



Scaling Financial Technology Applications with agentic, agnostic AI Platforms

Dr. Jochen Papenbrock

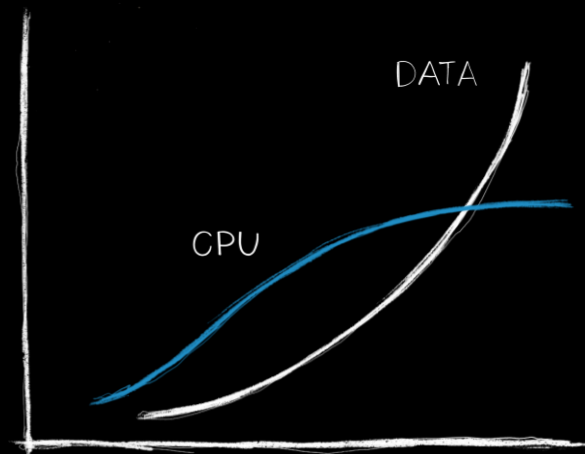
EMEA Head of Financial Technology

E-Mail: jpapenbrock@nvidia.com

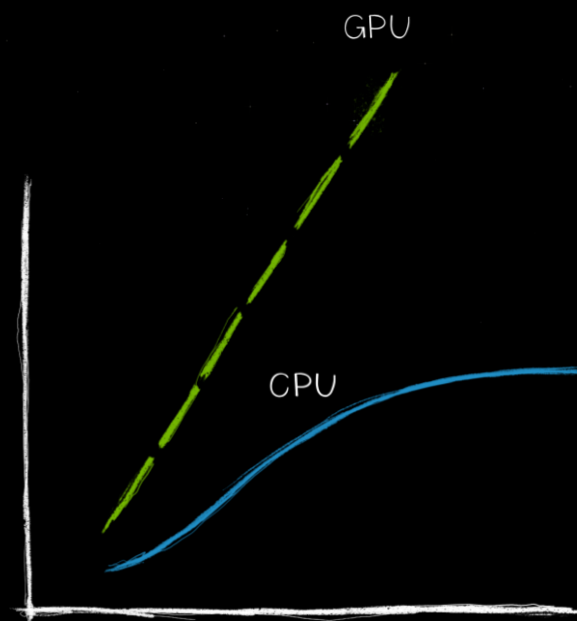




Industrial Revolution through Accelerated Computing



CPU SCALING SLOWS
... AND
COMPUTE DEMAND GROWS
EXPONENTIALLY



GPU-ACCELERATED COMPUTING
2006
CUDA

ACCELERATE EVERY APPLICATION

CPU



1t

100t



CPU



+

GPU



1t



= ~100X SPEED-UP

~3X POWER

~1.5X COST

60X PERF / \$ OR 98% SAVINGS

30X PERF / W OR 97% SAVINGS

A NEW INDUSTRIAL REVOLUTION

AI FACTORY

\$100T



TOKENS
"UNITS OF INTELLIGENCE"



SCIENCE



ENTERPRISE



HEALTHCARE



MANUFACTURING



FINANCIAL SERVICES

Accelerated Full-Stack Computing Platform

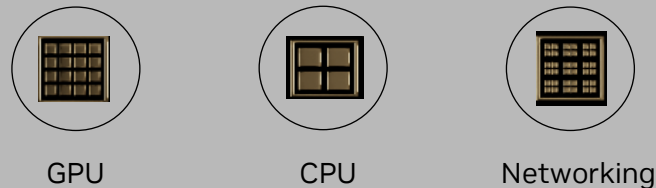
NVIDIA Application Frameworks



NVIDIA Platform Software



NVIDIA Accelerated Computing Infrastructure





NVIDIA Today

NVIDIA is one of the largest companies in the world — market cap (August 21 ~\$4 Trillion)

- 3 Chips — CPU, GPU, DPU
- NVIDIA Networking
- 30+ Million CUDA Downloads
- 5 + Million Developers
- 3,000+ Applications Accelerated
- 700 + SDKs & AI Models

“The announcements at GTC, which targets NVIDIA’s community of over four million developers, come in the context of NVIDIA’s continued AI dominance, particularly in the latest era of generative AI.....” — Dean Takahashi, [VentureBeat](#)





The NVIDIA AI Platform for GenAI / Agentic AI

Enterprises Face Challenges Experimenting With Generative AI

Organizations must choose between ease of use and control

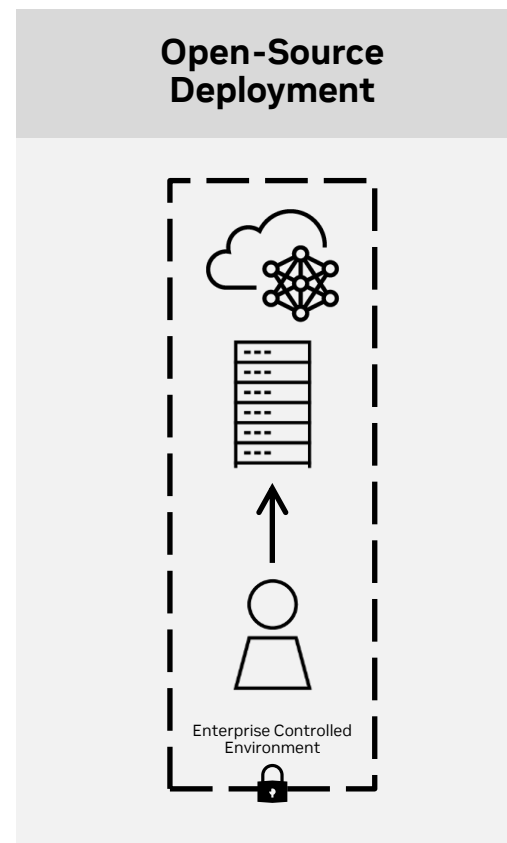
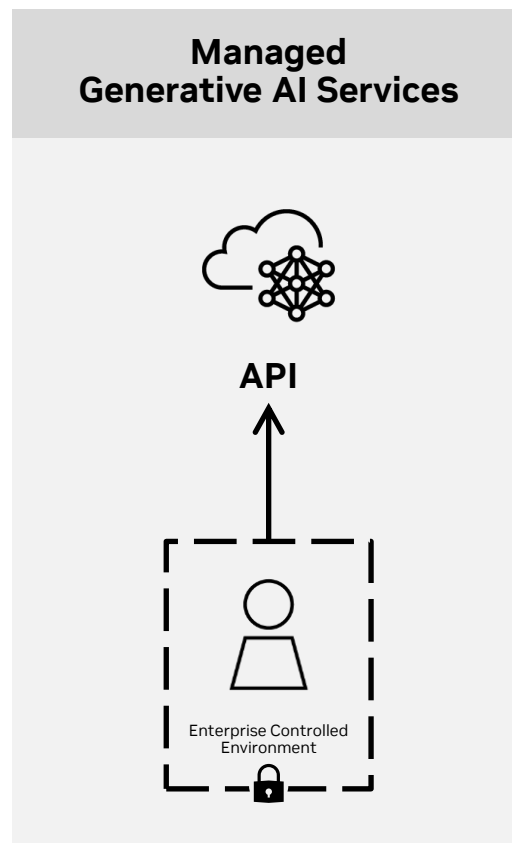
Easy to use
APIs for development

