

2024: Das Jahr der KI?

Ludwig von Reiche | NVIDIA GmbH | 1. Februar 2024

2024: Das Jahr der KI?

10 Leaders

The Economist January 27th 2024

Technology

AI for all

Artificial intelligence holds a tantalising promise of prosperity for the emerging world

NEW TECHNOLOGY brings with it both the sweet hope of greater prosperity and the cruel fear of missing out. Satya Nadella, the boss of Microsoft, says he is haunted by the fact that the Industrial Revolution left behind India, his country of birth. (Indian manufacturers hardly enjoyed a level playing-field—Britain was then both their rival and their ruler.) Many technologies, such as online-education courses, have generated more hype than economic growth in the emerging world. Some people worry that generative artificial intelligence (AI), too, will disappoint the global south. The big winners so far seem to be a bunch of Western early adopters, as well as startups in San Francisco and America's "magnificent seven" tech firms, which include Microsoft and have together added an astonishing \$4.6trn to their market value since ChatGPT's launch in November 2022.

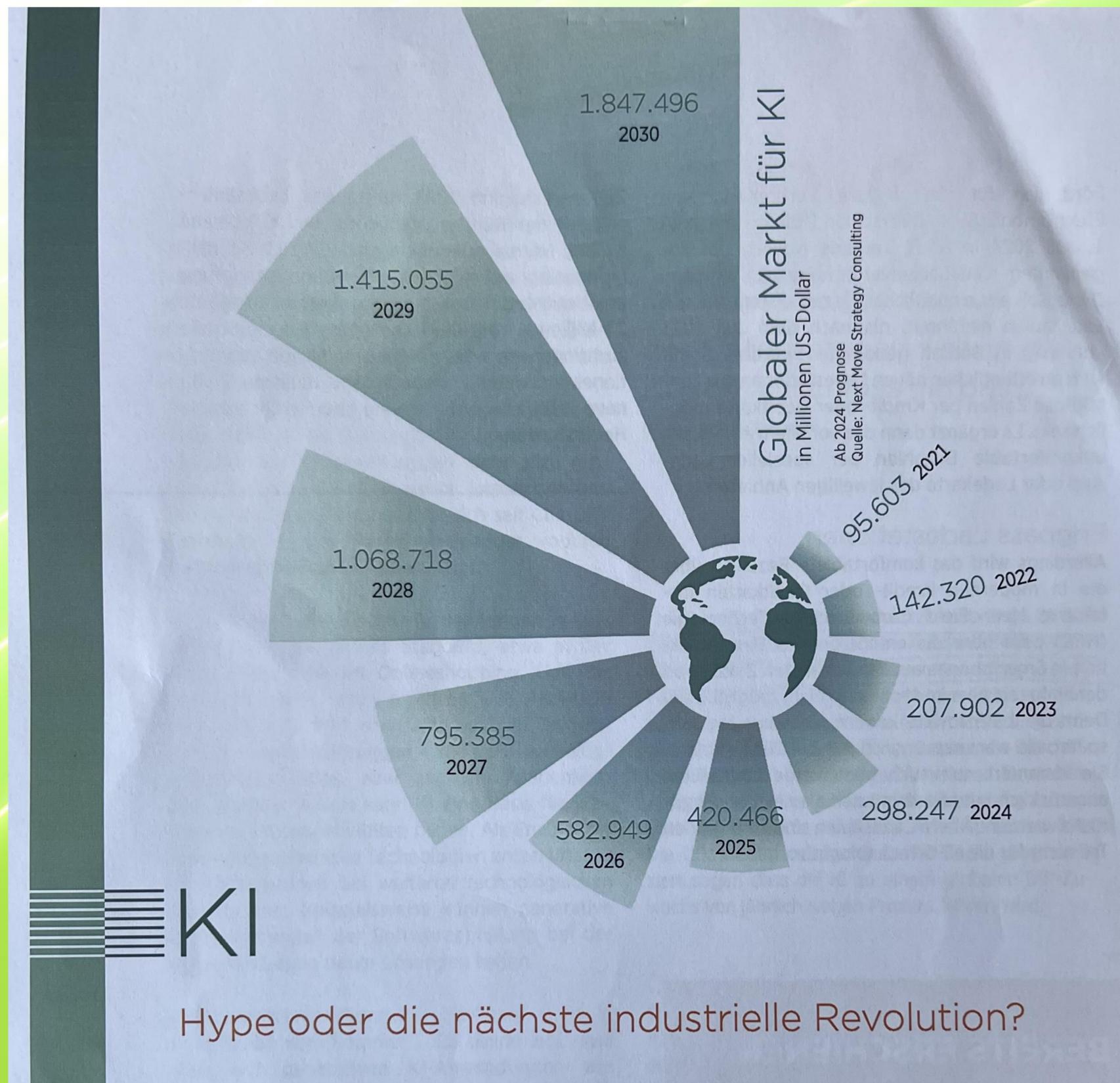
Yet AI stands to transform lives in the emerging world, too. As

across the emerging world already have, and many more soon will: the phone in their pockets. In time, chatbots will become much cheaper to provide and acquire.

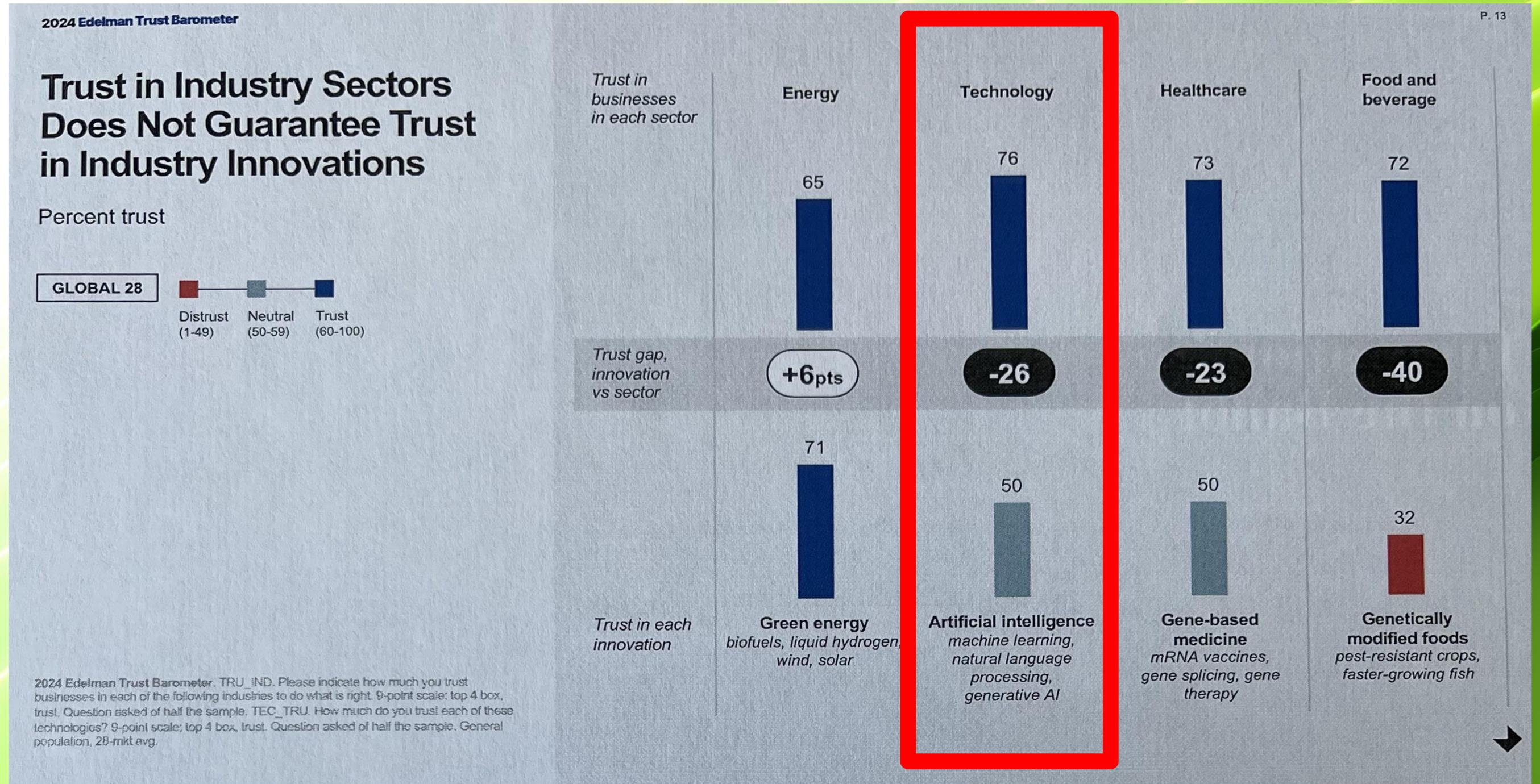
Moreover, the technology can be tailored to local needs. So far there is little sign that AI is ruled by the winner-takes-all effects that benefited America's social-media and internet-search firms. That means a variety of approaches could prosper. Some developers in India are already taking Western models and fine-tuning them with local data to provide a whizzy language-translation service, avoiding the heavy capital costs of model-building (see Science & technology section).

