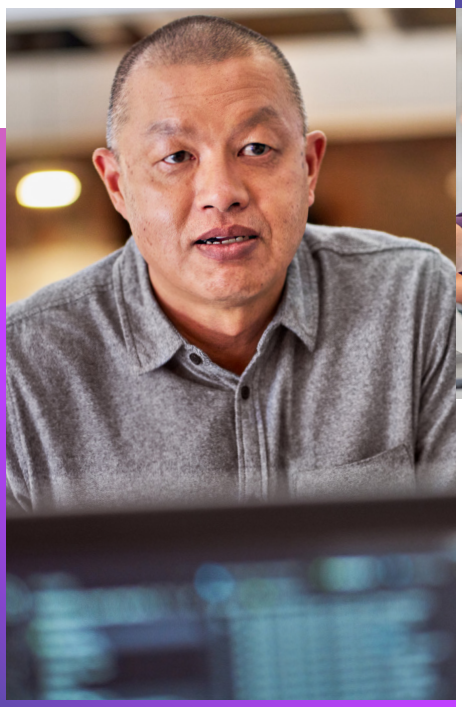
Hewlett Packard Enterprise

So kürzen Sie Ihren Weg zu KI ab

Die Implementierung von KI im großen Maßstab kann entmutigend sein. NVIDIA AI Computing by HPE kann die Komplexität reduzieren und die Einführung in Unternehmen beschleunigen



Künstliche Intelligenz (KI) ist heute ein wichtiges Ziel auf praktisch jeder Technologie-Roadmap von Unternehmen. Heute setzen 72 % der Unternehmen irgendeine Form von KI ein, wobei GenAI mit Abstand am weitesten verbreitet ist.[1] Der Stand der KI-Entwicklung in den Unternehmen ist jedoch sehr unterschiedlich.

Die meisten Unternehmen befinden sich im Hinblick auf GenAI noch in der Kennenlernphase und ermutigen ihre Mitarbeiter in der Regel, öffentlich zugängliche Chatbots zu nutzen. Eine kleinere Gruppe von Unternehmen führt Machbarkeitsstudien mit begrenzten Datensätzen durch, um die besten Anwendungsfälle für die Technologie zu ermitteln. Weniger als ein Drittel der meisten GenAI-Unternehmensprojekte haben den Übergang vom Experimentierstadium in die Produktion geschafft.[2]

„Viele dieser frühen Experimente finden in der Public Cloud statt“, sagt Peter Moser, Senior Distinguished Technologist bei HPE. Der Grund? Der Einstieg ist ganz einfach: Ein einfaches Modell entwickeln und dieses in einer Sandbox-Umgebung testen. Doch das kann sich schnell ändern.

„Sobald man mit der Skalierung beginnt und größere Datenmengen verarbeitet, wird KI schnell kompliziert und teuer“, sagt Moser. „Viele Kunden stellen fest, dass sie ihre Daten nicht in der Public Cloud haben möchten, wo sie nicht die Kontrolle darüber haben. Sie wollen sie vor Ort.“ Dieser Wandel ist bereits gut dokumentiert: Heute geben mehr als 60 % der Unternehmen an, dass sie planen, ihre KI-Anwendungen in einer Private Cloud vor Ort zu trainieren und bereitzustellen.[3]

[1] “The state of AI in early 2024: GenAI adoption spikes and starts to generate value,” McKinsey & Co., May 30, 2024

[2] “Now decides next: Moving from potential to performance,” Deloitte, August 2023

[3] “Essential Elements for Private Cloud Strategies,” IDC, August 2024

Ein wichtiger Grund für die Private Cloud ist die Datenhoheit und -sicherheit. Schließlich möchte kein Unternehmen riskieren, seine Kronjuwelen in der Public Cloud preiszugeben oder mit den immer häufiger auftretenden Data-Governance-Vorschriften in Konflikt zu geraten. Die Private Cloud bietet außerdem eine bessere Kontrolle über Anwendungsleistung, Latenz, Skalierbarkeit und Kosten, fügt Moser hinzu.

Doch der Übergang von begrenzten Experimenten in der Public Cloud zu produktionsreifen KI-Anwendungen vor Ort ist ein gewaltiger Sprung. Deshalb schaffen es auch neun von 10 Pilotprojekten nicht auf die andere Seite.[4] „Der überwiegenden Mehrheit der Unternehmen fehlt die notwendige technologische Infrastruktur und das Fachwissen, um KI im großen Maßstab einzusetzen", sagt Moser. Mit anderen Worten: Sie brauchen Hilfe.

Sie haben sich also für die Einführung einer Private KI-Cloud entschieden. Und nun?

