



Bessere Wirtschaftlichkeit und Leistung durch hybride KI

AUTOR

Steven Dickens

Chief Technology Advisor | The Futurum Group

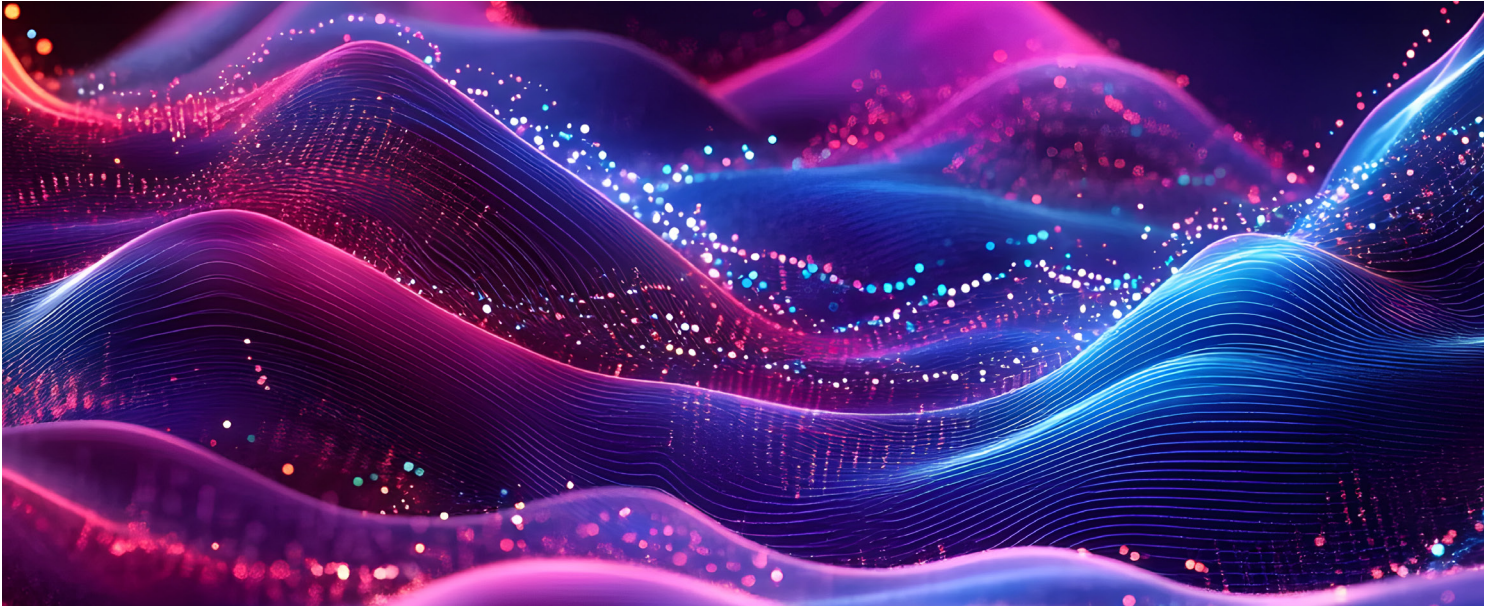
Ron Westfall

Research Director | The Futurum Group

AUGUST 2024

IN PARTNERSCHAFT MIT





Zusammenfassung

Entscheidungsträger, vor allem im IT-Bereich, haben heute die Aufgabe die KI-Reise ihres Unternehmens zu steuern. Es gibt eine große Anzahl an Optionen und Innovationen entwickeln sich rasant. Bei ihrer Suche nach dem besten Kurs für die KI-Reise entscheiden sich Entscheidungsträger zunehmend für ein hybrides KI-Modell. Warum hybride KI?

