DELL Technologies

Top 10 Cybersecurity Concerns for GenAI and LLMs

Get Started

Introduction

Artificial Intelligence (AI) is revolutionizing the way organizations operate, with Generative AI (GenAI) and Large Language Models (LLMs) becoming critical workloads in modern enterprise environments.

Just like any other workload, these applications come with their own set of complexities and vulnerabilities to address. As businesses continue to adopt AI to drive innovation, efficiency, and competitive advantage, ensuring the security of these applications becomes a foundational necessity. Good cyber hygiene is the foundation of securing any workload, and just as you prioritize security across all your workloads, it's essential to practice good cyber hygiene for AI too. This includes implementing practices such as proper system patching, multi-factor authentication, role-based access, and network segmentation. These measures are fundamental, but the key lies in understanding how these capabilities fit into the specific architecture and usage of your workload.

At Dell, we have a deep understanding of the AI workload and the unique security challenges it faces. By identifying the ways threat actors might target these workloads, Dell can help you create a robust security strategy. This includes addressing risks like: training data poisoning, model theft or manipulation, dataset reconstruction, and more.

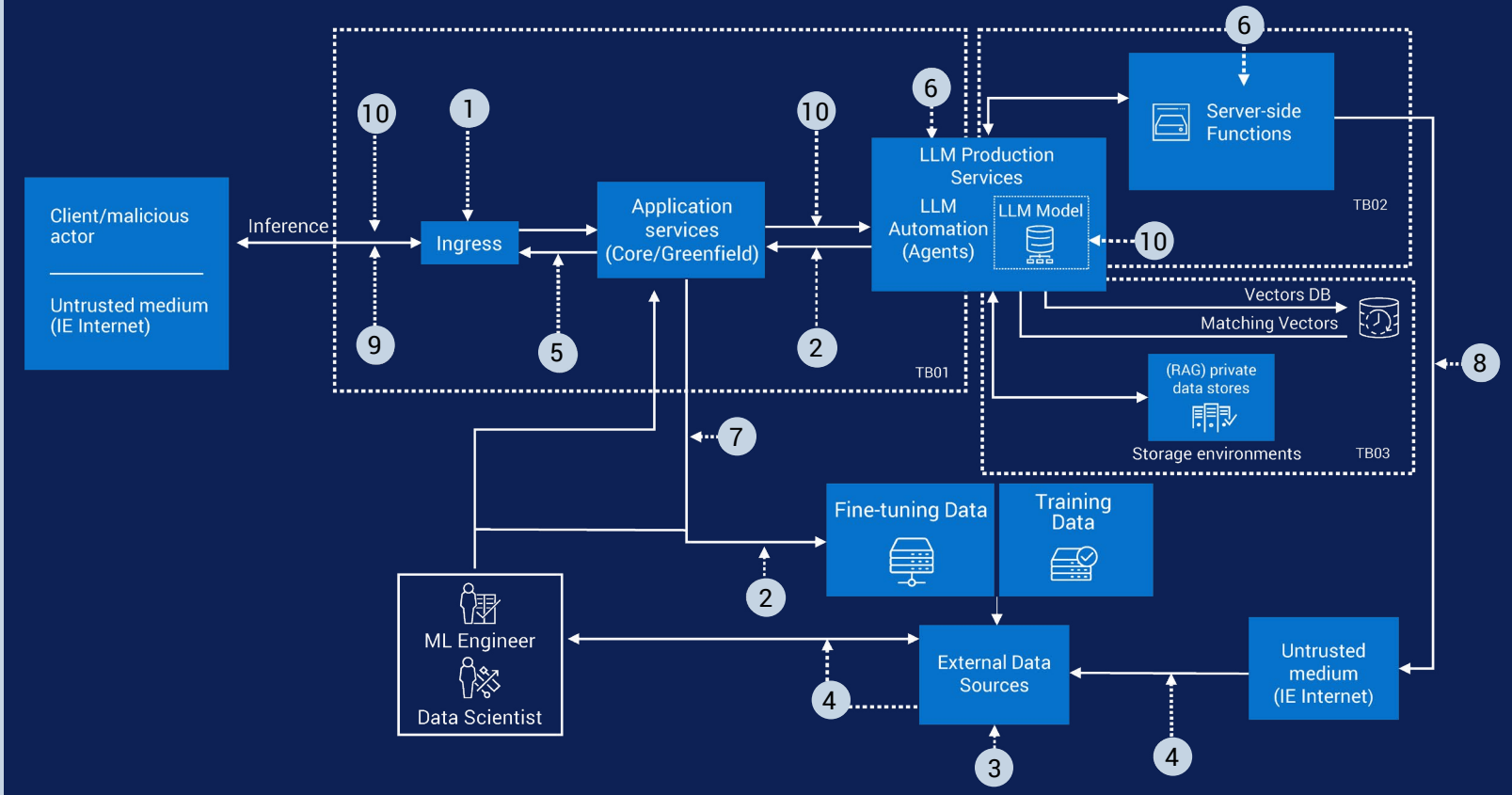
We also focus on managing challenges associated with the input to your AI model, such as preventing sensitive information disclosure, mitigating unsafe topics or bias, and ensuring compliance with regulations. On the output side, we help tackle issues like overdependence on the model and compliance-related risks.

