DELL Technologies

NVIDIA

Die 5 wichtigsten Sicherheitsüberlegungen für generative KI (GenAI)

Beschleunigen Sie die Einführung einer sicheren und skalierbaren Infrastrukturgrundlage mit Dell AI Factory with NVIDIA.

Das transformative Potenzial von GenAI

GenAI hat das Potenzial, eine Wende herbeizuführen, die sich VisionärInnen gerade einmal vorstellen können.

76 %

der IT- und Unternehmensführungskräfte sind der Ansicht, dass GenAI einen transformativen Mehrwert für ihr Unternehmen bringen wird.[1]

KI

Fortschrittliche Analysen und logikbasierte Techniken, um Ereignisse zu interpretieren und die Entscheidungsfindung sowie Aktionen zu unterstützen und zu automatisieren.

GenAI

Technologien und Techniken, die große Datenmengen nutzen, um neue Inhalte aus Prompts in natürlicher Sprache oder anderen nicht codierten und nicht herkömmlichen Eingaben zu generieren.

Simulation

- Digitaler Zwilling
- Synthetische Daten
- Design-Frameworks
- Prognose

Inhaltserstellung

- Kodierung
- Mathematik
- Schrift/Sprache
- Bild/Video
- Audio

Inhaltserkennung

- Suche in natürlicher Sprache
- Analyse großer Datasets
- Wissensmanagement
- Personalisierte Aus- und Weiterbildung

Nutzererlebnis

- Echtzeit-Übersetzungen für 70+ Sprachen
- Personalisierte Interaktionen mit natürlicher Mimik und Körpersprache

1 Dell Technologies Innovation Catalyst Study, Februar 2024.

Höheres Potenzial, erhöhtes Risiko

Es ist verlockend für Führungskräfte, schnell zu handeln und die Auswirkungen auf Daten, Compliance, Governance und andere Risiken zu vernachlässigen. Aber GenAI ist ein zweischneidiges Schwert, wenn es um Sicherheit geht.

Vorteile

- Verbesserte Bedrohungserkennung
- Verbesserte Betriebseffizienz
- Personalisierte Schulung zur Sicherheitssensibilisierung

Nachteile

- Immer raffiniertere Angriffe
- Erweitertes Social Engineering
- Schatten-KI

33 %

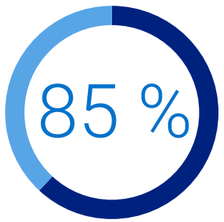
der Befragten stuften Cybersicherheit als das größte GenAI-Risiko ein, an dessen Eindämmung ihr Unternehmen arbeitet.[2]

2 Globale Umfrage von McKinsey zu KI: Der Stand der KI Anfang Mai 2024

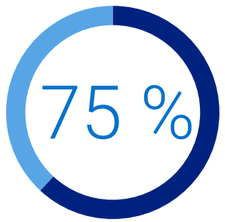
ÜBERLEGUNG 1

Die neue Bedrohungslandschaft

Mit dem Versprechen von GenAI geht eine ernüchternde Realität einher: AngreiferInnen entwickeln neue und komplexere Angriffe, die herkömmliche Abwehrmechanismen umgehen können, was es für Cybersicherheitsteams schwierig macht, Schritt zu halten.



