DELL Technologies

生成式 AI 和 LLM 十大网络安全问题

立即开始

简介

人工智能 (AI) 正在彻底改变组织的运营方式，生成式 AI (GenAI) 和大型语言模型 (LLM) 成为现代企业环境中的关键工作负载。

与任何其他工作负载一样，这些应用程序都面临着自己的一系列复杂性和漏洞需要解决。随着企业持续采用 AI 来推动创新、提升效率并获取竞争优势，确保这些应用程序的安全已成为基础性要求。良好的网络安全防护是保障各种工作负载安全的基础，正如您优先考虑所有工作负载的安全性一样，在 AI 领域践行良好的网络安全防护同样至关重要。这包括实施适当的系统修补、多因素身份验证、基于角色的访问控制以及网络分段等实践。这些衡量措施至关重要，但关键在于了解这些功能如何适应工作负载的特定体系结构和使用情况。

戴尔对 AI 工作负载及其面临的独特安全挑战有着深刻的了解。通过识别威胁行为者可能采用哪些途径攻击这些工作负载，戴尔可助您构建完善的安全战略。这包括解决以下风险，例如：训练数据中毒、模型被盗或被操纵、数据集重建等。

我们还专注于管理与 AI 模型输入相关的挑战，例如防止敏感信息泄露、规避不安全话题或偏见，以及确保遵守法规。在输出端，我们帮助解决过度依赖模型和合规性相关风险等问题。

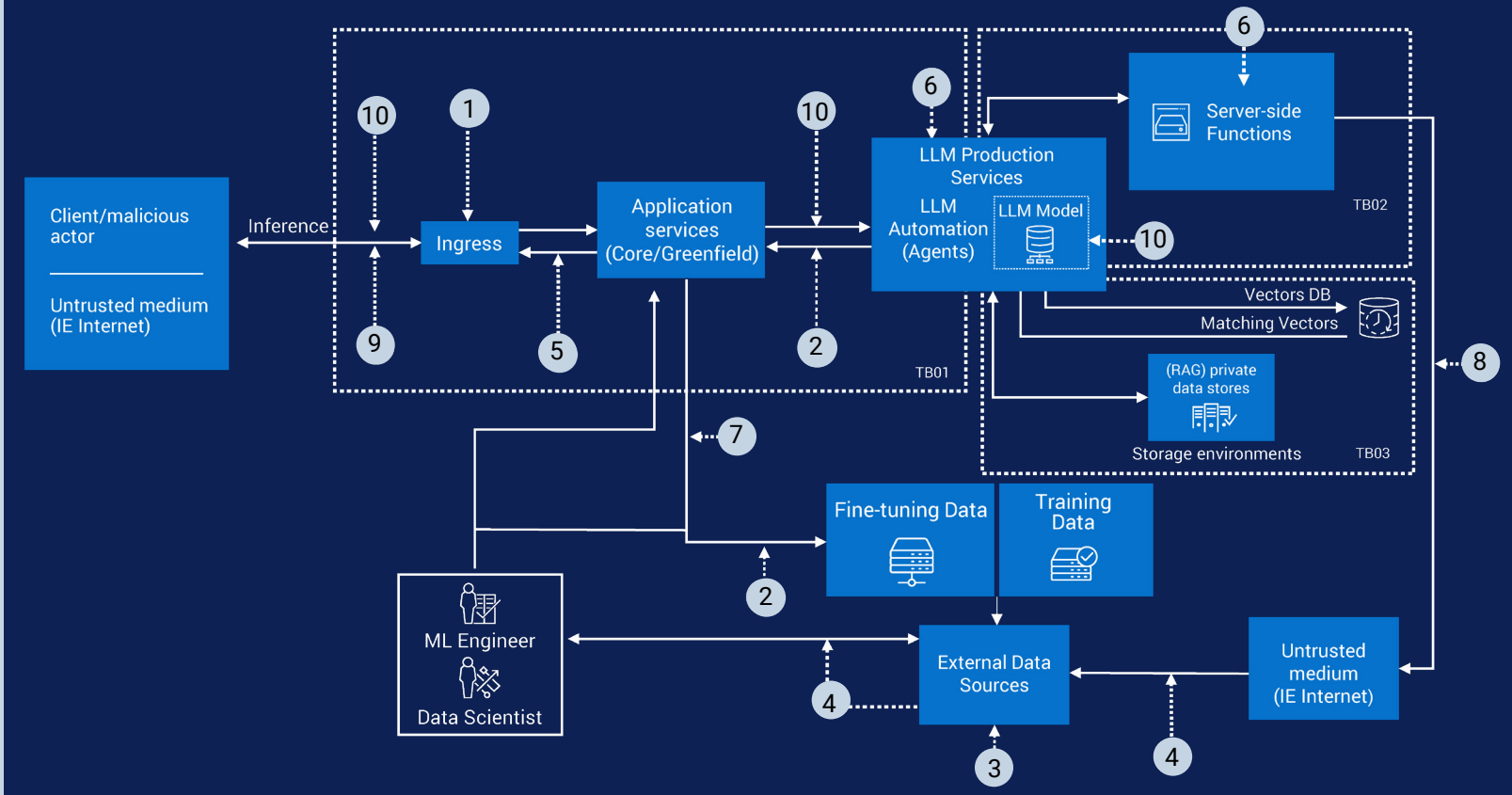
戴尔通过利用企业现有的网络安全解决方案或探索新工具和实践来保护其系统，帮助企业降低这些风险。我们的目标是确保安全保护不会阻碍您的创新。通过了解 AI 工作负载的工作原理及其面临的安全威胁，我们可以帮助您构建更强大的安全态势，从而提高您的环境弹性，让您能够信心十足地开展创新。凭借我们的专业知识，我们可帮助您自信地发挥 AI 的潜力，同时保持强大的安全性。