At Dell, we empower enterprises to mitigate these risks by leveraging their existing cybersecurity solutions or exploring new tools and practices to safeguard their systems. Our goal is to ensure that security doesn't hinder your innovation. By understanding how AI workloads function and the security threats they face, we can help you build a stronger security posture, making your environment more resilient while enabling you to innovate with confidence. With our expertise, we help you confidently harness the potential of AI while maintaining robust security.

Top-10 Cybersecurity Concerns for GenAI and LLM

These are the top concerns for protecting GenAI/LLM models, as outlined by OWASP.
Click on a concern to learn more:

1. Prompt Injection
2. Sensitive Info Disclosure
3. Supply Chain
4. Model Data Poisoning
5. Improper Output handling
6. Excessive Agency
7. System Prompt Leakage
8. Vector & Embedding Weaknesses
9. Misinformation
10. Unbounded Consumption

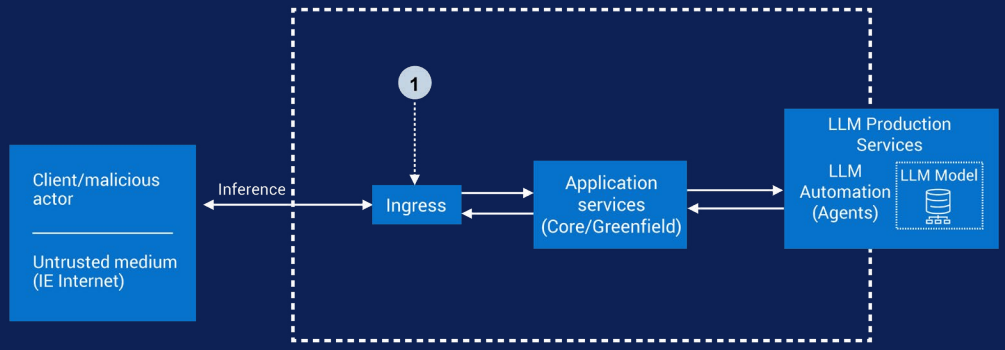


Concern # 1: Prompt Injection

Strategies for Mitigating Prompt Injection:

- **Data sanitization and input validation:** Screen user inputs thoroughly to remove harmful content. Use normalization and encoding to prevent misuse.
- **Natural Language Processing (NLP) and machine learning-based approaches:** Use NLP and machine learning to detect and block manipulated or malicious prompts.
- **Clear output formatting and response controls:** Set strict response boundaries to ensure outputs follow intended formats and prevent unauthorized actions. Use prompt filtering and response validation to maintain integrity.
- **Access restrictions and human oversight:** Apply role-based access control (RBAC), multi-factor authentication (MFA), and identity management to limit access. Use human review for critical decisions.
- **Monitoring, logging, and anomaly detection:** Continuously monitor and log AI system activities—using solutions like MDR/XDR/SIEM—to rapidly detect, investigate, and respond to unauthorized access, anomalies, or data leaks.
- **Secure prompt engineering:** Use secure prompt design and analysis as part of overall software security to protect input processing.
- **Model validation** Regularly validate ML models to ensure they haven't been tampered with before deployment, safeguarding their accuracy and integrity.
- **Prompt filtering, ranking and response validation:** Analyze and rank prompts to ensure only secure inputs are processed. Validate responses to prevent misuse.
- **Robustness checks:** Conduct regular evaluations to identify and fix vulnerabilities, keeping the AI secure and reliable.

Back to Architecture Diagram



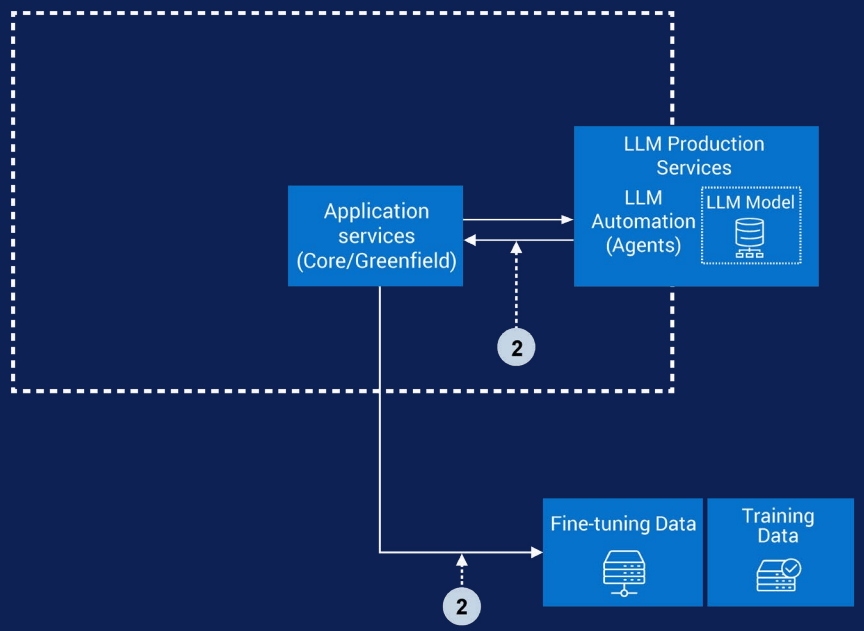
Prompt injection is an emerging challenge in the world of Generative AI (GenAI), where malicious inputs are crafted to manipulate the model's behavior or compromise its integrity. These attacks exploit vulnerabilities in the way AI systems process and respond to user inputs, potentially leading to unauthorized actions, misinformation, or the exposure of sensitive data. As GenAI becomes increasingly integrated into critical business workflows, addressing these risks is essential to maintaining trust and security