Hybride KI ist die Optimierung oder Erweiterung von KI-Modellen auf der Grundlage von maschinellem Lernen, Deep Learning und neuronalen Netzwerken mit menschlichem Fachwissen, um domänenspezifische KI-Modelle mit erhöhter Genauigkeit oder KI-anwendungsfallspezifische Modelle mit maximaler Genauigkeit oder Vorhersagewahrscheinlichkeit zu erstellen. Hybride KI steht für die Idee die Datenintensität zu wahren und KI-Lösungen und -Funktionen dorthin zu bringen, wo sie benötigt werden: auf Mobil- und Desktop-Geräten, in privaten Rechenzentren, im Edge Computing oder sogar in der Cloud. So können die Anwendungsfälle und die Datenverwaltungsstrategie bestimmen, wo und wie KI-Funktionen bereitgestellt werden sollen, wobei Faktoren wie Sicherheit, Kosten, Netzwerkinfrastruktur, Datengravitation und Zuverlässigkeit berücksichtigt werden.

Wir sehen, dass dieser Ansatz die Grenzen von Methoden mit einem einzelnen Modell oder einer einzelnen Infrastruktur überwindet, und den Kundenanforderungen nach flexiblem Datenmanagement und Anwendungssupport nachkommt. Hybride KI integriert Open-Source-KI und Closed-Source-Funktionen, wie die ChatGPT-Funktionen von OpenAI, und steigert so die Produktivität, Agilität, den Datenschutz und die KI-Demokratisierung.

Für wichtige Entscheidungsträger – einschließlich CIOs, CTOs, Chief Data Officers, Chief KI/Data Officer und Architekten – besteht der erste Schritt darin, Probleme zu definieren, Datenquellen zu identifizieren und Anwendungsfälle und KI-Anwendungen mit der richtigen Mischung aus Infrastruktur und Modellen in der richtigen Größe auszuwählen. Dieser Prozess umfasst in der Regel die Strategieentwicklung, den Aufbau eines KI-Stacks, einer KI-Fabrik oder eines Kompetenzzentrums (COE) und die Bereitstellung ausgewählter Lösungen.

Die Implementierung einer hybriden KI erfordert das Evaluieren, Testen und Skalieren neuer Technologien mit einem breiten Partner-Ökosystem. So können Sie eine Lösung entwickeln, die die richtige Kombination aus KI-Techniken, der richtigen Architektur und dem richtigen Modell für Ihr Unternehmen bietet.

Wir schauen uns an, warum das Portfolio und die Vision von Lenovo "Smarter AI for All", also smartere KI für alle, die einzigartigen Anforderungen bei der Auswahl und Implementierung von hybrider KI unternehmensweit erfüllen können.

Lenovos Vision "Smarter AI for All" priorisiert einen ergebnisorientierten Ansatz, indem es die richtige Mischung und Größe aus KI-PCs, Workstations, KI-fähigen Lösungen, Fachwissen sowie KI-optimiertem Speicher, Rechenleistung, Infrastruktur, Software und Partnerschaftsfunktionen mit einem validierten Partner-Ökosystem bietet. Dies beinhaltet die Möglichkeit, maßgeschneiderte, agile, skalierbare und energieeffiziente Industrie- und anwendungsfallbasierte KI-Lösungen für den persönlichen und industriellen Netzwerkrand (PCs, Workstations, Geschäftsstandorte), Rechenzentren (CoLos, am Standort) und High Performance Computing (HPC) zu liefern – und das alles mit Public-Cloud-Integration/Synergien.

Lenovo und NVIDIA bauen gemeinsam KI-Fabriken, die KI-optimierte und -fähige KI-Umgebungen für Unternehmen bieten, in denen KI-Anwendungen sicher, privat und skaliert entwickelt, bereitgestellt und verwaltet werden können.