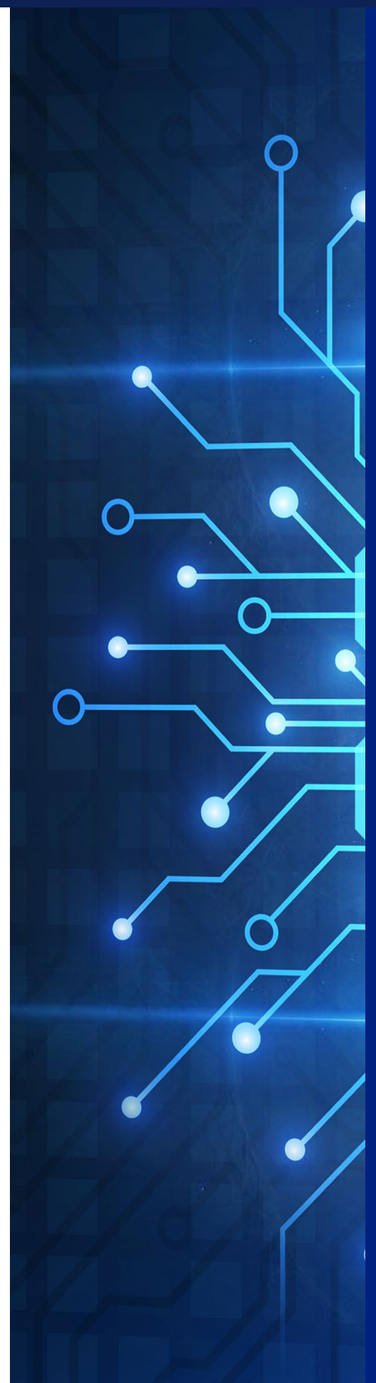
der Befragten sind der Meinung, dass Cybersicherheitsangriffe durch KI ausgefeilter geworden sind.[3]



der Sicherheitsexperten verzeichneten in den letzten 12 Monaten einen Anstieg der Angriffe.[4]

Um sich vor diesen neu auftretenden Bedrohungen zu schützen, müssen sich Unternehmen darauf konzentrieren, die Angriffsfläche zu minimieren, beispielsweise durch Penetrationstests, Monitoring und Überprüfung.

3 2024 Human Risk in Cybersecurity Survey, EY, Mai 2024
4 Voice-of-SecOps-Bericht „Generative AI and Cybersecurity: Bright Future or Business Battleground?" 2023



Neue Angriffsvektoren

Fortschrittliche Malware
Immer ausgefeiltere Malware, die GenAI nutzt, um „sich selbst weiterzuentwickeln" und ihren Code kontinuierlich so zu ändern, dass er von vorhandenen Sicherheitsvorkehrungen, wie z. B. signaturbasierter Erkennung, nicht erkannt wird.

Hochgradig personalisierte Phishing-E-Mails und -Kampagnen
Zunehmende Häufigkeit von authentisch aussehenden bösartigen E-Mails, denen die üblichen Betrugszeichen fehlen.

Überzeugende Deep-Fake-Daten
Identitätsdiebstahl, Finanzbetrug und Fehlinformationen werden durch die Möglichkeit, menschliche Handlungen wie Schrift, Sprache, Bilder oder Videos nachzuahmen, erleichtert.

Automatisierte Aufklärung
Informationsbeschaffung zur Identifizierung von Sicherheitslücken und Schwachstellen im Netzwerk oder System eines potenziellen Ziels, um gezieltere Angriffe zu ermöglichen.

ÜBERLEGUNG 2

Bereitstellungs- und Implementierungsrisiken

Unternehmen, die von den potenziellen Vorteilen von GenAI profitieren möchten, benötigen große Mengen hochwertiger Daten – Eingaben, die Modelle nutzen können, um die besten Ergebnisse zu erzielen. Aber Daten und Risiken gehen Hand in Hand. Bevor Unternehmen Informationen nutzen, müssen sie ihre individuellen Anforderungen, Eingaben und Risiken sorgfältig bewerten und berücksichtigen.

Schwachstellen von großen Sprachmodellen (LLMs)

GenAI-Services sind anfällig für Prompt-Injection-Angriffe, bei denen Angreifer die Ausgaben manipulieren, um Sicherheitsvorkehrungen zu umgehen oder unbefugten Zugriff auf Dateien zu erhalten, die möglicherweise zur Optimierung des Modells verwendet wurden.