Another idea that is also taking off in the West is to build smaller, cheaper models of your own. A narrower set of capabilities, rather than the ability to get every bit of information under the sun, can suit specific needs just fine. A medical AI is unlikely to need to generate amusing limericks in the style of William

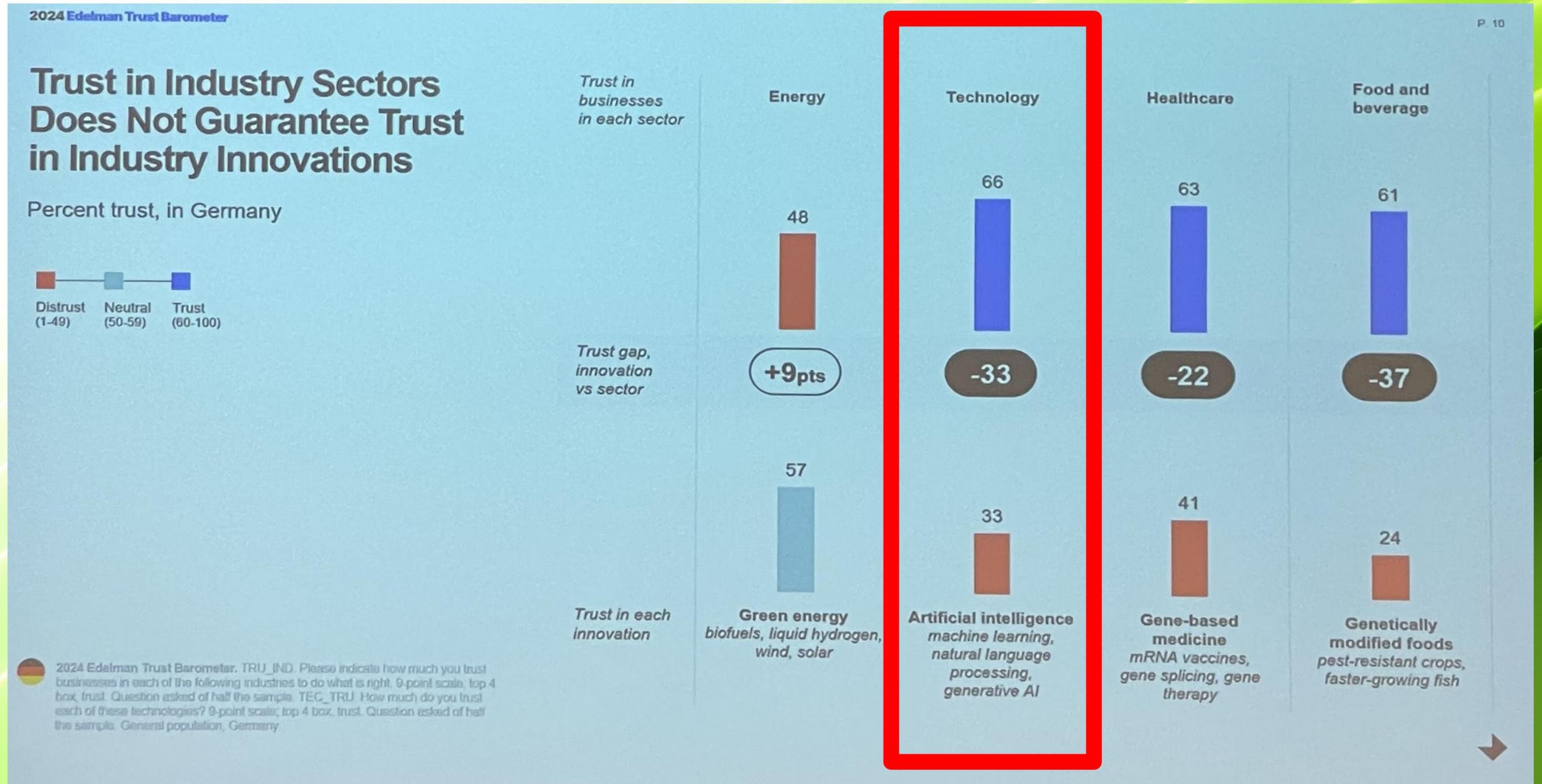
2024: Das Jahr der KI?



2024: Das Jahr der KI?



2024: Das Jahr der KI?



2024: Das Jahr der KI?

To Earn My Acceptance Show Me the Innovation Is Vetted by Scientists and Ethicists