Die hybride KI-Landschaft entwickelt sich schnell

Organisationen und ihre Mitarbeitenden möchten sich die transformative Kraft der Generativen KI (GenKI) zunutze machen und gehen über erste Experimente und KI/ML-Pilotprojekte hinaus, um ihre eigenen Daten zu nutzen. Eine breite Akzeptanz wird jedoch häufig durch rechtliche und regulatorische Herausforderungen, Wissenslücken, Datensilos, Sicherheitsbedenken, unzureichende KI-Infrastruktur und Komplexität des Ökosystems behindert. Budgetbeschränkungen erschweren die Dinge zusätzlich. Während 96 % der CIOs im kommenden Jahr mit mehr KI-Investitionen rechnen, erwarten nur 20 % einen Anstieg der IT-Budgets um mehr als 10 % (CIO-Studie 2024).

Hybride KI entwickelt sich zu einer wichtigen Komponente von KI-Geschäftsstrategien, die es Unternehmen ermöglicht, KI- und Datenlösungen zu entwickeln, die sich an ihren Zielen orientieren und KI näher an Datenquellen, Echtzeit-Interaktionen und Erlebnisgenerierungspunkte bringen. Obwohl kostenlose KI-Tools, öffentliche Basismodelle und vorhandene Cloud-Abonnements die Einführung erleichtern, sind viele Unternehmen in endlosen Proof-of-Concept-Zyklen bei steigenden Kosten gefangen. Dies liegt häufig am Fehlen einer soliden, hybriden KI-Strategie, die KI in privaten und öffentlichen Cloud-Umgebungen integriert, am Fehlen von benutzerdefinierten KI-Modellen, an begrenztem Zugriff auf hochwertige Unternehmensdaten und an unzureichender KI-Computing-Dimensionierung und -Planung.

Aus unserer Sicht wird die lokale Ausführung von generativer KI (GenAI) und KI am Edge sowie auf Geräten kritisch für Organisationen mit verteilten Daten und Edge-Standorten (z. B. Einzelhandel, Fertigung, Krankenhäuser, hybride Arbeitsplätze) – ebenso für diejenigen, die Echtzeitanalysen und Zugriff benötigen, sowie für Organisationen mit getrennten Computing-Anforderungen (z. B. beim Fernbetrieb in der Öl- und Gas-Branche oder in hochgradig geschützten Edge-Umgebungen).

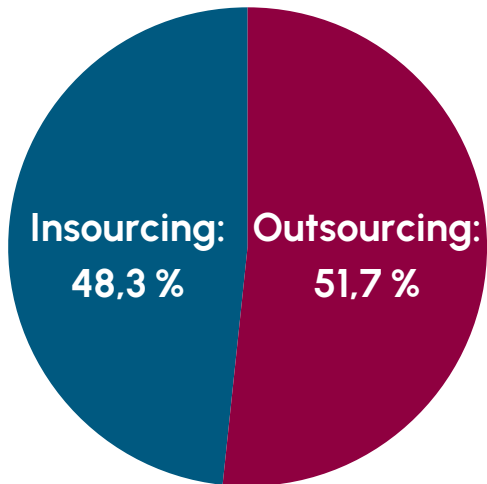
Edge-Standorte und -Geräte erzeugen große Datenmengen, von denen die meisten nicht analysiert werden. KI und GenAI bieten eine bedeutende Chance, diese ungenutzte Ressource zu erschließen. Mit KI-Anwendungen, die für Edge-Anwendungsfälle entwickelt wurden, immer effizienteren, kleineren Modellen und KI-fähigen Geräten wie KI-PCs wird Edge-Computing zur Realität. Dadurch müssen Daten oder KI-Modelle nicht mehr in Public Clouds oder zentralisierte Rechenzentren übertragen werden, um Echtzeitanalysen zu ermöglichen, sodass Sicherheit, Latenz und Bandbreitenbeschränkungen kein Hindernis mehr darstellen.

Lokale GenAI-Verarbeitung ermöglicht die Datenanalyse am Ort der Erstellung, reduziert die Latenz und verbessert die Effizienz – dies ist entscheidend für Echtzeitanwendungen wie Fertigungsprozesse und Robotik. Darüber hinaus erfordert der Datenschutz häufig eine lokale Verarbeitung, um sicherzustellen, dass sensible Daten auf dem Gerät bleiben.

