DELL Technologies

NVIDIA

生成式 AI (GenAI) 的 5 大安全考虑因素

借助 Dell AI Factory with NVIDIA，加快采用安全且可扩展的基础架构基础

生成式 AI 的变革潜力

生成式 AI 有潜力带来颠覆性改变，其改变方式连最具远见的人士也才刚刚开始构想。

76%

的 IT 和企业领导者认为，生成式 AI 将为其组织带来巨大的转型价值。[1]

AI

运用高级分析和基于逻辑的技术来解释事件，并提供支持，实现决策和行动自动化。

生成式 AI

利用海量数据，根据自然语言提示或其他非代码和非传统输入生成新内容的技术和方法。

模拟

- 数字孪生
- 合成数据
- 设计框架
- 预测

内容发现

- 自然语言搜索
- 大型数据集分析
- 知识管理
- 量身定制的教育和培训

内容创作

- 编码
- 数学
- 写作/演讲
- 图像/视频
- 音频

用户体验

- 实时翻译 70 多种语言
- 使用自然的面部表情和肢体语言进行个性化交流

1 Dell Technologies Innovation Catalyst 研究，2024 年 2 月

潜力越大，风险越大

企业领导者往往急于求成，容易忽略数据、合规、治理和其他风险方面的潜在影响。但就安全性而言，生成式 AI 是一把双刃剑。

优势

- 更先进的威胁检测
- 提高操作效率
- 量身定制的安全意识培训

缺点

- 攻击复杂性不断增加
- 先进的社会工程
- 影子 AI

33%

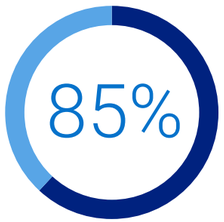
的受访者将网络安全列为其组织正在努力缓解的首要生成式 AI 风险。[2]

2 McKinsey 全球 AI 调查：AI 的早期状态，2024 年 5 月

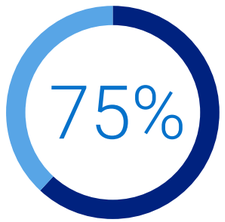
考虑因素 1

新的威胁局势

生成式 AI 带来希望的同时，也伴随着严峻的现实：攻击者正在创造更复杂的新型攻击，可以绕过传统防御措施，使网络安全团队难以应对。

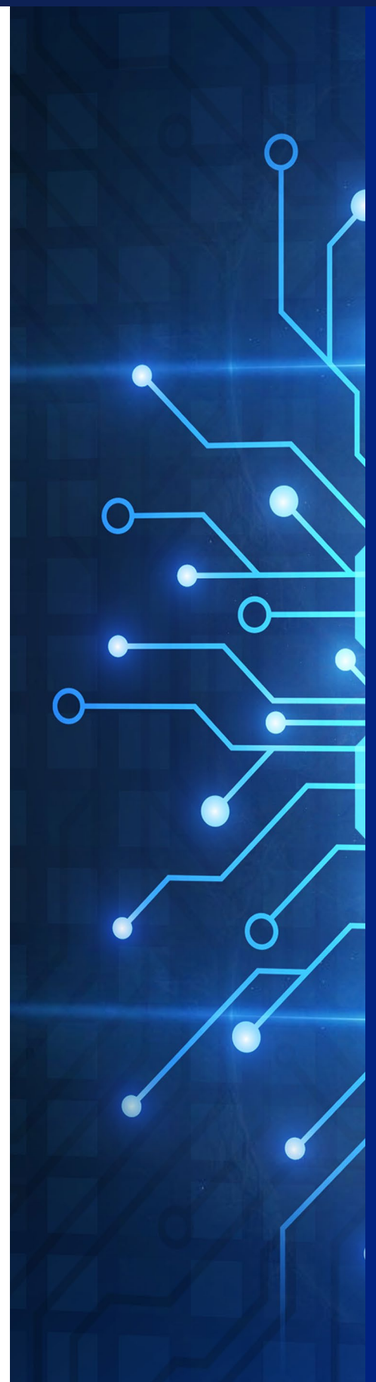


的受访者认为，AI 使网络安全攻击变得更加复杂。[3]



的安全专业人员发现，过去 12 个月攻击次数有所增加。[4]

