Vision For All: Unlocking Video Analytics With AI Agents

NVIDIA Metropolis Webinar

Physical AI Will Transform $50 Trillion Industries

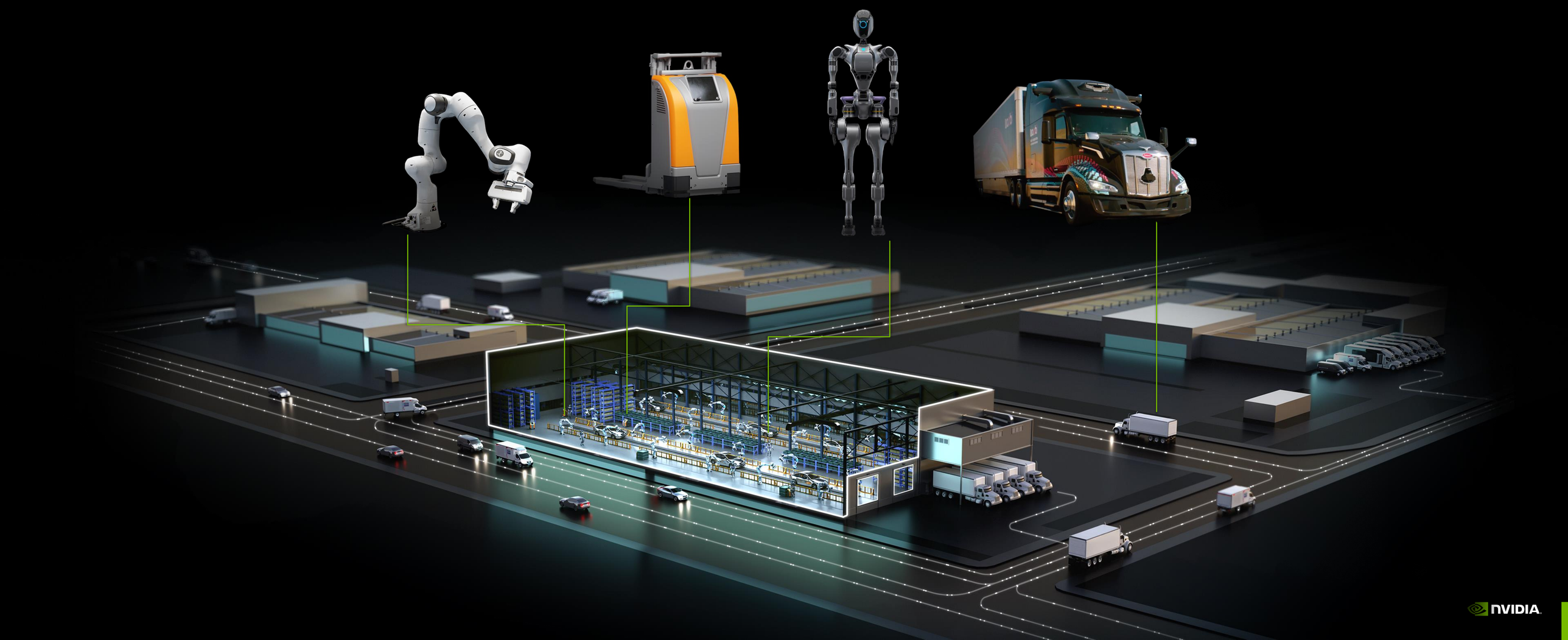
1.5 Billion
Commercial Cameras

10 Million
Factories

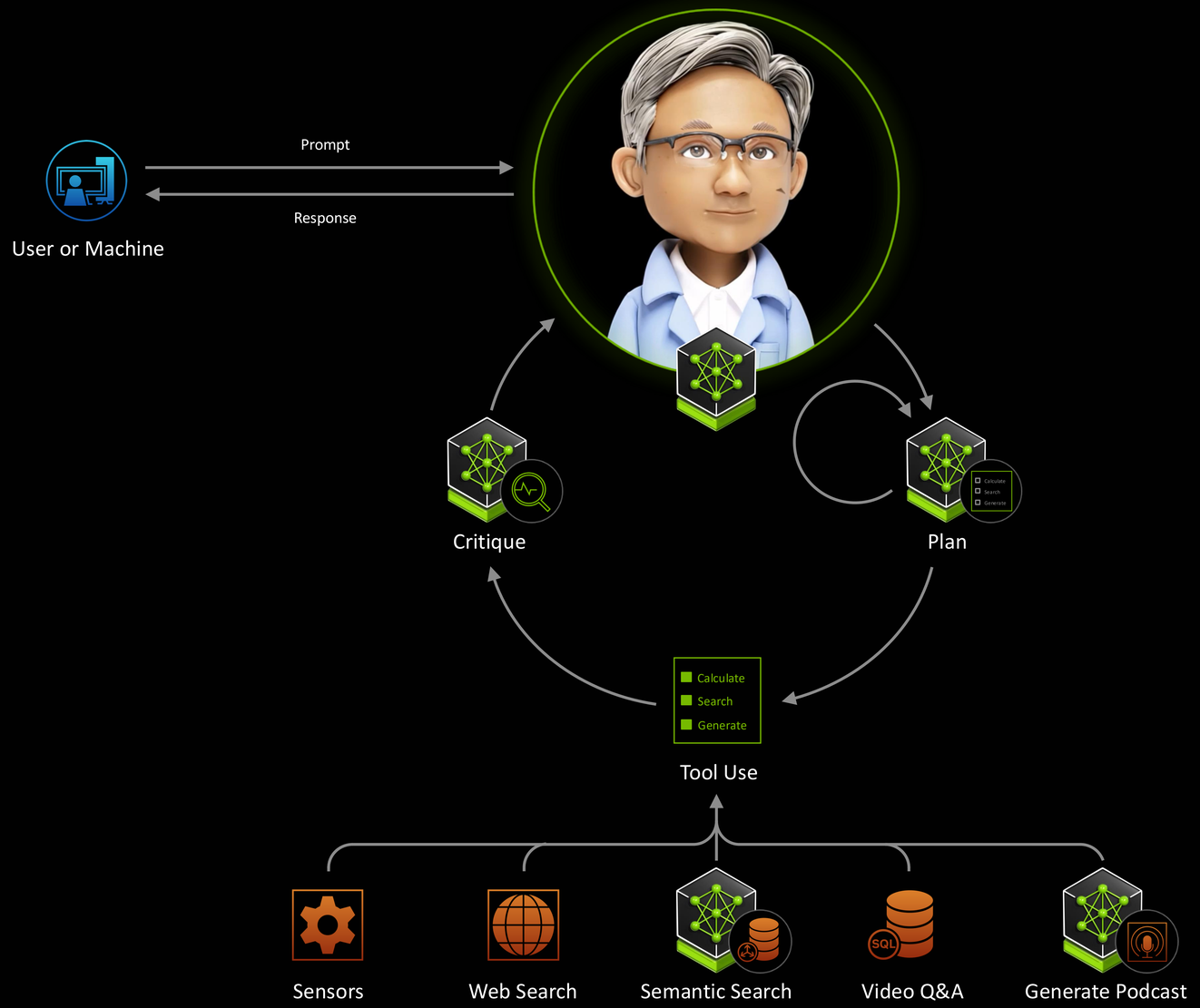
200K
Warehouses

1.5 Billion
Cars & trucks

Future Billion
Humanoid robots

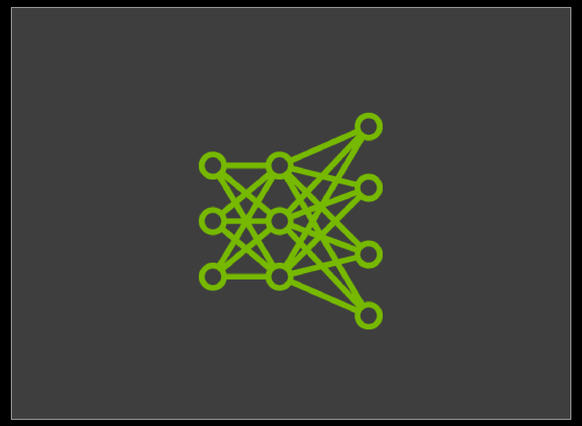


Agentic AI


Prompt
Response
User or Machine
Critique
Plan
Calculate
Search
Generate
Tool Use
Sensors
Web Search
Semantic Search
SQL
Video Q&A
Generate Podcast