Concern # 2: Sensitive Information Disclosure

Strategies for Mitigating Sensitive Information Disclosure:

- **Data sanitization and input validation:** Screen user inputs thoroughly to remove harmful content. Use normalization and encoding to prevent misuse.
- **Utilize homomorphic encryption** to process sensitive data securely without exposing its contents. This ensures that even while data is in use, it remains encrypted and protected from breaches.
- **Access restrictions and human oversight:** Apply role-based access control (RBAC), multi-factor authentication (MFA), and identity management to limit access. Use human review for critical decisions.
- **Leverage secure APIs and system interfaces** for AI data interactions, routinely reviewing configurations to minimize exposure and attack surface.
- **Secure data collection, storage, and policies** and enforce comprehensive data protection and governance policies that ensure regulatory compliance and minimize data risk.
- **Monitoring, logging, and anomaly detection:** Continuously monitor and log AI system activities—using solutions like MDR/XDR/SIEM—to rapidly detect, investigate, and respond to unauthorized access, anomalies, or data leaks.
- **Secure development, configuration, and audits:** Apply secure coding practices, use automated configuration management tools, and conduct regular reviews, audits, and updates to keep AI system configurations secure and current.
- **User education and security awareness:** Provide ongoing, AI-specific security awareness training to users and administrators to reduce unsafe usage and accidental data disclosure.

Back to Architecture Diagram



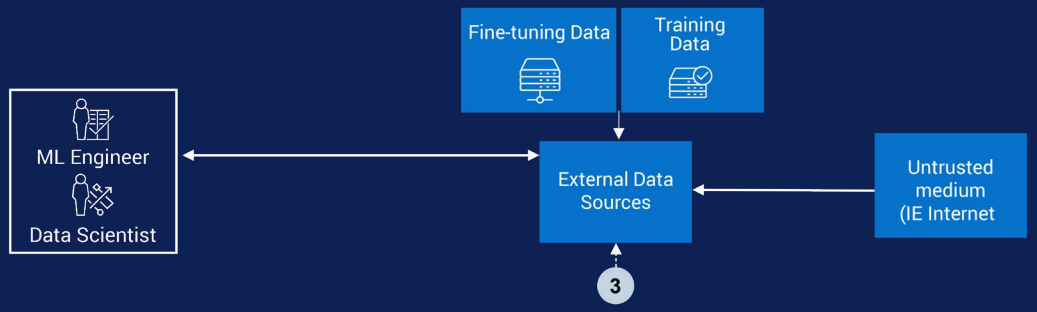
GenAI has brought incredible advancements, but it also comes with significant risks, particularly the unintentional exposure of sensitive information. Whether it's Personally Identifiable Information (PII) or proprietary business data, the misuse or mishandling of GenAI tools can lead to data leaks, regulatory non-compliance, or reputational damage. This makes it critical for organizations to understand these risks and proactively address them to ensure secure implementation and usage of AI systems.

Concern # 3: Supply Chain Vulnerabilities

Strategies for Mitigating Supply Chain Vulnerabilities:

- **Vet suppliers and ensure compliance with secure supply chain practices** Evaluate suppliers and establish agreements that prioritize supply chain security.
- **Implement a Software Bills of Materials** Track and verify the origins of software components, ensuring transparency and reducing the risk of compromised code.
- **Model validation** Regularly validate ML models to ensure they haven't been tampered with before deployment, safeguarding their accuracy and integrity.
- **Run containers and pods with least privileges** This reduces the potential impact in case of a compromise and limits unauthorized access.
- **Deploy firewalls** Block unnecessary network connectivity, reducing exposure to potential threats and limiting avenues for attackers.
- **Protect data and annotations** Secure your data and associated annotations to prevent tampering, unauthorized access, and corruption of critical information.
- **Secure hardware** Use hardware validated for security to prevent vulnerabilities that could arise from hardware-based attacks, ensuring a strong foundation for your infrastructure.
- **Secure ML software components** Use trusted and vetted ML software components to reduce vulnerabilities and enhance the overall security of your machine learning workflows.
- **Secure development, configuration, and audits:** Apply secure coding practices, use automated configuration management tools, and conduct regular reviews, audits, and updates to keep AI system configurations secure and current.

