



**DIGITAL
MAKERS**

CANCOM

DIE (R)Evolution der KI

Ein 360°-Blick auf den KI-Wandel



CANCOM



DIGITAL MAKERS



MANUEL MOSER

DIRECTOR DIGITAL INNOVATION – AI & SOFTWARE



manuel.moser@cancom.com



linkedin.com/in/m-moser/



ANDREAS FRÜHWIRT

LEAD IT ARCHITECT – SOFTWARE & PLATFORMS



andreas.fruehwirt@cancom.com



linkedin.com/in/andreas-fruehwirt/



FLORIAN BÖTTCHER

SOLUTION ARCHITECT – AI INFRASTRUCTURE



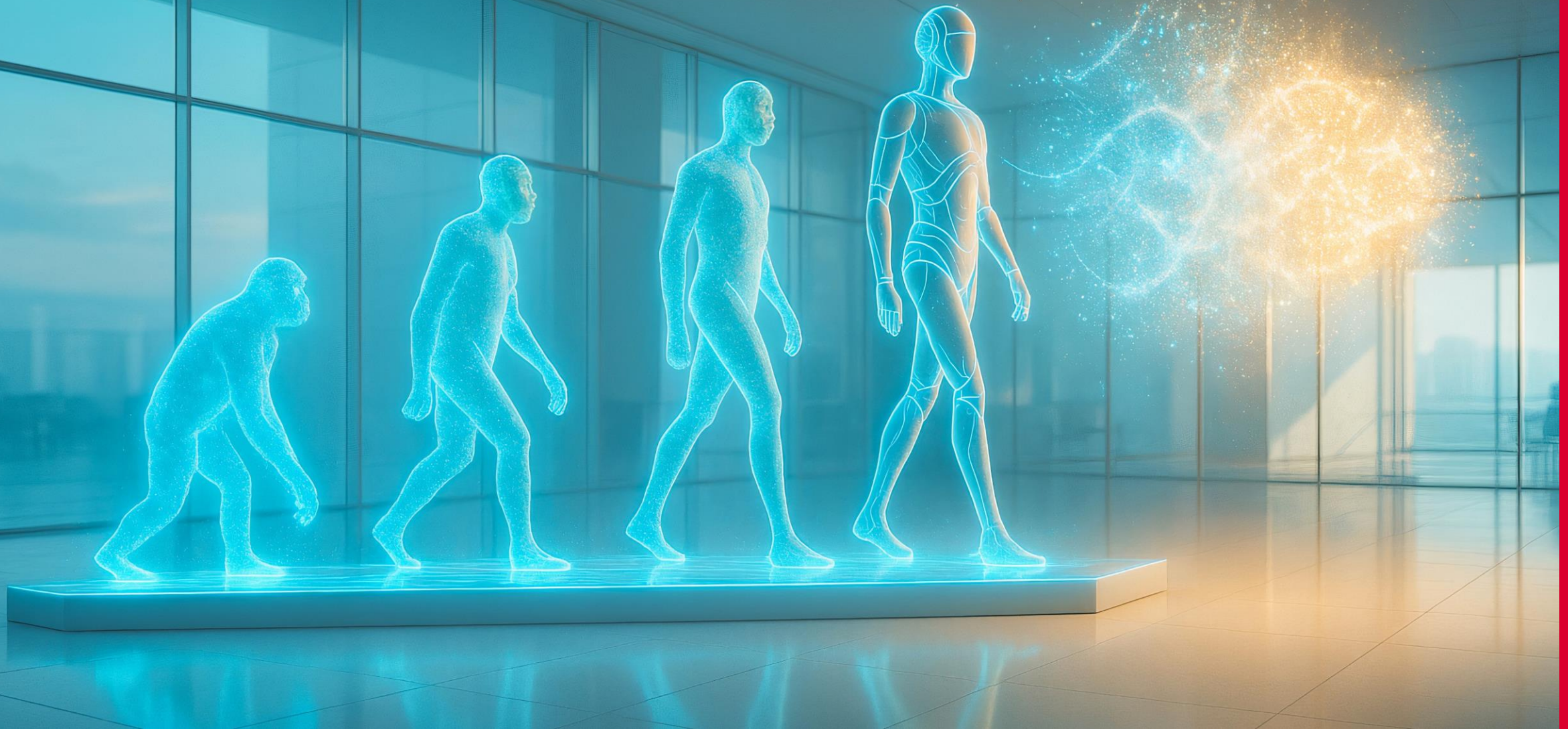
florian.boettcher@cancom.com



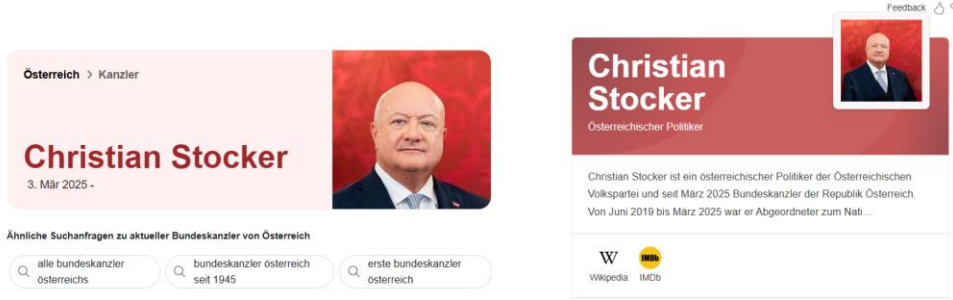
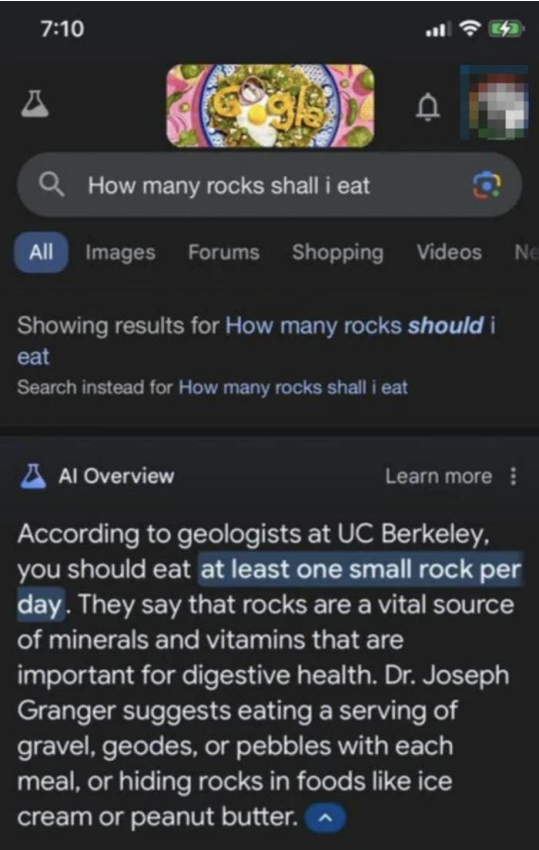
linkedin.com/in/florianboettcher/



VERLIERT DIE MENSCHHEIT IHR INTELLIGENZ-MONOPOL?



THE FUTURE?



AI beschleunigt den Wandel

Top Berufsgruppen, die generative KI einsetzen

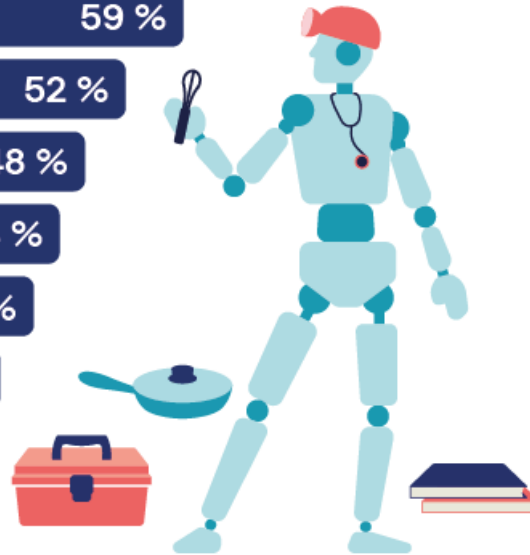
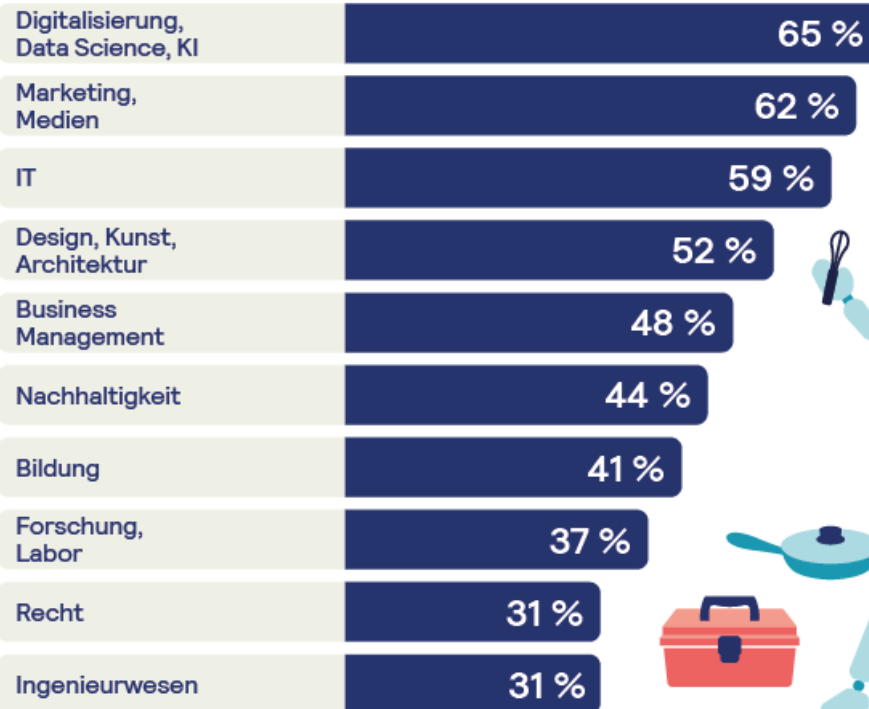
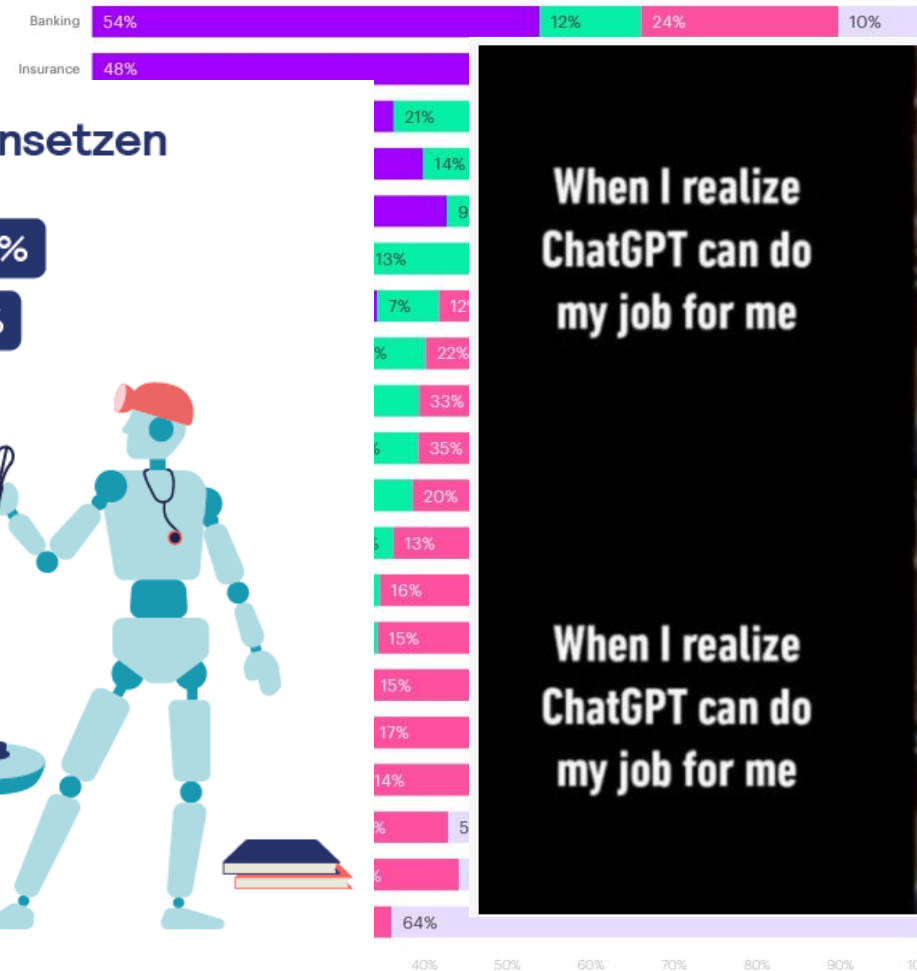