Average acceptance rating for each innovation

GLOBAL 28

Among those who have

○ ●
low high

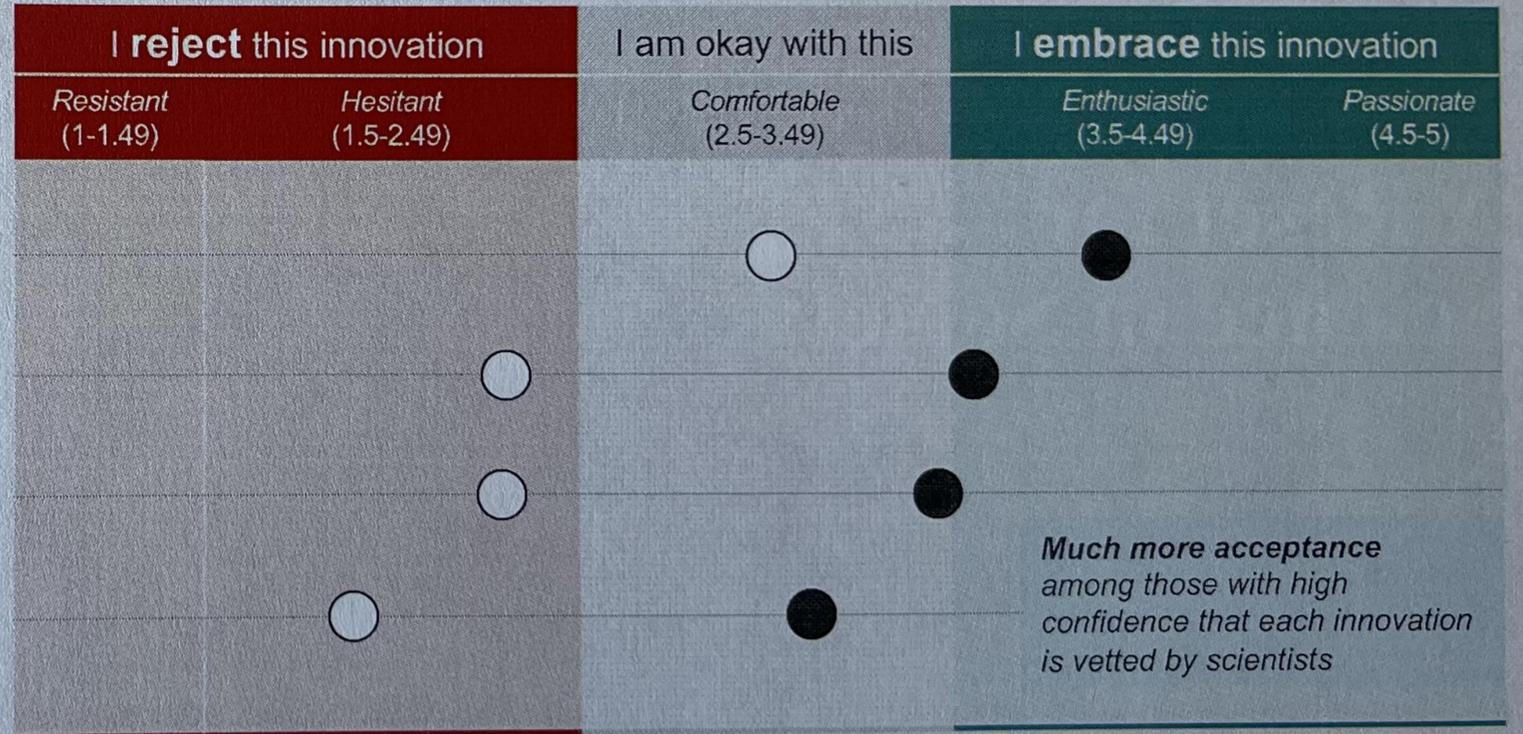
confidence that each innovation has been **evaluated by scientists and ethicists**

Green energy

AI

Gene-based medicine

GMO foods



2024 Edelman Trust Barometer, TEC_TST. How confident are you that these technologies have been adequately tested, evaluated, and assessed by scientists, ethicists, and other experts? 5-point scale; top 2 box, high confidence; bottom 2 box, low confidence. Question asked of half the sample. CHG_TEC_COM. How would you characterize your feelings about each of the following? 5-point scale; data plotted is an average acceptance rating for each innovation. General population, 28-mkt avg., by level of confidence. For a full explanation of how the Average Acceptance Rating was calculated, please see the Technical Appendix.



2024: **Ein besonders wichtiges** Jahr der KI.

- **Vertrauen schaffen***
- **Verantwortung wahrnehmen***
- **Rahmenbedingungen umsetzen**

* Was NVIDIA dazu beiträgt (und auch andere nutzen können), findet man [hier](#).

2024: Ein besonders wichtiges Jahr für KI.

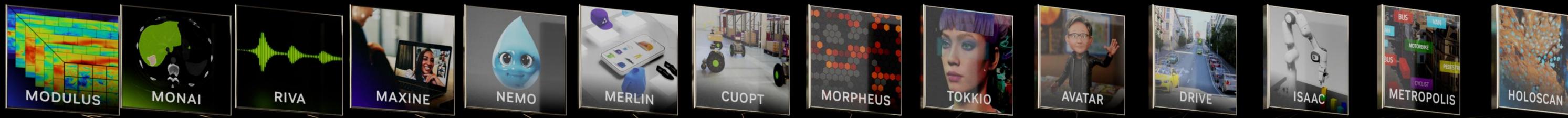
- **Get going!**

Von
Grafikkarten ...

... zu
NVIDIA DGX Systemen ...

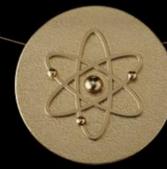
... und
"KI-Fabriken".