Fast path
to getting started with AI

Infrastructure limited
to managed environment

Data and prompts are
shared externally

Limited control
for overall generative AI strategy



Run anywhere
across data center and cloud

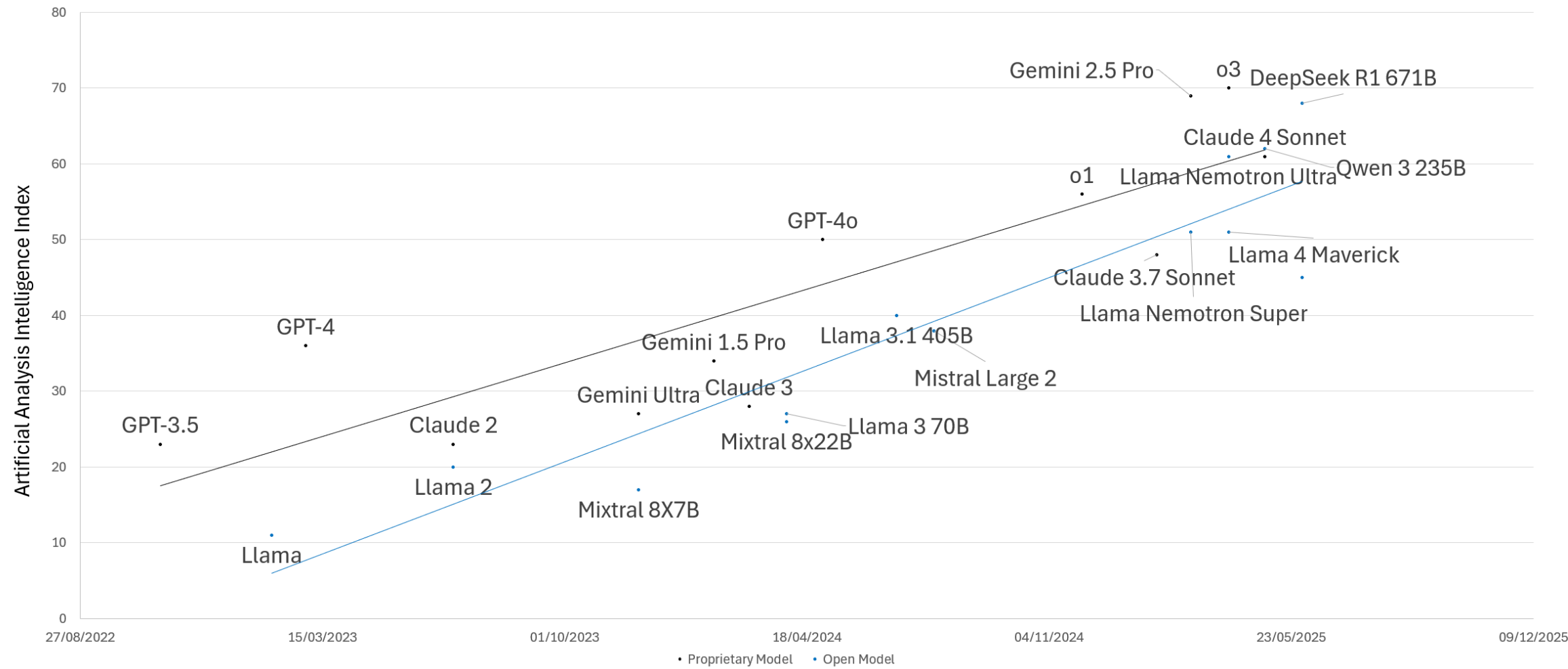
Securely manage
data in self-hosted environment