Figure 3: Generative AI will transform work across industries



Work time distribution by industry

When I realize
ChatGPT can do
my job for me

When I realize
ChatGPT can do
my job for me



he US in 2021

Non-language
tasks

cross
ted by
(LLMs)

total worked time
e high potential

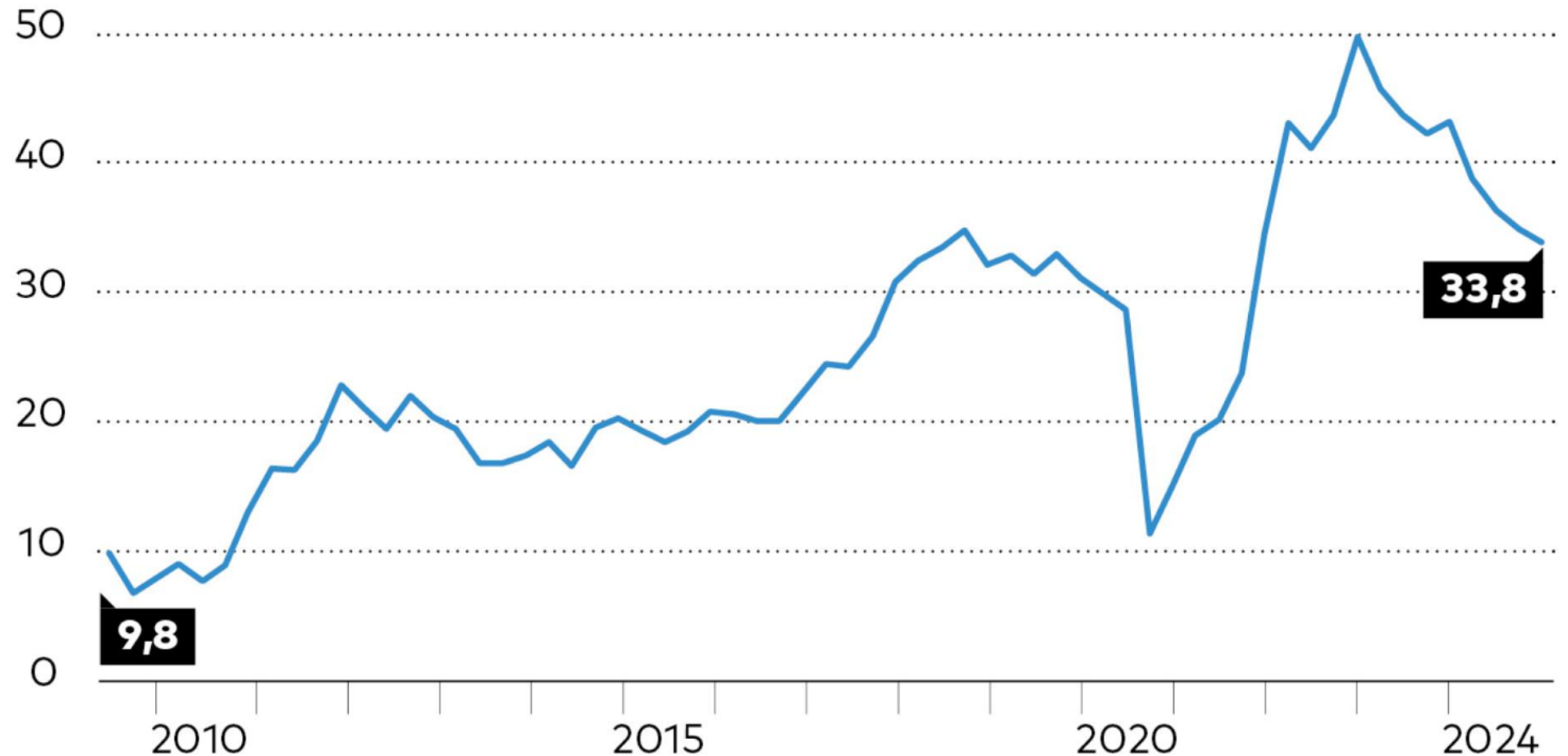
Occupational
S Bureau of

language (out
tries using their
ployment level
omation can
t from a human

worker. Tasks with higher potential for augmentation are those in which LLMs would need more involvement from human workers.

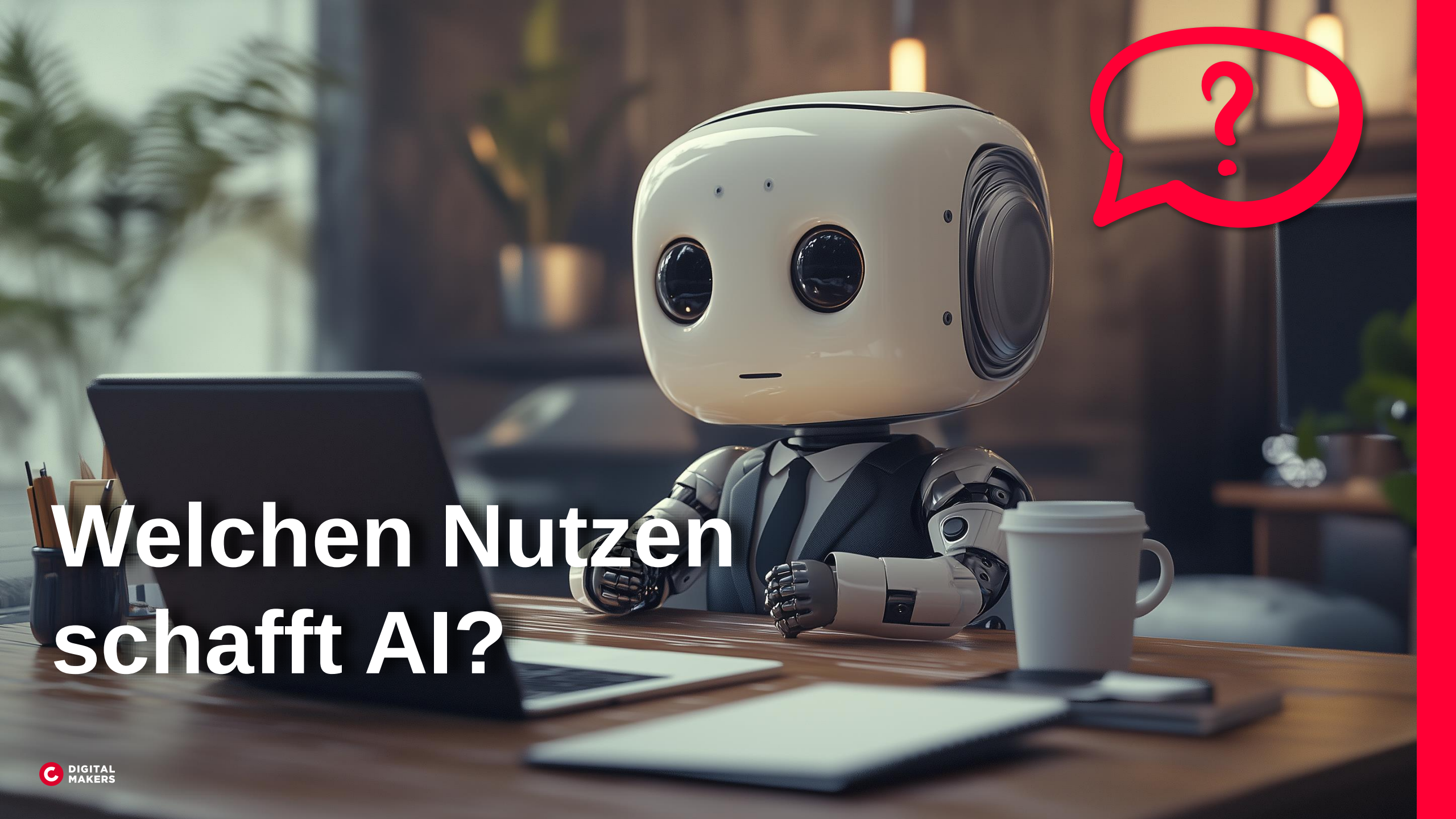
Einem Drittel fehlen die Spezialisten

Anteil der Unternehmen, die Fachstellen nicht besetzen können,
Gesamtwirtschaft, in Prozent