Back to Architecture Diagram



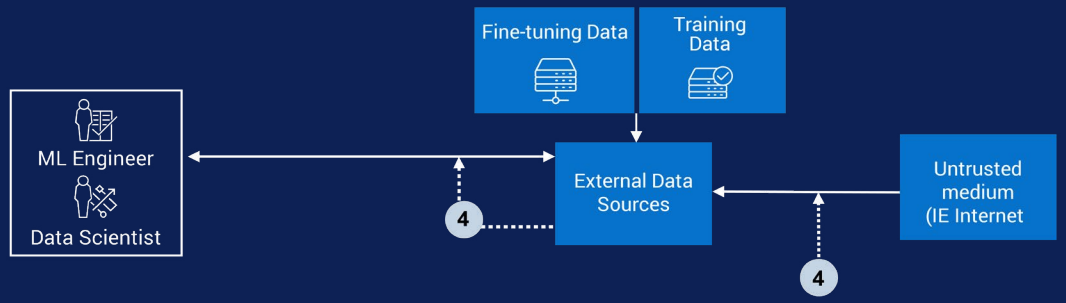
Explore vulnerabilities in the LLM supply chain, which can affect critical components such as pre-trained model integrity and third-party adapters. AI systems rely on both hardware and software that may be compromised long before deployment. Adversaries can exploit weaknesses at various stages of the machine learning supply chain, targeting GPU hardware, data and its annotations, elements of the ML software stack, or even the model itself. By compromising these unique portions, attackers can gain initial access to systems, posing significant risks to security and integrity. Understanding and mitigating these vulnerabilities is crucial for building robust, secure AI solutions.

Concern # 4: Model Data Poisoning

Strategies for Mitigating Model Data Poisoning:

- **Use anomaly detection and data validation during training** to identify and address inconsistencies in data and ensure only clean, high-quality data is used to train the model.
- **Isolate environments during fine-tuning phases** for fine-tuning to prevent unauthorized access or contamination of the model during critical stages of development.
- **Model validation** Regularly validate ML models to ensure they haven't been tampered with before deployment, safeguarding their accuracy and integrity.
- **Access restrictions and human oversight:** Apply role-based access control (RBAC), multi-factor authentication (MFA), and identity management to limit access. Use human review for critical decisions.
- **Data sanitization and input validation:** Screen user inputs thoroughly to remove harmful content. Use normalization and encoding to prevent misuse.
- **Secure development, configuration, and audits:** Apply secure coding practices, use automated configuration management tools, and conduct regular reviews, audits, and updates to keep AI system configurations secure and current.
- **Robustness checks:** Conduct regular evaluations to identify and fix vulnerabilities, keeping the AI secure and reliable.
- **Implement Network segmentation** to limit access to insecure interfaces and critical system components.
- **Monitoring, logging, and anomaly detection:** Continuously monitor and log AI system activities—using solutions like MDR/XDR/SIEM—to rapidly detect, investigate, and respond to unauthorized access, anomalies, or data leaks.

Back to Architecture Diagram



Explore Model data poisoning is a security threat in the AI lifecycle where adversaries intentionally contaminate training data with corrupt, misleading, or malicious inputs. This risk can affect critical components—ranging from raw data collection and annotation to the curation and integration of datasets used for machine learning or large language models. The reliability of AI systems depends on the integrity of their data sources, which may be exposed to manipulation prior to training, during preprocessing, or via external data pipelines.

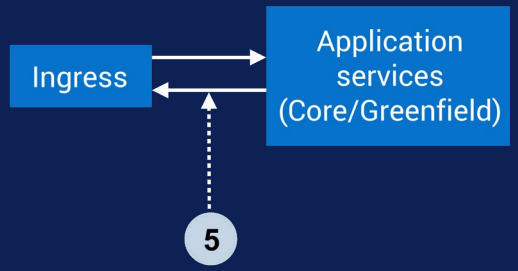
Attackers leverage data poisoning to degrade model accuracy, introduce vulnerabilities, or trigger harmful outputs. By targeting weaknesses in data provenance, annotation quality, or dataset ingestion processes, adversaries can undermine security, trustworthiness, and resilience. Recognizing and mitigating these data-centric threats is essential to building robust, dependable AI solutions.

Concern # 5: Improper Output Handling

Strategies for Mitigating Improper Output Handling:

- **Context-Aware Output Encoding:** Always apply encoding and escaping techniques that are tailored to the specific context where the output will be used, such as HTML, SQL, or API environments, to prevent vulnerabilities like injection attacks.
- **Output Sanitization:** Follow strict validation and sanitization practices for model outputs in alignment with Open Web Application Security Project (OWASP) Application Security Verification Standard (ASVS) guidelines to ensure safe downstream use and mitigate security risks.
- **Monitoring, logging, and anomaly detection:** Continuously monitor and log AI system activities—using solutions like MDR/XDR/SIEM—to rapidly detect, investigate, and respond to unauthorized access, anomalies, or data leaks.
- **Automated Output Security Testing:** Conduct regular security testing using automated tools to identify risks in outputs, such as cross-site scripting (XSS) or injection vulnerabilities, and address them proactively.
- **Access restrictions and human oversight:** Apply role-based access control (RBAC), multi-factor authentication (MFA), and identity management to limit access. Use human review for critical decisions.
- **Human-in-the-loop review:** For high-risk applications such as finance or healthcare, require human oversight and review of model outputs to ensure accuracy, security, and safety.
- **Privacy and Compliance:** Integrate privacy-preserving techniques into the output process and ensure compliance with relevant regulations and standards for the secure use of sensitive information.