AI APPLICATION
FRAMEWORK

PLATFORMS



NVIDIA HPC



NVIDIA AI



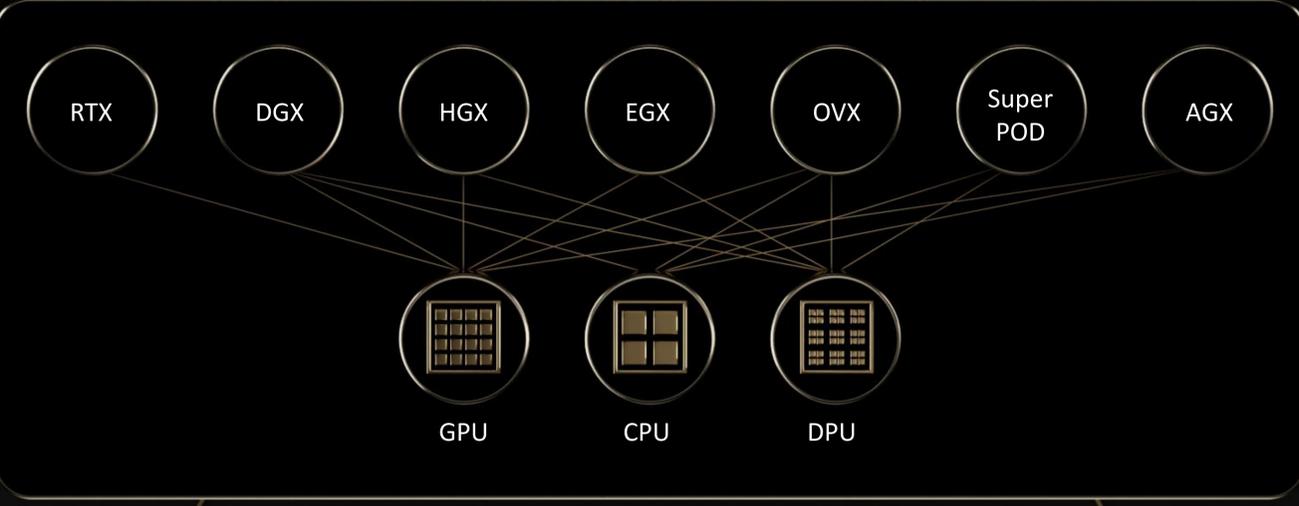
NVIDIA Omniverse

ACCELERATION
LIBRARIES



CLOUD-TO-EDGE
DATACENTER-TO-ROBOTIC SYSTEMS

3-CHIPS



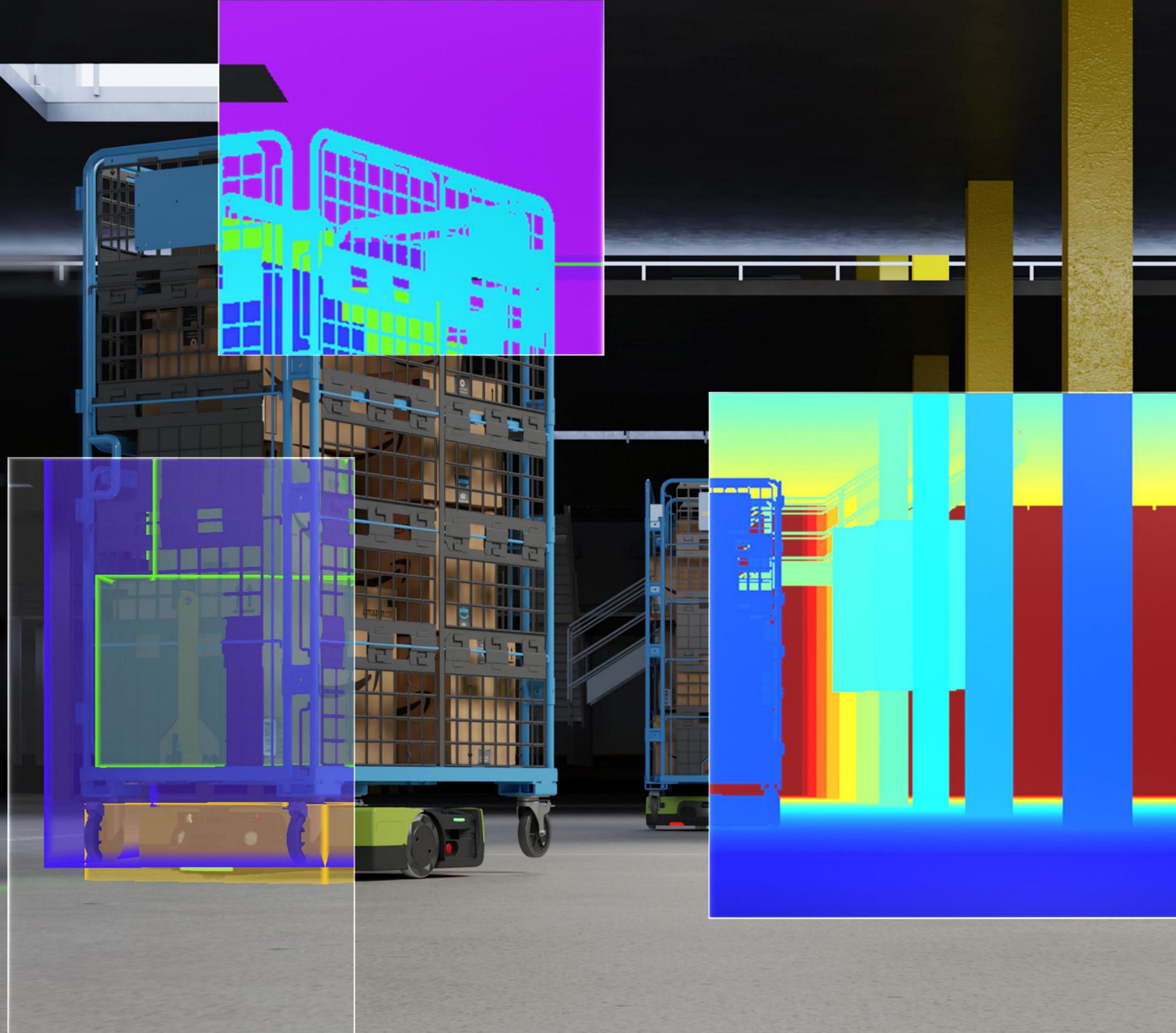
BMW verbindet Realität und virtuelle Welten beim Bau der Fabrik der Zukunft

Die BMW Group nutzt Omniverse, um einen voll funktionsfähigen digitalen Zwilling der Fabrik zu erstellen, bevor diese in der realen Welt gebaut wird. Durch den Einsatz von NVIDIA AI und Omniverse konnten 20 % der Kosten bei Planung und Umsetzung eingespart werden.

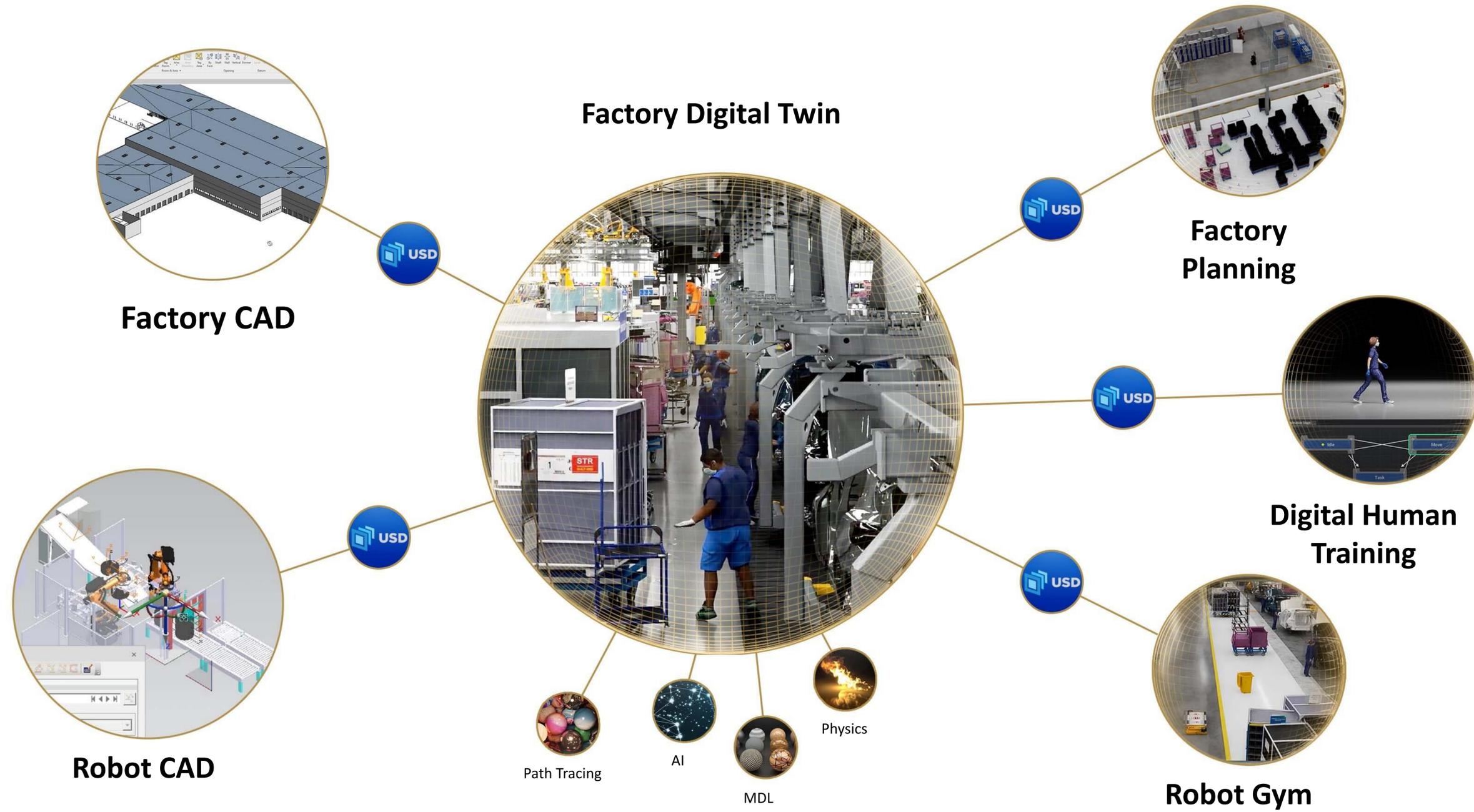


Amazon baut digitale Zwillinge von Lagerhäusern

Amazon verfügt über mehr als 200 Roboteranlagen, die täglich Millionen von Paketen bearbeiten. Mit NVIDIA Omniverse™ Enterprise und Isaac Sim™ erstellt Amazon Robotics KI-fähige digitale Zwillinge seiner Lagerhäuser, um Lagerdesign und -fluss zu optimieren und intelligentere Roboterlösungen zu trainieren.