Datenvergiftung

AngreiferInnen können während der Trainingsphase absichtlich veränderte Daten in ein LLM einspeisen. Dies kann dazu führen, dass das Modell durch in die Daten eingebettete Hintertüren anfällig für Angriffe wird. Ein Beispiel aus der Praxis ist der Angriff und die Ausnutzung von Spam-Filtern, indem diese mit Spam-E-Mails trainiert werden.

Komplexität behördlicher Auflagen

Regulierungsbehörden weltweit bemühen sich, GenAI zu verstehen, zu kontrollieren und die Sicherheit von GenAI zu gewährleisten. Während GenAI-Modelle den aktuellen Regeln zur Datenhoheit unterliegen, die vorschreiben, wie Daten gespeichert, verarbeitet und verwendet werden, sind die Leitungsgremien noch dabei, die Kontrolle über geistiges Eigentum und urheberrechtlich geschützte Informationen zu definieren. Die Einhaltung der Bestimmungen kann kostspielig sein, aber die Nichteinhaltung bestehender und neu entstehender Bestimmungen kann zu Geldbußen und anderen Strafen führen.

ÜBERLEGUNG 3

Schatten-KI

Viele MitarbeiterInnen nutzen bereits heute öffentliche Text-, Bild- und Videogeneratoren wie ChatGPT, um ihre täglichen Arbeitsabläufe zu optimieren. Wenn diese Tools jedoch ohne angemessene Governance eingesetzt werden, stellen sie eine kritische Bedrohung für Unternehmen dar, die versuchen, ihr geistiges Eigentum und ihre Daten zu schützen. Diese nicht autorisierte Nutzung von GenAI wird als Schatten-KI bezeichnet.

Verlust von geistigem Eigentum

Schon jetzt haben Unternehmen mit dem Verlust geistigen Eigentums zu kämpfen, weil MitarbeiterInnen vertrauliche Informationen in öffentlichen GenAI-Tools freigeben.

Quellcode-Datenverlust

EntwicklerInnen, die versuchen, den Quellcode mithilfe von ChatGPT zu optimieren, haben Datenlecks verursacht.

Um den Herausforderungen der Schatten-KI zu begegnen, sollten Unternehmen einen unternehmensweiten Rat oder Ausschuss einrichten, der befugt ist, Entscheidungen im Zusammenhang mit einer sicheren KI-Governance zu treffen.

Wo befinden sich Ihre Daten? Wo sollten Workloads platziert werden?

KI funktioniert am besten, wenn sie mit Ihren Daten verknüpft ist, unabhängig davon, wo diese gespeichert sind. Mit vollständiger Kontrolle über Infrastruktur und LLMs besteht kein Risiko von IP-Verlust oder Quellcode-Datenlecks.

Kosten

Durch die Nutzung von On-Premise-Implementierungen können die Gesamtbetriebskosten innerhalb von 3 Jahren um bis zu 75 % gesenkt werden.[5]

Sicherheit und Datenschutz

Gewährleisten Sie mit lokalen Workflows und Abläufen sichere KI-/GenAI-Umgebungen im gesamten Unternehmen. Üben Sie strenge Kontrolle über die Datensicherheit und die Einhaltung von Compliance-Bestimmungen aus, insbesondere in Branchen, die mit vertraulichen Daten umgehen.

5. Basierend auf einer von Dell in Auftrag gegebenen Studie der Enterprise Strategy Group, in der die On-Premise-Infrastruktur von Dell mit einer nativen Infrastruktur as a Service-Lösung in der Public Cloud verglichen wurde, April 2024. Analysierte Modelle zeigen, dass ein LLM mit 7 Mrd. Parametern mit RAG für ein Unternehmen mit 5.000 NutzerInnen bis zu 38 % kosteneffizienter ist, während ein LLM mit 70 Mrd. Parametern mit RAG für ein Unternehmen mit 50.000 NutzerInnen bis zu 75 % kosteneffizienter ist. Die tatsächlichen Ergebnisse können abweichen. Wirtschaftliche Zusammenfassung

