLM vulnerabilities for protection and compliancy.
- **Cost and Resource Savings –** Radware LLM Firewall is positioned in-line before the customer's own LLM and applications. Prompts blocked by Radware never even reach the customer's infrastructure. As a result, the organization saves LLM tokens, compute and network resources.
- **Faster Time to Market –** Onboarding the solution is immediate. It requires no changes to your infrastructure and little to no education of your personnel. Configuring the solution is intuitive and fast. This translates into super-fast time to market. Decide and start protection on the same day.
- **Continuous Control and Adaptation –** Radware LLM Firewall offers extensive visibility, allowing you to stay fully in control over your LLM and applications, adjust protections intuitively and quickly, and retain business continuity at all times.

Key Solution Capabilities

- **Inline, Pre-origin Protection –** Radware's LLM Firewall solution is uniquely positioned in that it is architecturally located even before the customer's own LLM engines and applications. Radware LLM Firewall catches the user prompt before it reaches the customer's own origin servers, thus blocking malicious usage early on. This architecture brings a considerable key advantage to customers in the form of economical savings (no prompt missuses), reduction in LLM costs, compute costs and application costs.
- **Zero-friction Onboarding and Assimilation –** Radware LLM Firewall requires virtually no integrations. There are no code exchanges, no application adaptations and no interruption to existing customer flows and behaviors. Simply configure and start using it.
- **Easy Configuration –** Radware's LLM Firewall solution is easy and intuitive to configure yet offers a wealth of customer-tailored configuration capabilities. One of its most important and unique capabilities is the option to create master-configuration templates. Each template is capable of serving multiple LLM models, prompts and applications, making LLM security easier to monitor and control.
- **Visibility With Tuning –** Radware's LLM Firewall solution offers extensive visibility with a full security event logging into user's actions, prompts, inputs and outputs. Dashboards also show the actions and interactions with each and every LLM module that the organization deploys. The solution can be activated in one of two modes: Blocking Mode or Report-only Mode, allowing false-positive tuning, safe rollout and continuous improvements and adjustments.

Figure 1:

Radware's LLM Firewall architecture

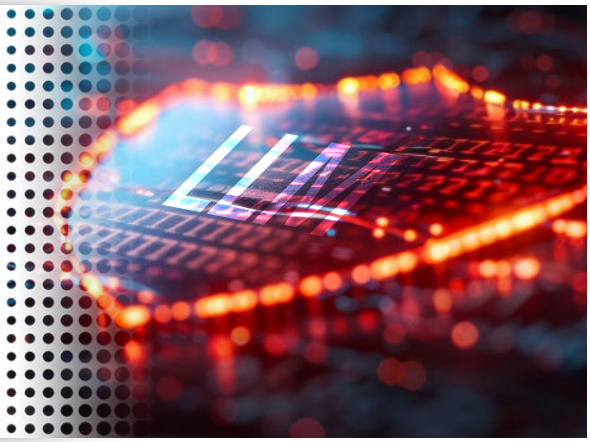


Summary

AI and LLM are already widely used by organizations to boost their businesses and services. The incorporation of LLM into one's business introduces new security risks that traditional security products and services cannot protect against. This calls for organizations to adopt and deploy new security solutions—ones that were born to detect and mitigate AI and LLM usage risks.

Organizations need to ensure that by utilizing AI and LLMs, they are preventing sensitive data from being exfiltrated and that LLM prompts are not hijacked to perform malicious actions that might cause brand damage, provide misleading information or cause potential lawsuits.

Radware LLM Firewall gives organizations the ideal solution to deal with these new and emerging security threats.



RWI 642 I 11.20.2025