Why Vision Language Models Are So Important?

Multimodal models capable of understanding and processing text, image and video



ZERO-SHOT LEARNING

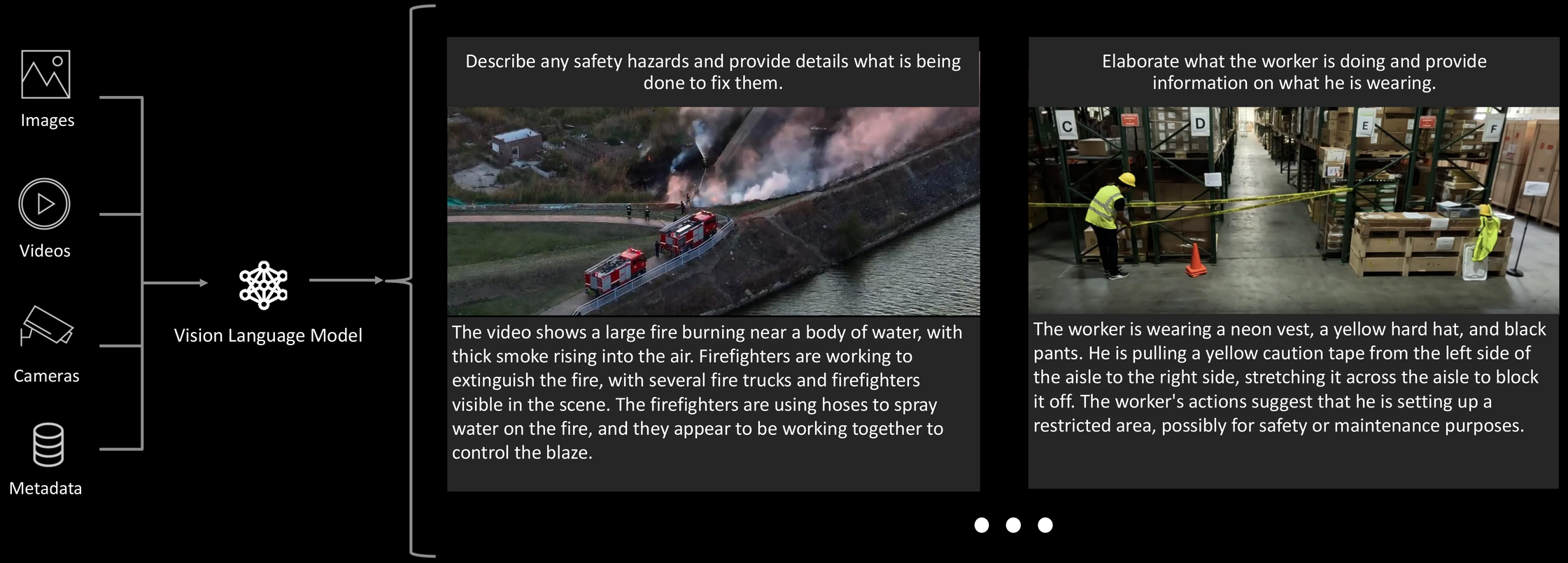
MULTIMODAL UNDERSTANDING

REASONING & COMPREHENSION

ENHANCED RETRIEVAL CAPABILITES

Vision Language Models For Insight Generation

Easily interact with your visual media to obtain valuable insights



Discover

MODELS

Reasoning
Vision
Visual Design
Retrieval
Speech
Biology
Simulation
Climate & Weather
Safety & Moderation

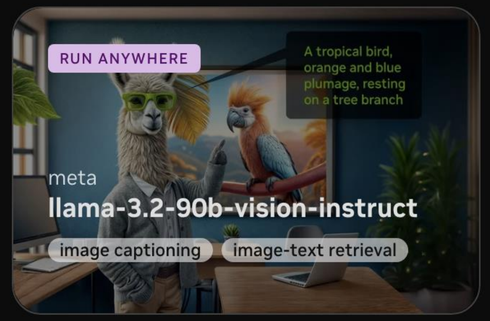
INDUSTRIES

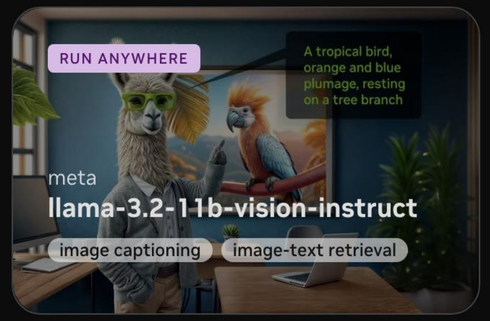
Gaming
Healthcare
Industrial
Robotics

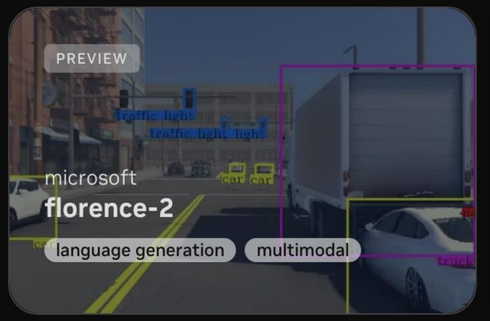
build.nvidia.com/explore/vision

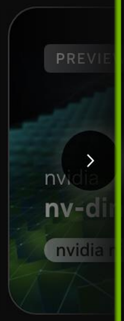
Vision Language Models (VLM)

Multimodal models that can reason against image and video inputs and perform descriptive language generation



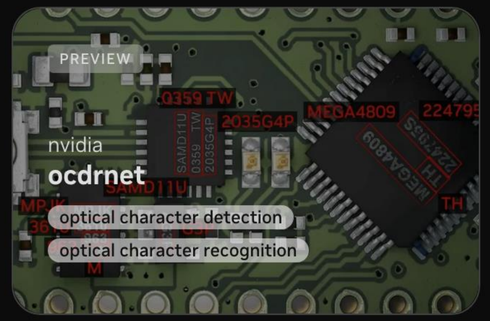


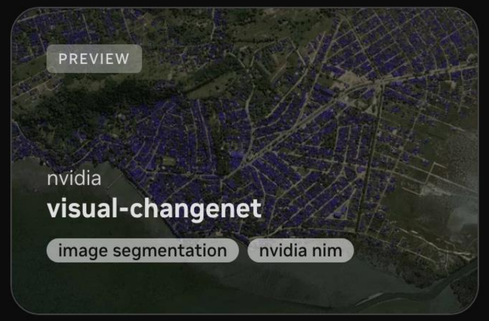


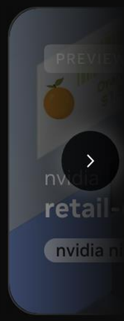


Specialized Foundation Models

Computer vision models that excel at particular visual perception tasks







Building A Virtual Workforce Of AI Agents

AI Agents & Co-Pilots

- **Increase productivity and reduce waste** by monitoring process / procedures
- **Boost asset management efficiency** through better space utilization
- **Reduce Labor Cost** through auto-generation of incident reports and summaries
- **Prevent accidents and production problems** by identifying anomalies