生成式 AI 和 LLM 十大网络安全问题

根据 OWASP 的界定，这些是保护生成式 AI/LLM 模型需重点关注的核心安全问题。

单击每个问题可了解详情：

1. 提示词注入
2. 敏感信息披露
3. 供应链
4. 模型数据中毒
5. 输出处理不当
6. 代理过度
7. 系统提示词泄露
8. 向量和嵌入弱点
9. 误导
10. 无限制消耗

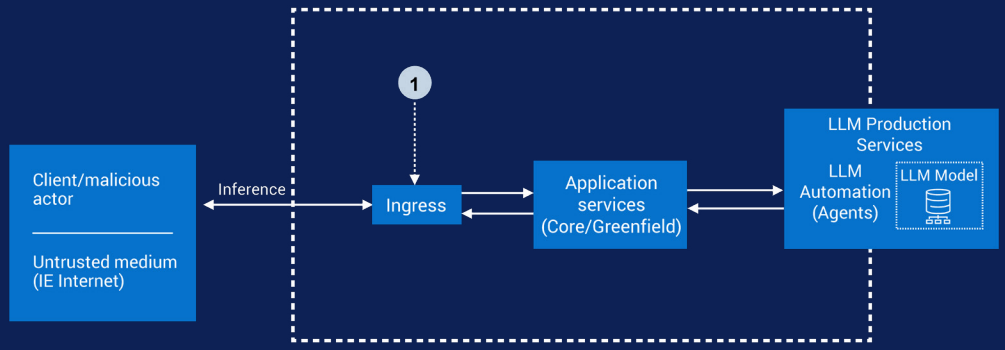


问题 1：提示词注入

提示词注入的应对战略：

- **数据清理和输入验证：**彻底筛选用户输入，以删除有害内容。使用标准化和编码，以防止滥用。
- **自然语言处理 (NLP) 和基于机器学习的方法：**使用 NLP 和机器学习，检测和阻止操纵或恶意提示词。
- **清晰的输出格式和响应控制：**设置严格的响应边界，以确保输出遵循预期格式并防止未经授权的操作。使用提示词筛选和响应验证来维护完整性。
- **限制访问和人工监督：**应用基于角色的访问控制 (RBAC)、多因素身份验证 (MFA) 和身份管理来限制访问。使用人工审查来制定关键决策。
- **监控、记录和异常检测：**使用 MDR/XDR/SIEM 等解决方案持续监控和记录 AI 系统活动，以快速检测、调查和响应未经授权的访问、异常或数据泄露。
- **安全的提示词工程设计：**使用安全的提示词设计和分析作为整体软件安全保护的一部分，以保障输入处理的安全性。
- **模型验证：**定期验证 ML 模型，以确保它们在部署之前未被篡改，从而保护其准确性和完整性。
- **提示词筛选、排名和响应验证：**分析提示词并排列顺序，以确保仅处理安全输入。验证响应，以防止滥用。
- **鲁棒性检查：**执行定期评估，以识别和修复漏洞，确保 AI 安全可靠。

返回架构图



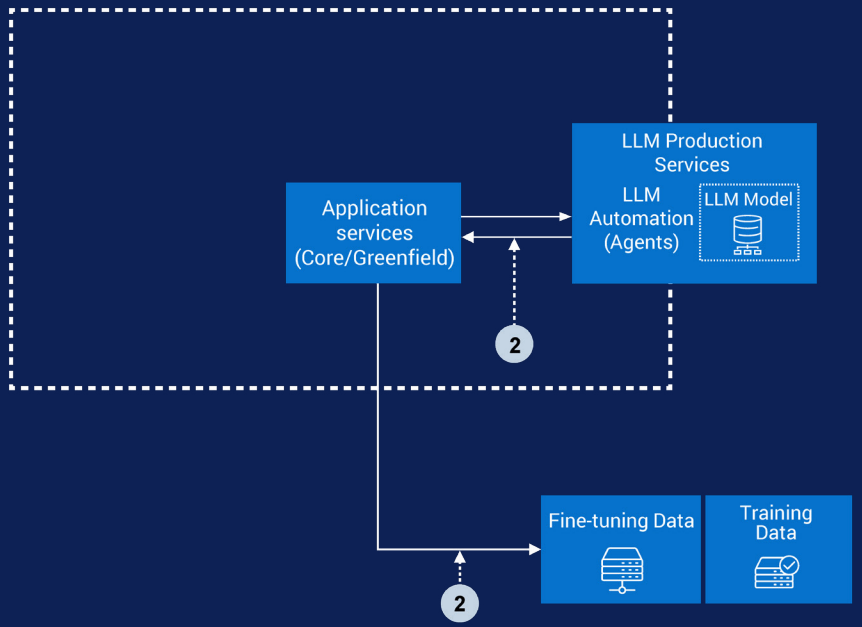
提示词注入是生成式 AI (GenAI) 领域涌现的新挑战，即通过精心构造的恶意输入来操纵模型行为或破坏其完整性。由于 AI 系统在处理和响应用户输入时存在漏洞，这些攻击可能导致未经授权的操作、误导或敏感数据泄露。生成式 AI 越来越多地集成到关键业务工作流中，因此应对这些风险对于维护信任和安全性至关重要