Back to Architecture Diagram



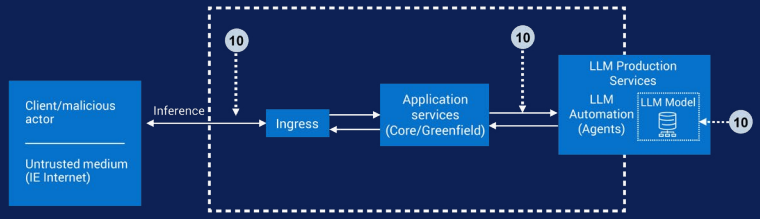
Insufficient validation or sanitization of AI model output can lead to serious security risks, including privilege escalation and data breaches. When AI models produce outputs that are not properly checked or filtered, malicious actors may exploit these vulnerabilities to gain unauthorized access or escalate their privileges within a system. This lack of oversight can result in compromised data, unauthorized actions, and significant security breaches, highlighting the importance of implementing robust validation and sanitization processes for any AI-generated outputs.

Concern # 6: Excessive Agency

Strategies for Mitigating Excessive Agency

- **Enforce least privilege:** Grant LLMs and agentic subsystems only the minimal permissions required to perform intended operations and regularly review access controls.
- **Access restrictions and human oversight:** Apply role-based access control (RBAC), multi-factor authentication (MFA), and identity management to limit access. Use human review for critical decisions.
- **Set operational boundaries:** Clearly define what LLMs/ agents can access or execute.
- **Human-in-the-loop review:** For high-risk applications such as finance or healthcare, require human oversight and review of model outputs to ensure accuracy, security, and safety.
- **Monitoring, logging, and anomaly detection:** Continuously monitor and log AI system activities—using solutions like MDR/XDR/SIEM—to rapidly detect, investigate, and respond to unauthorized access, anomalies, or data leaks.
- **Limit autonomy:** Restrict LLM capabilities to avoid unrestricted access or control.
- **Secure development, configuration, and audits:** Apply secure coding practices, use automated configuration management tools, and conduct regular reviews, audits, and updates to keep AI system configurations secure and current.
- **Deploy firewalls** Block unnecessary network connectivity, reducing exposure to potential threats and limiting avenues for attackers.
- **Robustness checks:** Conduct regular evaluations to identify and fix vulnerabilities, keeping the AI secure and reliable.

Back to Architecture Diagram



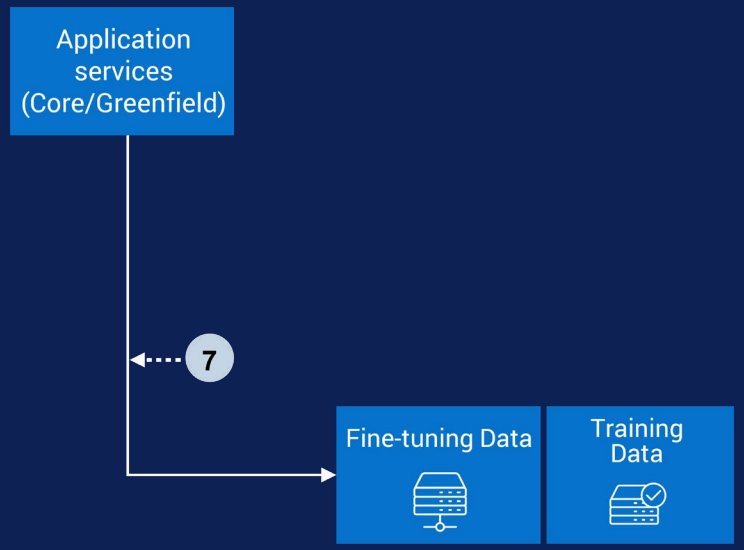
Granting AI agents or plugins excessive autonomy or unnecessary functionality within workflows can pose significant risks. When an AI system is given privileges or capabilities beyond what is required, it increases the likelihood of unintended consequences. This can happen when Large Language Model (LLM)-based systems are designed with excessive permissions, allowing them to take actions or access information they shouldn't. Such overreach can lead to errors, misuse of data, or even security vulnerabilities, emphasizing the importance of carefully limiting and monitoring AI capabilities to ensure safe and responsible use.

