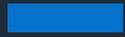
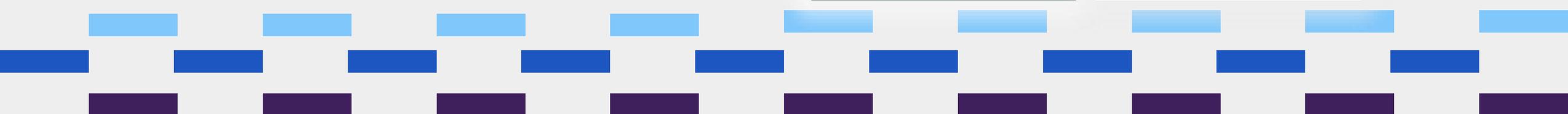
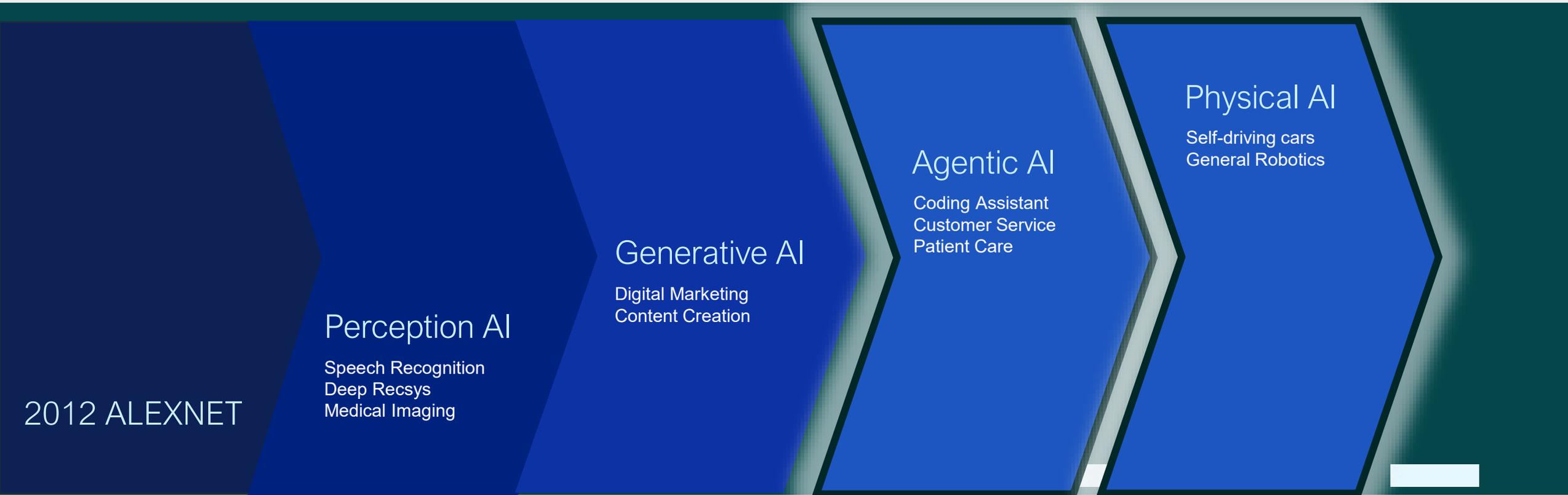