问题 2：敏感信息泄露

敏感信息泄露的应对战略：

- **数据清理和输入验证：**彻底筛选用户输入，以删除有害内容。使用标准化和编码，以防止滥用。
- **利用同构加密：**安全地处理敏感数据，而不会泄露其内容。即使数据在使用中，也能确保数据保持加密状态，并防止泄露。
- **限制访问和人工监督：**应用基于角色的访问控制 (RBAC)、多因素身份验证 (MFA) 和身份管理来限制访问。使用人工审查来制定关键决策。
- **利用安全的 API 和系统接口：**进行 AI 数据交互、定期审查配置，以尽可能减少风险和攻击面。
- **安全的数据收集、存储和策略：**同步实施全面的数据保护和治理策略，以确保合规性并尽可能降低数据风险。
- **监控、记录和异常检测：**使用 MDR/XDR/SIEM 等解决方案持续监控和记录 AI 系统活动，以快速检测、调查和响应未经授权的访问、异常或数据泄露。
- **安全的开发、配置和审核：**应用安全编码实践、使用自动化配置管理工具，并定期进行审查、审核和更新，以确保 AI 系统配置安全且最新。
- **用户培训和安全意识：**为用户和管理员提供持续的特定于 AI 的安全意识培训，以减少不安全使用和意外数据泄露。

返回架构图



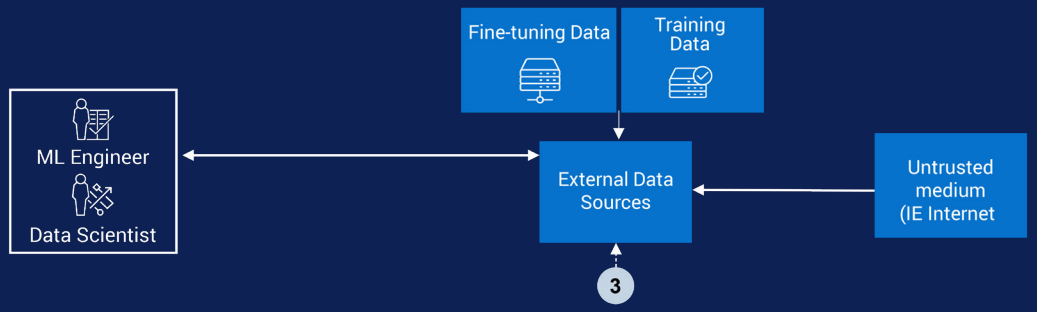
生成式 AI 带来了令人难以置信的进步，但它也带来了巨大的风险，尤其是敏感信息的无意泄露。无论是个人身份信息 (PII) 还是专有业务数据，对生成式 AI 工具的滥用或处理不当都可能导致数据泄露、监管违规或声誉损害。为了确保 AI 系统的安全实施和使用，组织必须充分了解这些风险并采取积极措施应对。

问题 3：供应链漏洞

供应链漏洞的应对战略：

- **审查供应商并确保遵守安全的供应链实践：**评估供应商并制定优先考虑供应链安全的协议。
- **实施软件物料清单：**跟踪并验证软件组件的来源，确保透明度并降低代码泄露的风险。
- **模型验证：**定期验证 ML 模型，以确保它们在部署之前未被篡改，从而保护其准确性和完整性。
- **以最低权限运行容器和 Pod：**这样可以减少泄露时的潜在影响，并限制未经授权的访问。
- **部署防火墙：**阻止不必要的网络连接，从而减少潜在威胁的风险并限制攻击者的入侵途径。
- **保护数据和注释：**保护您的数据和关联的注释，以防止篡改、未经授权的访问和损坏关键信息。
- **安全的硬件：**使用经过安全性验证的硬件，防止基于硬件的攻击产生的漏洞，从而确保为基础架构奠定坚实的基础。
- **安全的 ML 软件组件：**使用值得信赖且经过审查的 ML 软件组件来减少漏洞并增强机器学习工作流的整体安全性。
- **安全的开发、配置和审核：**应用安全编码实践、使用自动化配置管理工具，并定期进行审查、审核和更新，以确保 AI 系统配置安全且最新。

