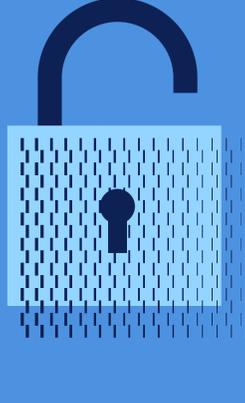


Erzielen besserer Ergebnisse



Dell Generative AI Solutions – alles, was Sie für das Inferencing großer Sprachmodelle benötigen

Generieren Sie qualitativ hochwertigere und schnellere Time-to-Value-Vorhersagen und -Ergebnisse und beschleunigen Sie gleichzeitig die Entscheidungsfindung mit einer leistungsstarken GenAI-Komplettlösung von Dell Technologies und NVIDIA. Diese gemeinsame Architektur bietet ein modulares, sicheres und skalierbares Design, das eine Vielzahl von Inferencing-Anwendungsfällen und -Rechenanforderungen unterstützt.

Jetzt können Sie die Einführung von GenAI in Ihrem Unternehmen vereinfachen und das Erreichen von Ergebnissen mit einer bewährten Lösung beschleunigen.



Schaffung neuer Werte mit einer sicheren Infrastruktur für Ihre geschäftskritischen Abläufe



Verbesserung des IT-Werts mit strategischer Anleitung



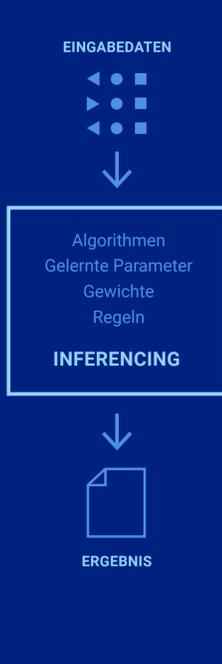
Anpassung der Größe Ihrer Infrastruktur mit flexiblen Nutzungsmodellen

Was ist Inferencing?

Inferencing in der KI bezieht sich auf den Prozess der Verwendung eines trainierten Modells, um Vorhersagen und Entscheidungen zu treffen oder Ergebnisse auf der Grundlage von Eingabedaten zu erzeugen. Dabei werden die gelernten Kenntnisse und Muster, die während der Trainingsphase des Modells erworben wurden, auf neue, ungesehene Daten angewendet.

Beim Inferencing nimmt das trainierte Modell Eingabedaten auf und verarbeitet sie durch seine Rechenalgorithmen oder die neuronale Netzwerkarchitektur, um ein Ergebnis zu erzeugen oder eine Vorhersage zu treffen. Das Modell wendet seine gelernten Parameter, Gewichte oder Regeln an, um die Eingabedaten in sinnvolle Informationen oder Aktionen umzuwandeln.

Das Inferencing ist eine entscheidende Phase im Lebenszyklus eines KI-Systems. Nach dem Training eines Modells mit markierten oder nicht markierten Daten, um Muster und Korrelationen zu erlernen, kann das Modell durch Inferencing sein Wissen verallgemeinern und Vorhersagen treffen oder Antworten auf reale oder ungesehene Daten generieren.



Wie nutzen Sie GenAI?

Generierung natürlicher Sprache:

Die Modelle können für Textgenerierungsaufgaben wie das Verfassen von Dokumenten und die Erstellung von Dialogen, Zusammenfassungen oder Inhalten genutzt werden.

Chatbots und virtuelle Assistenten:

Unterstützen Sie Conversational Agents, Chatbots und virtuelle Assistenten, indem Sie Antworten in natürlicher Sprache auf Basis von Nutzeranfragen oder -anweisungen generieren.

Codeentwicklung:

Erhalten Sie Hilfe bei der Softwareentwicklung mit Funktionen wie Codevervollständigung, der Möglichkeit, Einheitentests zu generieren, oder einer Chatfunktion zur Codeerklärung.

Dell Validated Design for Generative AI with NVIDIA - Inferencing

Beschleunigen Sie die Bereitstellung und reduzieren Sie Risiken mit vorab getesteten und bewährten Lösungen, um Probleme bei Design, Planung und Einführung zu vermeiden. Die Komponenten können gemischt und angepasst sowie je nach Anwendungsbedarf unabhängig skaliert werden.



Framework generativer KI

Dialogorientierte KI: NVIDIA Nemo

End-to-End-Enterprise-Framework für EntwicklerInnen zum Erstellen, Anpassen und Bereitstellen generativer KI-Modelle mit Milliarden von Parametern



AI Ops- und ML Ops-Plattformen

NVIDIA AI Enterprise

Partner AI Ops-Software für eine reibungslose Endnutzenerfahrung, einschließlich interaktiver Laptops, Experimentmanagement, Pipelines und mehr



Softwareinfrastruktur

NVIDIA Base Command Manager Essentials

Orchestrierungs- und Planungsebene für die Durchführung von KI-Trainings, einschließlich Aufträgen mit mehreren Nodes, und Skalierung der Inferenz



Infrastrukturmanagement

OpenManage Enterprise OneFS | CloudIQ

Vertraute Dell Managementtools mit proaktivem Monitoring und vorausschauenden Analysen zur Vereinfachung des Infrastrukturbetriebs



Hardwareinfrastruktur

Dell Server, Storage, Netzwerke | NVIDIA-GPUs

Compute mit Accelerators

Dell PowerEdge R760xa-Server mit NVIDIA A100- oder H100-GPUs

Storage

Support für Dell PowerScale, ECS und ObjectScale

Netzwerke

Dell PowerSwitch S5232F-ON oder S5248F-ON

Mit unserer Hilfe schneller Ergebnisse erzielen

Die Dell Services-ExpertInnen können Sie in jeder Phase Ihres Weges zu GenAI unterstützen:

Strategieentwicklung

Erstellung Ihrer Roadmap zum Erreichen der Innovationsziele Ihrer IT- und GeschäftsstakeholderInnen

Implementierung

Etablierung Ihrer Plattform unter Nutzung von Dell Validated ExpertIn Implementierung von Inferencing-Hardware und -Software für GenAI

Einführung

Beschleunigung des Werts Ihrer Anwendungsfälle durch die Implementierung eines vorab trainierten Inferencing-Modells

Skalierung

Verwaltung Ihres Innovationsportfolios mit ortsansässigen technischen ExpertInnen und Schulungsangeboten zur Entwicklung der GenAI-Fähigkeiten Ihres Teams

Dell Technologies und NVIDIA

Dell Technologies und NVIDIA arbeiten zusammen, um generative KI-Workloads zu ermöglichen und zu beschleunigen. Sie bieten technisch validierte Hardware und Software zur Beschleunigung von KI-, ML- und DL-Workloads, um die Kundenanforderungen in allen Unternehmen und Branchen zu erfüllen. Mit Dell Technologies und NVIDIA können Sie KI-Lösungen bereitstellen, um Ihre digitale Transformation durch Echtzeitdaten zu beschleunigen. Sie verbessern die wichtige Entscheidungsfindung mit Lösungen, die für eine schnelle Time-to-Value aus Ihren KI-Initiativen optimiert sind.



[Weitere Informationen zu den Lösungen von Dell](#)



Dell Technologies ExpertIn [kontaktieren](#)



[Weitere Ressourcen](#)



Reden Sie mit: [#PowerEdge @DellTech](#)