Concern # 7: Prompt Leakage

Strategies for Mitigating Prompt Leakage

- **Avoid embedding sensitive info in prompts** Never include credentials, API keys, or proprietary logic in prompts—manage them securely outside the system.
- **Separate security controls from prompts** Handle authentication, authorization, and session management in application logic, not in prompts.
- **Validate inputs and outputs** Sanitize prompts and responses with robust validation to block suspicious patterns or manipulations.
- **Access restrictions and human oversight:** Apply role-based access control (RBAC), multi-factor authentication (MFA), and identity management to limit access. Use human review for critical decisions.
- **Encrypt and secure prompts** Store prompts and configurations in encrypted, secure storage to prevent unauthorized access.
- **Monitoring, logging, and anomaly detection:** Continuously monitor and log AI system activities—using solutions like MDR/XDR/SIEM—to rapidly detect, investigate, and respond to unauthorized access, anomalies, or data leaks.
- **Regularly review prompts** Periodically review and sanitize prompts to remove sensitive data and ensure security compliance.
- **Test and red-team for weaknesses** Conduct adversarial testing to identify and fix vulnerabilities in prompt management or outputs.
- **Isolate prompts from user inputs** Design systems to prevent user queries from manipulating or exposing prompts.
- **Enforce rate limits** Limit API usage, throttle suspicious activity, and block automated prompt attacks.

Back to Architecture Diagram



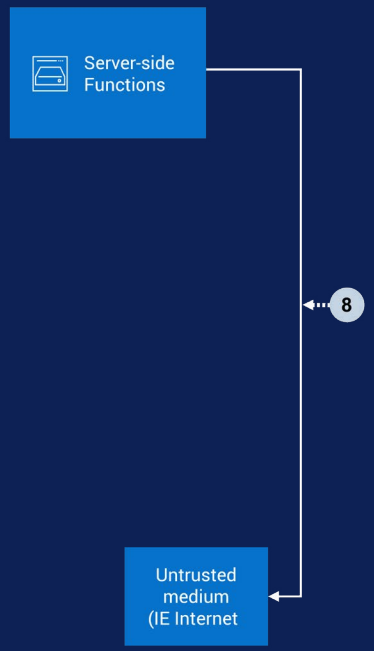
A system prompt leakage attack on a large language model (LLM) or AI system occurs when an attacker is able to extract or infer the hidden instructions—the "system prompts"—that guide the model's behavior and set operational boundaries. These prompts are typically not meant to be visible to end users, as they contain core rules, limitations, and sometimes sensitive operational logic. Through specially crafted inputs or exploiting vulnerabilities, an attacker may trick the LLM into revealing its system prompt, either in whole or in part. If leaked, this information can be used to reverse-engineer restrictions, bypass safety filters, or develop new targeted attacks, ultimately increasing the risk of prompt injection, privilege escalation, or misuse of the model and downstream systems that rely on its integrity.

Concern # 8: Vector & Embedding Weaknesses

Strategies for Mitigating Vector & Embedding Weaknesses

- **Access restrictions and human oversight:** Apply role-based access control (RBAC), multi-factor authentication (MFA), and identity management to limit access. Use human review for critical decisions.
- **Encryption** secure vector data in transit and at rest using robust encryption standards like AES.
- **Secure configuration and monitoring** harden systems, configure securely, and continuously monitor for misconfigurations, unauthorized access, or anomalies.
- **Vulnerability management** regularly update and patch all software, dependencies, and vector store engines to address security risks.
- **Data sanitization and input validation:** Screen user inputs thoroughly to remove harmful content. Use normalization and encoding to prevent misuse.
- **Leverage secure APIs and system interfaces** for AI data interactions, routinely reviewing configurations to minimize exposure and attack surface.
- **Monitoring, logging, and anomaly detection:** Continuously monitor and log AI system activities—using solutions like MDR/XDR/SIEM—to rapidly detect, investigate, and respond to unauthorized access, anomalies, or data leaks.
- **Secure hardware** Use hardware validated for security to prevent vulnerabilities that could arise from hardware-based attacks, ensuring a strong foundation for your infrastructure.
- **Secure development, configuration, and audits:** Apply secure coding practices, use automated configuration management tools, and conduct regular reviews, audits, and updates to keep AI system configurations secure and current.

Back to Architecture Diagram



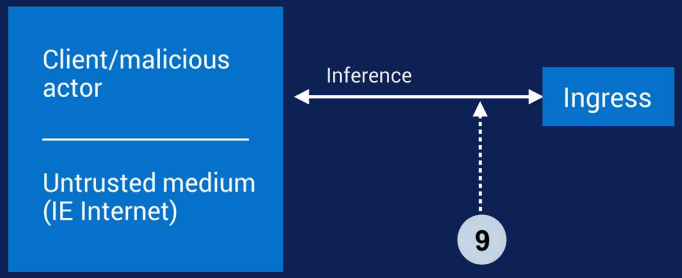
A vector and embedding weaknesses attack on a Large Language Model (LLM) or AI system—particularly those using Retrieval Augmented Generation (RAG)—targets vulnerabilities in how information is encoded, stored, and retrieved as numerical vectors and embeddings. Weaknesses in these mechanisms can be exploited through malicious actions such as embedding inversion (reconstructing sensitive data from embeddings), data poisoning (injecting harmful or biased content to manipulate model behavior), unauthorized access to vector databases (leading to data leaks), or manipulation of retrieval outputs. These attacks threaten privacy, integrity, and reliability by enabling attackers to disclose sensitive information, alter outputs, or undermine user trust in AI-driven applications. Proper access controls, data validation, encryption, and ongoing monitoring are critical in defending against these evolving threats.