返回架构图



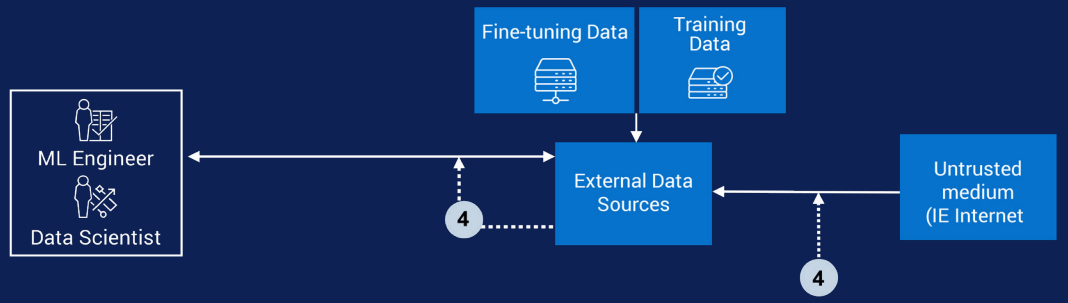
探索 LLM 供应链中的漏洞，这些漏洞可能会影响关键组件，例如预先训练的模型完整性和第三方适配器。AI 系统依赖于硬件和软件，这些硬件和软件可能在部署之前就会受到损害。机器学习供应链的各个阶段都存在潜在的弱点，攻击者可以利用这些弱点，攻击 GPU 硬件、数据及其注释、ML 软件堆栈的元素，甚至直接攻击模型本身。通过攻陷这些独特的部分，攻击者可以获得对系统的初始访问权限，从而对安全性和完整性构成重大风险。了解并缓解这些漏洞对于构建强大且安全的 AI 解决方案至关重要。

问题 4：模型数据中毒

模型数据中毒的应对战略：

- **在训练期间使用异常检测和数据验证：**识别和解决数据中的不一致性，并确保仅使用干净、高质量的数据来训练模型。
- **在微调阶段隔离环境：**以防止在关键开发阶段出现未经授权访问或模型污染。
- **模型验证：**定期验证 ML 模型，以确保它们在部署之前未被篡改，从而保护其准确性和完整性。
- **限制访问和人工监督：**应用基于角色的访问控制 (RBAC)、多因素身份验证 (MFA) 和身份管理来限制访问。使用人工审查来制定关键决策。
- **数据清理和输入验证：**彻底筛选用户输入，以删除有害内容。使用标准化和编码，以防止滥用。
- **安全的开发、配置和审核：**应用安全编码实践、使用自动化配置管理工具，并定期进行审查、审核和更新，以确保 AI 系统配置安全且最新。
- **鲁棒性检查：**执行定期评估，以识别和修复漏洞，确保 AI 安全可靠。
- **实施网络分段：**限制对不安全接口和关键系统组件的访问。
- **监控、记录和异常检测：**使用 MDR/XDR/SIEM 等解决方案持续监控和记录 AI 系统活动，以快速检测、调查和响应未经授权的访问、异常或数据泄露。

返回架构图



探索：模型数据中毒是 AI 生命周期中的一种安全威胁，攻击者向训练数据中输入破坏性、误导性或恶意内容，故意污染数据。这种风险可能会影响关键组件 — 从原始数据收集和注释到用于机器学习或大型语言模型的数据集的整理和集成。AI 系统的可靠性取决于其数据源的完整性，而这些数据源可能会在训练前的预处理阶段或通过外部数据管道被操纵。

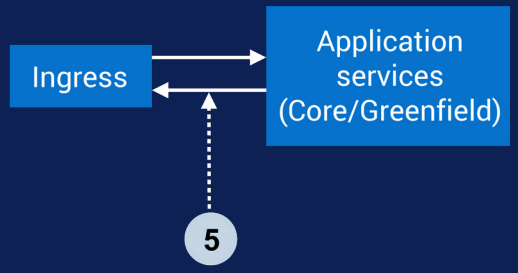
攻击者利用数据中毒来降低模型准确性、植入漏洞或触发有害输出。通过针对数据溯源、标注质量或数据集摄取流程中的薄弱环节，攻击者能够破坏系统的安全性、可信度和弹性。识别和缓解这些以数据为中心的威胁对于构建强大且可靠的 AI 解决方案至关重要。

问题 5：输出处理不当

输出处理不当的应对战略：

- **上下文感知输出编码：**根据具体在哪种上下文（如 HTML、SQL 或 API 环境）使用输出内容，始终采用针对性的编码和转义技术，从而防范注入攻击等安全漏洞。
- **输出清理：**根据开放 Web 应用程序安全项目 (OWASP) 应用程序安全验证标准 (ASVS) 准则，对模型输出遵循严格的验证和清理实践，确保下游使用的安全性并降低安全风险。
- **监控、记录和异常检测：**使用 MDR/XDR/SIEM 等解决方案持续监控和记录 AI 系统活动，以快速检测、调查和响应未经授权的访问、异常或数据泄露。
- **自动输出安全测试：**使用自动化工具定期执行安全测试，以识别输出中的风险，例如跨站点脚本 (XSS) 或注入漏洞，并主动解决这些风险。
- **限制访问和人工监督：**应用基于角色的访问控制 (RBAC)、多因素身份验证 (MFA) 和身份管理来限制访问。使用人工审查来制定关键决策。
- **人工审查：**对于金融或医疗等高风险应用程序，需要对模型输出进行人工监督和审查，以确保准确性、安全性和可靠性。
- **隐私和合规性：**将隐私保护技术集成到输出流程中，并确保遵守安全使用敏感信息的相关法规和标准。