Tuning required
for different infrastructure

Custom code
for APIs and fine-tuned models

Ongoing maintenance
and updates

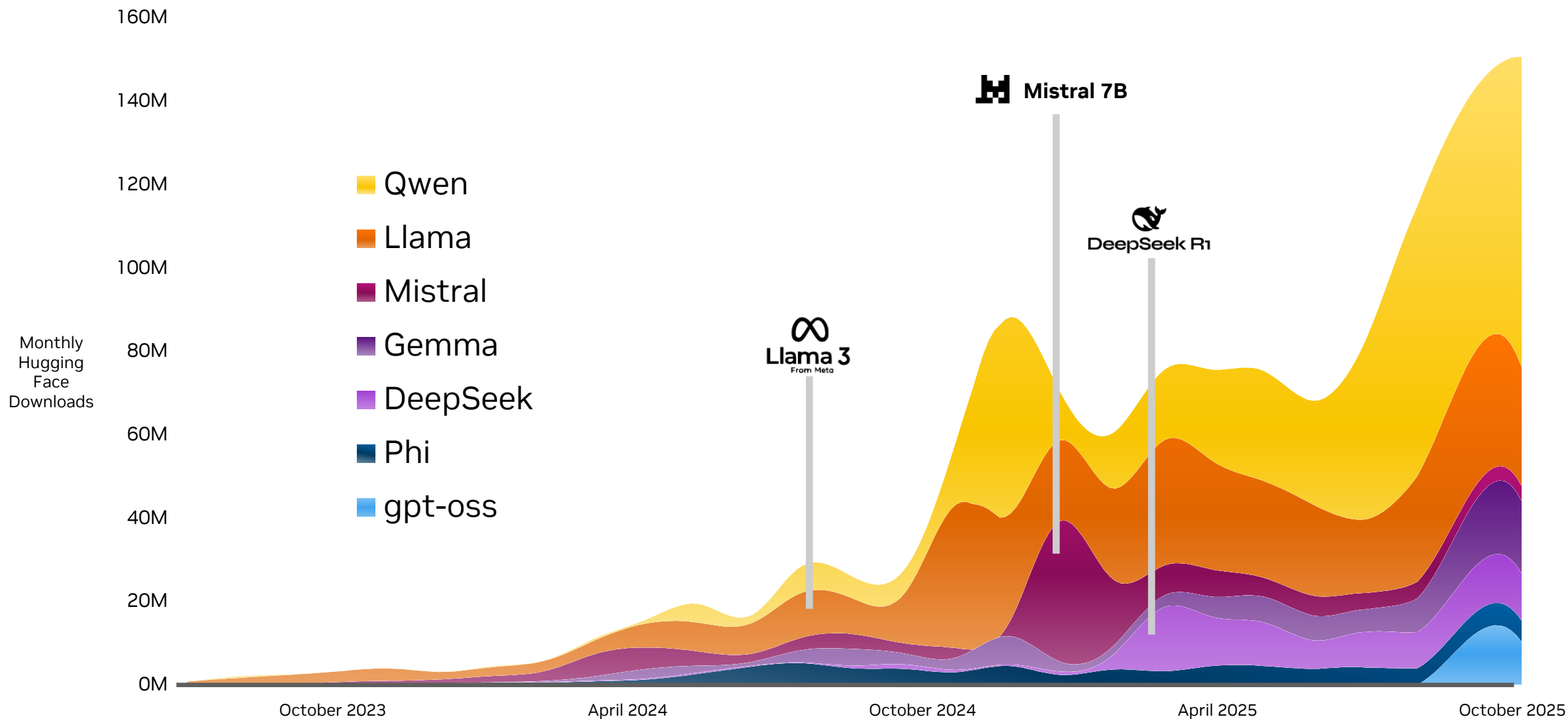
Open-Source AI Models Are Unlocking Enterprise AI Agents



Intelligence Index: MMLU-Pro, GPQA Diamond, Humanity's Last Exam, LiveCodeBench, SciCode, AIME, MATH-500.
Source: Artificial Analysis



Open-Source AI Models Are Unlocking Enterprise AI Agents



Two Ways to Build An AI Platform

Build from Open Source

Deploy on Current
NVIDIA GPU Architecture

Self Service Support

Do It Yourself (DIY)

Buy Secure, Scalable, & Reliable Platform

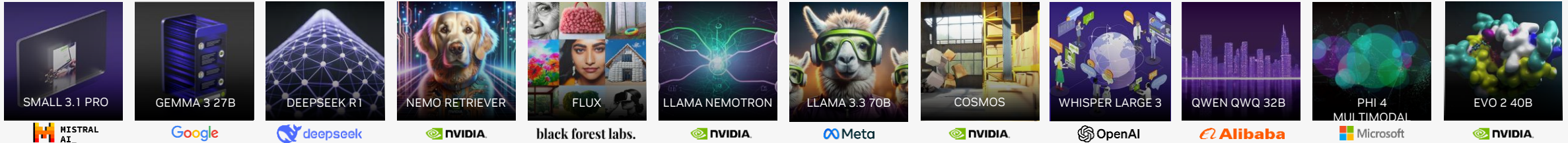
Deploy on Past, Current, & Future
NVIDIA GPU Architecture

NVIDIA Enterprise Support

NVIDIA AI Software

NVIDIA NIM Streamlines the Path to Production

170+ NIM, 40+ model types, 40+ partners driving AI model innovation



NVIDIA NIM



Speed time to market with prebuilt, continuously maintained microservices

Boost accuracy by tuning custom models from proprietary data sources

Deploy anywhere with security and control of AI applications and data

Empower developers with the latest AI models, standard APIs and enterprise tools

Deploy in production with API stability, security patching, and enterprise support