Concern # 9: Misinformation

Strategies for Mitigating Misinformation

- **Retrieval-Augmented Generation (RAG) with authoritative sources:** use RAG to retrieve and integrate information from verified, trusted databases and knowledge repositories, reducing hallucinations.
- **Model tuning & output calibration:** fine-tune models with diverse datasets and apply techniques to minimize bias and misinformation.
- **Automated fact-checking:** cross-reference outputs with reliable sources and flag false information automatically.
- **Uncertainty monitoring:** flag low-confidence responses for human review in critical cases.
- **Human-in-the-loop review:** For high-risk applications such as finance or healthcare, require human oversight and review of model outputs to ensure accuracy, security, and safety.
- **User feedback:** enable users to report errors for continuous model improvement and rapid correction of misinformation pathways.
- **Access restrictions and human oversight:** Apply role-based access control (RBAC), multi-factor authentication (MFA), and identity management to limit access. Use human review for critical decisions.
- **Secure development, configuration, and audits:** Apply secure coding practices, use automated configuration management tools, and conduct regular reviews, audits, and updates to keep AI system configurations secure and current.
- **Risk communication:** educate users on AI limitations and encourage independent verification.
- **Intentional UI and API design:** highlight AI-generated content and guide users on responsible use.

Back to Architecture Diagram



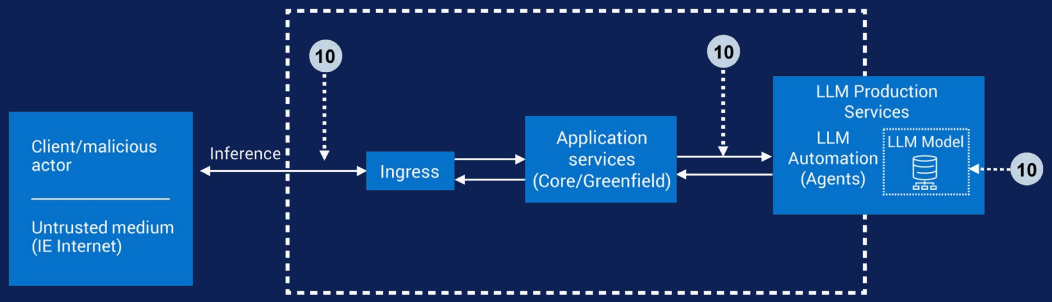
A misinformation attack on an LLM or AI system is an intentional effort to get the model to generate or spread false, misleading, or seemingly credible—but incorrect—information through its outputs. This vulnerability stems from several factors: the model's tendency to "hallucinate" (generating fabricated but plausible-sounding content), biases or gaps present in the training data, and the influence of adversarial prompts. Hallucinations occur because LLMs statistically generate text that fits a pattern, rather than truly understanding facts, leading to answers that appear authoritative but are actually unfounded. The risks of such attacks include security breaches, reputational harm, and even legal liability, especially in environments where users over-rely on LLM responses without verifying their accuracy or validity, potentially embedding errors or misinformation into critical decisions and processes.

Concern # 10: Unbounded Consumption

Strategies for Unbounded Consumption

- **Enforce rate limiting and user quotas** set strict limits on requests, tokens, or data per user, API key, or app to prevent abuse.
- **Require authentication and user segmentation** use strong authentication (e.g., API keys, OAuth) and assign roles or tiers to process only authorized requests.
- **Input validation and size restrictions** validate prompt size and structure, blocking or trimming large or malformed queries.
- **Apply processing timeouts and resource throttling set** timeouts and resource caps for each request to avoid long-running operations and resource drain.
- **Deploy smart caching and deduplication** cache responses for duplicate or similar queries to reduce unnecessary processing.
- **Monitoring, logging, and anomaly detection:** Continuously monitor and log AI system activities—using solutions like MDR/XDR/SIEM—to rapidly detect, investigate, and respond to unauthorized access, anomalies, or data leaks.
- **Budget tracking and spend controls** use dashboards and alerts to monitor costs and block usage at budget thresholds.
- **Sandboxing and isolation techniques** run workloads in isolated environments with limited permissions to reduce risks.
- **Limit call depth and conversation turns** impose limits on recursive calls or conversation steps to prevent exploitation.
- **Apply tiered model or resource allocation** route high-priority requests to premium models and low-priority traffic to cost-effective ones.

Back to Architecture Diagram



An unbounded consumption threat on an LLM or AI system refers to a security vulnerability where the application permits users—malicious or otherwise—to submit excessive, uncontrolled inference requests or prompts without effective rate limiting, authentication, or usage restrictions. Because LLM inference is computationally expensive, this lack of control can be exploited in several ways: attackers may cause denial of service (DoS) by overwhelming system resources, generate unforeseen economic losses in pay-per-use or cloud-hosted deployments, or systematically query the model to clone its behavior and steal intellectual property. The consequences include service disruption, degraded performance for other users, financial strain, and increased risk of sensitive model leakage. In essence, unbounded consumption occurs when resource usage is not properly governed, leaving LLM-based applications exposed to both accidental and deliberate exploitation.