返回架构图



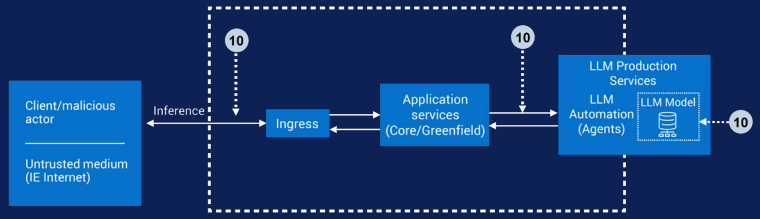
AI 模型输出未经充分验证或清理可能引发严重安全风险，包括权限提升和数据泄露。当 AI 模型产生未正确检查或筛选的输出时，恶意攻击者可能会利用这些漏洞来获得未经授权的访问权限或在系统中提升其权限。这种缺乏监督可能会导致数据泄露、未经授权的操作和严重的安全漏洞，凸显了对任何 AI 生成的输出实施强大的验证和清理流程的重要性。

问题 6：代理过度

代理过度的应对战略

- **实施最低权限：**仅授予 LLM 和代理子系统执行预期操作所需的最低权限，并定期查看访问控制。
- **限制访问和人工监督：**应用基于角色的访问控制 (RBAC)、多因素身份验证 (MFA) 和身份管理来限制访问。使用人工审查来制定关键决策。
- **设置运营边界：**明确定义 LLM/工程师可以访问或执行的内容。
- **人工审查：**对于金融或医疗等高风险应用程序，需要对模型输出进行人工监督和审查，以确保准确性、安全性和可靠性。
- **监控、记录和异常检测：**使用 MDR/XDR/SIEM 等解决方案持续监控和记录 AI 系统活动，以快速检测、调查和响应未经授权的访问、异常或数据泄露。
- **限制自主权：**限制 LLM 功能，以避免不受限制的访问或控制。
- **安全的开发、配置和审核：**应用安全编码实践、使用自动化配置管理工具，并定期进行审查、审核和更新，以确保 AI 系统配置安全且最新。
- **部署防火墙：**阻止不必要的网络连接，从而减少潜在威胁的风险并限制攻击者的入侵途径。
- **鲁棒性检查：**执行定期评估，以识别和修复漏洞，确保 AI 安全可靠。

返回架构图



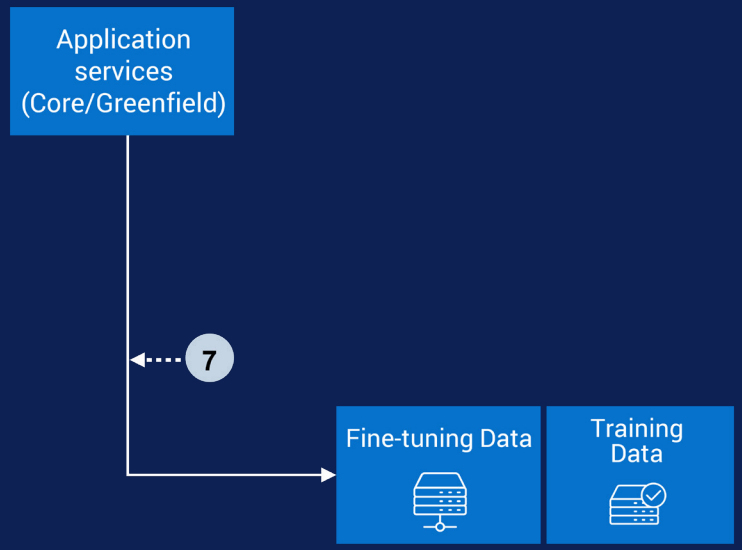
为 AI 代理或插件在工作流中授予过多的自主权或不必要的功能可能会带来重大风险。当 AI 系统被授予超出必要范围的权限或能力时，其引发意外后果的风险将显著增加。这种情况常见于基于大型语言模型 (LLM) 的系统在设计时被赋予过多权限，使其能够执行不应采取的操作或访问不应触及的信息。此类权限越界可能导致错误、数据滥用甚至安全漏洞，因此必须谨慎限制并持续监控 AI 能力，以确保其安全可靠地运行。

问题 7：即时泄漏

即时泄露的应对战略

- **避免在提示词中嵌入敏感信息：**如凭据、API 密钥或专有逻辑 — 在系统外部安全地管理这些信息。
- **将安全控制与提示词分离：**在应用程序逻辑中处理身份验证、授权和会话管理，而非通过提示词实现。
- **验证输入和输出：**通过鲁棒性验证检查，清理提示词和响应，以拦截可疑模式或操作。
- **限制访问和人工监督：**应用基于角色的访问控制 (RBAC)、多因素身份验证 (MFA) 和身份管理来限制访问。使用人工审查来制定关键决策。
- **加密和保护提示词：**将提示词和配置存储在加密的安全存储中，以防止未经授权的访问。
- **监控、记录和异常检测：**使用 MDR/XDR/SIEM 等解决方案持续监控和记录 AI 系统活动，以快速检测、调查和响应未经授权的访问、异常或数据泄露。
- **定期查看提示词：**定期查看和清理提示词，以删除敏感数据并确保安全合规性。
- **弱点测试和红队演练：**执行对抗性测试，以识别和修复即时管理或输出中的漏洞。
- **将提示词与用户输入隔离：**精心设计系统，以防止用户查询被操纵或提示词泄露。
- **实施频率限制：**限制 API 使用情况、限制可疑活动并阻止自动提示词攻击。