Quelle: Ifo-Institut

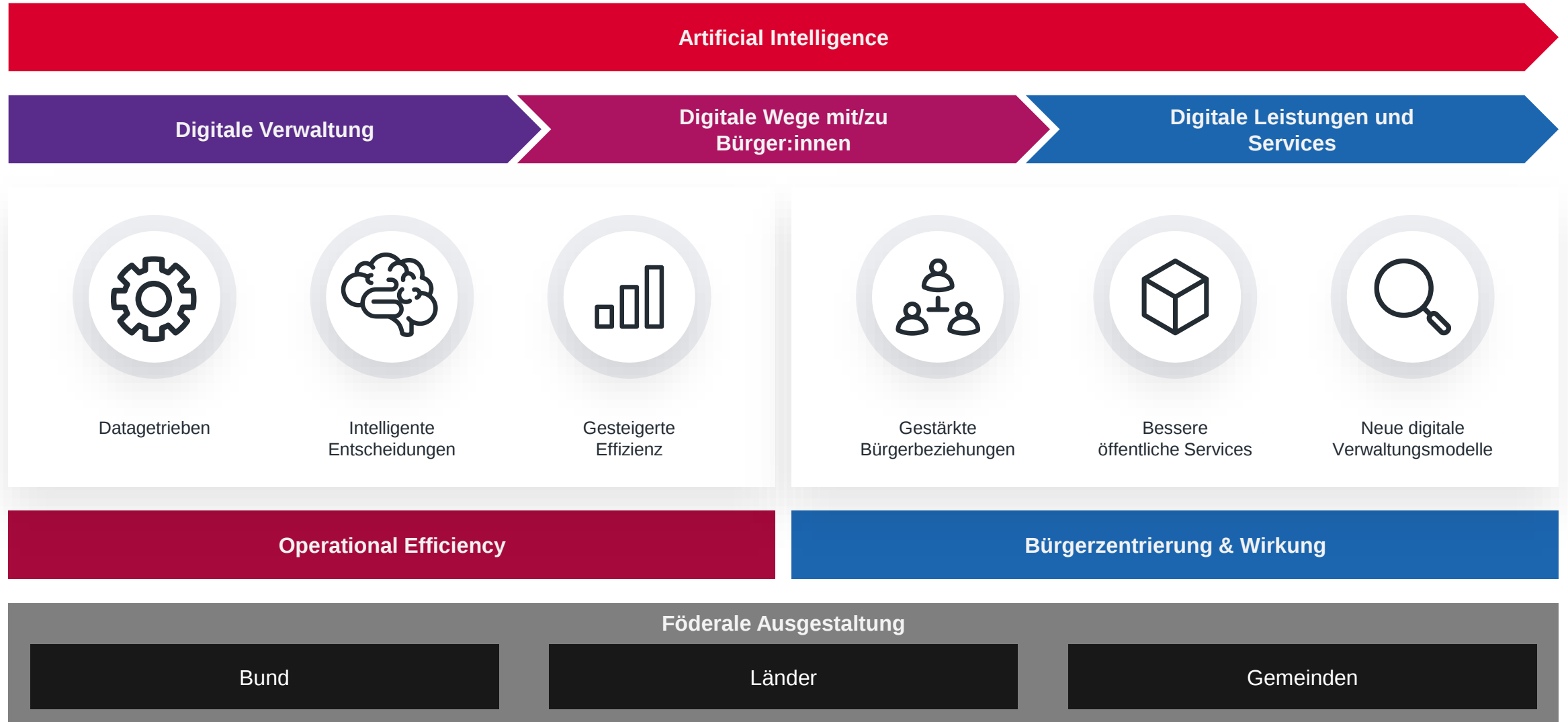
welt



Welchen Nutzen schafft AI?

UNSER ZIEL IST DIGITAL

Das (perfekte) digitale Verwaltung



KI IM E-GOVERNMENT: MEHR ALS NUR EIN CHATBOT?

ANWENDUNGSFÄLLE IN DER VERWALTUNG

Bürgerkommunikation & Service

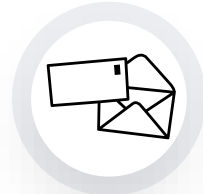
Allgemeine Verwaltung & Prozesse

Fachverwaltung & Spezialbereiche

Heute möglich



Service-
Chatbots



Automatisierte Briefe &
Bescheide



Zusammenfassung von
Fachtexten (Urteile, Akten, ...)

Pilotfähig



Formular Assistant

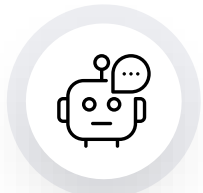


Erklärbare
Entscheidungsschreiben



Fachliche Beratung durch
Agent (Förderungen, Pflege, ...)

Visionär



Persönlicher Verwaltungs-
Agent



Verwaltungs-
Wissensagent



Adaptive Lern- &
Fachinhalte

CHANCEN DURCH KÜNSTLICHE INTELLIGENZ

Effizienz durch Prozess-Automatisierung

Entlastung von Fachkräften durch Automatisierung

Bessere Entscheidungen durch Wissenszugang

Analyse großer Datenmengen als Entscheidungsgrundlage

Verwaltung als Plattform

KI als Schnittstelle für die Vernetzung von Behörden für eine gemeinsame Plattform.



Verbesserte Bürgerkommunikation

Schnellere, verständlichere u. 24/7 Schnittstelle zum Bürger

Transparenz, Nachvollziehbarkeit u. Vertrauen

Chance zur besserung Dokumentation u. Begründung von Entscheidung

Vorausschauende Politikgestaltung

Frühzeitige Erkennung von Bürgertrends u. Simulation von Auswirkungen

ARBEITNEHMERVERTRETUNG

Bot-gestützte Mitglieder-Beratung



AUSGANGSLAGE

- Vielzahl an Mitglieder-Anfragen via E-Mail
- Berater:innen beantworten triviale, wiederkehrende Fragen
- Wissensaufbau für Themen notwendig, um beraten zu können (Steuerrecht, Arbeitsrecht, Familienrecht, etc.)
- Recherche in internen Systemen notwendig (Dokumente, Wiki, Datenbanken)



LÖSUNG

- Entlastung der Mitarbeiter:innen durch automatisierte Recherche und Validierung
- Vorgeschlagene Antwort an die Mitglieder
- Berater:in kann immer noch manuell eingreifen
- Mehr Zeit für qualitative Beantwortung von Spezialanfragen



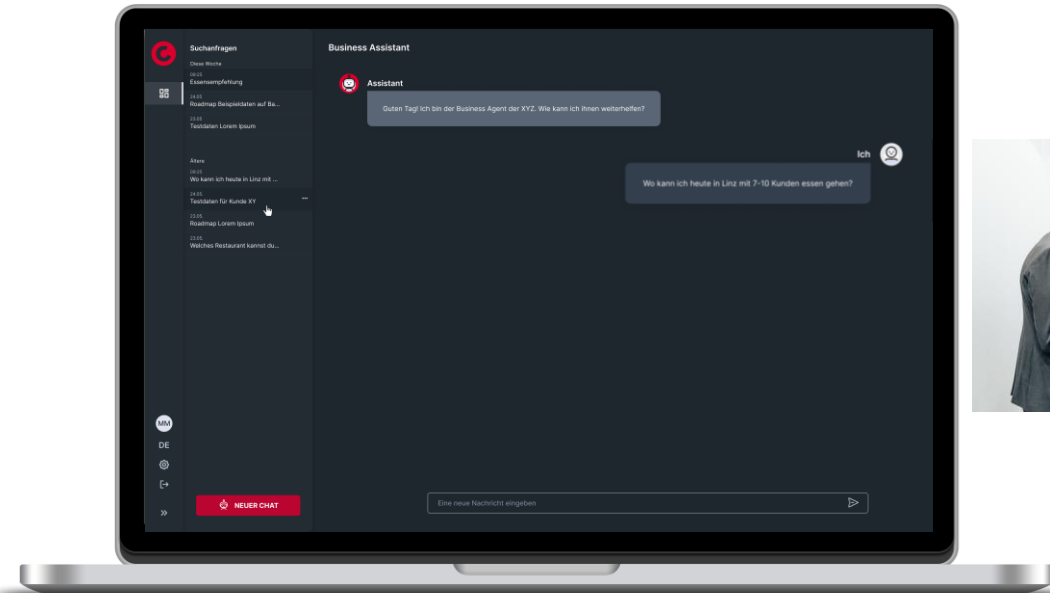
Generative AI



Service Chatbot



Cloud
Transformation



A man in a white shirt is sitting at a desk, looking at a large computer monitor. A white robot with glowing blue eyes and internal circuitry is sitting next to him, also looking at the monitor. The monitor displays various data visualizations, including line graphs, circular radar-like charts, and text-based data. The background is a blurred office environment with blue lighting.

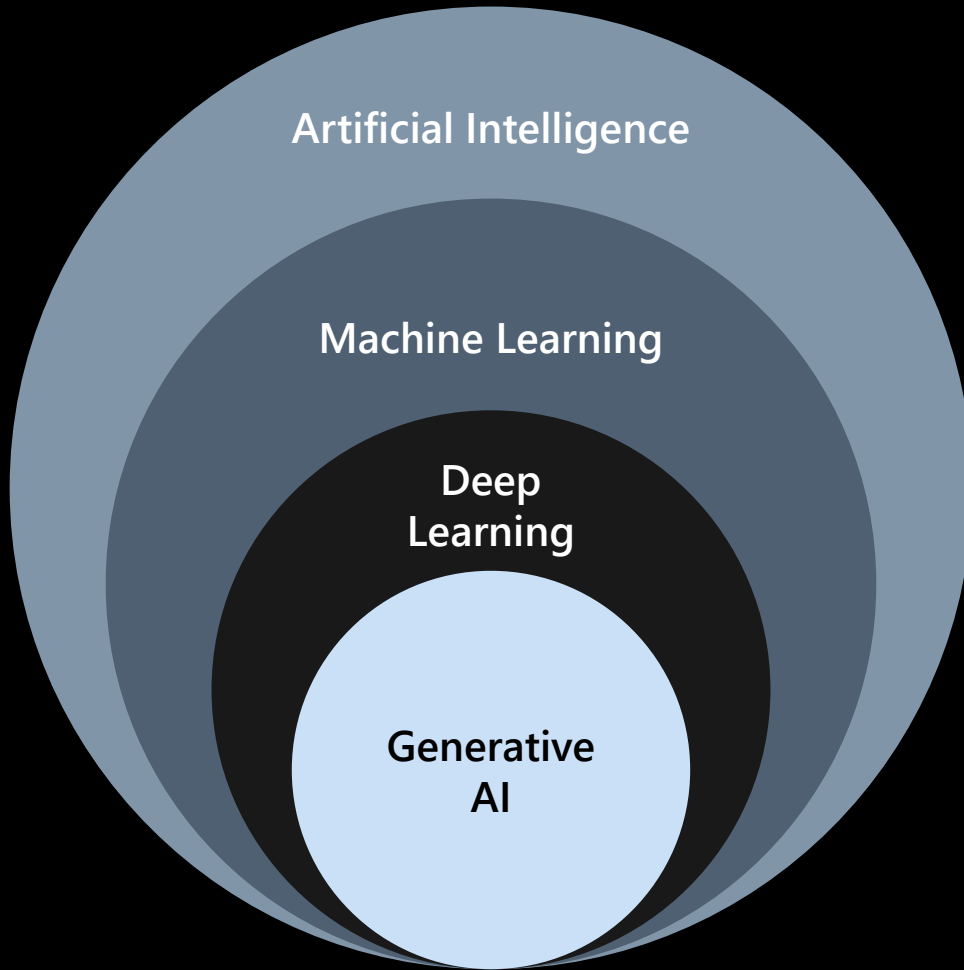
Human in the Loop

Auch KI braucht User Experience

A woman with long dark hair tied back is shown in profile, reaching out with both hands towards a large, glowing blue digital network visualization. The visualization consists of numerous bright blue nodes connected by thin lines, forming a complex, cloud-like structure. The scene is set in a dark room with a bookshelf visible in the background. The overall color palette is dominated by dark blues and blacks, with the bright blue of the digital network providing a strong contrast. A solid red vertical bar is visible on the far right edge of the image.