Optimized for greatest throughput to maximize TCO



Cloud



Data Center



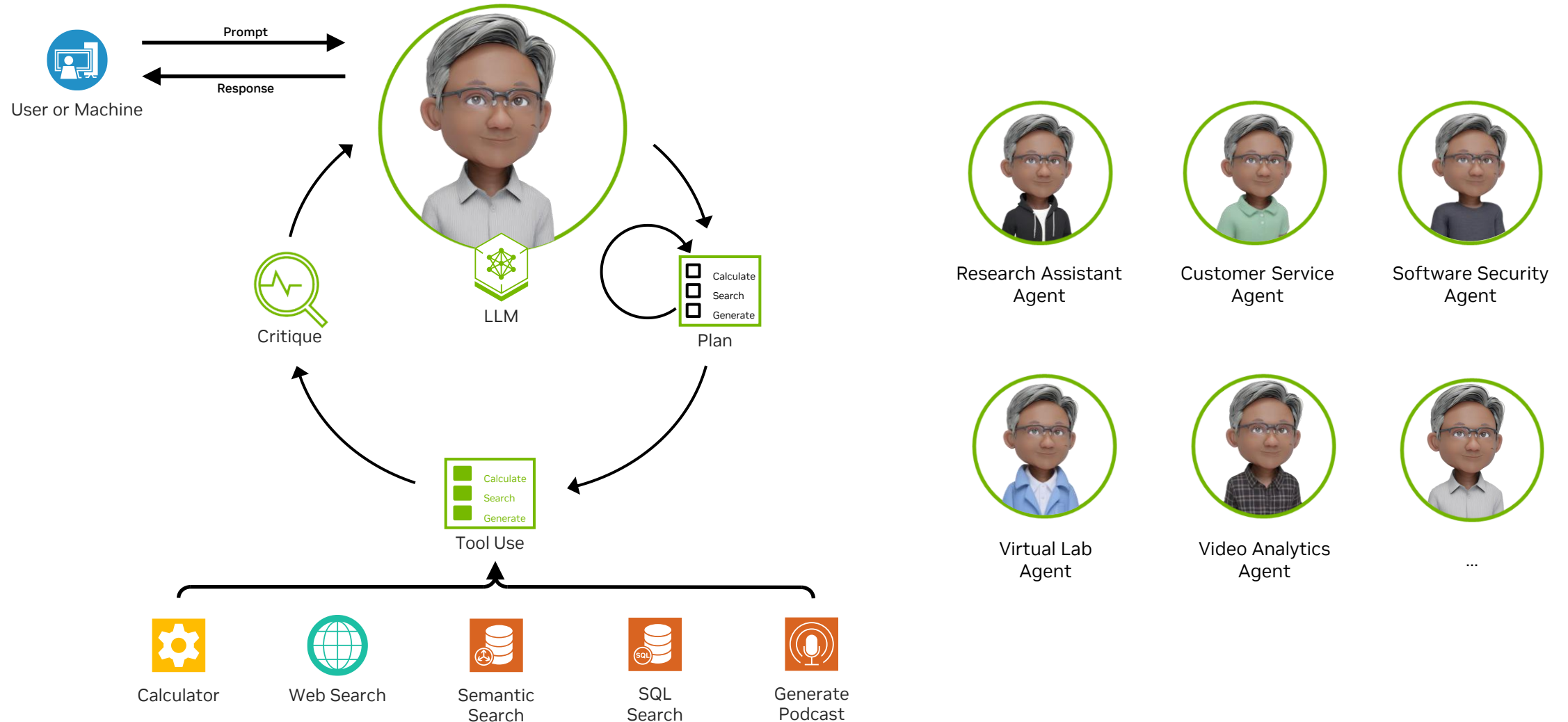
Edge



Workstation

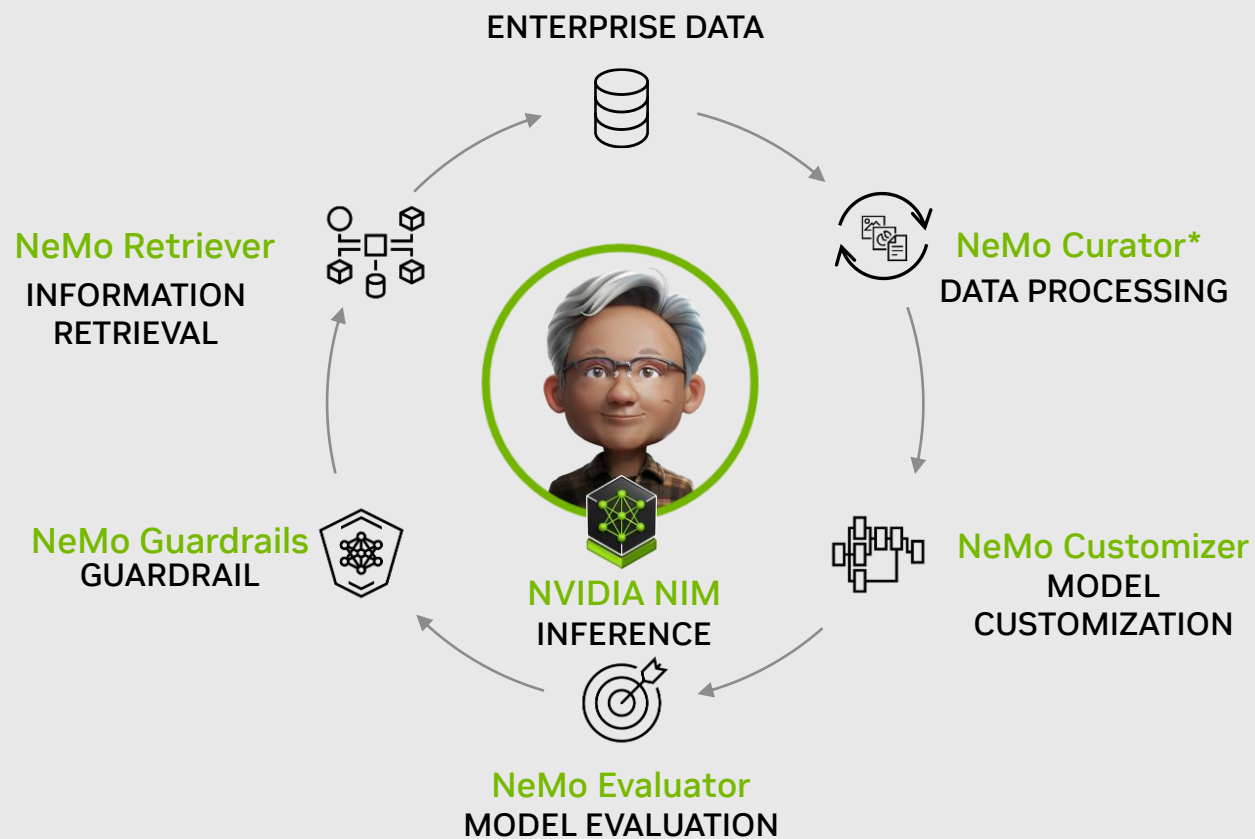
AI Agents: The New Digital Employees

They help you get more done but require effective on-boarding



NVIDIA NeMo Microservices

Modular End-to-End Platform to easily build Data Flywheels



Easy Setup

Modular microservices, deployable with standard APIs



Broad Ecosystem Support

Integrated in popular open frameworks and AI software platforms



Enterprise-grade

Secure, stable, and supported software



Run Anywhere

Provides higher security, privacy, and flexibility

NVIDIA Nemotron Models Available Today

Reasoning / Agentic

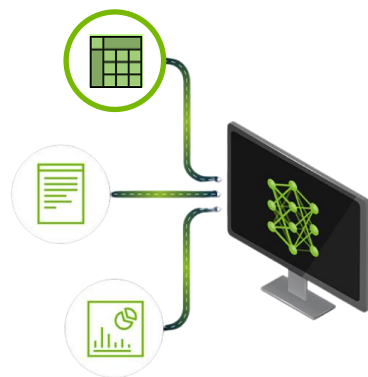


Nemotron Nano 2
High efficiency SLM