返回架构图



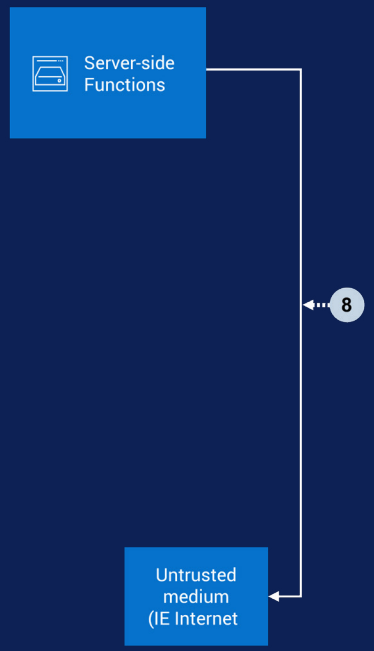
大型语言模型 (LLM) 或 AI 系统上的系统提示词泄露攻击是指攻击者能够提取或推断隐藏指令（即“系统提示词”），以指导模型的行为并设置操作边界。这些提示词通常不应向终端用户公开，因其包含核心规则、限制条件，有时甚至涉及敏感操作逻辑。攻击者可能通过精心构造的特别输入或利用漏洞，诱使 LLM 完整或部分泄露其系统提示词。若发生泄露，此类信息可能被用于逆向工程破解限制、绕过安全过滤器或开发新型定向攻击，最终将增加提示词注入、权限提升、模型及依赖其完整性的下游系统滥用等风险。

问题 8：向量和嵌入弱点

向量和嵌入弱点的应对战略

- **限制访问和人工监督：**应用基于角色的访问控制 (RBAC)、多因素身份验证 (MFA) 和身份管理来限制访问。使用人工审查来制定关键决策。
- **加密：**使用强大的加密标准（如 AES），保护传输中和静态的向量数据。
- **安全配置和监控：**强化系统、安全配置并持续监控配置错误、未经授权的访问或异常情况。
- **漏洞管理：**定期更新和修补所有软件、依赖项和向量存储引擎，以解决安全风险。
- **数据清理和输入验证：**彻底筛选用户输入，以删除有害内容。使用标准化和编码，以防止滥用。
- **利用安全的 API 和系统接口：**进行 AI 数据交互、定期审查配置，以尽可能减少风险和攻击面。
- **监控、记录和异常检测：**使用 MDR/XDR/SIEM 等解决方案持续监控和记录 AI 系统活动，以快速检测、调查和响应未经授权的访问、异常或数据泄露。
- **安全的硬件：**使用经过安全性验证的硬件，防止基于硬件的攻击产生的漏洞，从而确保为基础架构奠定坚实的基础。
- **安全的开发、配置和审核：**应用安全编码实践、使用自动化配置管理工具，并定期进行审查、审核和更新，以确保 AI 系统配置安全且最新。

返回架构图



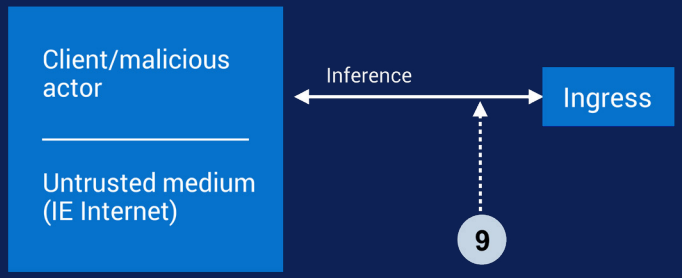
针对大型语言模型 (LLM) 或 AI 系统（尤其是采用检索增强生成 [RAG] 技术的系统）的向量和嵌入弱点攻击，主要利用信息在作为数值向量和嵌入内容进行编码、存储及检索过程中存在的弱点。攻击者可通过恶意手段利用这些机制中的弱点，例如嵌入逆向（从嵌入数据重构敏感信息）、数据中毒（注入有害或偏见内容以操纵模型行为）、未经授权访问向量数据库（引发数据泄露）或操纵检索结果。这些攻击使攻击者能够披露敏感信息、更改输出或破坏用户对 AI 驱动型应用程序的信任，从而威胁隐私、完整性和可靠性。正确的访问控制、数据验证、加密和持续监控，对于防范这些不断演变的威胁至关重要。