# Transforming AI Workflows with Dell AI Factory with NVIDIA



# The AI journey so far: It's evolving rapidly

Agentic AI is the next big thing in AI development



# Agentic AI is the next big thing in AI development



Agentic AI



Coding Assistant



Customer Service



Patient Care



Cybersecurity



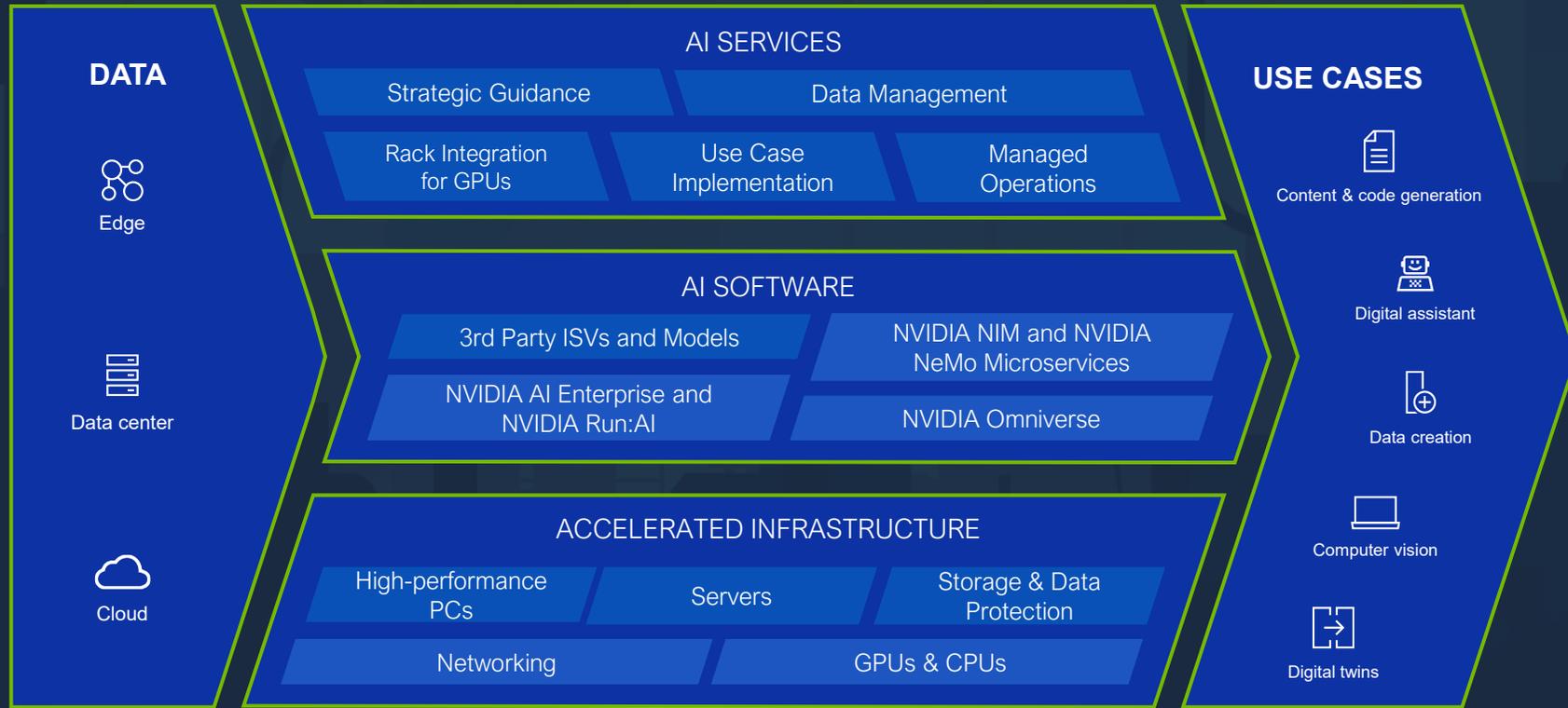
Supply Chain Management

# Accelerate the AI Journey

Fast forward to results with the next-generation Dell AI Factory with NVIDIA



AI  
-  
Generative AI  
-  
Agentic AI  
-  
Physical AI



Sustainability | Security

Simplify deployment and boost AI productivity to fast-track innovation

Adopt a comprehensive AI ecosystem to adapt to changing business needs

Simplify infrastructure management and free IT teams to focus on urgent priorities