„Nur die weltweit fortschrittlichsten Unternehmen, von denen viele schon seit Jahren mit maschinellem Lernen arbeiten, verfügen heute über die erforderlichen internen Tools und das Fachwissen, um GenAI-Anwendungen ohne Hilfe einzusetzen", sagt Moser. Den übrigen steht eine gewaltige Lernkurve bevor, und sie müssen beträchtliche Investitionen in Technologie und Zeit tätigen.

„High Performance Computing-Workloads wie KI nutzen parallele Dateisysteme", erklärt Moser. „Sie erfordern GPUs und komplexe Werkzeuge, über die die meisten IT-Unternehmen nicht verfügen. Das ist alles neu für sie."

Um diese KI-Modelle zu trainieren, müssen Unternehmen herausfinden, wie sie riesige Mengen unstrukturierter Daten konsolidieren können, die in mehreren Formaten und an mehreren Standorten gespeichert sind. Sie müssen Datenspezialisten einstellen, ihre IT-Teams im Umgang mit KI schulen, geeignete Data-Governance-Schutzmaßnahmen zur Risikominimierung implementieren und all diese neuen Systeme in ihre vorhandene Technologie-Infrastruktur integrieren.

In der Zwischenzeit tickt die Uhr. Der Marktdruck, GenAI rasch einzuführen und einzusetzen, wird weiter zunehmen.

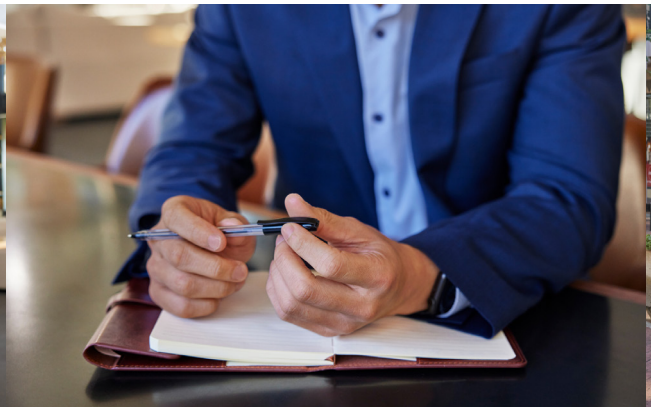
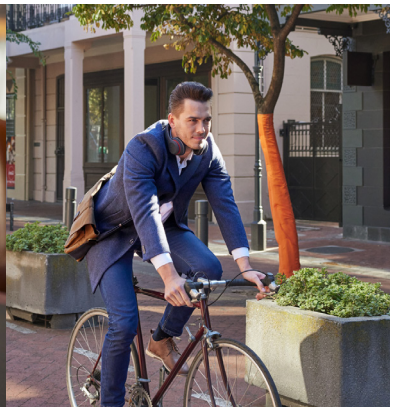
„Der Versuch, das alles selbst zu machen, kostet enorm viel Zeit und birgt große Risiken", sagt Moser. „Die meisten Unternehmen wissen einfach nicht genug über KI, um dies zeitnah zu tun."

HPE Private Cloud AI: Daten zugeben und umrühren

All diese Faktoren haben Hewlett Packard Enterprise dazu bewogen, gemeinsam mit NVIDIA eine einzigartige, sofort nutzbare Lösung für Unternehmen zu entwickeln: HPE Private Cloud AI.[5] Das ist mehr als bloße Infrastruktur. Die skalierbare, vorkonfigurierte Full-Stack-Private-Cloud-Option ist Teil des Portfolios von NVIDIA AI Computing by HPE. Diese vollständig integrierte KI-Lösung umfasst Rechenleistung und Speicher, Design und Planung, Integration und Optimierung, Schulungen und Training, Beratung und mehr.