ÜBERLEGUNG 4

Bewertungskriterien

Im letzten Jahr hat sich die KI-Community zunehmend auf drei Schlüsselthemen konzentriert: verantwortungsvolle Entwicklung und Bereitstellung, Bewertung der Auswirkungen und Risikominderung. Bei der Bewertung von GenAI-Modellen müssen Unternehmen einige wichtige Einschränkungen berücksichtigen:

Keine einheitlichen Berichtsanforderungen

Führende EntwicklerInnen testen ihre Modelle in erster Linie anhand verschiedener Benchmarks für verantwortungsvolle KI. Aufgrund dieses erheblichen Mangels an Standardisierung bei der Berichterstattung ist es schwierig, die Risiken und Grenzen der führenden KI-Modelle methodisch zu vergleichen.

Urheberrechtlich geschütztes Material in Ausgaben

Die Ausgaben beliebter LLMs können urheberrechtlich geschütztes Material enthalten, was möglicherweise gesetzeswidrig ist und zu Strafen für Unternehmen führen kann, die das Material verwenden.

Sicherheitslücken werden immer komplexer

Forscher finden weniger offensichtliche Strategien, die dazu führen, dass LLMs schädliches Verhalten aufzeigen, wie z. B. die Aufforderung an Modelle, zufällige Wörter unendlich oft zu wiederholen.

EntwicklerInnen mangelt es an Transparenz

In vielen Fällen sind KI-EntwicklerInnen nicht bereit, ihre Trainingsdaten und -methoden offenzulegen. Dies behindert die Bemühungen, die Robustheit und Sicherheit von KI-Systemen besser zu verstehen.

ÜBERLEGUNG 5

Sicherheitsvorteile

Neben den Sicherheitsrisiken von GenAI gibt es auch potenzielle Sicherheitsvorteile. GenAI wird zu einem wichtigen Verbündeten in der Cybersicherheit, der neue Möglichkeiten zum Schutz vor Cyberangriffen ermöglicht.

Sie können jetzt skalierbare Sicherheitsabläufe mit schnellerem Zugriff auf umfassendere Erkenntnisse und automatische Bedrohungserkennung aufbauen – und so für mehr Effizienz sorgen und unterbesetzte Sicherheitsteams ergänzen.

Threat Detection and Response

Durch die Analyse historischer Daten und die Identifizierung von Mustern und Anomalien kann GenAI neue und sich entwickelnde Bedrohungen in Echtzeit erkennen. Sie kann den Netzwerkverkehr, Systemprotokolle und das Nutzerverhalten kontinuierlich überwachen und Unregelmäßigkeiten, die auf Sicherheitsbedrohungen hindeuten können, umgehend erkennen.

Das Ergebnis ist eine leistungsstarke adaptive Bedrohungserkennung, die eine schnelle Reaktion auf sich ändernde Angriffsvektoren und einen proaktiven Abwehrmechanismus gegen aufkommende Cyberbedrohungen ermöglicht.

Bedrohungssimulation und -training

Mit GenAI können Unternehmen eine Vielzahl von Cybersicherheitsbedrohungen und Angriffsszenarien in einer kontrollierten Umgebung simulieren. Dadurch sind Teams besser darauf vorbereitet, Cyberbedrohungen zu erkennen, darauf zu reagieren und sie zu minimieren, wenn es schnell gehen muss.

Detaillierte Analyse und Zusammenfassung

GenAI ermöglicht es Teams, Daten aus unterschiedlichen Quellen oder Modulen zu untersuchen, sodass sie traditionell zeitintensive, mühsame Datenanalysen schneller und genauer durchführen können. Teams können auch Zusammenfassungen von Vorfällen und Bedrohungsbewertungen in natürlicher Sprache erstellen, wodurch die Effizienz verbessert und die Teamleistung gesteigert werden.