问题 9：误导

误导的应对战略

- **基于权威来源的检索增强生成 (RAG)：**运用 RAG 技术从经过验证的可信数据库和知识库中检索并整合信息，从而减少模型幻觉。
- **模型调整和输出校准：**使用多样化的数据集微调模型，并应用技术以尽可能减少偏见和误导。
- **自动事实检查：**使用可靠来源交叉引用输出，并自动标记虚假信息。
- **不确定性监控：**在关键场景中对低置信度响应启动人工审查。
- **人工审查：**对于金融或医疗等高风险应用程序，需要对模型输出进行人工监督和审查，以确保准确性、安全性和可靠性。
- **用户反馈：**使用户能够报告错误，以便持续改进模型并快速纠正误导路径。
- **限制访问和人工监督：**应用基于角色的访问控制 (RBAC)、多因素身份验证 (MFA) 和身份管理来限制访问。使用人工审查来制定关键决策。
- **安全的开发、配置和审核：**应用安全编码实践、使用自动化配置管理工具，并定期进行审查、审核和更新，以确保 AI 系统配置安全且最新。
- **风险沟通：**向用户普及 AI 的局限性，并鼓励他们进行独立验证。
- **意向性 UI 和 API 设计：**突出显示 AI 生成的内容，并指导用户负责任地使用。

返回架构图



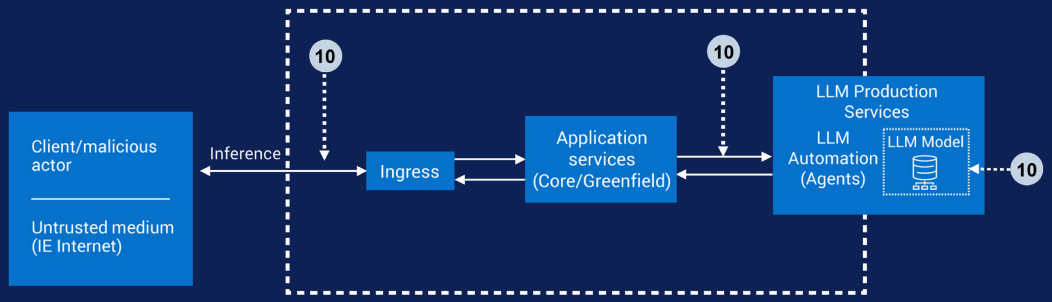
对 LLM 或 AI 系统的虚假信息攻击，是指通过恶意诱导使模型生成或传播虚假、误导性、看似可信但实为错误信息的行为。这种脆弱性源于多重因素：模型固有的"幻觉"倾向（生成虚构但看似合理的内容）、训练数据中存在的偏见或空白，以及对抗性提示词的影响。幻觉现象的产生，是因为 LLM 基于统计规律生成符合模式的文本，而非真正理解事实，导致其回答看似权威实则缺乏依据。此类攻击的风险包括安全漏洞、声誉损害甚至法律追责，尤其在用户过度依赖 LLM 响应却未验证其准确性或有效性的场景中，可能会在关键决策和流程中嵌入错误或误导。

问题 10：无限制消耗

无限制消耗的应对战略

- **实施频率限制和用户配额：**为每个用户、API 密钥或应用程序设置严格的请求数、令牌数或数据量上限，以防止滥用。
- **需要身份验证和用户分段：**使用强身份验证（例如 API 密钥、OAuth）并分配角色或层以仅处理授权请求。
- **输入验证和规模限制：**验证提示词规模和结构，对大型或格式异常的查询进行拦截或截断处理。
- **应用处理超时和资源限制集：**为每个请求设置超时和资源上限，以避免长时间运行的操作和资源消耗。
- **部署智能高速缓存和重复数据消除：**针对重复或类似查询的高速缓存响应，以减少不必要的处理。
- **监控、记录和异常检测：**使用 MDR/XDR/SIEM 等解决方案持续监控和记录 AI 系统活动，以快速检测、调查和响应未经授权的访问、异常或数据泄露。
- **预算跟踪和支出控制：**使用控制面板和警报来监控成本，并在达到预算阈值时自动阻断使用。
- **沙盒和隔离技术：**在权限有限的隔离环境中运行工作负载，以降低风险。
- **限制呼叫深度和对话轮次：**对递归调用和对话步骤设置严格上限，防止恶意利用。
- **应用分层模型或资源分配：**将高优先级请求路由到高端模型，将低优先级流量路由到经济高效的模型。

返回架构图



LLM 或 AI 系统的无限制消耗威胁，是指应用程序允许用户（无论是否恶意）在缺乏有效频率限制、身份验证或使用管控的情况下提交过量且不受控的推理请求或提示词的安全漏洞。由于 LLM 推理计算成本高昂，这种管控缺失可能通过以下多种方式被利用：攻击者可通过耗尽系统资源引发拒绝服务攻击 (DoS)，在按量付费或云托管部署中造成意外经济损失，或系统性查询模型以克隆行为和窃取知识产权。由此产生的后果包括服务中断、其他用户性能下降、财务压力以及敏感模型泄露的风险增加。在本质上，当资源使用未得到适当控制时，就会发生无限制消耗，使基于 LLM 的应用程序面临意外和蓄意利用的风险。

为什么选择戴尔实现 AI 安全保护

戴尔通过涵盖硬件、软件和托管服务的全面方法，助力企业保护 AI 模型和 LLM 的安全。基于零信任原则，安全防护贯穿从供应链到终端设备、基础架构、数据和应用程序的各个环节。在整个产品组合中，戴尔的解决方案旨在通过 MFA、RBAC、最低权限和持续验证等功能提升网络安全。这种"设计即安全"的全面方法，确保组织能够安心开展 AI 和 LLM 创新，从而更大限度地降低模型窃取、数据泄露、对抗性攻击以及其他高级网络威胁带来的风险。