Warum der Hype?

AI breakthroughs



1956

Artificial Intelligence

The field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence.

1997

Machine Learning

Subset of AI that enables machines to learn from existing data and improve upon that data to make decisions or predictions.

2012

Deep Learning

A machine learning technique in which layers of neural networks are used to process data and make **decisions**.

2021

Generative AI

Create new written, visual, and auditory content given prompts or existing data.

Warum ist Generative KI anders?

Deep Learning

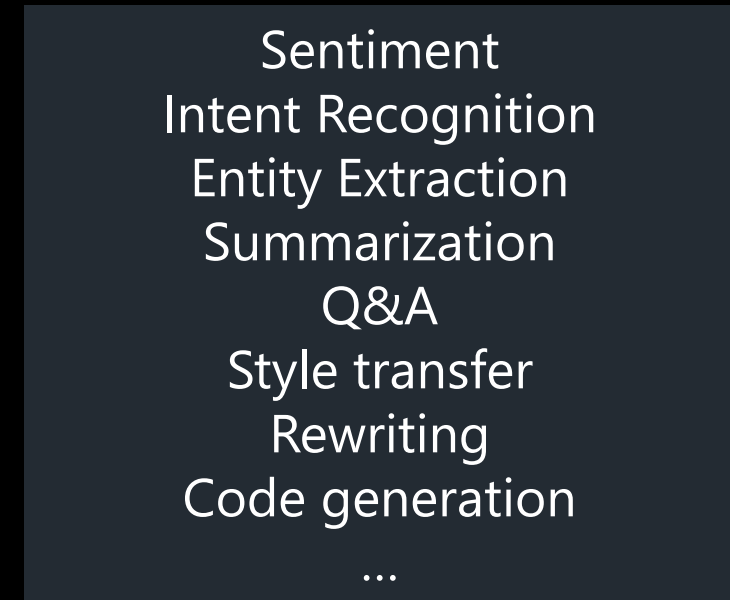
Typisches Machine Learning für Natural Language Processing (NLP)



- **Ein Modell pro Funktion**
- **Viele Trainingsdaten**
- **Hoch optimiert**

Generative AI

z.B. OpenAI ChatGPT-4



- **Ein einziges Modell für alle Workloads**
- **Formulierung der Aufgabe in natürlicher Sprache**

Die (R)Evolution der KI

ChatGPT

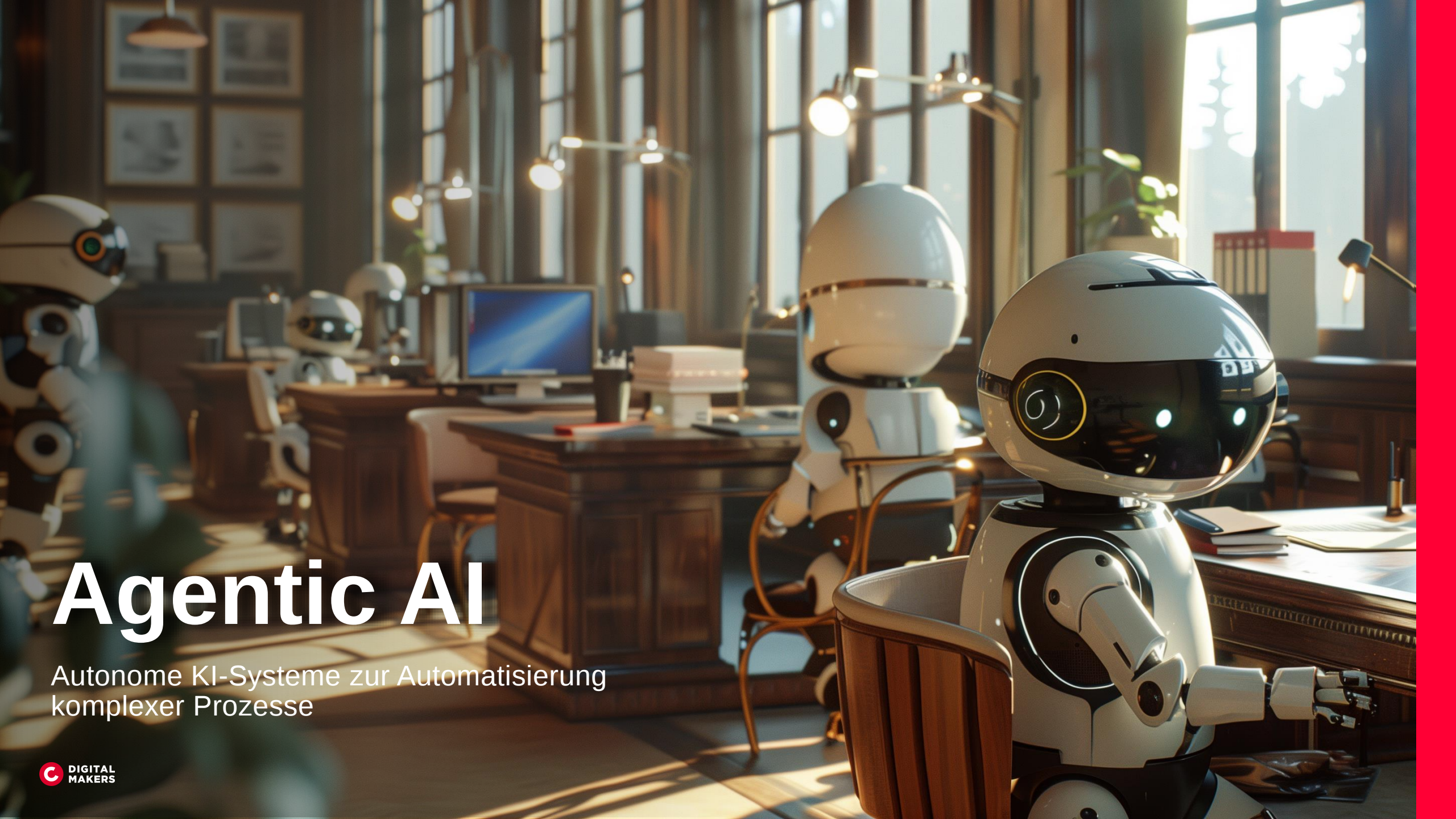


Assistant



Agenten





Agentic AI

Autonome KI-Systeme zur Automatisierung komplexer Prozesse



Let's ~~code~~
Let's prompt

AGENTIC AI

Building an agentic world

WAS SIND AI AGENTS?



AUTONOME KI-SYSTEME



INTELLIGENTE
ENTSCHEIDUNGSFINDUNG



AGIEREN NACH
UNTERNEHMENS-REGELN

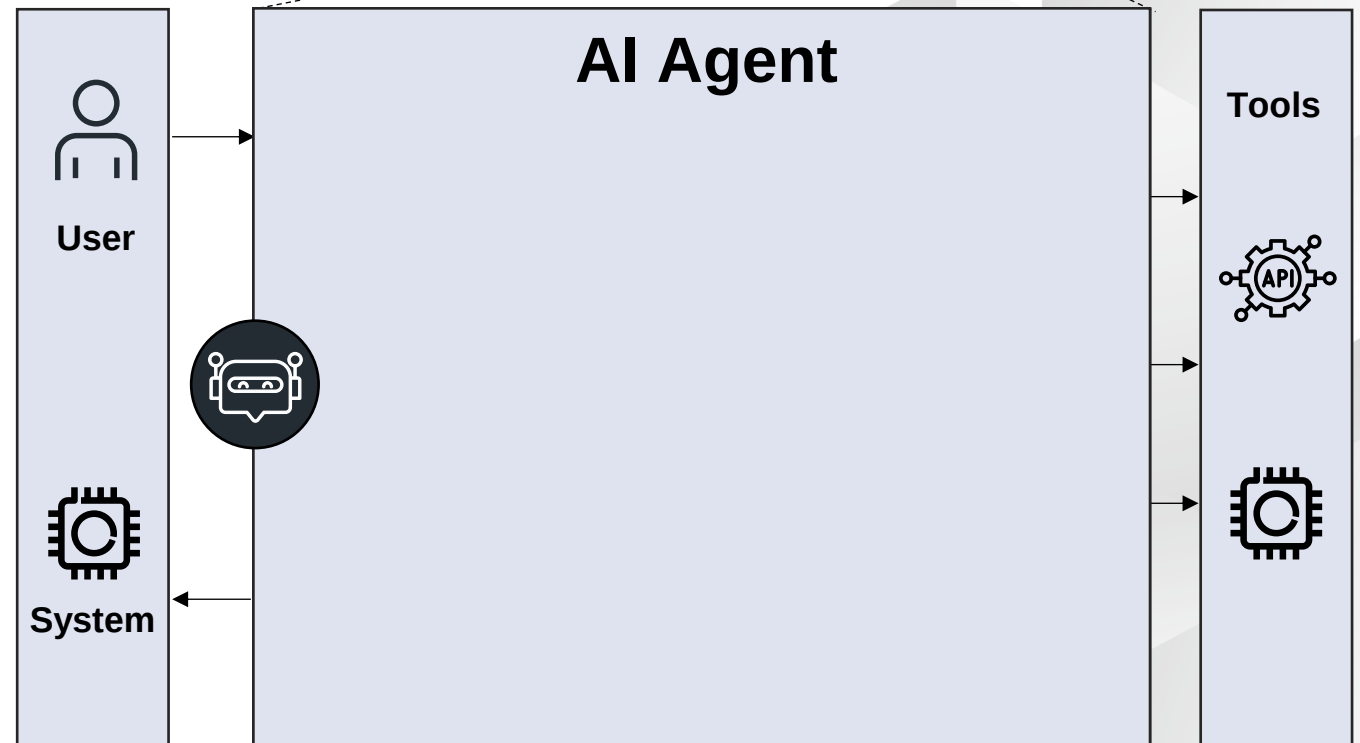
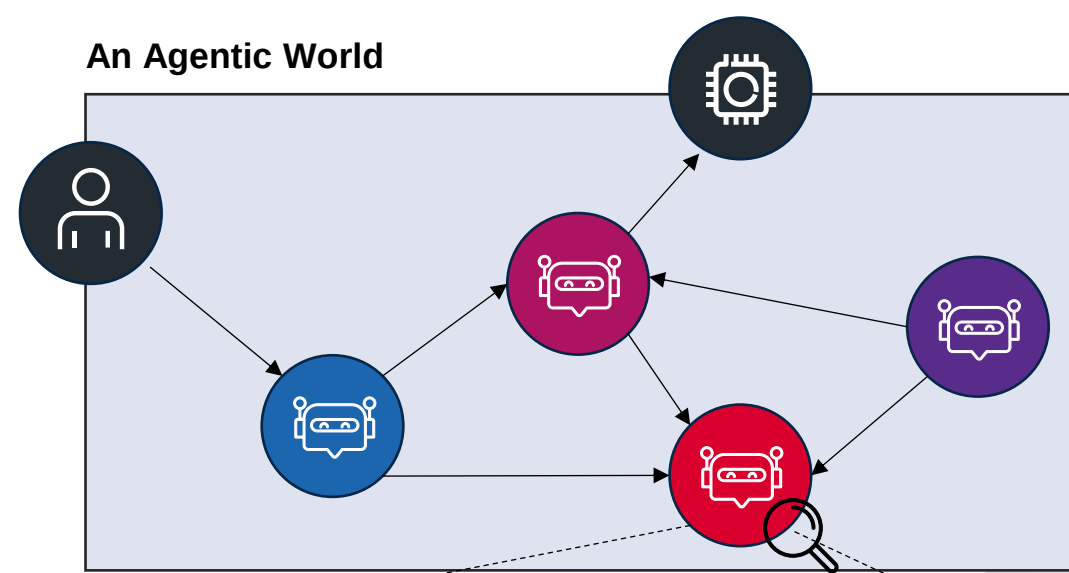