Personalisierte Schulung zur Sicherheitssensibilisierung

Durch die Kombination von konversationeller KI mit GenAI und die Integration eines KI-Avatars in die Benutzeroberfläche können Unternehmen personalisierte Interaktionen (24/7 verfügbar) mit natürlicher Mimik und Körpersprache bereitstellen. Dies kann für Schulungen und Weiterbildungen im Bereich Sicherheit genutzt werden und bietet eine natürlichere, individuellere und interaktivere Lernerfahrung, automatisierte Bewertungen und vieles mehr.

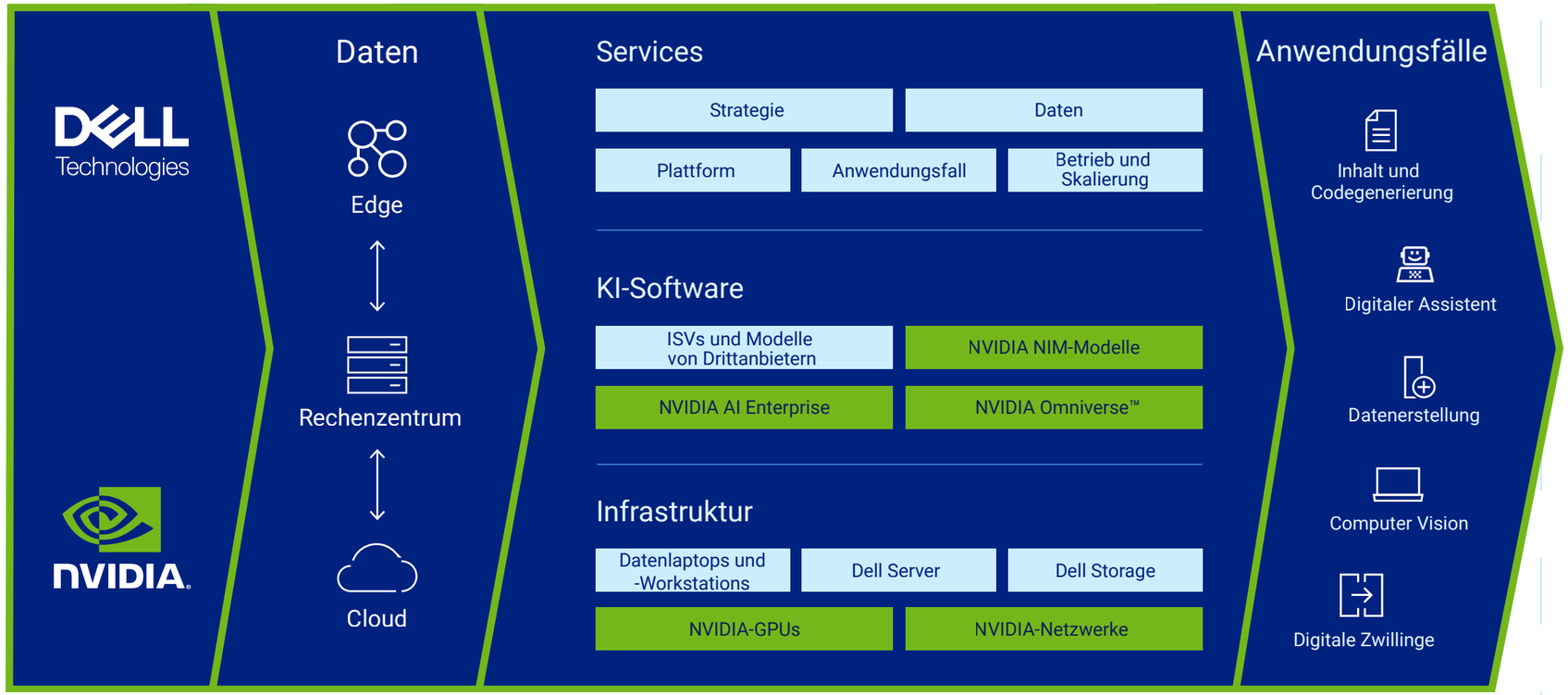
Dell AI Factory with NVIDIA

Beschleunigen Sie Ihre KI-Reise und transformieren Sie Ihre Daten sicher in Erkenntnisse – mit der branchenweit ersten umfassenden, gebrauchsfertigen KI-Lösung. Dell AI Factory with NVIDIA erfüllt die komplexen Anforderungen von Unternehmen, die KI und GenAI nutzen möchten. Mit führender Infrastruktur und Services in Verbindung mit KI-Software können Sie die Time-to-Value für Ihre Projekte verkürzen, indem Sie die Entwicklung und die Bereitstellung vereinfachen.

- Verringern Sie das Risiko von Datenkompromittierungen durch eine Infrastruktur, die über intrinsische Sicherheitvorkehrungen verfügt, einschließlich Root of Trust-Sicherheit und anderer wichtiger Funktionen.
- Schützen Sie sich vor Datenlecks, die zum Verlust von geistigem Eigentum führen könnten, mit einer On-Premise-KI-Lösung, die Sie kontrollieren.
- Erfüllen Sie strenge Anforderungen an Compliance und Datenhoheit, indem Sie KI mit sicherem Zugriff auf Ihre Daten bereitstellen.
- Schützen Sie die Privatsphäre Ihrer Stakeholder, indem Sie kontrollieren, wer und wo Zugriff auf Ihre Daten hat.

Dell AI Factory with NVIDIA

BRANCHENWEIT ERSTE UMFASSENDE ENTERPRISE-KI-LÖSUNG



Daten sind der Antrieb für die AI Factory und Ihre Anwendungsfälle