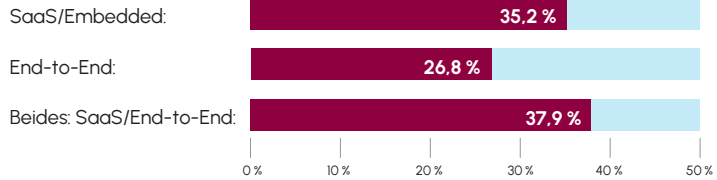
Cloud-basierte, große Sprachmodelle (LLM) stellen Herausforderungen dar, die Unternehmen zunehmend vermeiden möchten. Dazu gehören das Risiko von Halluzinationen – logisch inkonsistente, aber plausibel klingende Ergebnisse – und die Verbreitung von Ungenauigkeiten aus Trainingsdaten. Darüber hinaus haben Unternehmen nur begrenzte Kontrolle und Transparenz über die Infrastruktur und Implementierung von Cloud Managed Services, was Sicherheitsprobleme aufwirft.

Ein bemerkenswertes Beispiel dafür gab es im März 2023, als ChatGPT von OpenAI aufgrund einer Schwachstelle in einer Open-Source-Bibliothek ausfiel, wodurch Zahlungsinformationen von Kunden potenziell offengelegt wurden. Dieser Vorfall unterstreicht, wie wichtig rigorose Tests zum Schutz von LLMs und robuste Mechanismen zur Abwehr von Cyberbedrohungen sind, wenn cloudbasierte KI-Services eingesetzt werden.

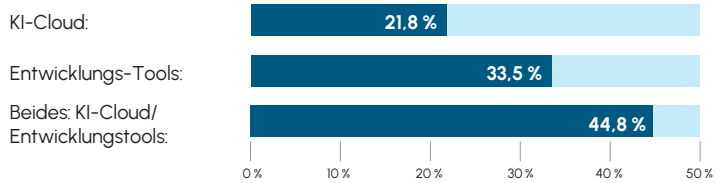
Darüber hinaus können wir feststellen, dass Unternehmen einen ausgewogenen Ansatz bei der Bereitstellung von KI-Services verfolgen:



Outsourcing



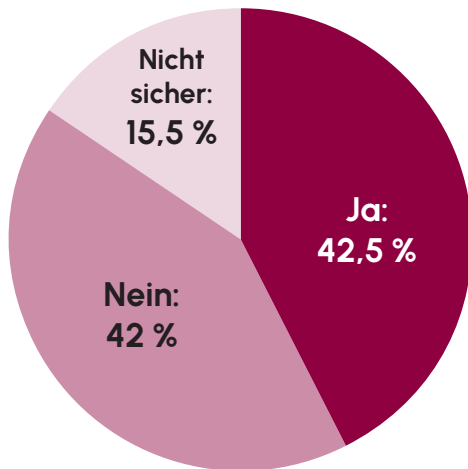
Insourcing



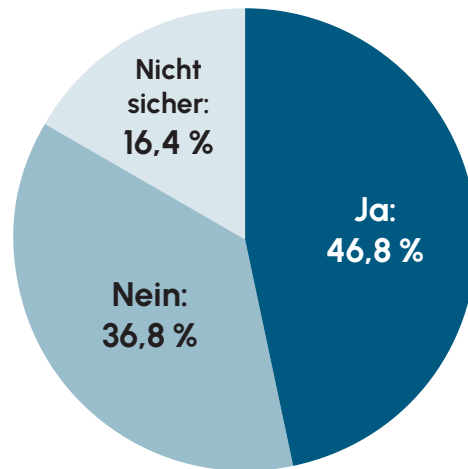
Quelle: The Futurum Group

Das schnell wachsende Interesse an hybrider KI schafft neue Wettbewerbsdynamik, da Unternehmen zunehmend Pläne zur Änderung oder Hinzufügung neuer Anbieter prüfen (2023-24):