Llama Nemotron Super
Optimal accuracy-efficiency LLM

Llama Nemotron Ultra
Highest accuracy powered by world knowledge

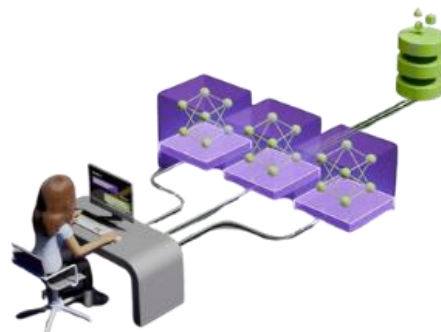
Vision Language



Nemotron Nano 2 VL
Video understanding and Image reasoning

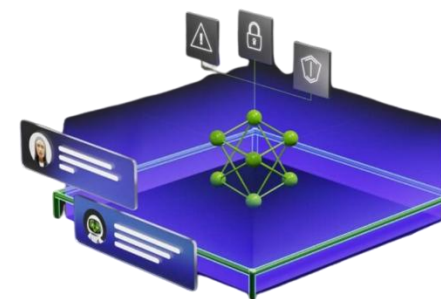
Nemotron Parse 1.1
Document extraction

Information Retrieval



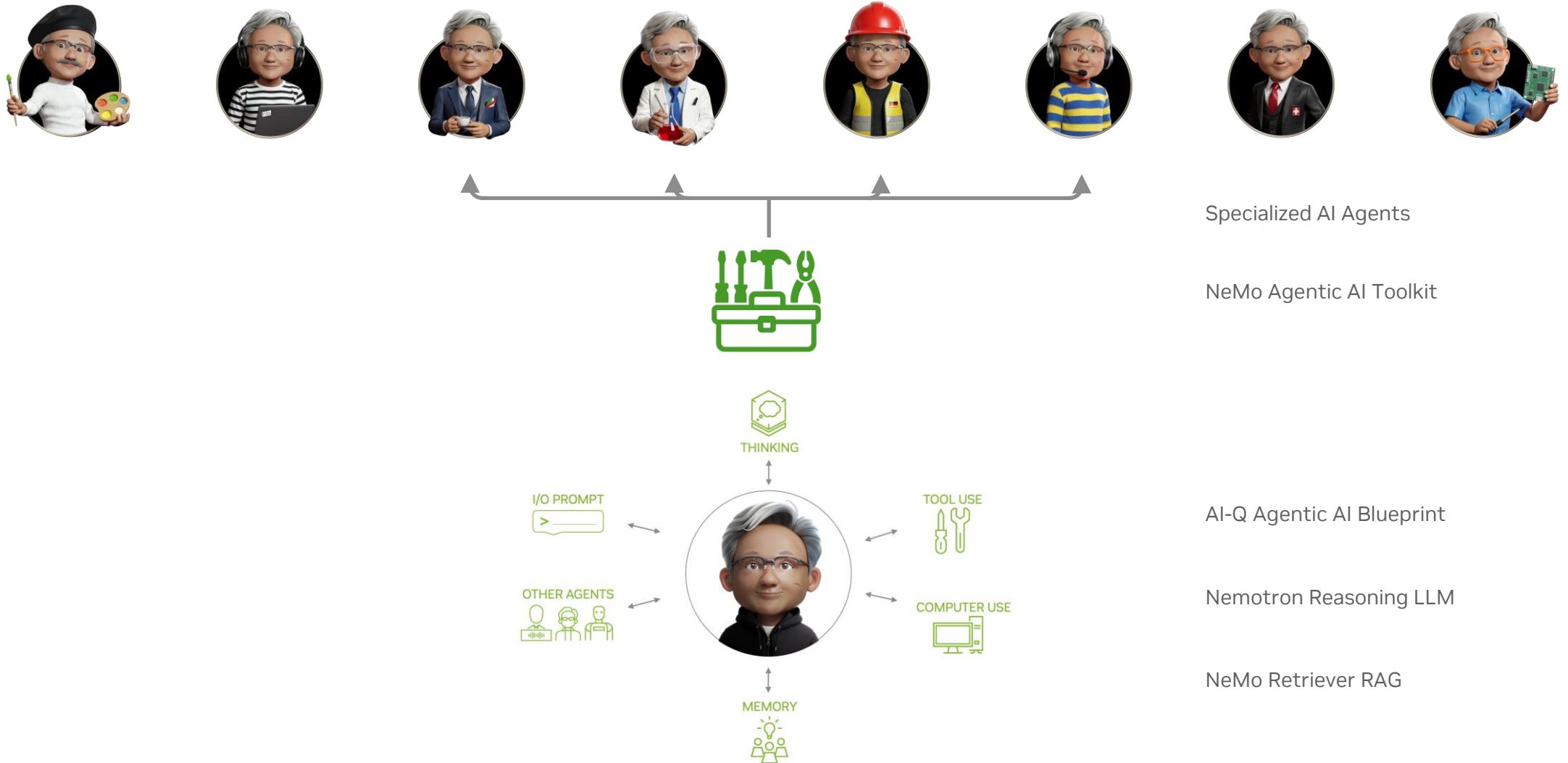
Nemotron RAG
Extraction, embedding, and reranking

Content Safety



Llama Nemotron Safety Guard
Multilingual content safety

NVIDIA Enterprise AI Agent Platform



Try the Models on build.nvidia.com

← ↻ 🔒 https://build.nvidia.com/models

 Explore **Models** Blueprints GPUs Docs

🔍 Search

Models

Deploy and scale models on your GPU infrastructure of choice with NVIDIA NIM inference microservices