AUTOMATISCHE
ADAPTIERUNG AN KONTEXT



AUTOMATISIEREN
KOMPLEXER PROZESSE

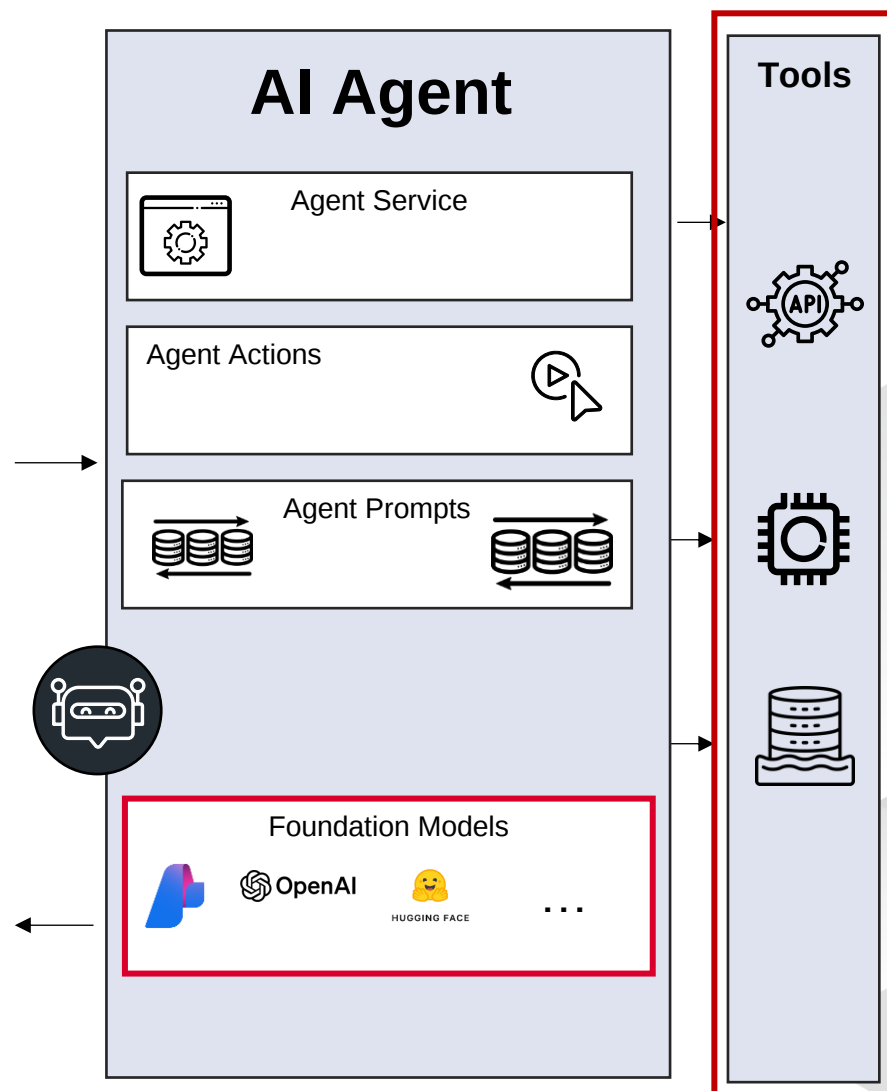
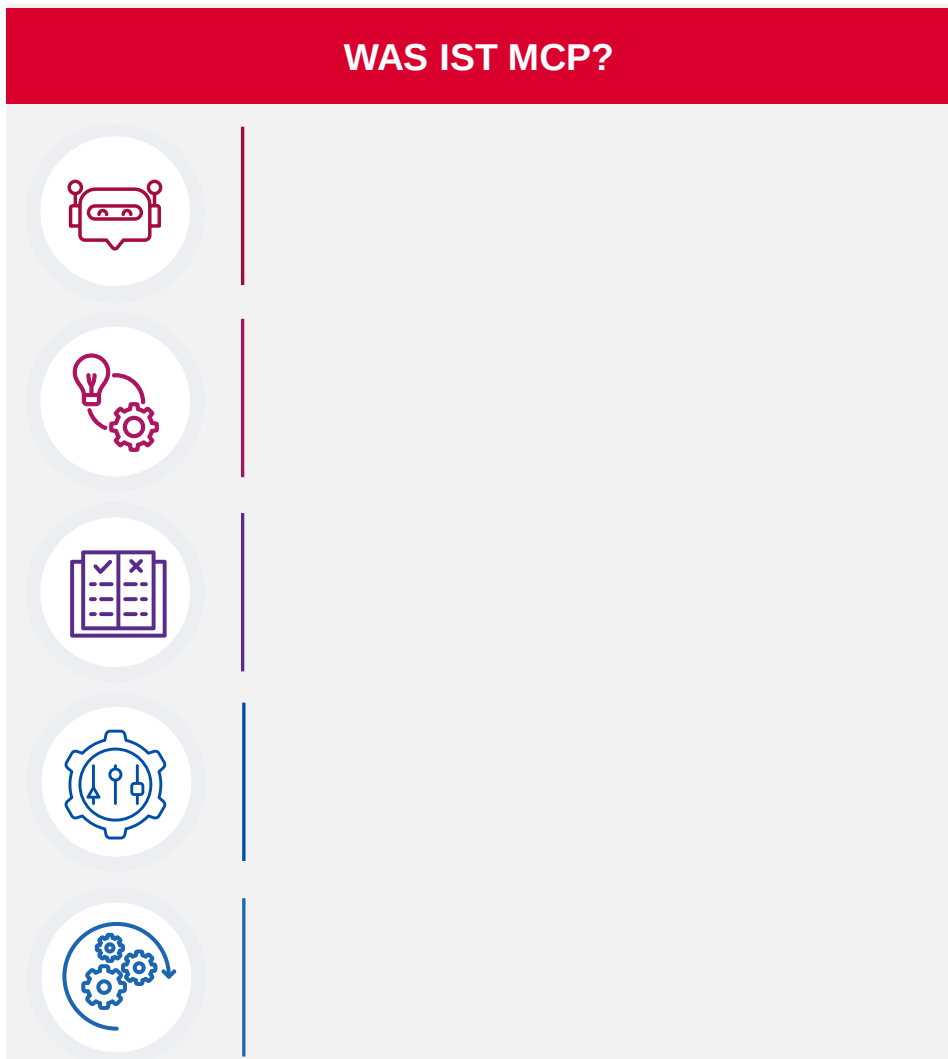
An Agentic World



CONNECT YOUR AGENT WITH DATA

Unified Agentic Data Access via Model Context Protocol

WAS IST MCP?



CONNECT YOUR AGENT WITH DATA

Unified Agentic Data Access via Model Context Protocol

WAS IST MCP?