Utilize AI Factories and streamlined workflows to accelerate time to value



# Client Solutions

# Why Dell Pro Max for your AI Workloads



## Exceptional performance

Engineered and configurable to power your AI tasks effortlessly



## Certified Reliability

100+ ISV applications tested for seamless performance



## Unmatched Security

The industry's most secure & manageable commercial AI PCs



## Scalability for your AI goals

Foundational infrastructure within the Dell AI Factory with NVIDIA

## Tailored for professionals with data-intensive workflows

Power Users & Collaborators

Creators & Designers

Specialists & AI Developers



# Right size for your use case and scale with Dell Pro Max

Purpose-built for AI Development

Dell Pro Max  
with GB10



Dell Pro Max  
with GB300



Consistent hardware &  
software environment  
within

Dell AI Factory  
with NVIDIA

Prototype & develop  
locally and scale from  
desktop to data center

Broader solutions for all your AI Needs



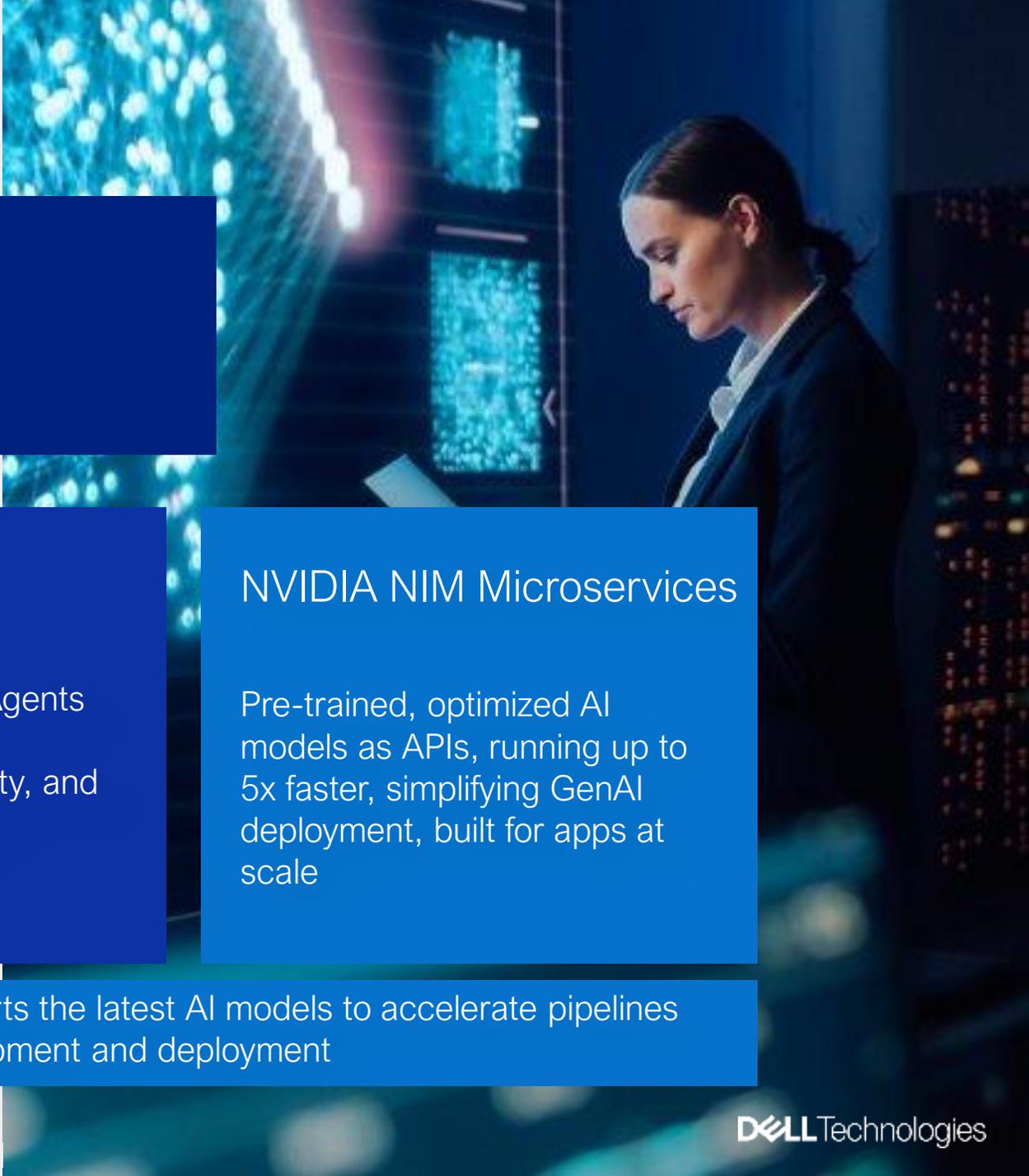
Dell Pro Max  
Laptops



Dell Pro Max  
Desktops

← NVIDIA AI Enterprise →  
One consistent experience with the software layer across the Dell Pro Max portfolio

# Building blocks of Agentic AI



It doesn't matter what you're building.

This powerful toolkit empowers you to create context-aware AI agents powered by your own data.

## NVIDIA AI Blueprints

Pre-trained, customizable reference AI workflows or “building blocks” to build AI-powered apps and agents for various use cases

## NVIDIA NeMo

Onboard and train AI Agents on your unique data, providing security, safety, and topical relevance with guardrails