Ein Beispiel ... mit NVIDIA Omniverse für Digital Twins



Wie kann KI (auch) dem Mittelstand nutzen?

Frage an ChatGPT 3.5 (12. November 2023)*

Vorausschauende Wartung	KI kann eingesetzt werden, um den Zustand von Maschinen und Ausrüstungen zu überwachen und frühzeitig auf potenzielle Ausfälle oder Wartungsbedarf hinzuweisen. Dies hilft, ungeplante Stillstandzeiten zu minimieren und die Effizienz der Anlagen zu maximieren.
Qualitätskontrolle	Durch den Einsatz von Bildverarbeitungsalgorithmen kann KI dazu verwendet werden, Qualitätsprüfungen während des Produktionsprozesses durchzuführen. Dies trägt dazu bei, fehlerhafte Produkte zu erkennen, bevor sie die Endfertigung erreichen, und verbessert die Gesamtqualität.
Automatisierung von Prozessen	KI-gesteuerte Robotic Process Automation (RPA) kann repetitive und zeitaufwändige Aufgaben automatisieren. Dies ermöglicht es den Mitarbeitern, sich auf komplexere Aufgaben zu konzentrieren, was die Effizienz steigert.
Optimierte Produktionsplanung	KI kann historische Produktionsdaten analysieren, um zukünftige Produktionspläne zu optimieren. Dies verbessert die Nutzung von Ressourcen, reduziert Überproduktion und minimiert Abfall.
Energieeffizienz	Durch die Überwachung und Optimierung des Energieverbrauchs kann KI dazu beitragen, die Energieeffizienz in industriellen Prozessen zu steigern. Dies führt nicht nur zu Kosteneinsparungen, sondern auch zu einer nachhaltigeren Betriebsweise.
Lieferkettenoptimierung	KI kann bei der Vorhersage von Nachfrageänderungen, der Optimierung von Lagerbeständen und der Verbesserung der Lieferkettenlogistik unterstützen. Dies trägt dazu bei, die Effizienz der gesamten Lieferkette zu steigern.
Kundenservice und Personalisierung	KI kann im Kundenservice eingesetzt werden, um Anfragen zu bearbeiten, Probleme zu lösen und sogar personalisierte Interaktionen mit Kunden zu ermöglichen. Chatbots und virtuelle Assistenten können dabei helfen, den Kundenservice effizienter zu gestalten.
Innovationsförderung	KI kann dazu verwendet werden, innovative Lösungen und Produkte zu entwickeln. Durch die Analyse großer Datenmengen kann KI Trends erkennen und Unternehmen dabei unterstützen, wettbewerbsfähige Produkte auf den Markt zu bringen.
Arbeitssicherheit	KI kann zur Überwachung von Arbeitsumgebungen eingesetzt werden, um potenzielle Sicherheitsrisiken zu identifizieren. Dies trägt dazu bei, Unfälle zu verhindern und die Sicherheit am Arbeitsplatz zu verbessern.
Datensicherheit	KI kann dazu verwendet werden, Unternehmensdaten vor Cyberbedrohungen zu schützen. Durch den Einsatz von maschinellem Lernen kann die KI anomale Aktivitäten erkennen und frühzeitig auf Sicherheitsverletzungen reagieren.

*ChatGPT kann Fehler machen. Bitte überprüfen Sie wichtige Informationen.

Wie kann KI (auch) dem Mittelstand nutzen?

Frage an ChatGPT 3.5 (12. November 2023)*

Vorausschauende Wartung	KI kann eingesetzt werden, um den Zustand von Maschinen und Ausrüstungen zu überwachen und frühzeitig auf potenzielle Ausfälle oder Wartungsbedarf hinzuweisen. Dies hilft, ungeplante Stillstandzeiten zu minimieren und die Effizienz der Anlagen zu maximieren.
Qualitätskontrolle	Durch den Einsatz von Bildverarbeitungsalgorithmen kann KI dazu verwendet werden, Qualitätsprüfungen während des Produktionsprozesses durchzuführen. Dies trägt dazu bei, fehlerhafte Produkte zu erkennen, bevor sie die Endfertigung erreichen, und verbessert die Gesamtqualität.
Automatisierung von Prozessen	KI-gesteuerte Robotic Process Automation (RPA) kann repetitive und zeitaufwändige Aufgaben automatisieren. Dies ermöglicht es den Mitarbeitern, sich auf komplexere Aufgaben zu konzentrieren, was die Effizienz steigert.
Optimierte Produktionsplanung	KI kann historische Produktionsdaten analysieren, um zukünftige Produktionspläne zu optimieren. Dies verbessert die Nutzung von Ressourcen, reduziert Überproduktion und minimiert Abfall.
Energieeffizienz	Durch die Überwachung und Optimierung des Energieverbrauchs kann KI dazu beitragen, die Energieeffizienz in industriellen Prozessen zu steigern. Dies führt nicht nur zu Kosteneinsparungen, sondern auch zu einer nachhaltigeren Betriebsweise.
Lieferkettenoptimierung	KI kann bei der Vorhersage von Nachfrageänderungen, der Optimierung von Lagerbeständen und der Verbesserung der Lieferkettenlogistik unterstützen. Dies trägt dazu bei, die Effizienz der gesamten Lieferkette zu steigern.
Kundenservice und Personalisierung	KI kann im Kundenservice eingesetzt werden, um Anfragen zu bearbeiten, Probleme zu lösen und sogar personalisierte Interaktionen mit Kunden zu ermöglichen. Chatbots und virtuelle Assistenten können dabei helfen, den Kundenservice effizienter zu gestalten.
Innovationsförderung	KI kann dazu verwendet werden, innovative Lösungen und Produkte zu entwickeln. Durch die Analyse großer Datenmengen kann KI Trends erkennen und Unternehmen dabei unterstützen, wettbewerbsfähige Produkte auf den Markt zu bringen.
Arbeitssicherheit	KI kann zur Überwachung von Arbeitsumgebungen eingesetzt werden, um potenzielle Sicherheitsrisiken zu identifizieren. Dies trägt dazu bei, Unfälle zu verhindern und die Sicherheit am Arbeitsplatz zu verbessern.
Datensicherheit	KI kann dazu verwendet werden, Unternehmensdaten vor Cyberbedrohungen zu schützen. Durch den Einsatz von maschinellem Lernen kann die KI anomale Aktivitäten erkennen und frühzeitig auf Sicherheitsverletzungen reagieren.

*ChatGPT kann Fehler machen. Bitte überprüfen Sie wichtige Informationen.

AUTOMOBILHERSTELLER NUTZEN METROPOLIS ZUR VERBESSERUNG DER PRODUKTIONSLINIEN

Produktionslinien auf der ganzen Welt werden zunehmend mit dem Ziel organisiert, Montagefehler zu minimieren und die Kontrollen durch Automatisierung von Montageaufgaben zu verschärfen.