为了防范这些新出现的威胁，公司必须集中精力通过渗透测试、监测和审计等方式尽可能减少受攻击面。



新出现的攻击载体

高级恶意软件
日益复杂的恶意软件利用生成式 AI“自我进化”，不断更改代码以躲避现有安全措施的检测，例如基于签名的检测。

高度个性化的网络钓鱼电子邮件和活动
看似真实却带有恶意的电子邮件越来越多，这些电子邮件往往没有常见的诈骗迹象。

逼真的深度伪造数据
模仿人类行为（例如写作、语音、图像或视频）的能力使身份盗窃、金融欺诈和虚假信息传播变得更加容易。

自动侦测
收集信息以识别潜在目标网络或系统中的漏洞和弱点，从而发动更有针对性的攻击。

3 2024 年网络安全领域的人为风险调查，EY，2024 年 5 月
4 Voice of SecOps 报告“生成式 AI 和网络安全：光明的未来还是商业战场？”2023

考虑因素 2

部署和实施风险

想要利用生成式 AI 潜在优势的组织需要大量的优质数据，模型可以利用这些输入来产出更好的结果。但是，数据和风险形影相随。在利用任何信息之前，公司必须仔细评估并考虑其独特的要求、输入和风险。

大型语言模型 (LLM) 的漏洞

生成式 AI 服务容易受到提示注入攻击，攻击者可以通过操纵输出来绕过安全防护或未经授权地访问可能用于优化模型的文件。

数据投毒

攻击者可能在训练阶段故意将经过篡改的数据提供给 LLM。这可能导致模型容易受到通过数据中嵌入的后门发起的攻击。一个实例是利用垃圾邮件训练垃圾邮件过滤器，从而攻击和利用垃圾邮件过滤器。

监管复杂性

世界各地的监管机构都在争相了解、控制和保障生成式 AI 的安全性。虽然生成式 AI 模型受当前数据主权规则的约束，这些规则规定了数据的存储、处理和使用方式，但监管机构仍在界定对知识产权和版权信息的监管。遵守法规可能成本高昂，但不遵守现行和新兴法规可能会导致罚款和其他处罚。

考虑因素 3

影子 AI

如今，许多员工已在使用 ChatGPT 等公共文本、图像和视频生成器为日常工作流赋能。但是，如果在没有适当治理措施的情况下使用这些工具，会对试图保护企业知识产权和数据的组织构成严重威胁。这种未经授权使用生成式 AI 的行为被称为影子 AI。

知识产权损失

许多公司已经面临员工在公共生成式 AI 工具中共享敏感信息造成的知识产权损失。

源代码数据泄露

开发人员试图使用 ChatGPT 优化源代码，却导致了数据泄露。

为了应对影子 AI 带来的挑战，公司应设立一个公司级的委员会或董事会，授权其就安全的 AI 治理做出决策。

您的数据位于什么位置？工作负载应该放在哪里？

无论数据存储在哪里，将 AI 与数据结合使用才能获得最佳效果。完全控制基础架构和 LLM，即可消除知识产权损失或源代码数据泄露的风险。

成本

利用本地实施，可在 3 年内将 TCO 降低多达 75%。[5]

安全和隐私

在整个组织内创建安全的 AI/生成式 AI 环境，并在本地执行工作流和运营。严格控制数据安全并遵守合规性法规，尤其是在处理敏感数据的行业。

5 基于戴尔委托 Enterprise Strategy Group 进行的研究，比较了本地戴尔基础架构与原生公有云基础架构即服务，2024 年 4 月。分析的模型显示，对于拥有 5,000 名用户的组织，利用 RAG 的 70 亿参数 LLM 的成本效益提高了 38%；而对于拥有 50,000 名用户的组织，利用 RAG 的 700 亿参数 LLM 的成本效益提高了 75%。实际结果可能有所不同。**经济性摘要**