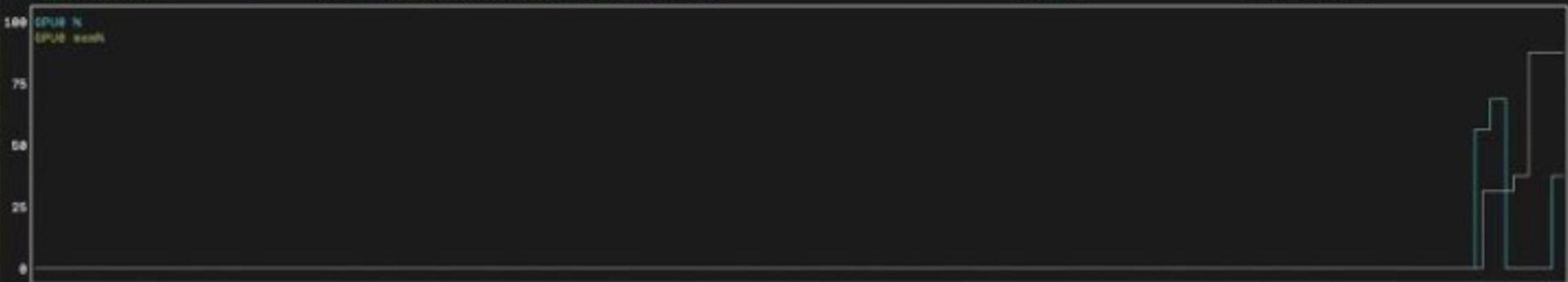
## NVIDIA NIM Microservices

Pre-trained, optimized AI models as APIs, running up to 5x faster, simplifying GenAI deployment, built for apps at scale

## NVIDIA AI Enterprise support

AI software platform that supports the latest AI models to accelerate pipelines and simplify AI workflow development and deployment

```
nvidia@nvidia-System-Product-Name: -- ssh nvidia@192.168.50.211 -- zsh
Device 0 [NVIDIA RTX 4000 Ada Generation] PCIe 00x 40 Ex RK: 300.0 KIB/s TK: 300.0 KIB/s Device 1 [NVIDIA RTX 4000 Ada Generation] PCIe 00x 40 Ex RK: 300.0 KIB/s TK: 300.0 KIB/s
GPU 2500MHz MEM 10001MHz TEMP 43°C FAN 30% POW 102 / 300 W GPU 2500MHz MEM 9601MHz TEMP 45°C FAN 30% POW 64 / 300 W
GPU[|||||] 37% MEM[|||||] 42.55001/47.98801 GPU[|||||] 0% MEM[|||||] 1.38401/44.98801
```



PID	USER	DEV	TYPE	GPU	GPU MEM	CPU	HOST MEM	Command
124738	nvidia	0	Compute	100%	42928MiB	87%	181%	4688MiB /opt/nix/lin/venv/bin/python3 sh bin/llm_sdk_entrypoints/gpu_server --serve-model-name meta/llama-3.1-8b-instruct --async-engine
184618	root	1	Compute	0%	588MiB	1%	0%	1133MiB /usr/bin/python3 /usr/local/bin/wvicorn main:app --host 0.0.0.0 --port 8807
2565	nvidia	1	Graphic	0%	181MiB	0%	0%	286MiB /usr/bin/gnome-shell
2424	nvidia	1	Graphic	0%	64MiB	0%	0%	96MiB /usr/lib/xorg/Xorg vt2 -displayfd 3 -auth /run/user/1000/gdm/Xauth
2424	nvidia	0	Graphic	0%	4MiB	0%	0%	96MiB /usr/lib/xorg/Xorg vt1 -displayfd 3 -auth /run/user/1000/gdm/Xauth

The NVIDIA AI Workbench host loads Llama 3.1-Instruct 8B into VRAM.

linux-worky pdf-to-podcast

local\_run Current branch Commit 2 file changes Publish Publish branch to Git server Open Frontend





# Infrastructure Solutions

# Expand with AI: Push past the limits of innovation

The next generation Dell AI Factory with NVIDIA enables AI adoption, faster

## COMPUTE STORAGE NETWORKING



### PowerEdge XE Servers

- ✓ Maximize AI acceleration
- ✓ NVIDIA Blackwell and RTX Pro 6000 GPUs



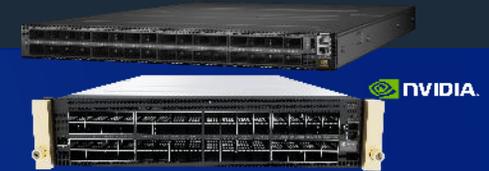
### Dell Pro Max

- ✓ Right sized power for developers
- ✓ NVIDIA RTX PRO 6000 Blackwell GPUs and Dell Pro Max with GB10 & GB300



### PowerScale & Dell AI Data Platform with NVIDIA

- ✓ Seamless structured and unstructured data access



### Dell PowerSwitch SN5600

- ✓ High-performance AI fabric

## NVIDIA AI SOFTWARE & DELL SERVICES



NVIDIA AI  
Enterprise



Reasoning  
model support



AI-Q  
blueprints



NeMo Agent  
Toolkit



AI Workload  
management

- ✓ Create AI agents and boost accuracy with reasoning models



Dell Managed  
Services

- ✓ Simplify AI lifecycle management

# Unlock the value of video

## Video search and summarization for unstructured video

### The opportunity

- GB's to TB's to PB's of stored video
- Moments, Actions, trends, patterns to be interpreted
- Unlock new insights and prompt existing video



### Frame analysis

Break down video into frames to capture key visuals.