Smart Spaces

CPG

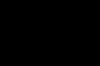
Electronics

Automotive

Pharma & Health

Energy

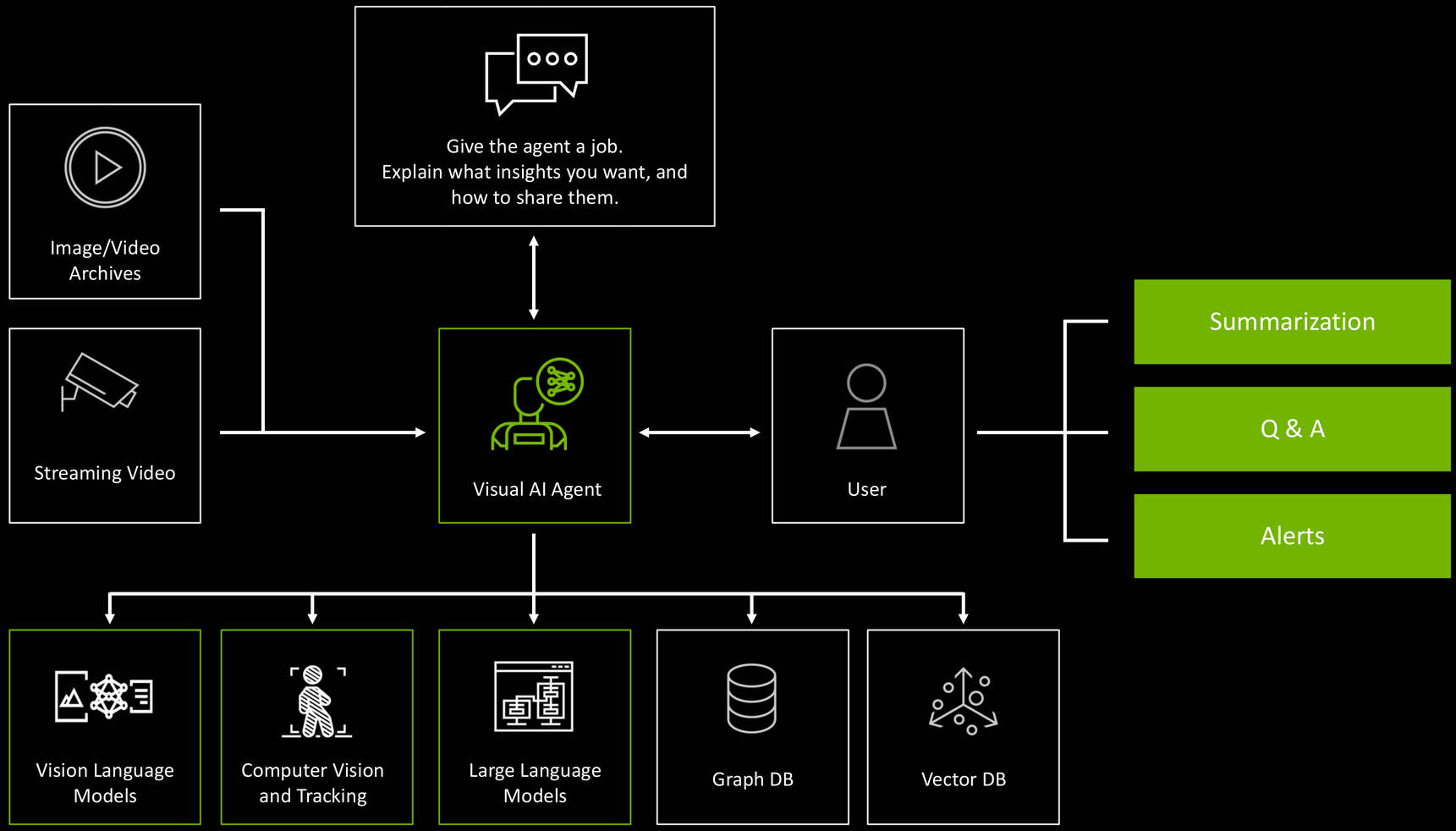
Retail




4 8 MIL
3 7 BOS
Chevron
Athletics
93

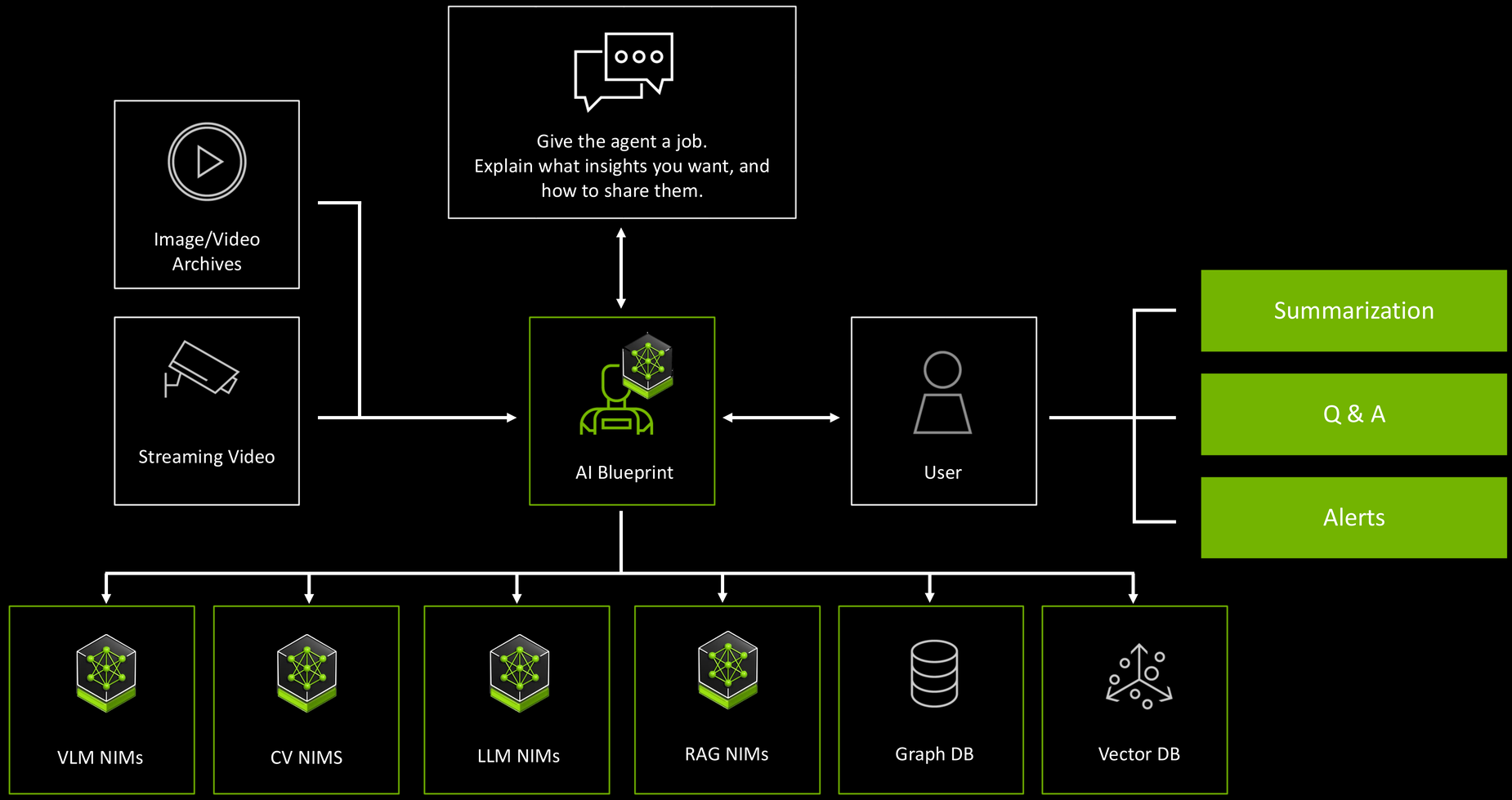
Video Analytics AI Agents – an Entirely New Class of Applications