XXII hat eine Echtzeit-Videoanalyselösung entwickelt, die bei der Überwachung von Montagelinien und der Verbesserung von Produktionsprozessen hilft. Die Erkenntnisse aus der Analyse helfen dabei, die Teileauswahl und -montage straffer und präziser zu gestalten und Verstöße rechtzeitig zu erkennen.

Durch den Einsatz der XXII-Anwendung senken große Automobilhersteller ihre Montagekosten und erhöhen die Zuverlässigkeit ihrer Produktionslinien und -prozesse.



OPTIMIERUNG DES LAGERBETRIEBS MIT KI

Quantiphi nutzt NVIDIA-Produkte - von Edge-GPUs bis hin zu optimierten Deep-Learning-Tools - zur Entwicklung einer Echtzeit-Videoanalyseanwendung zur Optimierung von Lagerabläufen.

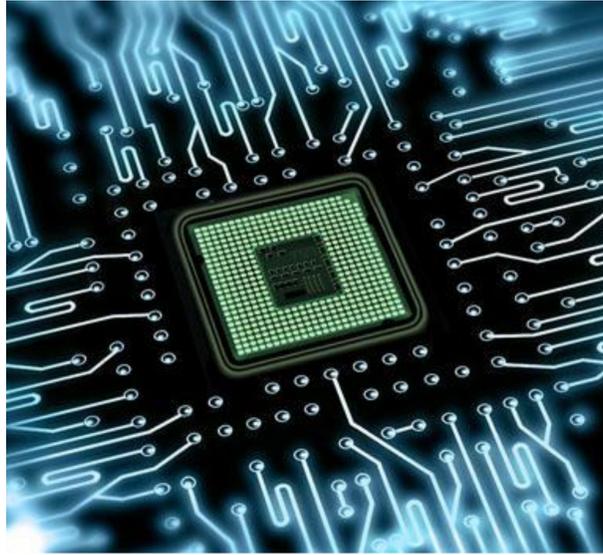
Quantiphi nutzt das TAO-Toolkit, um benutzerdefinierte Objekterkennungs- und Klassifizierungsmodelle zu trainieren, die die Versandgenauigkeit validieren. Mit DeepStream kann die KI-Anwendung problemlos auf Jetson NX Geräten im gesamten Lager eingesetzt werden.

Verbesserte Genauigkeit für Echtzeit-Computer-Vision-Anwendungen mit TAO-Toolkit.

 **NVIDIA.**  **quantiphi**



Beispiele des Nutzens durch Inspektion von Fertigungsschritten



Halbleiter

1% zusätzlicher Ausbeute bringt 60 Mio. \$ Gewinn/Jahr



Elektronik

64% weniger Testausfälle führen zu einer Ertragssteigerung von über 1,5% durch bessere Kapazitätsauslastung



Metallkomponenten

Verdoppelung des Durchsatzes von 20-30 Sekunden auf weniger als 9 Sekunden für Defekte bis zu 50 µm



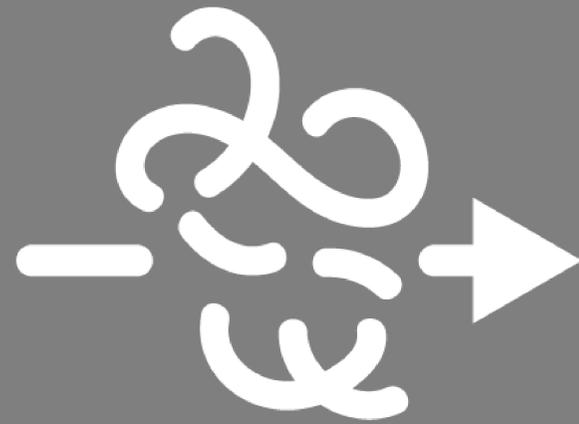
Montage

Jede Linie hat im Durchschnitt 1 Testausreißer pro Tag - die Reduzierung der Fehlerrate spart 50.000 Dollar pro Linie und Tag

KI-Anwendungsfelder in der Industrie

	Entwurf, Simulation, Produktionsplanung	Lieferkette und Logistik	Produktion	Vertrieb und Marketing	Service
	(Generative) Mechanical / Electrical Design / Additive MFG	Document Analytics & Proc.	Quality Control / Metrology	Customer Transparency & Analytics	Product Analytics
	Fluid & Aerodynamics Sim	Supply Chain Optimization / Recommendations	Meter Reading	Digital Customer Journey	Predictive Maintenance
	Structural Mechanics Sim	Routing Optimization	Anomaly Detection (machine, line, factory)	Customer Recommendations	Asset Tracking / Mgmt
	Materials & Chemical Science	Capacity / Demand Forecast	Predictive Maintenance	Marketing Mix Optimization / Targeted Marketing	Technician / Decision Support
	Production Planning & Commissioning	Inventory / Asset Mgmt	Root Cause Analysis		Spare Parts Handling
	Process Optimization	Materials Handling	Thermographic Plant Analysis		Document Analytics & Proc.
	Chemical Process Optim.	Production Network Optimization / Site Balancing	Robotics / AMR / AGV		
			Energy Optimization		
			Process Optimization		
			Safety & Security		
			Document Analytics & Proc.		
	Digital Thread – Product, Supply Chain, Manufacturing Digital Twin(s)				
Verwaltung / Finanzen	Process / Task Mining Document Analytics & Proc.	Financial Monitoring Anomaly Detection	Secure Transactions Risk Management		
Beschaffung	Supplier Identification and Benchmarking	Strategic Sourcing Contract Management	Supplier Risk Management Payment Tracking	Anomaly Detection Spend Analysis	
Personalwesen	Candidate Sourcing Job Recommendations	Candidate Matching Candidate Experience	Candidate Benchmarking Individualized Offers	Document Analytics & Proc.	

Herausforderungen bei der KI-Entwicklung



Komplexe Prozesse



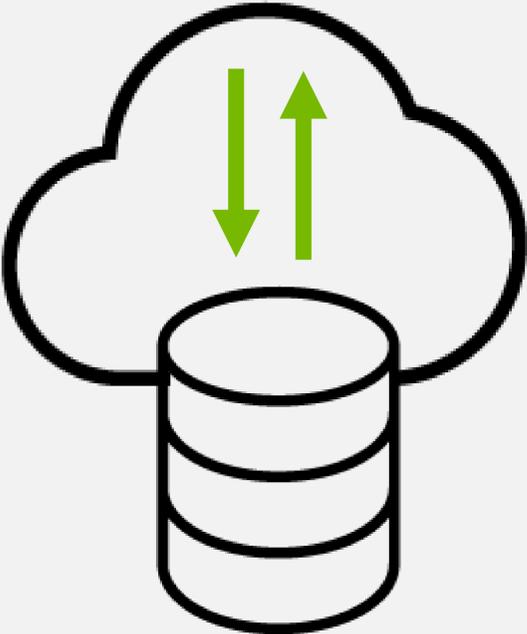
Begrenzte Ressourcen
und Verständnis



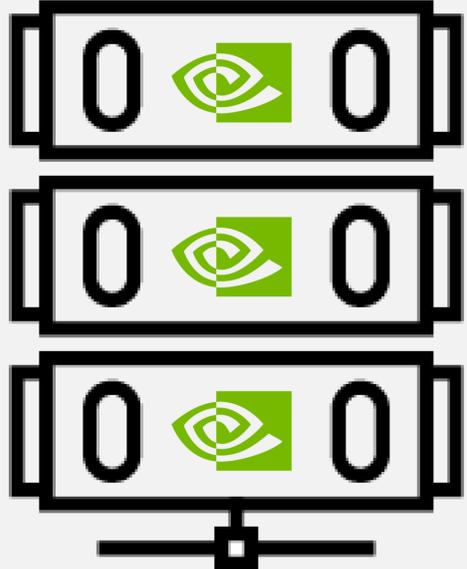
Vom Prototypen zum Produktivsystem
unter hohem Zeitdruck