Agenten Information über sich und sein Environment



Contract zwischen Agent und Data Environment



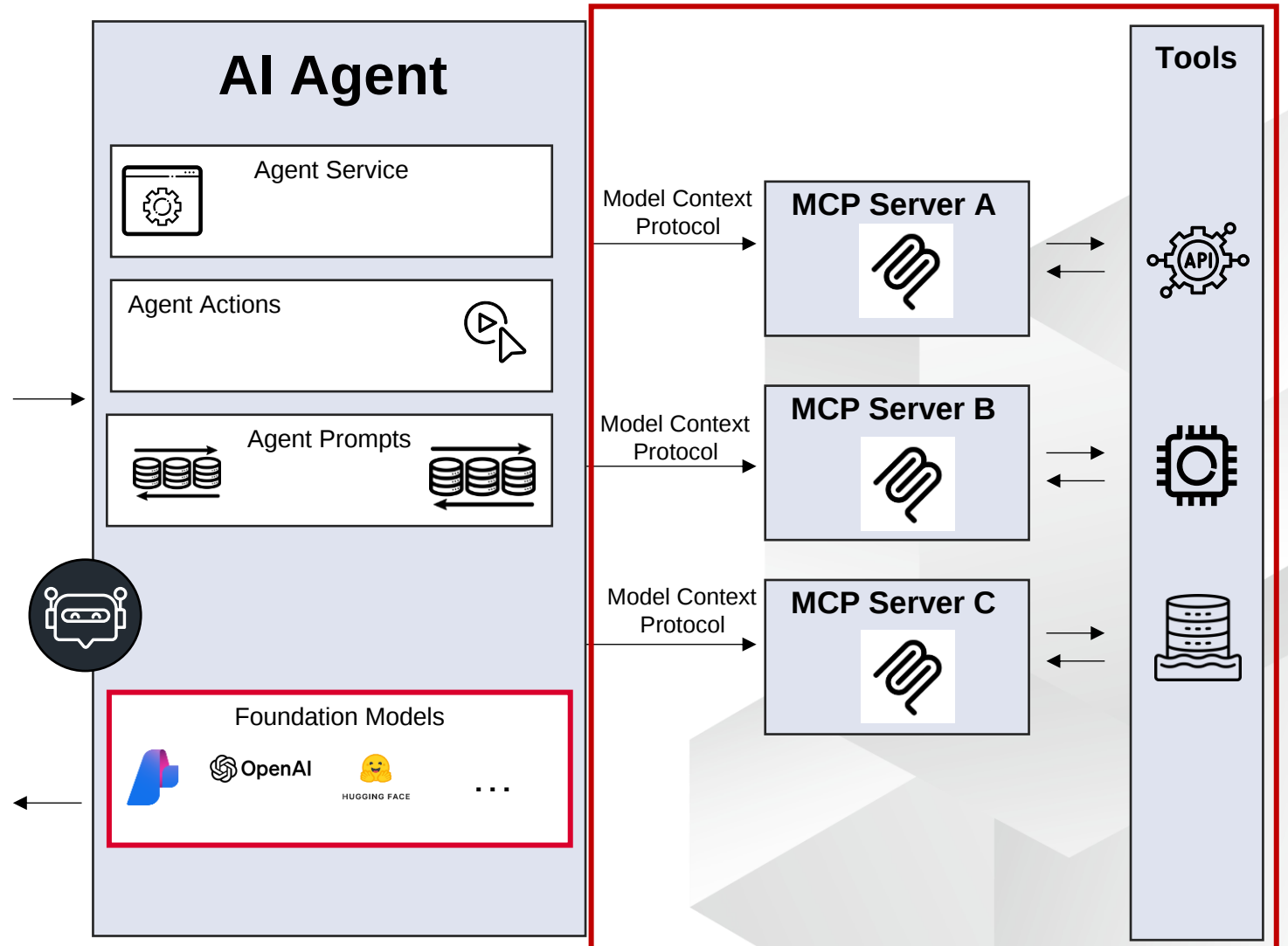
Austausch von Fähigkeiten und Wissen



Dynamische Selbstbeschreibung von Fähigkeiten und Kontextwissen



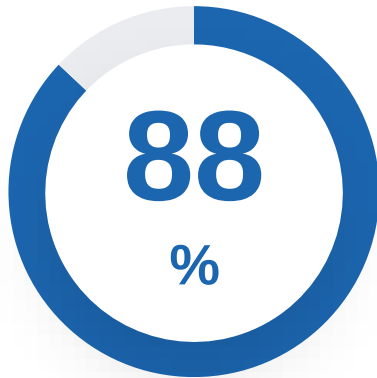
Schlüsselbaustein für interoperable Agentensysteme



AGENTIC AI

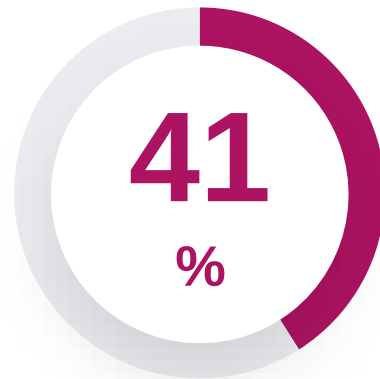
Der Schritt von operativer Exzellenz zur intelligenten Autonomie.

Technologische Infrastruktur



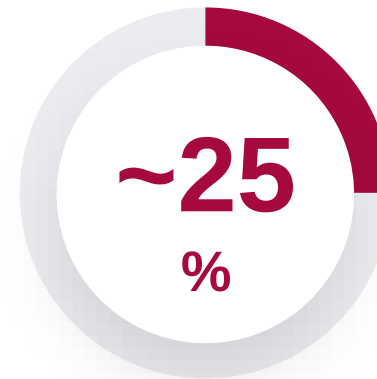
... der Führungskräfte sehen fehlende technologische Infrastruktur für KI als zentrales Hindernis für die Skalierung.*

Jährliches Wachstum



... beträgt die prognostizierte durchschnittliche jährliche Wachstumsrate bis 2030.**

Produktivität



... steigt die Produktivität durch den Einsatz von KI und den Einsatz von Agentic AI.***



AI Agents im Unternehmen

Laut Gartner werden bis 2028 15 % der alltäglichen Entscheidungen autonom von AI Agents getroffen