Unlock knowledge and insights from camera streams and archived videos



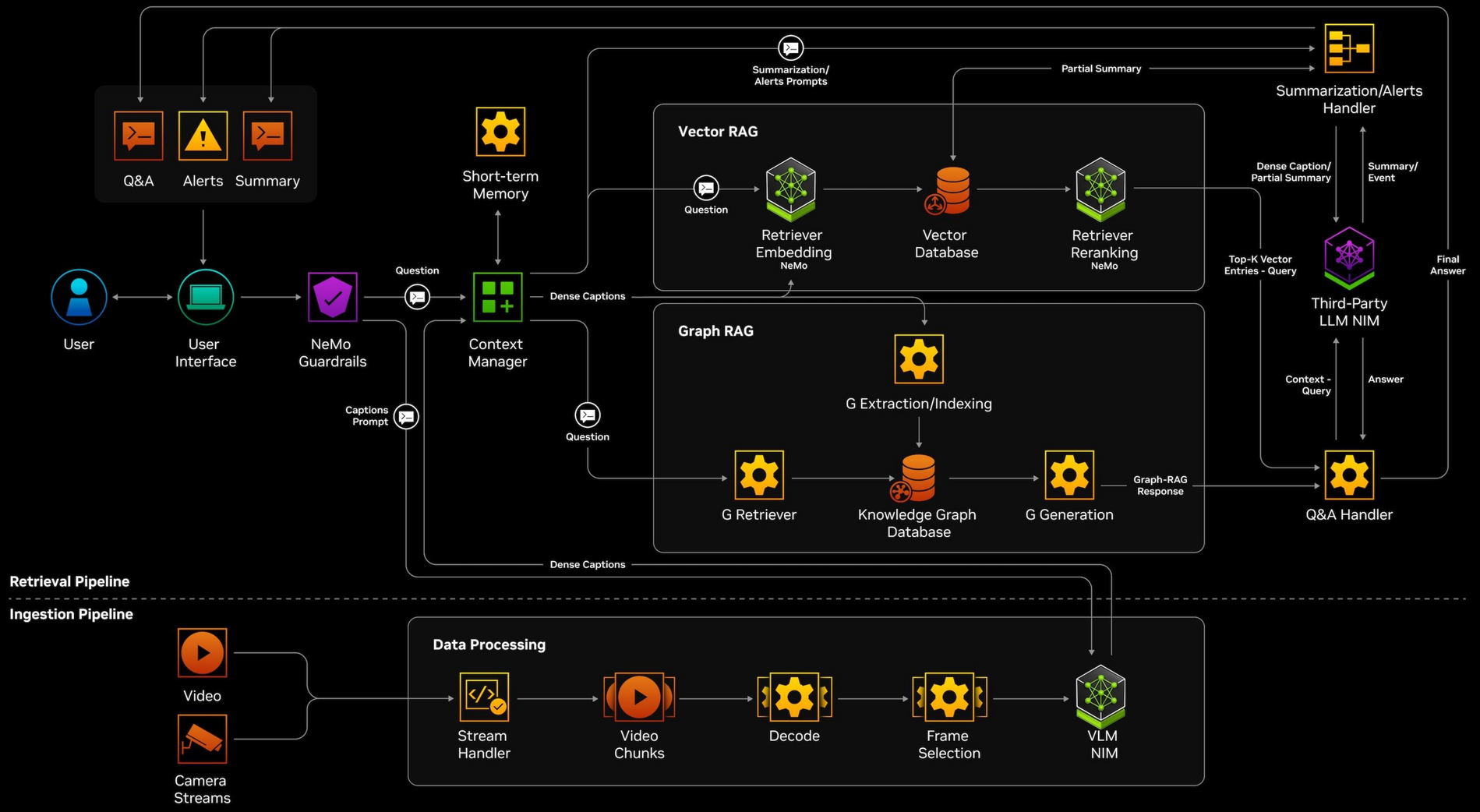
NVIDIA AI Blueprint For Video Search and Summarization (VSS)