Anforderungen für die Erstellung benutzerdefinierter LLMs

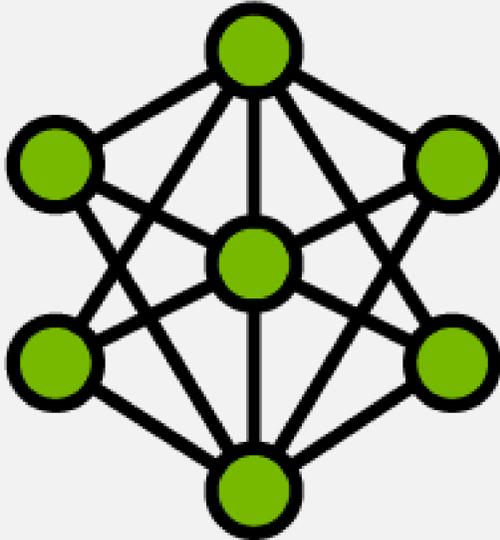
Trainingsdaten



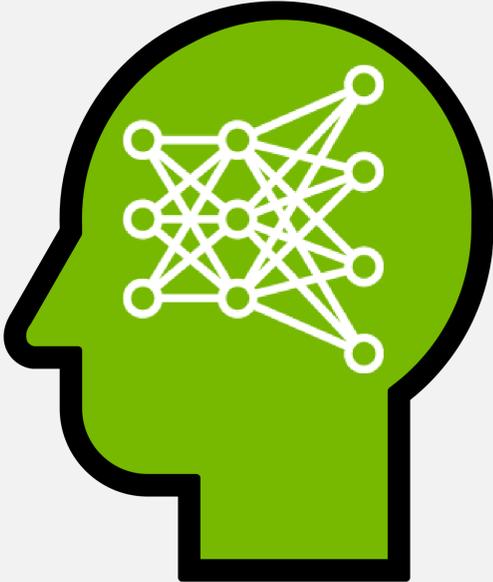
Rechner-Infrastruktur



Trainings- und Inferenz-Werkzeuge



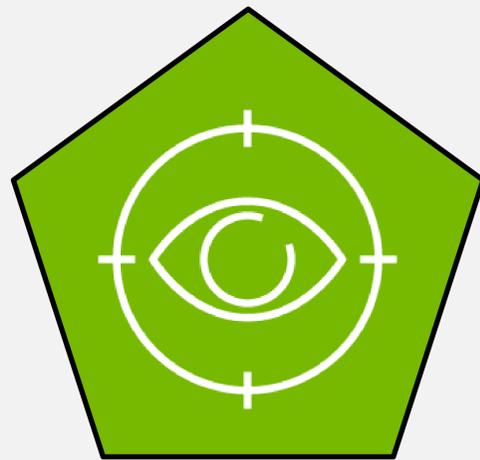
KI Fachwissen



Schritte für den Einstieg in die generative KI

Nutzen Sie maßgeschneiderte LLMs, um Ihr Unternehmen zu differenzieren

Geschäftschancen identifizieren



Zielen Sie auf Anwendungsfälle ab, die einen bedeutenden Einfluss auf das Geschäft haben und mit einzigartigen Daten angepasst werden können.

Aufbau von Domain- und KI-Teams



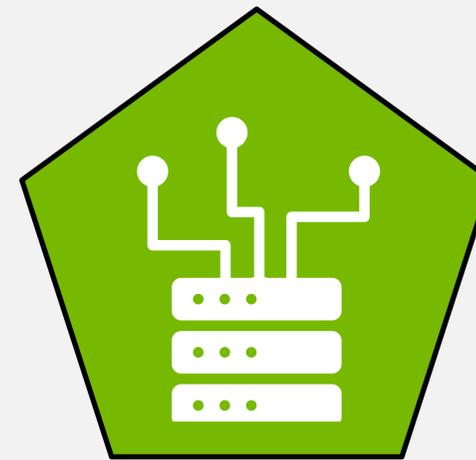
Identifizieren Sie interne Ressourcen und ergänzen Sie diese mit KI-Fachwissen von Partnern und Anwendungsanbietern.

Datenanalyse für Training & Anpassung



Erfassen, verfeinern und sichern Sie Daten, um entweder datenintensive Grundmodelle zu erstellen oder bestehende Modelle anzupassen.

Beschaffung der Infrastruktur



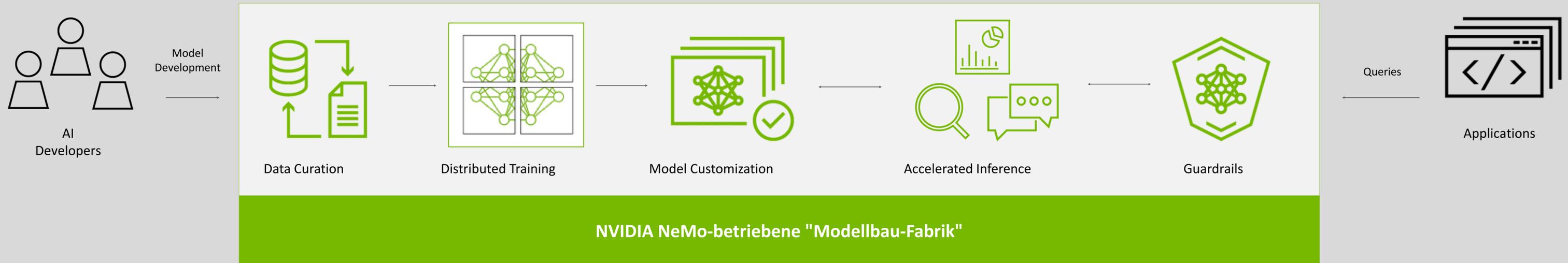
Bewertung der Infrastruktur, der Architektur und des Betriebsmodells unter Berücksichtigung der Kosten und des Energieverbrauchs.