Why Dell for AI Security

Dell helps organizations secure AI models and LLMs through a comprehensive approach that spans hardware, software, and managed services. Security is embedded from the supply chain through to device, infrastructure, data, and applications, all aligned to Zero Trust principles. Across the portfolio, Dell's solutions are built to advance cyber hygiene with features like MFA, RBAC, least privilege, and continuous verification. This comprehensive, "secure by design" approach ensures organizations can confidently innovate with AI and LLMs, minimizing risk from model theft, data leakage, adversarial attacks, and other advanced cyber threats.

Supply Chain

Dell's Secure Supply Chain provides foundational protection for AI models and LLMs by embedding security throughout every stage of product development, manufacturing, and delivery. Through cryptographically-signed BIOS and firmware updates, Secured Component Verification, AI-focused software bill of materials (SBOM), dataset lineage tracking, integrated security software and configuration, and rigorous vendor risk assessments aligned with global standards, Dell minimizes risks from tampering, unauthorized access, and supply chain attacks—ensuring organizations can deploy trusted, resilient AI workloads with full transparency, integrity, and regulatory compliance.

AI PCs

Dell offers foundational security for on-device AI workloads. Dell Trusted Devices--the world's most secure commercial AI PCs*---are designed with security in mind. Supply chain security minimizes the risk of product vulnerabilities and tampering. Unique defenses built directly into hardware and firmware keep the PC and end-user protected in use. Dell SafeBIOS provides deep, BIOS-level visibility and tamper detections, while Dell SafeID enhances credential security and enables password-less authentication. Partner software provides advanced protection across endpoint, network and cloud environments.

Cyber Resilience

Dell's PowerProtect cyber resilience solutions secure AI Data with encrypted, immutable backups, rapid restoration, and isolated cyber recovery vaults. These capabilities prevent destruction, mitigate impact from malicious updates, and support compliance and recovery after an attack.

Servers

PowerEdge servers feature confidential computing to isolate and secure AI/LLM prompts and embeddings, trusted retrieval-augmented generation (RAG) solutions anchored in authoritative sources, along with MFA, RBAC, silicon root of trust, signed firmware, and continuous monitoring to protect critical AI workloads.

Storage

Dell's storage portfolio ensures secure, encrypted storage for sensitive AI data with robust AES-256 encryption for data at rest and in transit. Advanced encryption designed to be resilient against future quantum threats is available on select offerings. The portfolio includes high-speed NVMe performance, FIPS-compliant encryption modules to secure data—including those utilized in AI workloads—immutable snapshots, and air-gapped cyber recovery vaults to counter ransomware attacks. Zero trust architecture, supply chain security, and tamper-proof audit capabilities enhance governance. Built-in anomaly detection and AIOps ML models protect workloads without using customer data for training, thereby minimizing input-based attack risks.

AIOps

Dell AIOps provides automated, continuous monitoring to detect misconfigurations, vulnerabilities (including CVEs), and supports supply chain risk awareness impacting AI/LLM workloads. Real-time CVE scanning, smart alerts, and AI-powered dashboards empower rapid intervention by flagging anomalies and tracking resolution workflows. Built-in compliance features, role-based access controls, and automated reporting help maintain secure operations across workloads, while seamless EDR/XDR integration and AI-driven operational insights—including generative capabilities in supported solutions—further enhance IT efficiency.

Networking

Dell Networking solutions protect AI/LLM environments through robust network segmentation, minimizing lateral movement. Encrypted network paths and integrated firewall controls block unauthorized access to AI data.

AI Security and Resilience Services

Dell's AI Security and Resilience services are crafted to address novel risks associated with integrating AI into your organization. Built to work alongside your teams as you onboard AI as quickly as possible, our services provide expertise to guide in strategic planning, solution implementation and managed security services to ease operational burdens so you can innovate securely with AI. Each is tailored to help organizations address evolving AI risks and optimize secure AI deployments.

Dell AI Factory

An integrated portfolio of purpose built security such as Dell's secure supply chain, zero trust capabilities to enforce least privilage, and AI MDR solutions designed to keep your model safe and secure.

*Based on Dell internal analysis, October 2024 (Intel) and March 2025 (AMD). Applicable to PCs on Intel and AMD processors. Not all features available with all PCs. Additional purchase required for some features. Intel-based PCs validated by Principled Technologies, July 2025.

Conclusion

To build resilient AI frameworks, a collaborative approach between organizations and security experts is essential. As AI and LLMs continue to reshape industries, it's critical to address the risks they bring, including data security, model integrity, and compliance challenges. Organizations must prioritize proactive strategies that integrate security into every stage of their AI journey.

Dell Technologies stands as a trusted partner in this mission, offering end-to-end GenAI customization, security consulting, and integrated solutions tailored to your unique needs. By leveraging Dell's robust cybersecurity solutions, enterprises can effectively mitigate AI and LLM risks while maximizing the potential of their existing security investments. Dell empowers organizations to protect their AI infrastructure by seamlessly integrating advanced security into their current frameworks, ensuring a future-ready, secure environment.

Learn how Dell's comprehensive AI solutions can secure your GenAI and LLM environments: Dell.com/CyberSecurityMonth