Outsourcing



Insourcing



Quelle: The Futurum Group

Der Erfolg hybrider KI hängt von der entscheidenden Rolle ab, die Ökosysteme bei der Kuratierung von KI-Lösungen spielen, die auf die Anforderungen des Unternehmens zugeschnitten sind. Ökosysteme vereinen eine Vielzahl von Interessengruppen der Wertschöpfungskette, um die nahtlose Integration von KI-Systemen in hybriden Umgebungen zu ermöglichen, was zur Schaffung maßgeschneiderter KI-Lösungen führt. KI-Lösungen erfordern häufig eine Kombination aus Technologien und Fachwissen und sind auf Ökosysteme angewiesen, um Zugriff auf die breite Palette an Software, Anwendungen, Tools und Funktionen zu erhalten, damit Lösungen entwickelt werden können, die die geschäftlichen KI-Herausforderungen überwinden können.

Darüber hinaus ist die Flexibilität von hybrider KI unerlässlich, um die vielfältigen Bedürfnisse der Personen zu erfüllen, die KI im Unternehmen nutzen. KI-Entwickler und Datenwissenschaftler verwenden vertraute KI-Tools und Frameworks, unabhängig von ihrer Verteilung. Am Edge erwarten und benötigen Branchen wie Einzelhandel und Fertigung Echtzeit-Reaktionen von KI-Systemen sowie die Fähigkeit, Daten schnell zu verarbeiten.

Die effektive Nutzung hybrider KI erfordert jedoch eine sorgfältige Planung und Ausführung, wobei Technologie, Mitarbeiter, Prozesse und Ansätze für verantwortungsvolle KI berücksichtigt werden müssen. Unternehmen kämpfen mit isolierten Initiativen, falsch ausgerichteten KI-Lösungsoptionen und Budgets, während IT-Teams versuchen, die Infrastruktur zu modernisieren und verschiedene KI-Computing-Anforderungen, Compliance und Sicherheitsstandards zu erfüllen und einen hybriden Betrieb zu leiten. Datenmanagement und Public-Cloud-Nutzung werden unkontrollierbar, sodass Sicherheit, Governance und steigende Cloud-Kosten zu drängenden Problemen werden. Diesen Herausforderungen müssen die Fähigkeiten und das Change Management gegenüberstehen, die erforderlich sind, um mit KI sinnvolle Geschäftsergebnisse zu erzielen.

Data Scientists und ML-Ingenieure sollten beispielsweise nicht die meiste Zeit mit Datenqualität, Vorschriften und Compliance verbringen. Unternehmen sollten IT&O-Teams in KI-bezogene IT-Entscheidungen einbeziehen. Ein Bottom-up-Ansatz von IT-Teams, um KI im Unternehmen einzuführen, kann zu einer Über- oder Unterversorgung führen, ohne dass die Geschäftsanforderungen und Anwendungsfälle genau verstanden werden.



Lenovo und NVIDIAs Co-Engineered Portfolio

Das Wachstum der KI-Branche wird von mehreren Faktoren getragen, wobei die steigende Nachfrage nach Automatisierung und Optimierung branchenübergreifend von größter Bedeutung ist. Wie in unserem Artikel beschrieben, nutzen Unternehmen die Fortschritte bei der Rechenleistung und der Cloud-Infrastruktur, um effizientere KI-Anwendungen zu implementieren. Dieser Wandel hat zu einer wachsenden Anzahl von Anfragen nach kundenspezifischen KI-Lösungen geführt, einschließlich Beratungs-, Design-, Bereitstellungs- und Wartungsservices, mit denen dieser Prozess vereinfacht werden kann.

