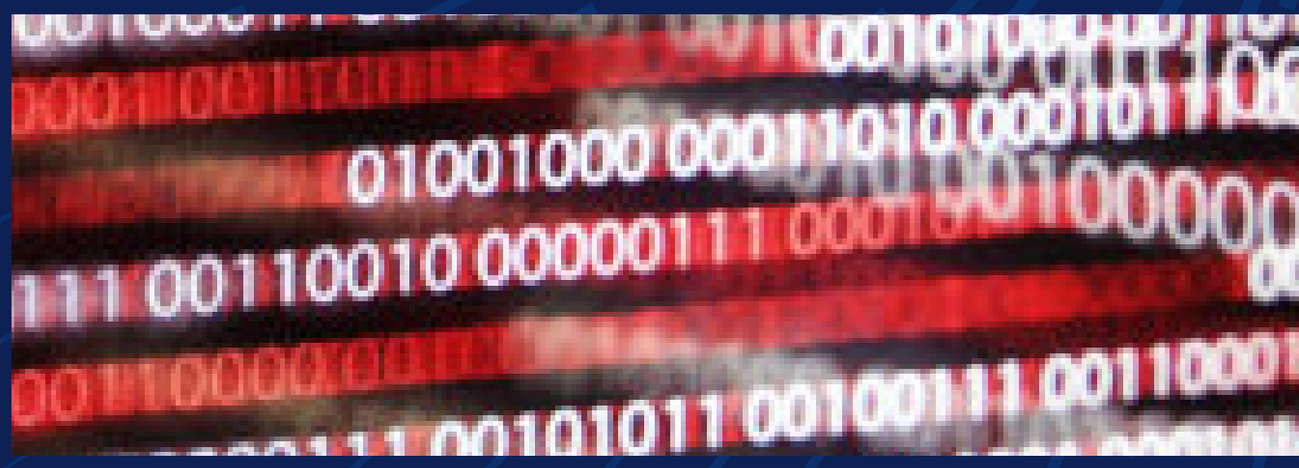
网络安全误区终结者：

打破 AI 安全误解



AI 正在改变各行各业，但在保护 AI 安全方面，许多组织被各种误区所困，误以为这比实际情况复杂得多。真相是什么？保护 AI 系统并不需要完全从零开始 – 将现有的网络安全原则应用于 AI 的独特挑战大有裨益。

Dell Technologies 了解 AI 背后的体系结构，可以帮助您调整当前的解决方案，以适应这一新框架。让我们来打破围绕 AI 安全性的常见误解，揭示真相，帮助您有效保护系统。

误解 1：“AI 系统过于复杂，无法保护。”

真相：AI 确实会带来新的网络安全风险，例如提示注入、数据操纵和敏感信息泄露等。此外，代理 AI 系统还具有更广泛的攻击面，因为它们可能会被用来操纵结果或提升权限。

但关键在于：既要识别这些漏洞并实施安全措施以保护 AI 系统来应对传统威胁与 AI 特定的威胁，也要认识到这些风险可控，AI 模型能够被有效保护。必须明确：AI 系统既需要海量数据输入，又会生成大规模数据输出。这使数据保护成为核心安全战略之一，与之并列的还包括：

- 零信任原则，例如身份管理、基于角色的访问和持续验证。
- 定期渗透测试和漏洞管理，以识别弱点。
- 日志记录和审核，以验证数据输入和输出

误解 2：“我现有的任何工具都无法保护 AI。”

真相：AI 安全并非要推倒重来 – 关键在于更智能地运用现有工具。大多数现有的网络安全工具，在经过调整后都能有效地保护 AI 系统。究其本质，AI 不过是驱动业务发展的又一工作负载 – 尽管具有独特属性。基础网络安全实践（如身份管理、网络分段和监控、端点保护和数据保护）对于保护 AI 环境仍然至关重要。关键是要调整这些实践以应对特定的 AI 挑战，例如保护训练数据、加密算法以及降低对抗性输入等风险。

强大的防御体系始于良好的网络卫生习惯，例如系统修补、访问控制和漏洞管理。关键在于量身定制这些实践，以应对 AI 特定的风险。通过将以 AI 为中心的战略融入到您现有的安全策略中，并使用合适的工具，AI 安全变得易于管理和有效。

然而，需要特别指出，更新后的硬件可以在打击网络攻击方面发挥关键作用。例如，现代 AI PC 为抵御主要攻击媒介（即端点）提供了强大的第一道防线。Windows 10 支持结束后，过时的 PC 会成为风险。除此之外，Windows 11 还需要可信平台模块 (TPM) 版本 2.0，这是一种安全芯片，它有助于加密、安全启动和防范固件攻击。许多较旧的 PC 根本没有 TPM 或仅支持较旧版本。戴尔提供内置这些增强功能的安全商用 AI PC。

服务器和存储等 AI 基础架构也是如此。Dell AI Factory 包含专为 AI 安全而优化的硬件，并且具备一系列内置安全特性，从安全供应链到数据不可变性再到隔离和加密，一应俱全。

误解 3：“AI 安全只是保护数据。”

真相：AI 安全不仅限于基本的数据保护，它还涉及到保护整个 AI 生态系统，包括模型、API、输出、系统和设备。随着 AI 越来越多地集成到关键应用程序中，与滥用或利用相关的风险也会升级。没有强大的安全措施，AI 模型可能会被篡改以生成有害或误导性的输出、API 可能会被利用来获得对敏感系统的未经授权访问，并且输出可能会无意中泄露私密或机密信息。