供应链

戴尔的安全供应链通过在产品开发、制造和交付的每个阶段嵌入安全性，为 AI 模型和 LLM 提供基础保护。通过加密签名的 BIOS 和固件更新、安全组件验证、以 AI 为中心的软件物料清单 (SBOM)、数据集沿袭跟踪、集成的安全软件和配置，以及符合全球标准的严格供应商风险评估，戴尔可更大限度地降低由于篡改、未经授权访问和供应链攻击导致的风险，确保组织能够部署值得信赖的弹性 AI 工作负载，实现完全透明、完整性和合规性。

AI PC

戴尔为设备端 AI 工作负载提供基础安全性。作为安全可靠的商用 AI PC*，戴尔可信设备在设计之初就将安全性放在首位。供应链安全性可更大限度地降低产品漏洞和篡改风险。硬件和固件中直接构建的独特防御功能可确保 PC 和终端用户在使用时受到保护。Dell SafeBIOS 提供 BIOS 级别的深度可见性和篡改检测，而 Dell SafeID 增强了凭据安全性并支持无密码身份验证。合作伙伴软件提供跨端点、网络和云环境的高级保护。

网络弹性

Dell PowerProtect 网络弹性解决方案，通过不可变的加密备份、快速恢复和隔离的网络恢复存储区，为保护 AI 数据保驾护航。 这些功能可防止销毁、减轻恶意更新的影响，并在攻击后支持合规性和恢复。

服务器

PowerEdge 服务器配备机密计算功能，可隔离和保护 AI/LLM 提示词和嵌入内容；其可信检索增强生成 (RAG) 解决方案基于权威数据源构建，并集成 MFA、RBAC、硅信任根、签名固件和持续监控，以便保护关键 AI 工作负载。

存储

戴尔存储产品组合为敏感 AI 数据提供具备可靠安全保障的加密存储，对静态数据和传输中数据均采用强大的 AES-256 加密。部分产品还提供能抵御未来量子威胁的先进加密技术。该产品组合包括高速 NVMe 性能，符合 FIPS 标准的加密模块（用于保护 AI 工作负载等数据）、防篡改快照以及可抵御勒索软件攻击的安全隔离 Air Gap 网络恢复存储区。零信任体系结构、供应链安全性和防篡改审核功能，可增强治理。内置异常检测和 AIOps ML 模型可保护工作负载，而无需使用客户数据进行培训，从而尽可能降低基于输入的攻击风险。

AIOps

戴尔 AIOps 提供自动化持续监控，能检测错误配置、漏洞（包括 CVE），并支持影响 AI/LLM 工作负载的供应链风险意识。通过实时 CVE 扫描、智能警报和 AI 驱动的控制面板，可标记异常并跟踪处置流程，从而助力快速干预。内置合规性功能、基于角色的访问控制和自动化报告可帮助跨工作负载维护安全运营，而无缝的 EDR/XDR 集成和AI 驱动型运营见解（包括受支持解决方案中的生成功能）可进一步提高 IT 效率。

网络产品

Dell Networking 解决方案通过强大的网络分段，保护 AI/LLM 环境，从而尽可能减少横向移动。密网络路径和集成防火墙控制，可阻止未经授权访问 AI 数据。

AI 安全性和弹性服务

戴尔的 AI 安全性与弹性服务旨在应对贵组织集成 AI 所带来的新风险。我们的服务旨在与您的团队并肩合作，助您尽快启用 AI。我们提供专业知识，指导战略规划、解决方案实施以及托管安全服务，以减轻运营负担，让您能够安全地利用 AI 进行创新。它们都经过量身定制，可帮助组织应对不断变化的 AI 风险并优化安全的 AI 部署。

Dell AI Factory

专门构建的安全保护的集成产品组合，例如戴尔的安全供应链、零信任功能（用于实施最低权限），以及 AI MDR 解决方案（旨在确保模型安全可靠）。

* 基于戴尔在 2024 年 10 月（英特尔）和 2025 年 3 月 (AMD) 进行的内部分析。适用于搭载英特尔和 AMD 处理器的 PC。并非所有 PC 都提供所有功能。某些功能需要额外购买。搭载英特尔处理器的 PC 经 Principled Technologies 验证，2025 年 7 月。

结论

为了构建弹性 AI 框架，组织与安全专家之间的协作方法至关重要。随着 AI 和 LLM 持续重塑各行业格局，有效应对其带来的数据安全性、模型完整性和合规性挑战至关重要。组织必须采取主动防御战略，将安全防护深度融入 AI 之旅的每个阶段。

Dell Technologies 是这一使命中值得信赖的合作伙伴，提供端到端生成式 AI 自定义、安全咨询和集成式解决方案，以满足您的独特需求。利用戴尔强大的网络安全解决方案，企业可以有效降低 AI 和 LLM 风险，同时更大限度地发挥现有安全投资的潜力。通过将高级安全性无缝集成到现有框架中，戴尔为组织提供了一种保护其 AI 基础架构的方法，确保了未来就绪、安全可靠的环境。

了解戴尔全面的 AI 解决方案如何为您的生成式 AI 和 LLM 环境保驾护航：Dell.com/CyberSecurityMonth