Ihre wertvollsten Daten befinden sich in On-Premise-Umgebungen und am Edge. Dell Technologies unterstützt Sie dabei, KI für diese wertvollen Daten zu nutzen, und ist führend beim Speichern, Schützen und Managen dieser Daten.

Vom Anwendungsfall bis zu den Ergebnissen
Die AI Factory erzeugt Geschäftsergebnisse, die auf Ihren Anwendungsfällen mit der höchsten Priorität basieren. Dell vereinfacht die Bereitstellung Ihrer wichtigsten KI-Anwendungsfälle mit validierten Lösungen und maßgeschneiderten Services.

Lassen Sie nicht zu, dass Sicherheitsrisiken Innovationen ausbremsen

Wir helfen Ihnen, sich in der Welt von KI und GenAI zurechtzufinden, damit Sie die Früchte Ihrer Arbeit ernten können.

STRATEGIEPLANUNG

Kostenloser Accelerator Workshop für generative KI

- Der erste Schritt zur Entwicklung einer erfolgreichen Strategie
- Herausforderungen und Lücken erkennen, Ziele priorisieren und Chancen identifizieren
- Eine Bereitschaftsbewertung für einen tieferen Einblick in die Anforderungen an die Infrastruktur, KI-Modelle, betriebliche Integrationen und vieles mehr

TECHNISCHE VORBEREITUNG

Sofort einsatzbereites mobiles Labor

Starthilfe für Ihren Weg zum Erfolg Umfasst eine Dell Precision Mobile Workstation 5690/7780 mit NVIDIA-Grafikprozessoren sowie zwei Tage Beratungsservices für die ersten Schritte.

- Portable Sandbox-Umgebung für GenAI-Tests und -Demonstrationen
- Vorab validiert mit der NVIDIA AI Workbench-Plattform, bereit für EntwicklerInnen
- Erster Chatbot-Anwendungsfall, der mit Ihren Daten implementiert wird
- Kostengünstiger, risikoarmer Ansatz zum Experimentieren und Entwickeln von GenAI-Kompetenzen

DELL PRECISION MOBILE WORKSTATION 5690/7780 MIT NVIDIA-GRAFIKPROZESSOREN

Beginnen Sie noch heute!

DELL Technologies AI Factory WITH NVIDIA