Development platform for building video analytics AI agents



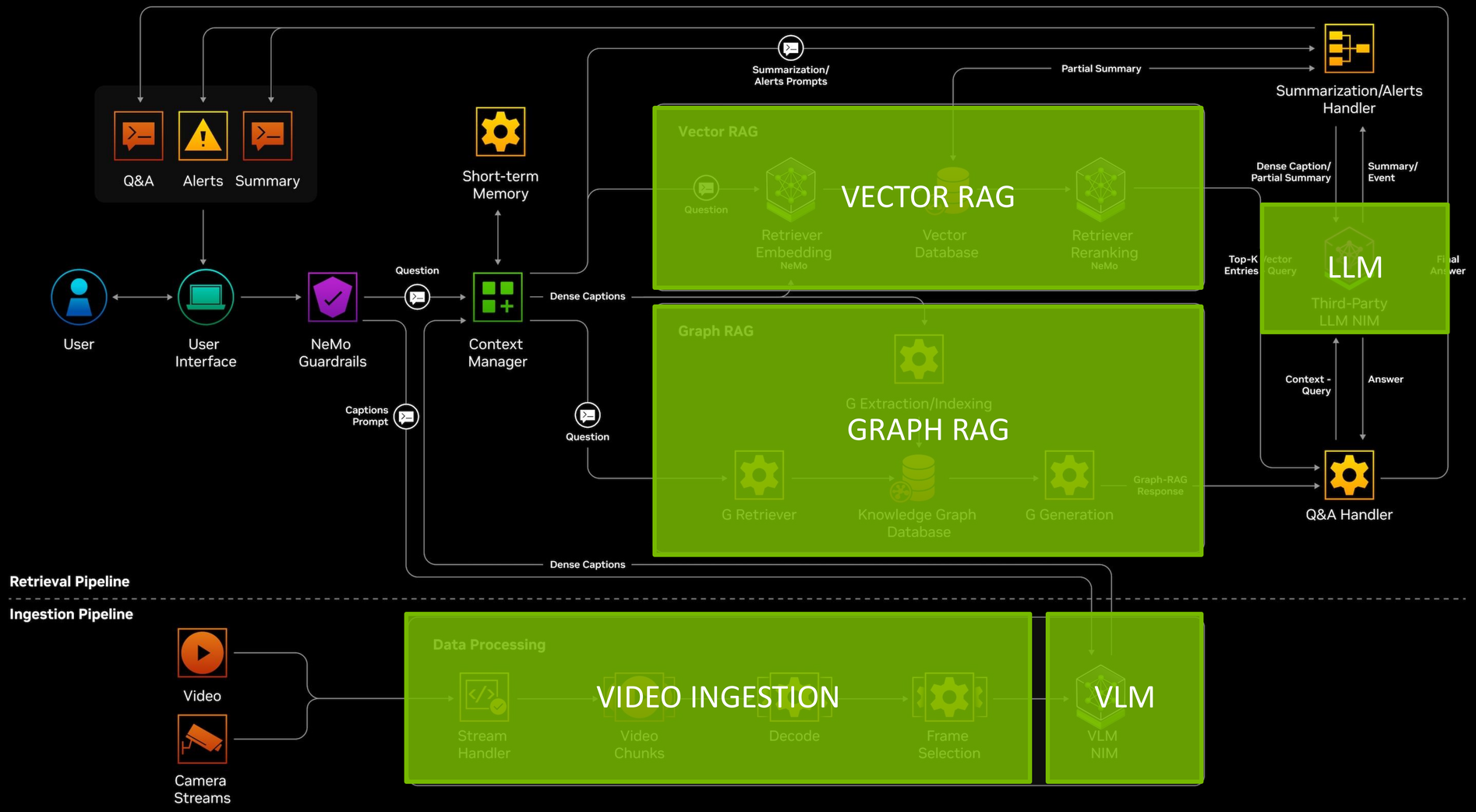
NVIDIA AI Blueprint For Video Search And Summarization

Unlocks Knowledge & Insights from Billions of Camera Streams and Petabytes of Archived Video



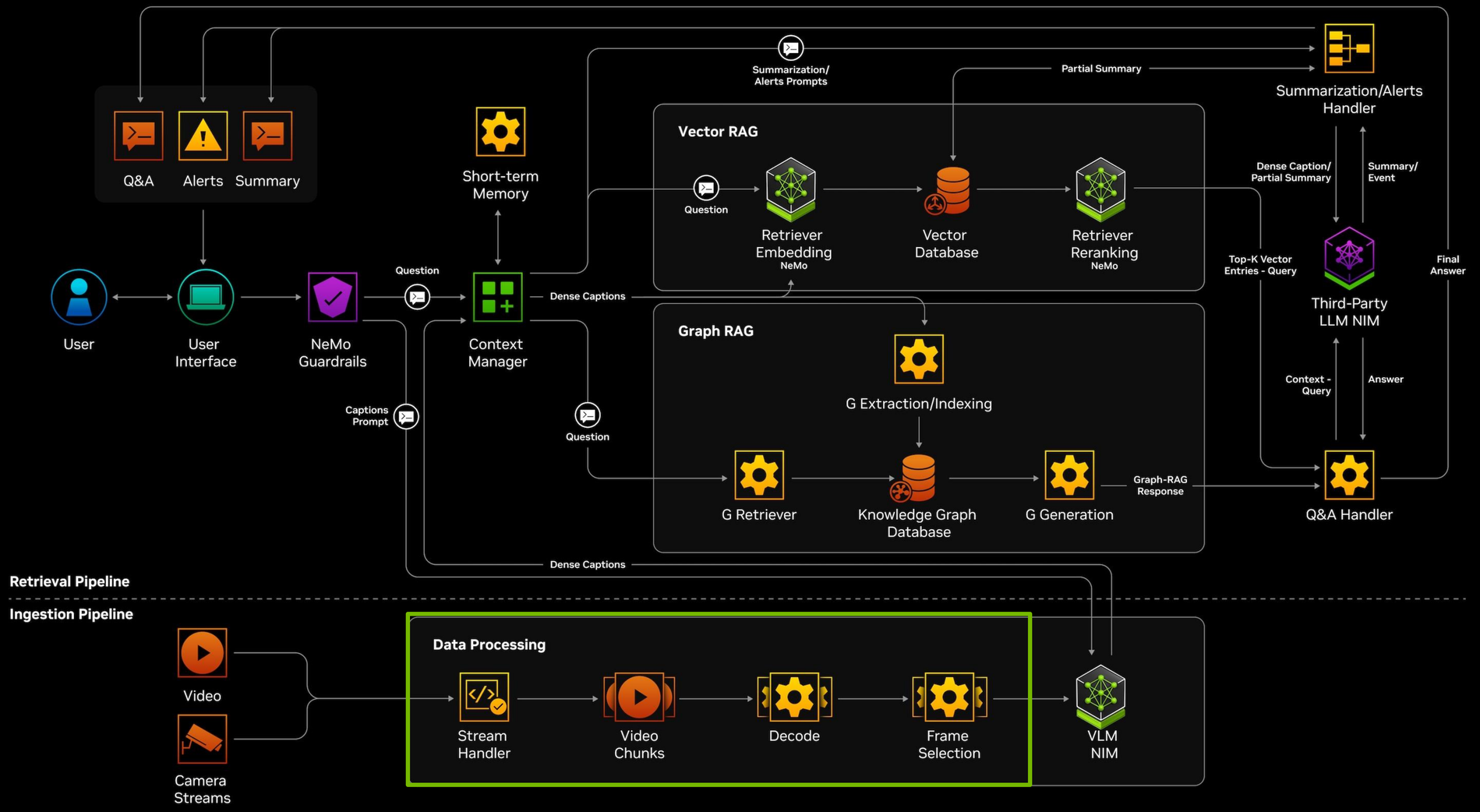
NVIDIA AI Blueprint For Video Search And Summarization

Unlocks Knowledge & Insights from Billions of Camera Streams and Petabytes of Archived Video