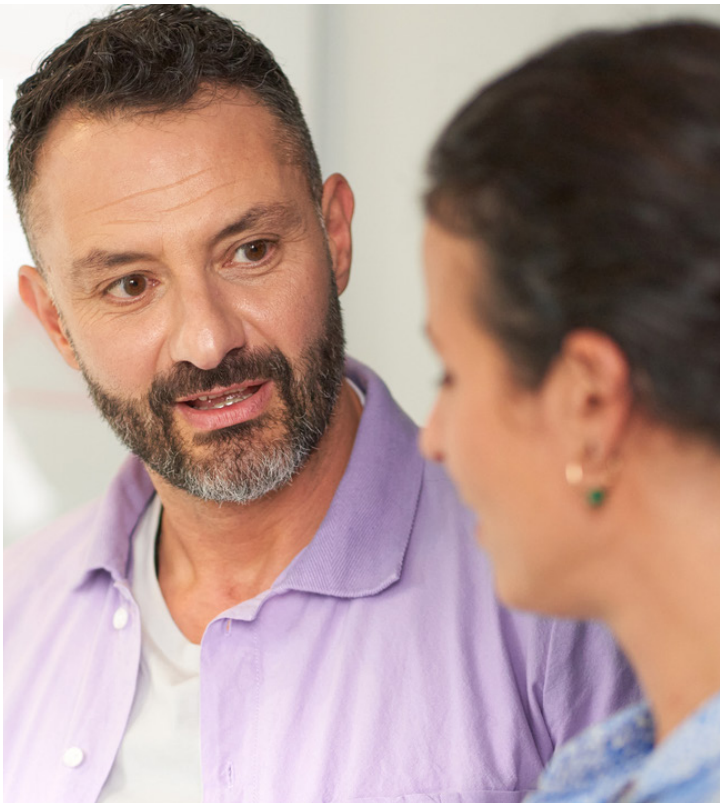
Der Prozess beginnt mit einem ausführlichen Gespräch über die Möglichkeiten, die KI für Ihr Unternehmen bietet, einer Bewertung, inwieweit Ihr Unternehmen für KI bereit ist, einer Diskussion über Best Practices und der Entwicklung von auf Ihre Bedürfnisse abgestimmten Anwendungsfällen. „HPE bietet Workshops für Unternehmen an, die ihre KI-Reise beginnen möchten, aber nicht wissen, wo sie anfangen sollen oder welche Fragen sie stellen sollen", sagt Moser.

„Einer der großen Fehler, den viele Unternehmen begehen, ist, dass sie nicht wissen, was KI alles kann",

[4] "Reasons Why Generative AI Pilots Fail To Move Into Production," Forbes, Jan 8, 2024

[5] "Hewlett Packard Enterprise introduces one-click-deploy AI applications in HPE Private Cloud AI," HPE, Sep 5, 2024



sagt er. „Sie beschränken sich von Anfang an auf die Kunst des Möglichen. Diese Workshops helfen ihnen, sich auf die Prioritäten zu konzentrieren, die für das Unternehmen am wichtigsten sind. Sie verändern das ganze Gespräch."

Der nächste Schritt ist die technische Konzeption und Planung. Nicht alle Unternehmen haben für ihre GenAI-Projekte die gleichen Anforderungen oder Ambitionen. Aus diesem Grund ist HPE Private Cloud AI in vier vorkonfigurierten Serverangeboten erhältlich. Diese hängen von den Rechen- und Speicheranforderungen des jeweiligen Anwendungsfalls ab, und davon, ob sie für Training, Feinabstimmung oder Inferenzen eingesetzt werden.

„Sie erhalten etwas, das bereits von den Unternehmen entworfen und zusammengestellt wurde, die all diese Werkzeuge herstellen und zertifizieren, dass alles zusammen funktioniert", sagt Moser. „Am nächsten Tag kommen wir und installieren es. Innerhalb von 48 Stunden ist es einsatzbereit."

Jeder von HPE konfigurierte HPE ProLiant Server ist für die Arbeit mit 70 großen Sprachmodellen (LLMs) zertifiziert, die zusammen mit gängigen geschäftlichen Anwendungsfällen zum Download zur Verfügung stehen. HPE Ezmeral Data Fabric ermöglicht es Unternehmen, ihre verteilten Datenspeicher zu einem zusammenhängenden Ganzen zu verknüpfen, wodurch die Aufnahme von unstrukturierten Trainingsdaten im Petabyte-Bereich für LLMs erleichtert wird. Unternehmen müssen lediglich das richtige LLM für ihre Geschäftsanforderungen auswählen, es mit ihren KI-fähigen Daten füttern und mit der Gewinnung von Einblicken beginnen.

Unterstützung des gesamten KI-Lebenszyklus

Sobald die Systeme in Produktion gehen, können Unternehmen sie entweder selbst verwalten oder sich von HPE und seinen Partnern beraten lassen, wie sie die sofort nutzbaren Lösungen von HPE und NVIDIA in ihre bestehenden Sicherheits- und Data-Governance-Frameworks integrieren können. Das Fachwissen steht während des gesamten KI-Lebenszyklus zur Verfügung – von der Implementierung bis zur Optimierung.

HPE kann Unternehmen auch bei der Bewältigung heikler Fragen rund um die ethische und transparente Nutzung von Daten, die Vermeidung von Datenschutz- und Urheberrechtsverletzungen und die Einhaltung einer zunehmenden Zahl von KI-Vorschriften unterstützen.

„Die richtige KI einzuführen ist unglaublich komplex", fügt Moser hinzu. „Die Leute wissen nicht, was sie nicht wissen, weil sie es noch nie getan haben. Und diejenigen, die versucht haben, alles selbst zu machen, stellen fest, dass es viel komplizierter ist als gedacht."

Weitere Informationen unter
hpe.com/AI

HPE.com besuchen

Jetzt chatten

Hewlett Packard Enterprise

a50011771DEE, Rev. 1