Plan für verantwortungsvolle KI entwickeln



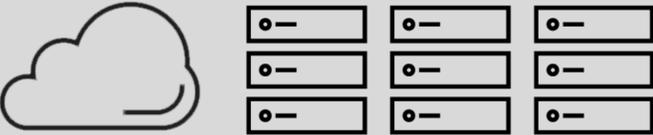
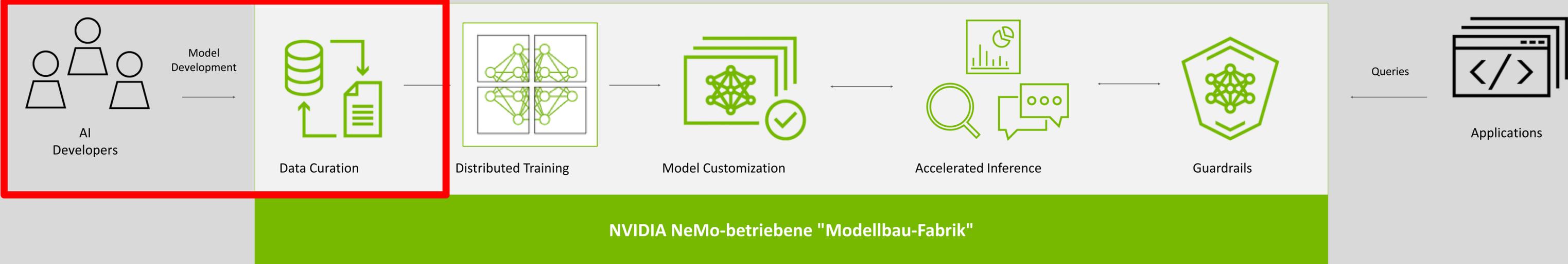
Nutzung von Tools und Best Practices, um sicherzustellen, dass Grundsätze der verantwortungsvollen KI im gesamten Unternehmen umgesetzt werden.

NVIDIA NeMo: Ein Werkzeugkasten zur Erstellung benutzerdefinierter großer Sprachmodelle (LLMs)



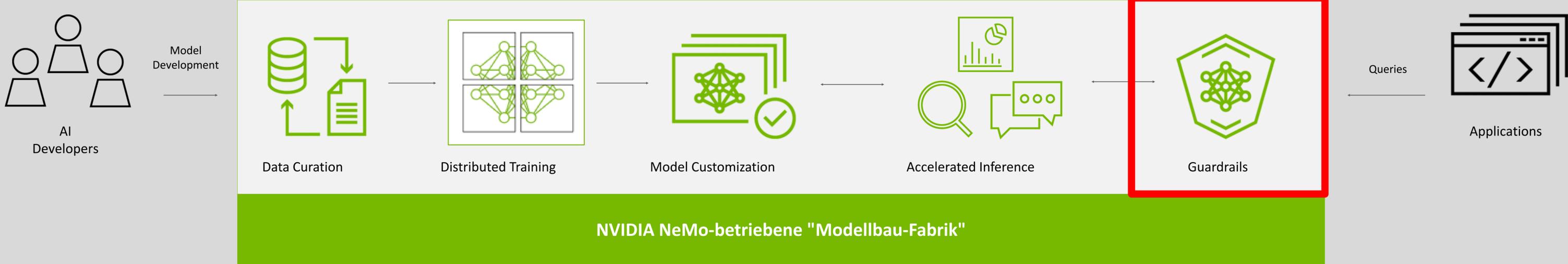
Vertrauen < > Kompetenz

"Werkzeugkasten" zur Erstellung benutzerdefinierter großer Sprachmodelle



Vertrauen < > Technik

"Werkzeugkasten" zur Erstellung benutzerdefinierter großer Sprachmodelle



2024: Ein besonders wichtiges Jahr für KI.

- **Für Hilfe, Anregungen und Training empfohlen:
Die GTC Technology Conference '24 (17.-21. März 2024)**



Ludwig von Reiche
Geschäftsführer
NVIDIA GmbH

lreiche@nvidia.com

NVIDIA in Brief (1/2)

NVIDIA in Brief



NVIDIA is tackling challenges no one else can solve. Our work in AI and digital twins is transforming the world's largest industries and profoundly impacting society. [Learn more.](#)

Company History

Since its founding in 1993, NVIDIA has been a pioneer in accelerated computing. The company's invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined computer graphics, ignited the era of modern AI, and is fueling industrial digitalization across markets. NVIDIA is now a full-stack computing company with data center-scale offerings that are reshaping industry.

Key Stats

- › Founded in **1993**
- › Founder and CEO: **Jensen Huang**
- › **27,000+** employees in **50+** locations
- › **\$27 billion** revenue in FY23
- › **7,500+** granted and pending patent applications worldwide
- › **\$1 trillion** available market opportunity
- › **4 million** developers in the [NVIDIA Developer Program](#)
- › **14,000** global startups in [NVIDIA Inception](#)
- › **"Best Places to Work in 2023"** – *Glassdoor*
- › **"World's Best Performing CEO"** – *Harvard Business Review*

Impact by Industry



Automotive

NVIDIA DRIVE® powers all 30 of the 30 top autonomous vehicle data centers.



AI Factories

More than 40,000 companies use NVIDIA AI technology to power AI factories.



Digital Twins

NVIDIA Omniverse™ has 300,000 individual users and 700 companies in the pipeline.



Gaming

More than 200 million gamers and creators use NVIDIA GeForce® GPUs.



Healthcare

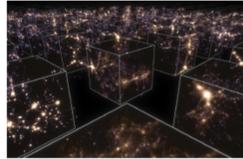
More than 1 million developers have downloaded the MONAI framework for AI in healthcare imaging.



Robotics

More than 1 million developers use the NVIDIA Jetson™ platform for AI at the edge.

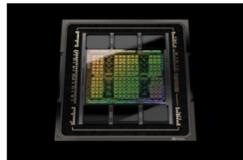
NVIDIA in Brief (2/2)



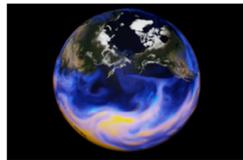
NVIDIA is the world's engine for AI. Services from **Alibaba, Amazon, Google, Meta, Microsoft, Snap, Spotify, Tencent**, and 40,000 other companies are built and run on NVIDIA AI technologies.



ChatGPT, powered by an **NVIDIA DGX™** AI supercomputer, reached 100 million users in just two months, making it the fastest-growing app in history and marking the "iPhone moment for AI."



NVIDIA technologies are behind the recent breakthroughs in **large language models** used to build **generative AI**, the most important AI models today. The **NVIDIA Hopper™** GPU architecture's Transformer Engine supercharges both.



Accelerated computing is sustainable computing. If we switched accelerated computing workloads from CPU-only servers to DPU- and GPU-accelerated systems worldwide, we estimate nearly 20 trillion watt-hours of energy savings per year.



NVIDIA DGX Cloud, through partnerships with **Microsoft Azure, Google Cloud, and Oracle Cloud Infrastructure**, makes it possible for every enterprise to access its own AI supercomputer using a simple web browser.



[For more information](#)

Latest NVIDIA News



AI

- › [NVIDIA H100 Tensor Core GPU](#) now offered by cloud giants.
- › [NVIDIA DGX Quantum](#) accelerates quantum-classical computing.
- › [NVIDIA cuOpt™ software](#) achieves world record in route optimization.
- › [Generative AI cloud services](#) advance text, visual content, and biology.



RTX/Graphics

- › [GeForce RTX™ 40 Series GPUs for laptops](#) unveiled.
- › [400+ RTX games and applications](#) have been created—including over 280 games and applications that support AI-powered DLSS.
- › [DLSS 3](#), supported in nearly two dozen games, delivers up to 4X rendering performance.
- › [GeForce NOW™ library](#) now has over 1,600 games.



Omniverse/Industrial Digitalization

- › BMW Group started a [global rollout of NVIDIA Omniverse™](#).
- › [NVIDIA DRIVE Hyperion™ architecture](#) achieves new safety milestones.
- › [Omniverse Cloud](#) SaaS offering coming to Microsoft Azure.
- › Full lineup of [Jetson Orin™](#) modules available for edge AI and robotics.
- › [Mercedes-Benz](#) using NVIDIA Omniverse to design next-gen factories.

