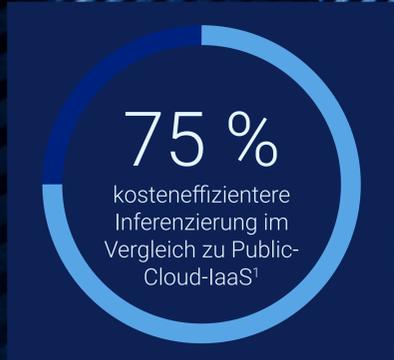


# Dell AI Factory

Generative KI-Lösung von Dell mit AMD

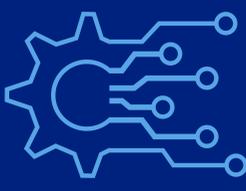
Beschleunigen Sie Innovationen, senken Sie Kosten und schützen Sie Daten mit einer skalierbaren und modularen Architektur für komplexe GenAI.



Wichtige Anwendungsfälle erfordern Leistung, Flexibilität und Skalierbarkeit



AssistentInnen, Chatbots und Inhaltserstellung



Accelerator as a Service



Multimodale Retrieval Augmented Generation (RAG)



## Vereinfacht

Optimieren Sie GenAI-Bereitstellungen mit bewährten, validierten Lösungen, die durch mehr als 340.000 Entwicklungsstunden unterstützt werden.

### Leistungsoptimierung

Leistungsstarker Accelerator, offene Architektur und KI-optimierte Fabrics

### KI überall

Überall und für alle Daten mit Multi-Cloud-Storage-Flexibilität

### Gebrauchsfertige Multi-Node-Lösung

Bewährte Full-Stack-KI-Grundlagen für schnellere Ergebnisse



## Maßgeschneidert

AMD ROCm™-Open-Source-Software und offene Ökosysteme fördern die Entwicklung und den Betrieb von KI.

### Schnellere Innovationen

Verwenden Sie Open-Source-Software und -Ökosysteme, um einzigartige Anwendungen zu entwickeln.

### Beschleunigte Entwicklung

Nutzen Sie branchenübliche Frameworks mit flexiblen Technologiestacks.

### Aktiviere Daten

Führen Sie gleichzeitig mehrere KI-Anwendungsfälle effizient aus.



## Vertrauenswürdig

82 % der ITDMs bevorzugen ein On-Premise- oder Hybridmodell.<sup>3</sup> Ihre Daten bestimmen Ihre Ergebnisse. Schützen Sie sie.

### Schneller Start

On-Premise-Grundlagen mit Root of Trust-Sicherheit und vollständiger Kontrolle

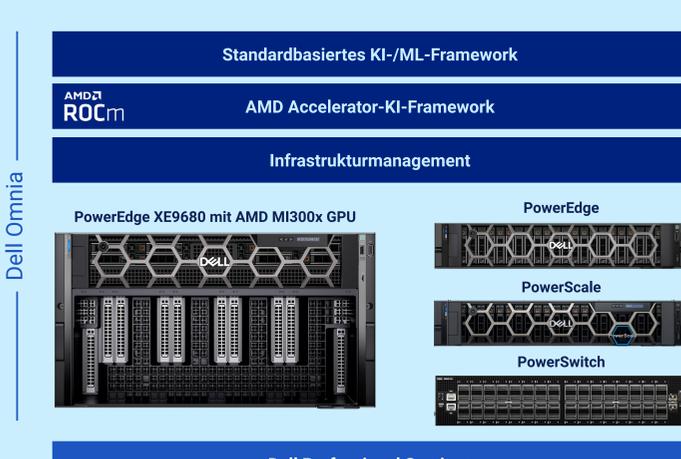
### Optimierte Konnektivität

Sichere, funktionsreiche Fabrics mit Skalierbarkeit und optimiertem Datenverkehr

### Automatische Bereitstellung

Open-Source-Grundlage für die Bereitstellung und Verwaltung von leistungsfähigen Clustern

## Dell GenAI-Lösungen mit AMD



### Inferenz

Führen Sie ein 70-B-Parametermodell auf einem einzelnen AMD Instinct™ MI300X Accelerator aus.<sup>4</sup>

### Anpassung

Stellen Sie acht gleichzeitige 70-B-Modelle auf einem einzigen Dell PowerEdge XE9680 bereit und optimieren Sie sie.<sup>4</sup>

### Erweiterung

Beziehen Sie Ihre Daten in den generativen Prozess ein.

Verschaffen Sie sich einen Wettbewerbsvorteil mit einer bewährten, offenen Lösung, die sichere On-Premise-KI-Anwendungen in großem Umfang bereitstellt.

Mehr erfahren

<sup>1</sup> Enterprise Strategy Group, Maximierung des KI-ROI: Die On-Premise-Inferenzierung mit Dell Technologies kann 75 % kosteneffizienter sein als Public Cloud, April 2024.

<sup>2</sup> Schätzung basierend auf einer Analyse von Dell vom Mai 2024, in der die Zeit für die Einrichtung eines Kubernetes-Clusters mit 2 Nodes für ein allgemeines LLM mithilfe automatisierter Skripte mit der manuellen

Bereitstellung eines gemeinsamen Designs verglichen wurde. Die Bereitstellungszeit umfasst nur die Basisinstallation. Die tatsächliche Setup-Zeit variiert je nach Lösungskonfiguration.

<sup>3</sup> Dell Technologies, Generative AI Pulse Survey, August und September 2023.

<sup>4</sup> Dell Technologies Blog, Silicon Diversity: Bereitstellung von GenAI auf dem PowerEdge XE9680 mit AMD Instinct MI300X Accelerators, Mai 2024.