Das Lenovo Smarter AI for All Portfolio erfüllt diese sich wandelnden Anforderungen mit KI-fähigen Geräten, intelligenter Infrastruktur und robusten Cloud-Lösungen, die eine nahtlose KI-Integration über verschiedene Plattformen hinweg ermöglichen. Die strategische Partnerschaft zwischen Lenovo und NVIDIA beschleunigt Innovationen im KI-Sektor. Das co-engineered Portfolio vereinfacht die KI-Akzeptanz, liefert schnellere Ergebnisse und verbessert die Leistung und Effizienz in hybriden KI-Umgebungen, was das Wachstum der Branche weiter vorantreibt.

Wichtige Informationen zu diesem Bericht

MITWIRKENDE

Steven Dickens

Chief Technology Advisor | The Futurum Group

Ron Westfall

Research Director | The Futurum Group

HERAUSGEBER

Daniel Newman

CEO | The Futurum Group

ANFRAGEN

Kontaktieren Sie uns gerne, wenn Sie über diesen Bericht sprechen möchten. The Futurum Group wird Ihnen umgehend antworten.

ZITIERUNGEN

Dieser Artikel kann von akkreditierten Medien und Analysten zitiert werden, muss jedoch im Kontext zitiert werden, den Namen des Autors, den Titel des Autors und „The Futurum Group“ enthalten. Alle anderen Personen müssen sich vorher von The Futurum Group eine schriftliche Genehmigung einholen.

LIZENZIERUNG

Dieses Dokument, einschließlich aller unterstützenden Materialien, gehört The Futurum Group. Diese Publikation darf ohne vorherige schriftliche Genehmigung von The Futurum Group in keiner Form reproduziert, verbreitet oder geteilt werden.

OFFENLEGUNG

The Futurum Group bietet vielen Hightech-Unternehmen ihre Forschungs-, Analysen und Beratungsdienste an, einschließlich der in diesem Artikel genannten Unternehmen. Mitarbeitende des Unternehmens halten keine Aktienpositionen bei den in diesem Dokument genannten Unternehmen.



ÜBER LENOVO UND NVIDIA

Lenovo bringt die neue Ära KI-gestützter Innovationen für alle. Unser Full-Stack-Portfolio bietet leistungsstarke, flexible und verantwortungsvolle KI-Lösungen, um Branchen zu transformieren und Einzelpersonen zu unterstützen. Wir schaffen eine Zukunft mit intelligenterer KI für alle. Bei Lenovo sind wir davon überzeugt, dass KI für die Öffentlichkeit und KI für Unternehmen in Zukunft koexistieren wird. Lenovo bringt KI zu Ihnen und Ihren Daten.

In Partnerschaft mit NVIDIA werden hybride KI-Lösungen durch technische Zusammenarbeit speziell entwickelt, um KI effizient für Kundendaten zu nutzen, wo und wann Benutzer sie am meisten benötigen. Dies wird die Vision von Lenovo weiterentwickeln, KI für alle zu ermöglichen und Unterstützung für die Einführung bahnbrechender Technologien und Architektur für die nächste Generation generativer KI zu bieten. Lenovo Hybridlösungen, die bereits für die Ausführung von NVIDIA KI Enterprise-Software für sichere, unterstützte und stabile KI-Produktion optimiert sind, bieten Entwicklern Zugang zu NVIDIA Microservices, einschließlich NVIDIA NIMs und NeMo Retriever.



ÜBER THE FUTURUM GROUP

[TheFuturum Group](#) ist ein unabhängiges Forschungs-, Analyse- und Beratungsunternehmen, das sich auf digitale Innovationen und marktverändernde Technologien und Trends spezialisiert hat. Unsere Analysten, Forscher und Berater helfen Führungskräften aus der ganzen Welt dabei, tektonische Veränderungen in ihren Branchen zu antizipieren und disruptive Innovationen zu nutzen, um sich einen Wettbewerbsvorteil in ihren Märkten zu verschaffen oder zu erhalten.



KONTAKTINFORMATIONEN

The Futurum Group LLC | futurumgroup.com | (833) 722-5337 |