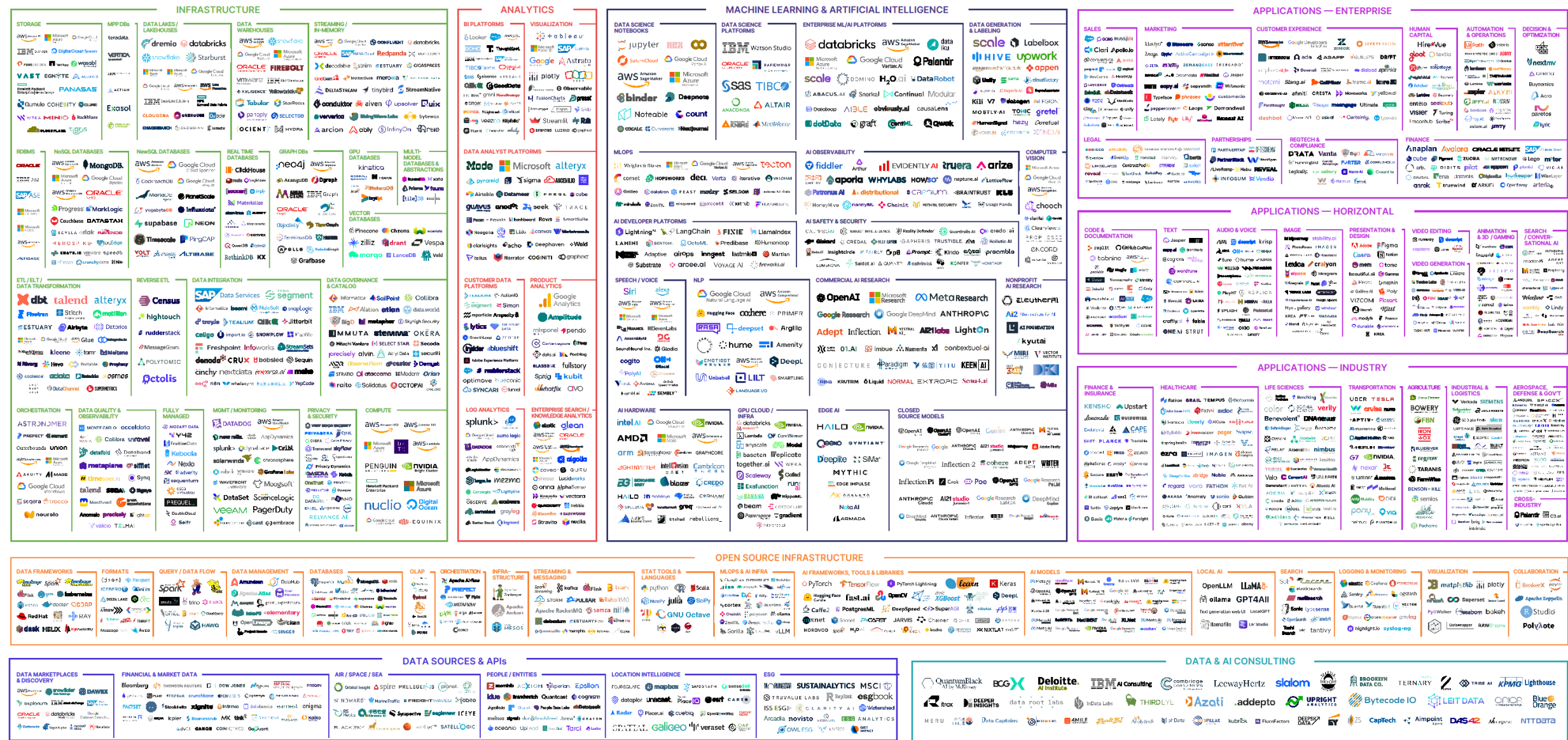


Let's ~~code~~
Let's prompt

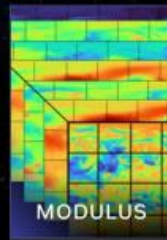


AI PLATFORM

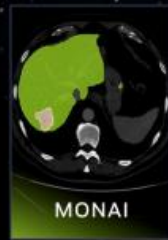
THE 2024 MAD (MACHINE LEARNING, ARTIFICIAL INTELLIGENCE & DATA) LANDSCAPE



APPLICATION
FRAMEWORKS



MODULUS



MONAI



MAXINE



NEMO



AVATAR



DRIVE



ISAAC



METROPOLIS

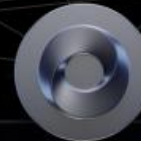


HOLOSCAN

PLATFORM

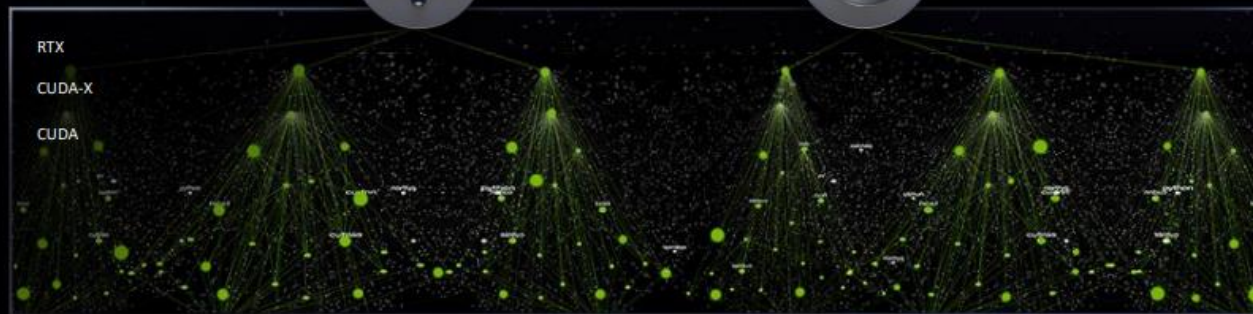


NVIDIA
AI

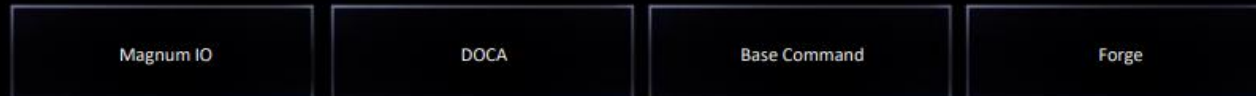


NVIDIA OMNIVERSE

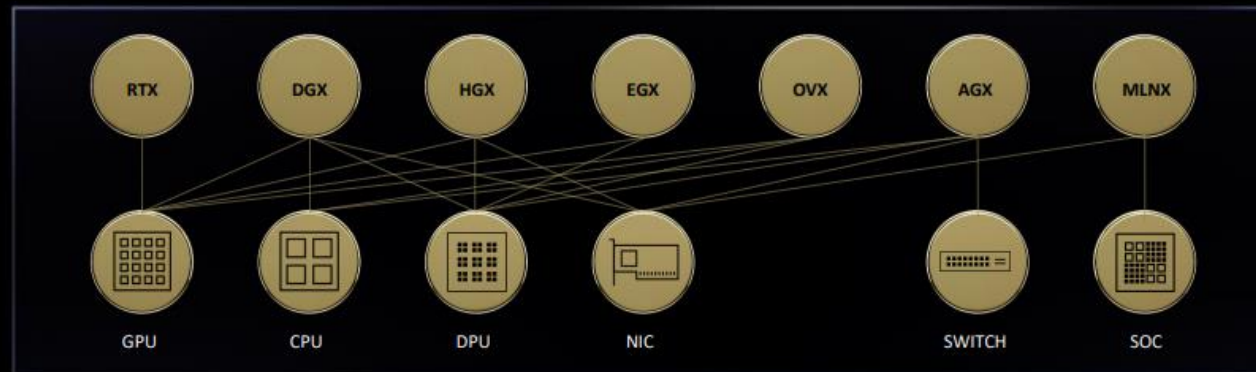
ACCELERATION
LIBRARIES



SYSTEM
SOFTWARE



HARDWARE

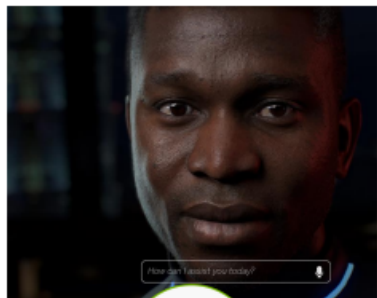


NVIDIA Blueprints

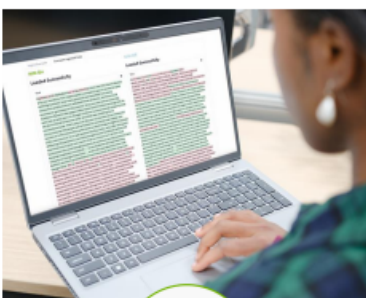
Available on build.nvidia.com



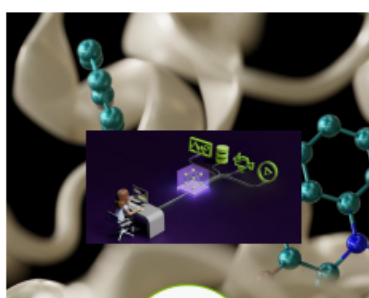
Digital Humans
for Customer Service



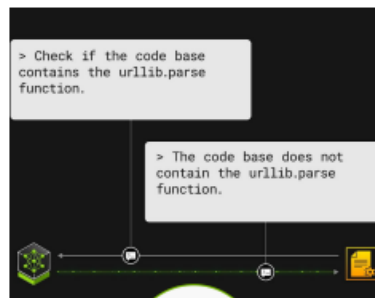
Multimodal PDF Data
Extraction for Enterprise RAG



Generative Virtual Screening
for Drug Discovery



Vulnerability Analysis
for Container Security



AI Virtual Assistants
for Customer Service



Visual AI Agent for Video
Search and Summarization



Reference Application



Sample Data



Reference Code



Architecture



Customization Tools



Orchestration Tools

Typical NVIDIA-Certified Systems



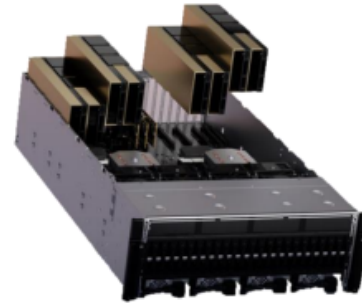
Large Model Training
& Inference

**HGX H100 8-GPU
(or 4-GPU)**



Inference, Visual Computing
(Omniverse)

NVIDIA OVX L40S



LLM Inference
Up to 175B Parameters

NVIDIA H100 NVL



2x Faster Inference
than HGX H100

NVIDIA HGX H200

RTX PRO 6000 Blackwell Server Edition

The Most Powerful Blackwell Data Center Platform for AI and Visual Computing

Breakthrough Multimodal AI Inference

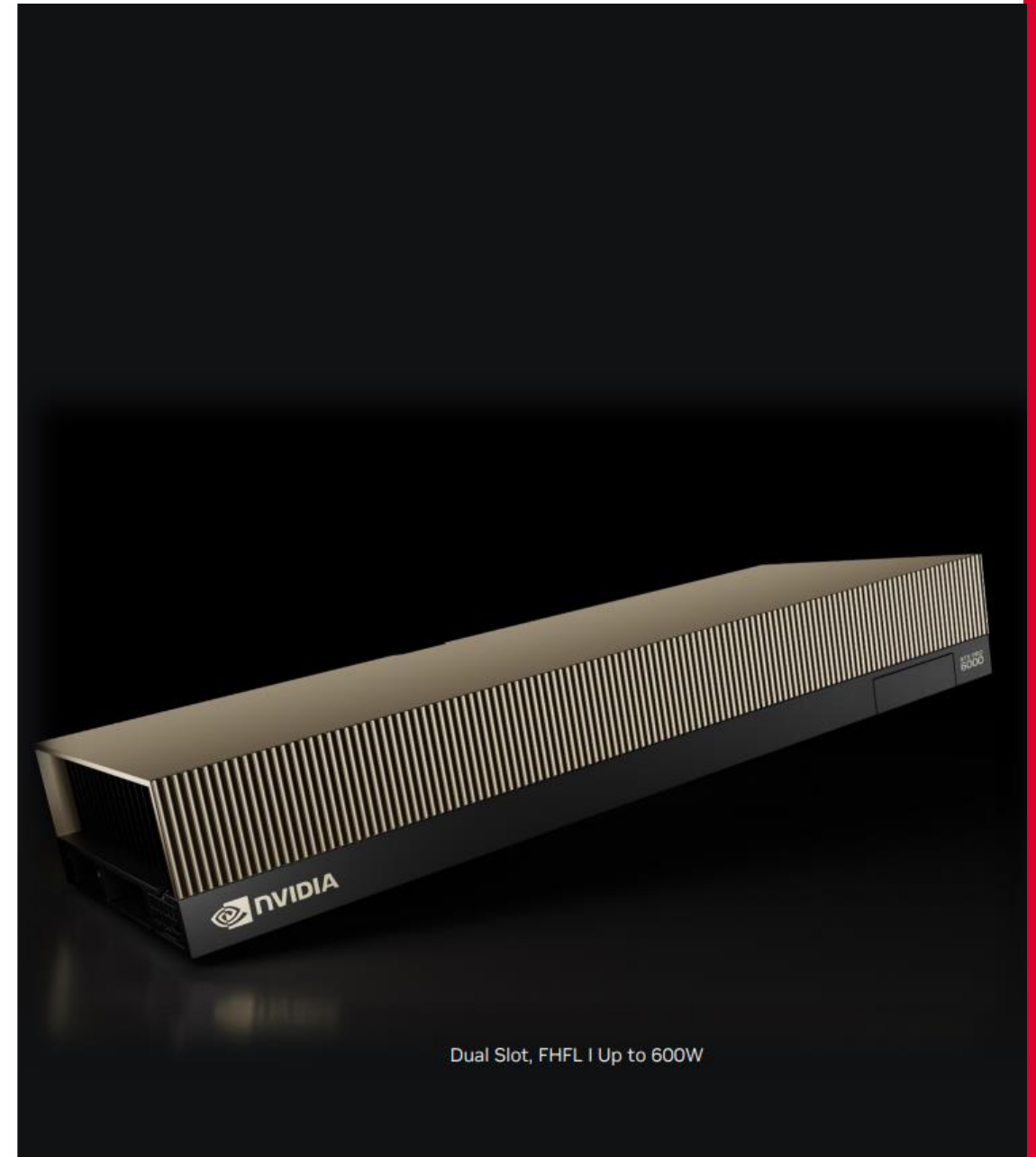
- 5th-Gen Tensor, 2nd-Gen Transformer Engine, FP4,
- Full Media Pipeline: 4 NVENC/ NVDEC/ NVJPEG

Powerful Graphics and Visual Computing

- 4th-Gen RTX, Neural Shaders, DLSS 4

Data Center Ready

- 96GB GDDR7, 1.6 TB/s Memory BW, 128MB L2 Cache
- Multi-Instance GPU (MIG), TEE Confidential Compute