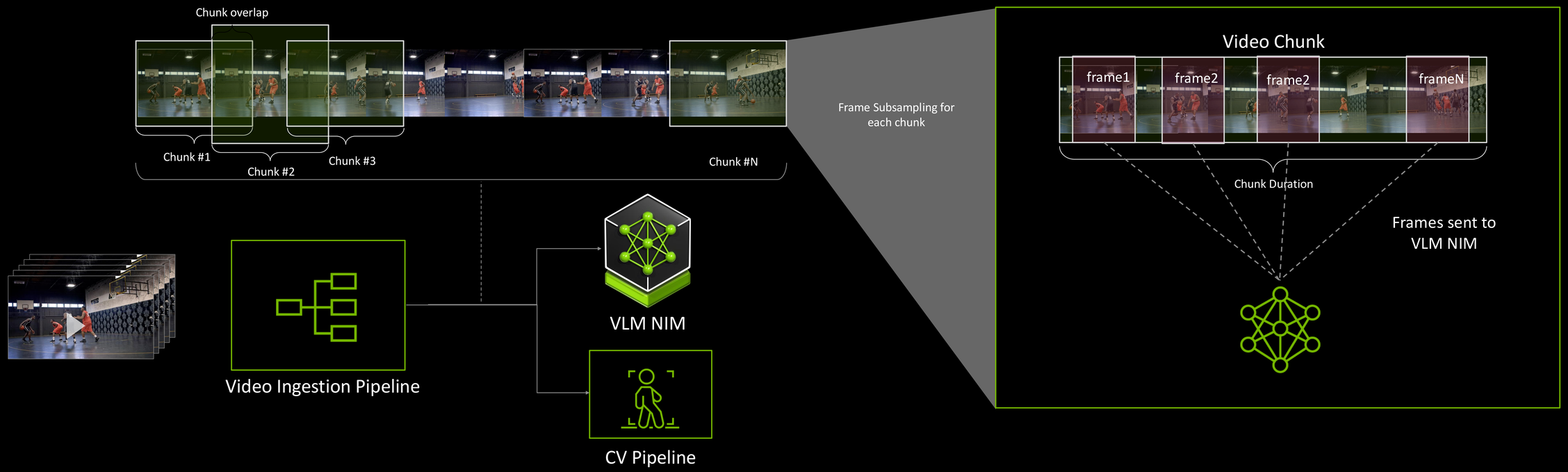


NVIDIA AI Blueprint For Video Search And Summarization

Unlocks Knowledge & Insights from Billions of Camera Streams and Petabytes of Archived Video



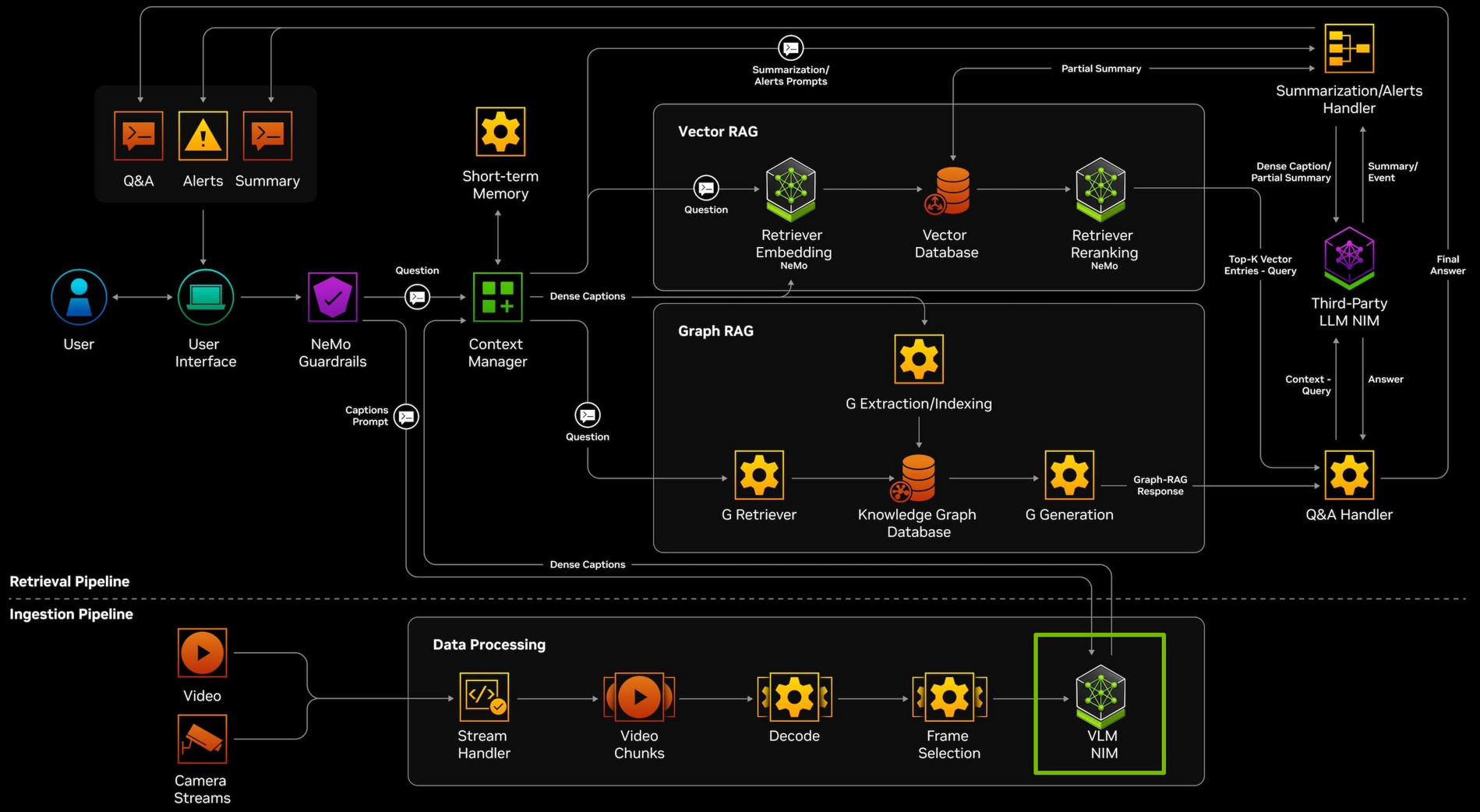
Data Ingestion And Processing



Configurable Parameter	Common Values	Impact on Accuracy	Impact of Performance
Chunk Duration	10s, 30s, 1m	Shorter chunk size works well for short actions	Higher chunk duration will yield higher performance
Chunk Overlap	10s		
Summary Duration	60s	Shorter duration can work better for more real-time update	Higher summary duration will yield higher performance
Number of frames for VLM inference	8, 10, 16, 32	More frames will yield better accuracy	More frames will reduce performance