NVIDIA VILA Model NIM

- Vision language model



### Relational Embedding

Map visual and context data into searchable relationships



NVIDIA NEMO Retriever text embedding NIM



### Search and Re-rank

Locate specific video moments (multimodal input – text output)



NVIDIA NEMO Retriever text reranking NIM



### Summarization

Generated summaries of video based on captions and prompts



LLM NIM



#### “Here is a summary of the video”

- Warehouse scene (0:00 – 25:00)
  - A person wearing a yellow vest walks down the warehouse aisle...

#### “Here is a summary of the video”

- Traffic report (10:00 AM – 10:05 AM)
  - A green car is seen driving from the left to right of the intersection
  - A yellow bus is seen making a turn at the intersection .....

#### “Golf Swing report”

- Setup: the foundation of a great swing
  - Strengths: your setup is solid
  - Recommendations: Maintain
- Impact: the critical moment
  - Areas for improvement: Your body is slightly leaning into the swing...

Safety and training | Sports analytics | Media and Journalism | Healthcare



Optimize  
Productivity

# Build your own outcomes

Blueprints and guides enable your AI proof-of concepts



Beyond the  
Frame: Video  
Search and  
Summarization



Accelerate  
software  
development  
with Tabnine



AI-driven data  
retrieval and  
decision-making



Accelerated  
Customer Issue  
Resolution with  
H2O.ai Complaints  
Summarizer

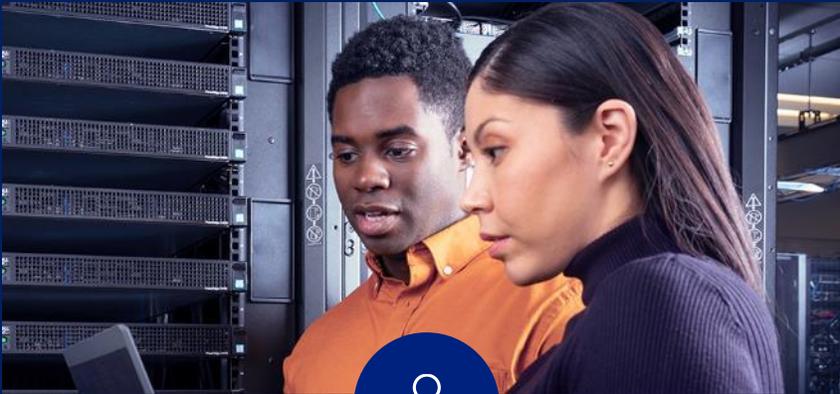


**DELL**Technologies  
**AI Factory**  
WITH  **NVIDIA**



# Transforming workflows with Agentic

## Benefits

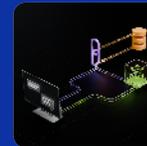


- Increased Productivity: Frees up time by handling routine work
- Enhanced Efficiency: Automates complex tasks
- Improved Decision-Making: uses reasoning models for higher accuracy

## Building Agentic workflows



- NVIDIA AI Blueprints
- NVIDIA NeMo
- NVIDIA Llama Nemotron NIM
- NVIDIA NeMo Agent Toolkit
- Dell AI Factory with NVIDIA

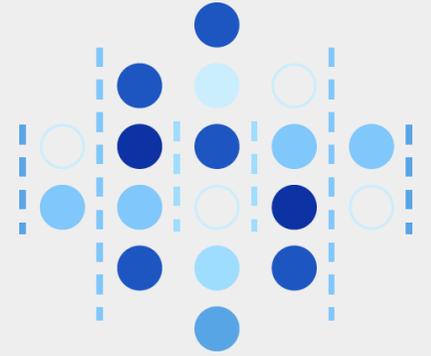


## Demonstrating capabilities



- RAG workflow accelerator
- Health assistant
- Financial, telecom, AI Agents for industry segments

# Next steps



## Learn



Deepen your knowledge of Agentic AI: Enroll in SalesU training and use SalesChat (Prompt: “Tell me about agentic AI” and view the resources)

## Educate



Half-day Dell Services strategy engagement on use cases, requirements, skills and processes (fee waived).

## Discover



Point customers to Customer Solutions Centers demos, HPC & AI Innovation labs, and test drives to make their AI outcomes real.

## Accelerate



Discuss how they can fast-track their workflows with AI partners, Managed Services and desktop-to-datacenter solutions.