Dual Slot, FHFL | Up to 600W

Blackwell Ultra | Blackwell for Every AI Use Case

Delivering the Highest Performance for AI Factories

Blackwell Ultra



GB300 NVL72

Maximum Performance &
Lowest TCO for AI
Reasoning
Liquid-Cooled



HGX B300 NVL16

Best Performance & TCO
for Air-cooled HGX

Blackwell



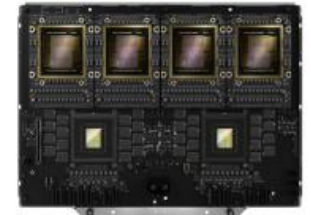
GB200 NVL72

Compute for at Scale
AI Training & Inference
Liquid-Cooled



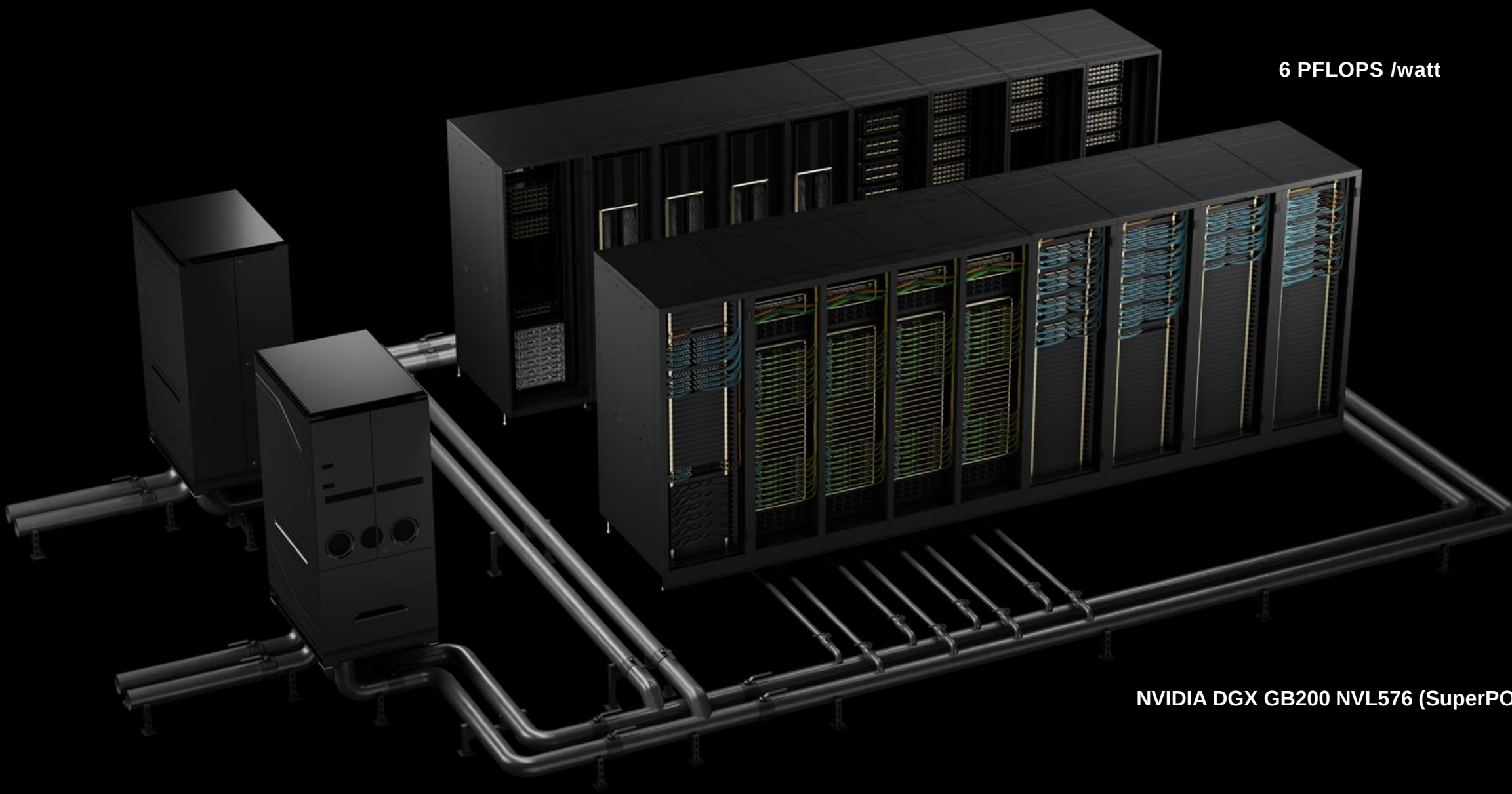
HGX B200

Optimized
Performance & TCO
for Air-cooled HGX



GB200 NVL4

Best HPC + AI
Performance in a single
server

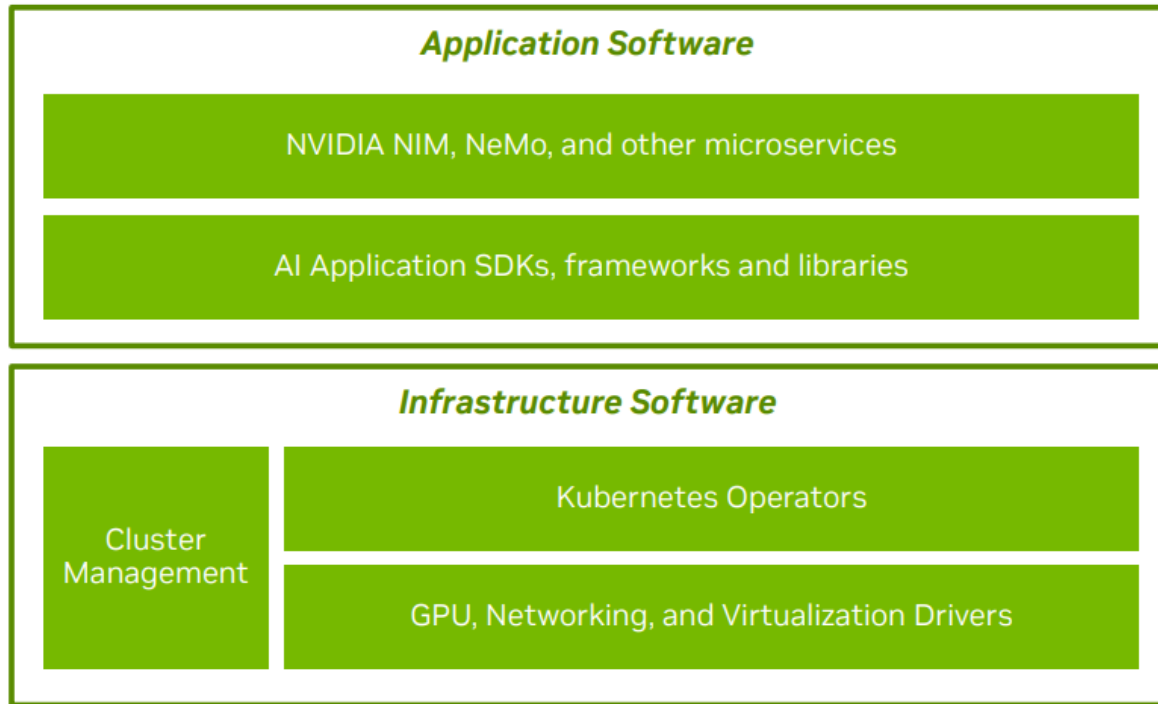


6 PFLOPS /watt

NVIDIA DGX GB200 NVL576 (SuperPOD)

NVIDIA AI Enterprise

What's Inside?



Cloud



Data Center



Edge



Workstation

Application Software

Optimized NIM and NeMo microservices enhance model performance and speed time to deployment for generative AI

SDKs, Frameworks, and libraries across many domains, including speech AI, route optimization, cybersecurity, and vision AI

Infrastructure Software

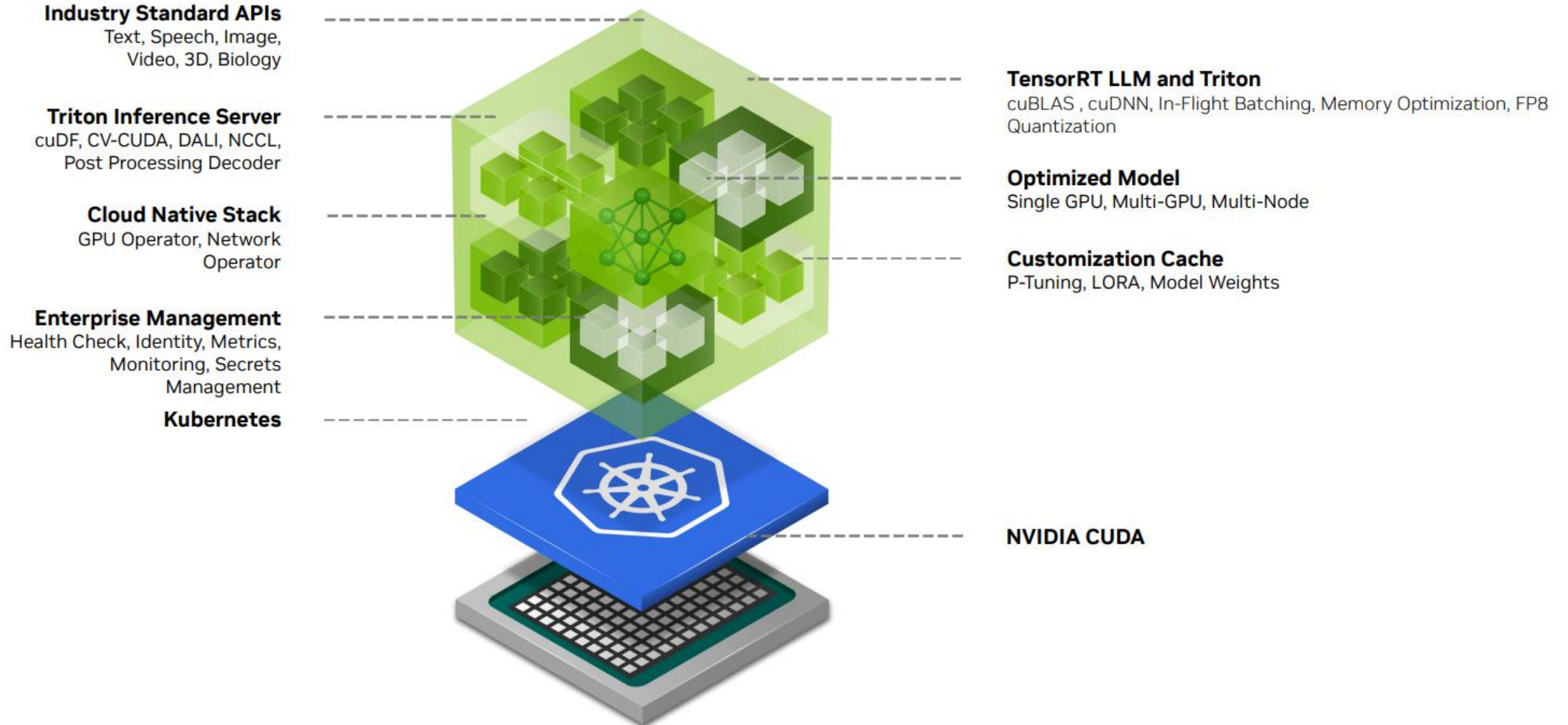
Kubernetes operators for managing GPU and networking in containers and the lifecycle of microservices and AI pipelines

Cluster Management software to provision and monitor servers at scale

Drivers to optimize utilization of NVIDIA GPUs and Networking in bare metal and virtualized environments.

NVIDIA NIM

Production Inference for Enterprise Generative AI



NVIDIA & Microsoft: Architecture Possibilities

Partnership: Platforms and toolset choices



Developer Productivity



Visual Studio



GitHub Copilot



Copilot Studio



Azure AI Foundry SDK

AI Platform



Azure AI Foundry



Azure OpenAI Service



Azure AI Services



Azure AI Content Safety



Model as a Service



Azure AI Search



Azure Machine Learning



Azure AI Agent Service



NVIDIA AI Foundry



NVIDIA NIM



NVIDIA DGX Cloud



NVIDIA AI Enterprise

1800+ Model Catalog
including NVIDIA models

App Platform



Azure Kubernetes Service



Azure Container Apps



Azure API Management



Azure Functions



Azure App Service



NVIDIA Omniverse APIs

Data Platform



Azure SQL Database



Azure Cosmos DB



OSS Databases

AI Infrastructure



Azure AI-optimized
Virtual Machines



Azure HPC



Azure Confidential
Computing



NVIDIA Blackwell
Platform GB200, B200



NVIDIA GPUs: H200,
H100, A100, A10

Safety, Governance, AI Security, AI-Ready Infrastructure

Better Together



Azure AI Foundry

&



NVIDIA

CANCOM

CANCOM

WIE STARTET MAN AM BESTEN?

Wie kann ich KI strategisch nutzen und operativ realisieren, ohne technische Hürden?

AI gezielt
verstehen u.
nutzen für echten
Business Value

AI Strategie

AI Roadmap

AI Konzeption

Von der Idee
zur
einsatzbereiten
Lösung

Use-Case Umsetzung

Individuelle AI Projekte

CANCOM AI Produkte

Technologische
Basis für
nachhaltige
Skalierung

Cloud AI-Plattform

On-Premises AI-Plattform

Komplexität
meistern.
Infrastruktur
optimieren.

Technisches Consulting

CANCOM



**DIGITAL
MAKERS**

LET'S TALK!





Modern AI Datacenter *Summit*

München Hoch5
24. Juni 2025

CANCOM