Filter by text Sort By: Most Recent Publisher Use Case NIM Type

Featured Models



qwen
qwen3-next-80b-a3b-thinking
80B parameter AI model with hybrid reasoning, MoE architecture, support for 119 languages.
reasoning text-to-text



nvidia
parakeet-ctc-0.6b-zh-cn
Record-setting accuracy and performance for Mandarin English transcriptions.
asr mandarin nvidia nim +2



nvidia
parakeet-ctc-0.6b-es
Accurate and optimized Spanish English transcriptions with punctuation and word...
asr nvidia nim spanish +2



microsoft
TRELLIS
MSFT TRELLIS is a 3D AI model that generates h quality 3D assets from text or image inputs.
run-on-rtx image-to-3d text-to-3d



qwen
qwen3-next-80b-a3b-thinking
80B parameter AI model with hybrid reasoning, MoE architecture, support for 119 languages.
reasoning chat text-to-text qwen



nvidia
parakeet-ctc-0.6b-zh-cn
Record-setting accuracy and performance for Mandarin English transcriptions.
asr streaming speech-to-text +3



nvidia
parakeet-ctc-0.6b-es
Accurate and optimized Spanish English transcriptions with punctuation and word...
asr streaming speech-to-text +3



microsoft
TRELLIS
MSFT TRELLIS is a 3D AI model that generates high-quality 3D assets from text or image inputs.
text-to-3d run-on-rtx +2



nvidia
parakeet-ctc-0.6b-vi
Accurate and optimized Vietnamese-English transcriptions with punctuation and word...
asr streaming speech-to-text +3



bytedance
seed-oss-36b-instruct
ByteDance open-source LLM with long-context, reasoning, and agentic intelligence.
thinking budget chat reasoning +2



microsoft
TRELLIS
MSFT TRELLIS is a 3D AI model that generates high-quality 3D assets from text or image inputs.
text-to-3d run-on-rtx +2



qwen
qwen3-coder-480b-a35b-instruct
Excels in agentic coding and browser use and supports 256K context, delivering top results.
agentic coding moe long context +3



deepseek-ai
deepseek-v3.1
DeepSeek V3.1 Instruct is a hybrid AI model with fast reasoning, 128K context, and strong tool use.
reasoning chat text-to-text +1



nvidia
nvidia-nemotron-nano-9b-v2
High-efficiency LLM with hybrid Transformer-Mamba design, excelling in reasoning...
mamba thinking budget slm +6



stabilityai
stable-diffusion-3.5-large
Stable Diffusion 3.5 is a popular text-to-image generation model
image generation text-to-image +1



black-forest-labs
FLUX.1-Kontext-dev
FLUX.1 Kontext is a multimodal model that enables in-context image generation and editing.
image generation text-to-image +2



nvidia
cosmos-reason1-7b
Reasoning vision language model (VLM) for physical AI and robotics.
video understanding +9



Try the Blueprints on build.nvidia.com

←

↺

↻

🔒

https://build.nvidia.com/blueprints

NVIDIA

Explore

Models

Blueprints

GPUs

Docs

🔍 Search

Build Your AI Application with Blueprints

Get started with workflows and code samples to build AI applications from the ground up

Filter by text

Sort By: Most Recent

Publisher

Blueprint Type

Launchable

Enterprise Blueprints

Customizable reference workflows with best practices for enterprise performance and scale

ENTERPRISE



nvidia

Refine AI Agents through Contin...

Build a data flywheel, with NVIDIA NeMo microservices, that continuously optimizes AI...

blueprint data flywheel nim +4

ENTERPRISE



nvidia

Build Digital Twins for AI Factory ...

Design, test, and optimize a new generation of intelligence manufacturing data centers usin...

ai factory blueprint +4

ENTERPRISE



nvidia

Build an AI Agent for Enterprise R...

Build a custom deep researcher powered by state-of-the-art models that continuously...

blueprint llama nemotron +7

ENTERPRISE



nvidia

AI Weather Analytics with Earth-2

Develop AI powered weather analysis and forecasting application visualizing multi...

ai weather prediction blueprint +5

ENTERPRISE



nvidia

Synthetic Manipulation Motion G...

Generate exponentially large amounts of synthetic motion trajectories for robot...

blueprint humanoids +10

Developer Examples

Interactive tutorials and example patterns to bootstrap AI development



nvidia

3D Object Generation

Transform your scene idea into ready-to-use 3D assets using Llama 3.1 8B, NV SANA, and Microsof...

blueprint run-on-rtx nvidia ai nvidia

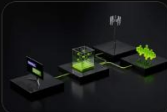


nvidia

Retail Shopping Assistant

Elevate Shopping Experiences Online and In Stores.

blueprint nemo retriever nim +4

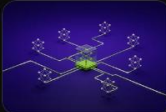


nvidia

Streaming Data to RAG

Sensor-captured radio enables real-time awareness, AI-driven analytics for actionable, searchable...

blueprint nim riva launchable +4

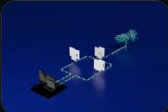


nvidia

Multi-LLM NIM

Use the multi-LLM compatible NIM container to deploy a broad range of LLMs from Hugging Face.

blueprint nvidia ai nvidia



nvidia

Biomedical AI-Q Research Agent Blu...

Build advanced AI agents within the biomedical domain using the AI-Q Blueprint and the BioNeMo...

launchable agent blueprint +4



wandb

AI Observability for Data Flywheel!

Streamline evaluation, monitoring, and optimization of AI data flywheel with Weights & Biases.

launchable ai agents +7



iguazio

AI Orchestration for Data Flywheel

Orchestrate AI agents for data flywheel with MLRun and NVIDIA NeMo microservices.

orchestration launchable +6



nvidia

Safety for Agentic AI

Improve safety, security, and privacy of AI systems at build, deploy and run stages.

security launchable blueprint +6

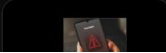


nvidia

AI Agent for Telecom Network Confi...