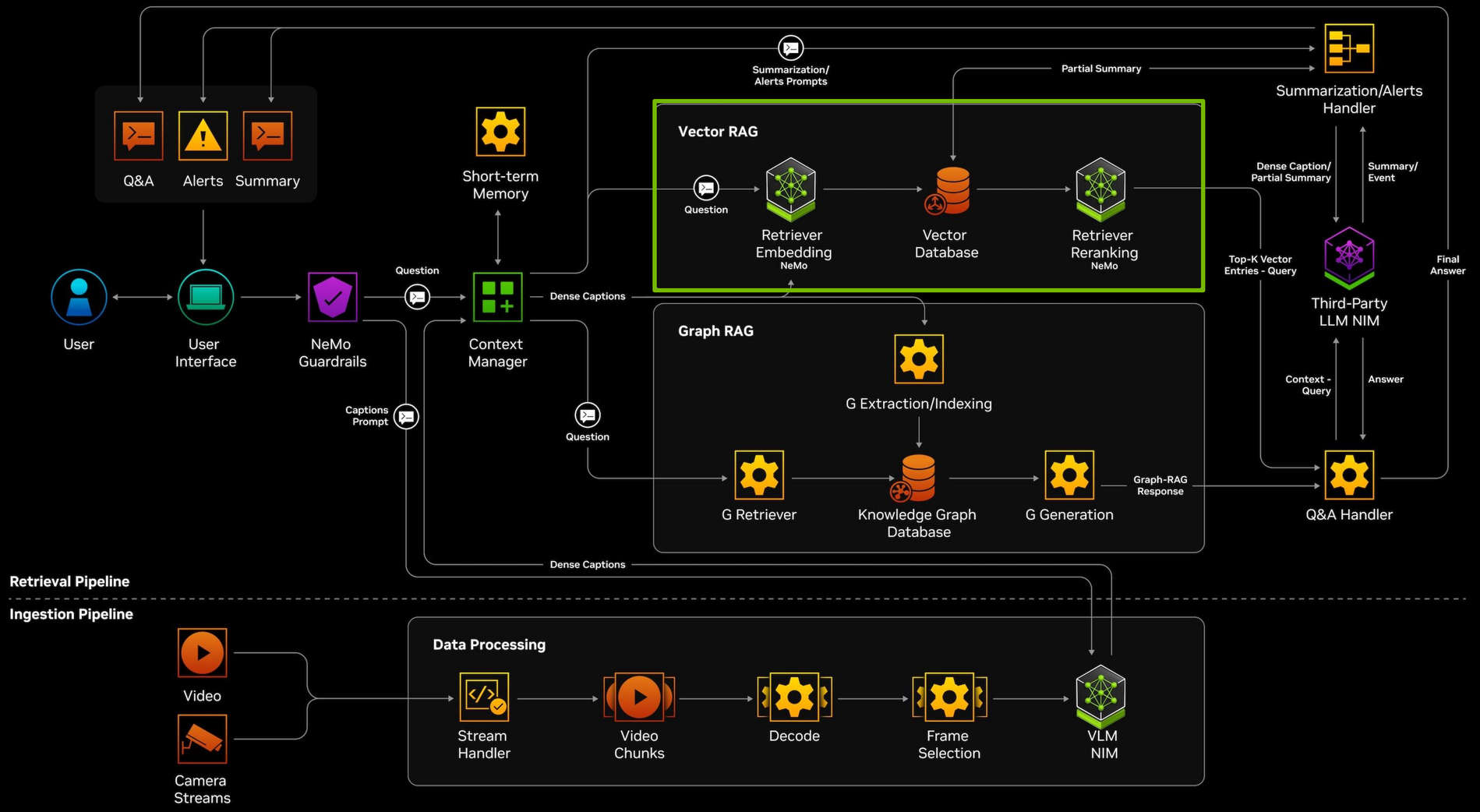
NVIDIA AI Blueprint For Video Search And Summarization

Unlocks Knowledge & Insights from Billions of Camera Streams and Petabytes of Archived Video



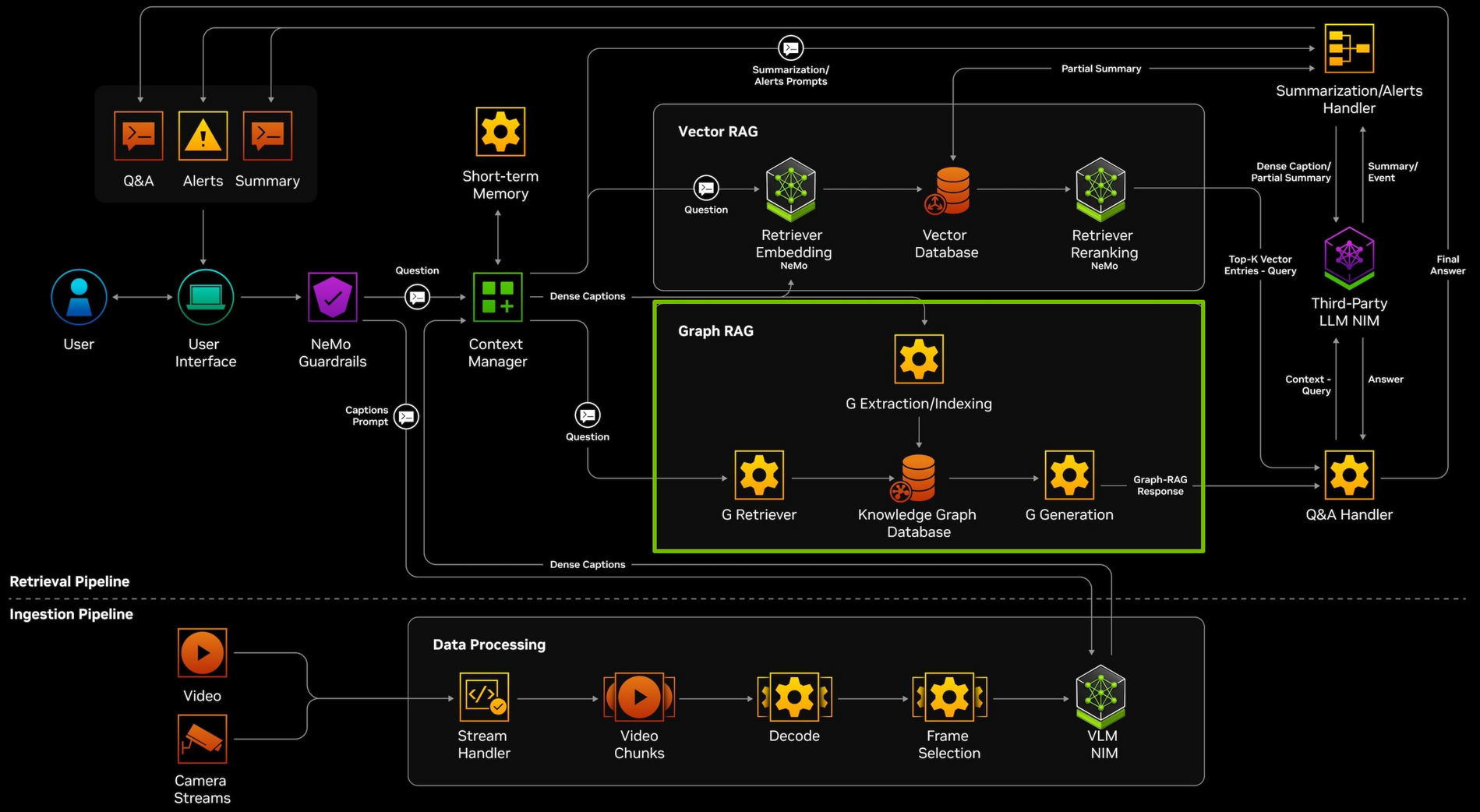
NVIDIA AI Blueprint For Video Search And Summarization

Unlocks Knowledge & Insights from Billions of Camera Streams and Petabytes of Archived Video