考虑因素 4

评估标准

过去一年，AI 社区越来越关注这 3 个关键问题：负责任的开发和部署、影响评估和风险缓解。公司在评估生成式 AI 模型时，必须考虑以下重要事项：

缺乏一致的报告要求

优秀的开发人员主要根据不同的负责任 AI 基准来测试他们的模型。由于报告标准化严重不足，很难系统地比较主流 AI 模型的风险和局限性。

输出中包含受版权保护的材料

主流 LLM 的输出可能包含受版权保护的材料，这可能违反法律，并导致使用这些材料的公司面临处罚风险。

漏洞日益复杂

研究人员发现有些不太明显的策略会导致 LLM 表现出有害行为，例如要求模型无限重复输出随机单词。

开发人员缺乏透明度

很多情况下，AI 开发人员不愿透露其训练数据和方法。这阻碍了人们进一步了解 AI 系统的稳健性和安全性。

考虑因素 5

安全优势

生成式 AI 除了带来安全风险，也蕴藏着潜在的安全优势。生成式 AI 正在成为网络安全领域的重要盟友，开辟新的保护途径。

现在，您可以开始建立可扩展的安全运营，更快地获得更丰富的见解，并自动检测相关威胁，从而提高效率，并弥补安全团队人员不足的问题。

威胁检测和响应

通过分析历史数据并识别相关模式和异常，生成式 AI 可以实时识别新的和不断演变的威胁。它可以持续监控网络流量、系统日志和用户行为，并及时识别可能预示安全威胁的异常活动。

由此，即可实现强大的自适应威胁检测，能够快速响应不断变化的攻击载体，并提供针对新兴网络威胁的主动防御机制。

威胁模拟和训练

借助生成式 AI，公司可以在受控环境中模拟各种网络安全威胁和攻击场景。因此，团队可以未雨绸缪，从而在时间紧迫的情况下识别、响应和缓解网络威胁。

深入分析和总结

生成式 AI 使团队能够调查来自不同来源或模块的数据，从而更快、更准确地执行传统上耗时且繁琐的数据分析。团队还可以创建事件和威胁评估的自然语言摘要，从而提高效率和团队产出。

量身定制的安全意识培训

通过在生成式 AI 基础上封装对话式 AI，并将 AI 虚拟形象整合到用户界面中，组织可以实现个性化的交互（可全天候大规模应用），并利用自然的面部表情和肢体语言。这可以用于安全培训和教育，提供更自然、定制化和交互式的学习体验、自动评估等。

Dell AI Factory with NVIDIA

借助业界较早推出的全面、全包式 AI 解决方案，加速 AI 之旅，并安全地将数据转化为见解。对于寻求利用 AI 和生成式 AI 的企业，Dell AI Factory with NVIDIA 可以满足其复杂的需求。借助卓越的基础架构、服务和 NVIDIA AI 软件，您可以简化开发和部署，从而加快项目价值实现速度。

- 通过具有固有安全性（包括信任根和其他关键功能）的基础架构降低入侵风险。
- 利用您所控制的本地 AI 解决方案，保护您的数据免遭泄露，避免知识产权损失。
- 通过为数据引入 AI 并实现安全访问，满足严格的合规性和数据主权要求。
- 通过控制有权访问数据的地点和人员来保护利益相关者的隐私。

Dell AI Factory with NVIDIA

业界率先推出的端到端企业 AI 解决方案

数据为 AI 工厂和您的应用场景提供动力

您更有价值的数据位于本地和边缘。Dell Technologies 可帮助您将 AI 应用于这些宝贵数据，是存储、保护和管理这些数据的先行者。

应用场景和成果

AI 工厂聚焦优先级更高的应用场景，以产生出色的业务成果。Dell Technologies 通过经验证的解决方案和量身定制的服务，简化了重要 AI 应用场景的部署。

不要让安全风险扼杀创新

让我们帮助您在 AI 和生成式 AI 领域如鱼得水，收获丰厚回报。

战略规划

适用于生成式 AI 的免费 Accelerator Workshop

- 开始制定制胜战略
- 解决挑战、弥补差距，确定目标的优先顺序，发现机会
- 获取就绪性评估，以便更深入地了解基础架构要求、AI 模型、运营集成等。

技术储备

即用型移动实验室

快速开启您的成功之旅。包括搭载 NVIDIA GPU 的戴尔 Precision 移动工作站 5690/7780 和为期两天的咨询服务，助您开启成功之旅。

- 用于生成式 AI 测试和演示的便携式沙盒环境
- 经过预验证的 NVIDIA AI Workbench 平台已向开发者开放
- 利用您的数据实施的初始聊天机器人应用场景
- 经济高效、低风险的生成式 AI 技能实验和培养方法

搭载 NVIDIA GPU 的戴尔 PRECISION 移动工作站 5690/7780

立即行动

DELL Technologies AI Factory WITH NVIDIA