Automate and optimize the configuration of radio access network (RAN) parameters using agentic AI...

nim launchable blueprint +4



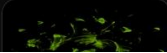
nvidia

Financial Fraud Detection



nvidia

3D Guided Generative AI

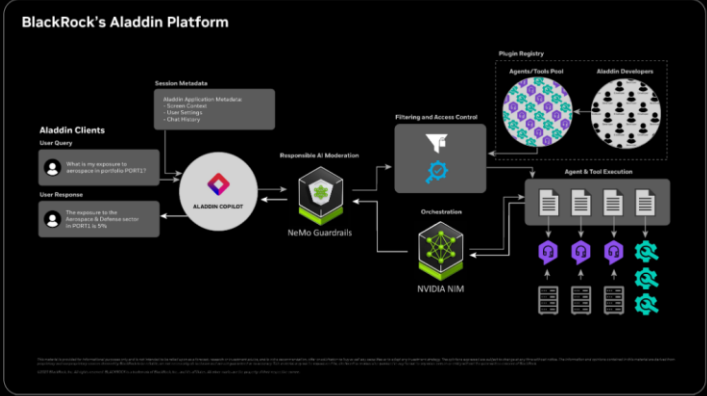


nvidia

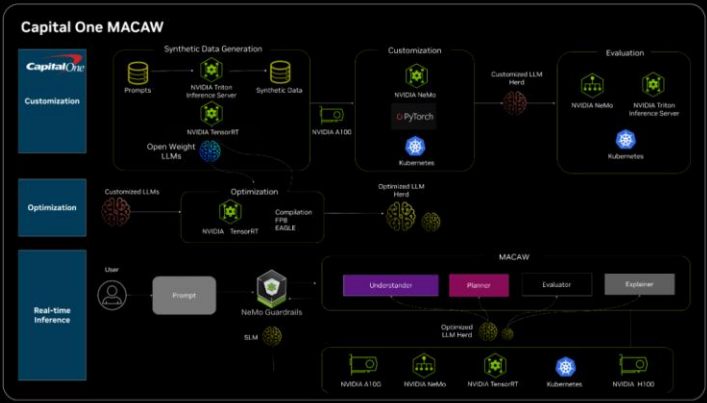
Single Cell Analysis



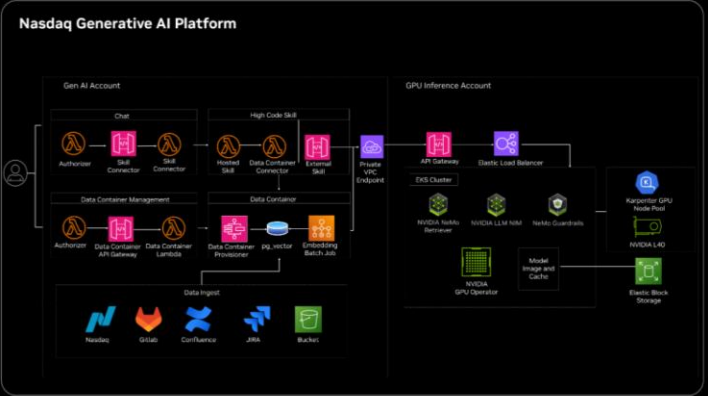
BlackRock



Capital One



Nasdaq



Accelerating Financial Insights with Agentic AI

RBC anticipates accelerating its \$5 billion top-line revenue growth goal by 1.5 years through this initiative

Report generation time was dramatically reduced from 40 hours to just 15 minutes, saving RBC 18 months of work

RBC analysts can now potentially cover 50 companies instead of 10, significantly increasing productivity



AI Factories in Banking



Financial Services Session



Jean-Michel Garcia
BNP Paribas

Julia Koch
Finanz Informatik

Malcolm deMayo
NVIDIA

Building Sovereign AI Factories
in Financial Services

Finanz Informatik treibt in Zusammenarbeit mit NVIDIA die Weiterentwicklung ihrer souveränen KI-Plattform für die Sparkassen-Finanzgruppe voran

European Bank Builds AI Factory to Develop and Scale Financial Services Applications

Across Europe, banks are building regional AI factories to enable the deployment of AI models for customer service, fraud detection, risk modeling and the automation of regulatory compliance.

In Germany, Finanz Informatik, the digital technology provider of the Savings Banks Finance Group, is scaling its on-premises AI factory and using NVIDIA AI Enterprise software for applications including an AI assistant to help its employees automate routine tasks and efficiently process the institution's banking data.



BNP PARIBAS



Financial Services Session



Jean-Michel Garcia
BNP Paribas

Emmanuel Salazard
BNP Paribas

Building an Optimized AI Platform
for the Banking Industry

INNOVATION & TECHNOLOGY

PRESS RELEASE

BNP Paribas provides its businesses with an LLM as a Service platform to accelerate the industrialization of generative AI use cases.

« With LLM as a Service, we are creating a common technological foundation that allows entities to focus on business use cases, while relying on a shared and secure infrastructure. It is a common lever to innovate and improve the performance of certain applications. » Marc Camus, Group Chief Information Officer, BNP Paribas

500M€
Additional Value

By end 2025, the Group estimates that it will be able to generate **€500 million of additional value** per year, through the generation of new income, retention of customers and partners, cost optimization and optimized risk management.

[Home](#) > [News](#) > FCA allows firms to experiment with AI alongside NVIDIA

FCA allows firms to experiment with AI alongside NVIDIA

Press Releases | First published: 09/06/2025 | Last updated: 09/06/2025

The Financial Conduct Authority (FCA) will launch a Supercharged Sandbox to help firms experiment safely with AI to support innovation.

Through a new collaboration, announced today, firms will have the opportunity to experiment with AI using NVIDIA accelerated computing and NVIDIA AI Enterprise Software.

This Supercharged Sandbox will give firms access to better data, technical expertise and regulatory support to speed up innovation. It is open to any financial services firm looking to innovate and experiment with AI.

The sandbox will help firms who are in the discovery and experiment phase with AI. An existing AI Live Testing service helps those further along in development and ready to use AI. In its [new strategy](#), the FCA has committed to supporting economic growth by enabling innovation and harnessing technological advances like AI.