全面的 AI 安全性需要采用多层次防护策略。这包括：保护模型免受试图通过篡改输入数据来欺骗 AI 系统的对抗性攻击；采用强认证方法保障 API 安全以防止未经授权使用；**持续监控输出**是否存在可能预示攻击或系统故障的异常可疑模式。有效的 AI 安全不仅可确保 AI 系统的完整性和可靠性，还可降低恶意使用或意外后果的风险，从而与用户及利益相关方建立信任。

误解 4：“AI 不需要人工监督”。

真相：治理和人工监督对于确保 AI 系统以合乎道德、可预测的方式运行，并与人类价值观保持一致至关重要。高级 AI 系统，特别是具有自主决策能力的代理 AI，带来了独特的挑战，需要强大的安全措施。如果缺乏适当监管，这些系统可能偏离预期目标或产生具有风险的非预期行为。

为了解决这一问题，必须建立明确的界限、实施分层控制机制，并确保人工持续参与关键决策过程。定期审核、AI 操作透明化及全面测试，能进一步提升问责制和信任度，从而预防滥用并促进负责任地部署 AI 技术。

增强 AI 安全性的最佳实践

为了弥补 AI 特定的安全漏洞，组织需要采取主动式战略性方法。保护 AI 系统的 10 项最佳实践：

分层安全体系结构：
使用分段、防火墙和强身份验证，在每一层保护您的基础架构、软件和数据。

监控和验证 AI 输出：
使用异常检测、日志记录和警报，监控 AI 输出中的异常模式或有害行为。

保障供应链安全：
实施强大的供应商管理计划。审核供应商和第三方组件、验证完整性，并依靠签名代码来防止 AI 开发生命周期中出现漏洞。

弹性计划：
定期备份数据，并测试灾难恢复计划，以更大限度地减少停机时间，并确保在发生漏洞时能够迅速恢复。

保护培训数据和模型：
通过监控数据完整性并应用强大的验证工具，防范中毒数据、对抗性输入和其他威胁。

实施强效加密方案：
采用高强度算法对静态和传输中的敏感数据加密，实施安全管理，并定期轮换加密密钥。

强化访问控制：
执行最低权限原则、实施基于角色的访问控制 (RBAC)、定期轮换凭据和审核权限，以防止未经授权的访问。

定期执行安全审核和渗透测试：
时常评估系统漏洞，并通过渗透测试提前发现潜在风险，以防止其被利用。

保障 API 安全：
使用强身份验证协议（如 OAuth 2.0），强制实施 HTTPS 加密，并定期更新 API 以消除潜在漏洞。

AI 安全最佳实践员工培训：
定期对团队开展安全开发、威胁识别及强化安全实践的培训，以防范漏洞入侵。

DELL Technologies

戴尔的价值主张：实用的 AI 安全解决方案。

AI 安全可能看起来很复杂，但实际上并没有那么令人畏惧。真相是什么？保护 AI 与保护现有工作负载并无二致，关键在于理解其体系结构并采取合适的策略。这正是 Dell Technologies 能够大显身手的地方。

通过利用您现有的解决方案并将其无缝集成到以 AI 为中心的体系结构中，我们揭示了 AI 安全性的奥秘。我们致力于应对即时注入、API 滥用和对抗性攻击等挑战，而无需彻底改造现有基础架构。

戴尔的专业知识能够破除与 AI 安全性相关的误解，并证明其可行性其实触手可及。无论您是刚刚启程 AI 之旅，还是想要增强防御能力，我们都将帮助您守护投资、保障系统安全，同时充满信心且高效地构建弹性数字未来。让我们携手，共同简化 AI 安全性。

戴尔产品和解决方案，可以助您一臂之力

精选戴尔解决方案	说明
Dell AI Factory	Dell AI Factory 通过安全的供应链保护 AI 工作负载，实现从开发到部署的全流程可信基础架构。数据不可变性、隔离和加密等功能可保护敏感模型和数据集，防范网络威胁，并在数据驱动的动态环境中实现可扩展、高效和无缝的 AI 运营。
网络弹性	PowerProtect 通过不可变性和隔离等高级功能保护 AI 工作负载，确保数据完整性并防范网络威胁。它提供端到端加密和异常检测，同时支持快速恢复，以更大限度地减少停机时间。
Dell Trusted Workspace（端点安全性）	内置功能和可选附加功能的组合，旨在保护商用 AI PC 及其运行的 AI 工作负载。基于安全供应链实践构建，内置功能包括采用 TPM 技术的 SafeBIOS 和 SafeID。可选附加组件包括安全组件验证、SafeID 与 ControlVault，以及合作伙伴软件 CrowdStrike 和 Absolute，以更大限度地提高工作区的安全性。
AI 安全咨询服务	一套可帮助您制定和实施全面的 AI 安全战略的服务组合。这些服务包括咨询服务、AI vCISO 和数据安全规划。
AI 托管安全运营	实现跨技术堆栈的深度可见性，以快速检测和应对威胁。这些功能包括 Managed Detection and Response、Managed AI Guard、AI 渗透测试以及 Incident Response and Recovery Services。
安全软件集成	设计、安装和配置安全工具，以保护访问管理、应用程序、网络、云等。

了解如何应对当今一些主要的网络安全挑战：**dell.com/cybersecuritymonth**

DELL Technologies