Jessica Rusu, the FCA's chief data, intelligence and information officer, said:

'This collaboration will help those that want to test AI ideas but who lack the capabilities to do so. We'll help firms harness AI to benefit our markets and consumers, while supporting economic growth.'

Dr Jochen Papenbrock, EMEA head of financial technology, NVIDIA added:

'AI is fundamentally reshaping the financial sector by automating processes, enhancing data analysis, and improving decision-making, which leads to greater efficiency, accuracy, and risk management across a wide range of financial activities.

'The FCA's Supercharged Sandbox provides firms with a secure environment to explore AI innovations using NVIDIA's full stack accelerated computing platform, supporting industry-wide growth and efficiency.'

Leveraging AI in Core Banking – scaling Innovation

...quick implementation of Use Cases – in less than 4 weeks

GenAI Agent

Key Features:

- Simplify support for routine questions
- Employ **AI model** for queries by translation **text-to-SQL**
- Use **KORDOBA** data integration

Personalised Portfolio Podcast

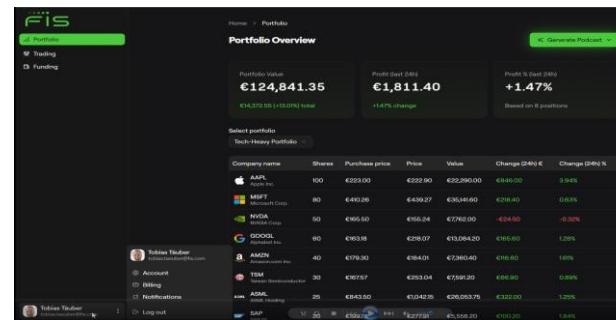
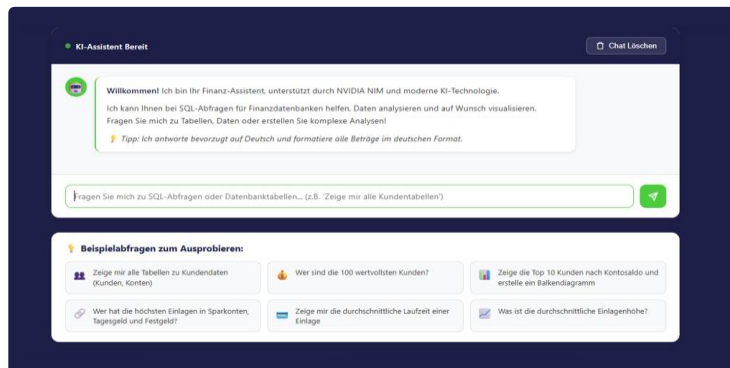
Key Features:

- Delivery of **personalized information**
- Use **Text-to-Speech** AI models
- Target insights based on **client portfolio** in combination with **external market updates**

Interactive Customer Communication

Key Features:

- Transition from eBanking chatbot to **AI Financial Assistant as proactive communication**
- Use AI-powered **offer cards** & deliver **personalized product recommendations**



The screenshot shows the 'Marketing Kampagnen' dashboard. It includes a search bar and a table of campaigns with columns for campaign name, type, status, and dates. The table lists various campaigns like 'Kampagne Produkt X' and 'Kampagne Produkt Y'.

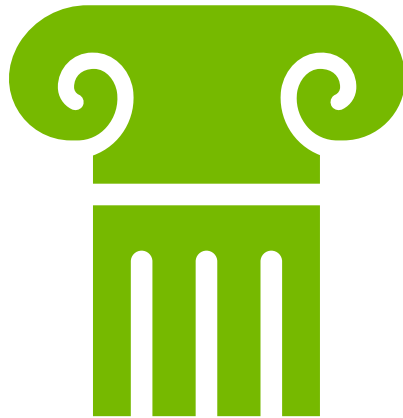
Technical Stack:

Infrastructure and PaaS-Services: Atos CloudPlatform & Atos Kubernetes
NVIDIA Stack: NIMs (e.g., QWEN for text-to-sql)

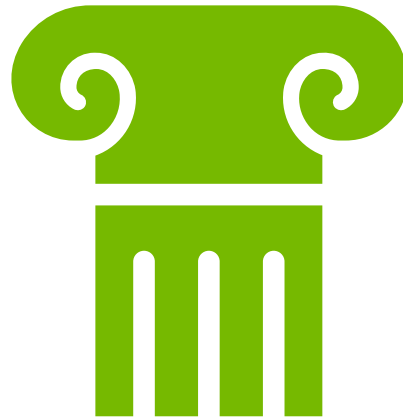
Pillars of Self-Enablement

You can become an expert on your own.

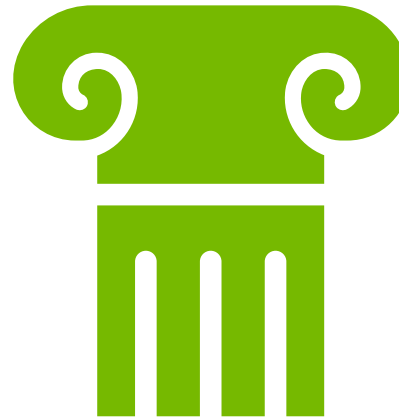
**Developer
Program**



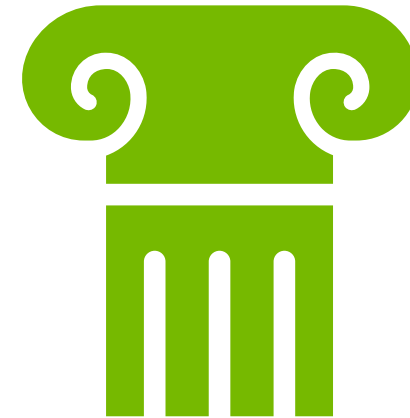
API Catalog
ai.nvidia.com



**Deep Learning
Institute (DLI)**



Launchpad



[NVIDIA FSI](#)

[Financial Service Developer Resources | NVIDIA Developer](#)

<https://resources.nvidia.com/en-us-financial-services-industry>

[NVIDIA On-Demand](#)

Contact:

jpapenbrock@nvidia.com