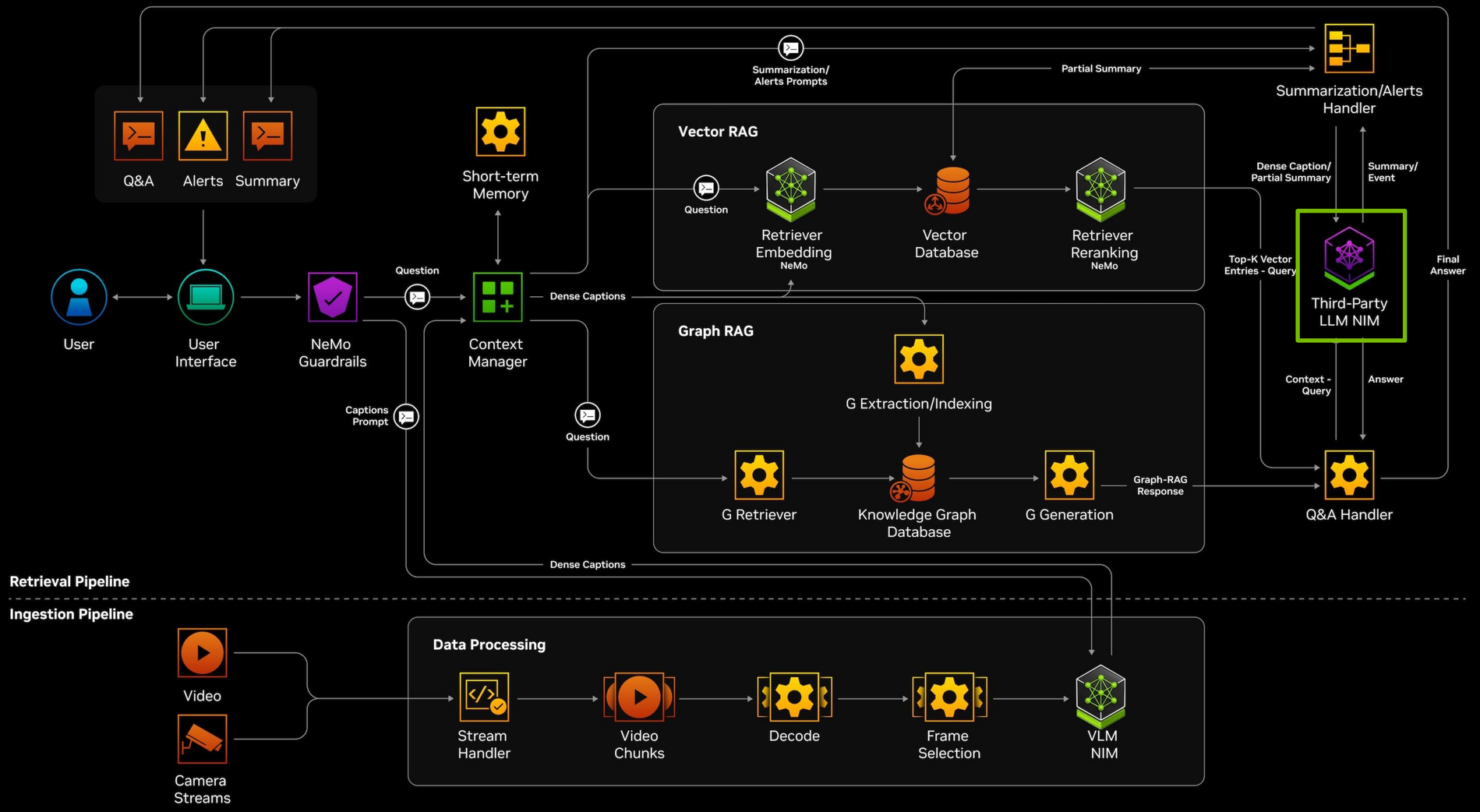
NVIDIA AI Blueprint For Video Search And Summarization

Unlocks Knowledge & Insights from Billions of Camera Streams and Petabytes of Archived Video

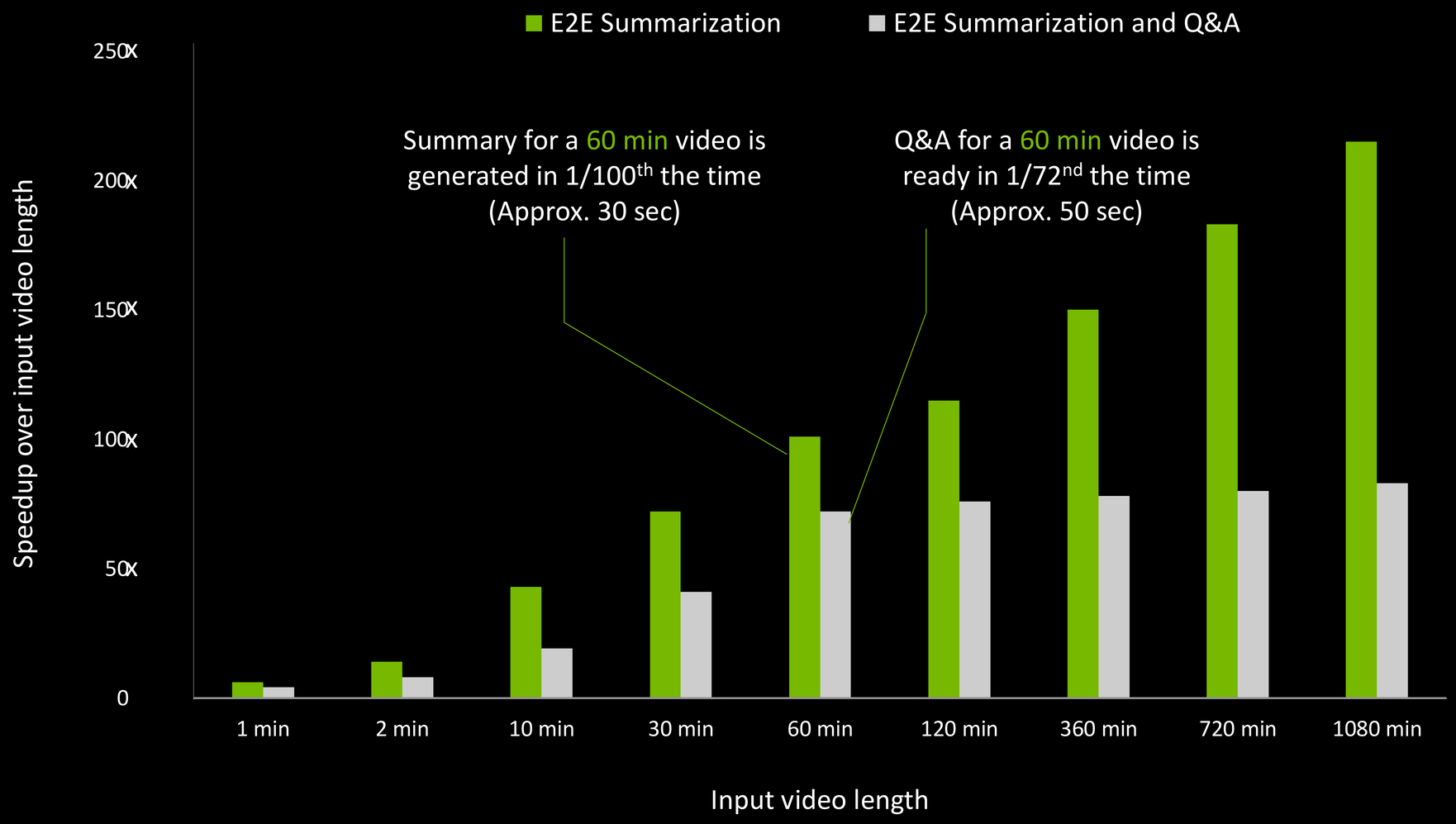


NVIDIA AI Blueprint For Video Search And Summarization

Unlocks Knowledge & Insights from Billions of Camera Streams and Petabytes of Archived Video



Speedup To Analyze Videos In VSS



Build Video Analytics AI Agents With AI Blueprint

ACCELERATED TIME TO MARKET

Pre-built workflows

Production-ready models

Easy to use APIs

COST EFFECTIVE AT SCALE

Efficient data ingestion and processing

High throughput with VLM and LLM NIMs

Optimized architecture

HIGH ACCURACY & CUSTOMIZABILITY

Long context understanding with videos

Flexible deployment from edge to any cloud

Choice of VLM/LLM NIMs and databases

Finetune models and prompts

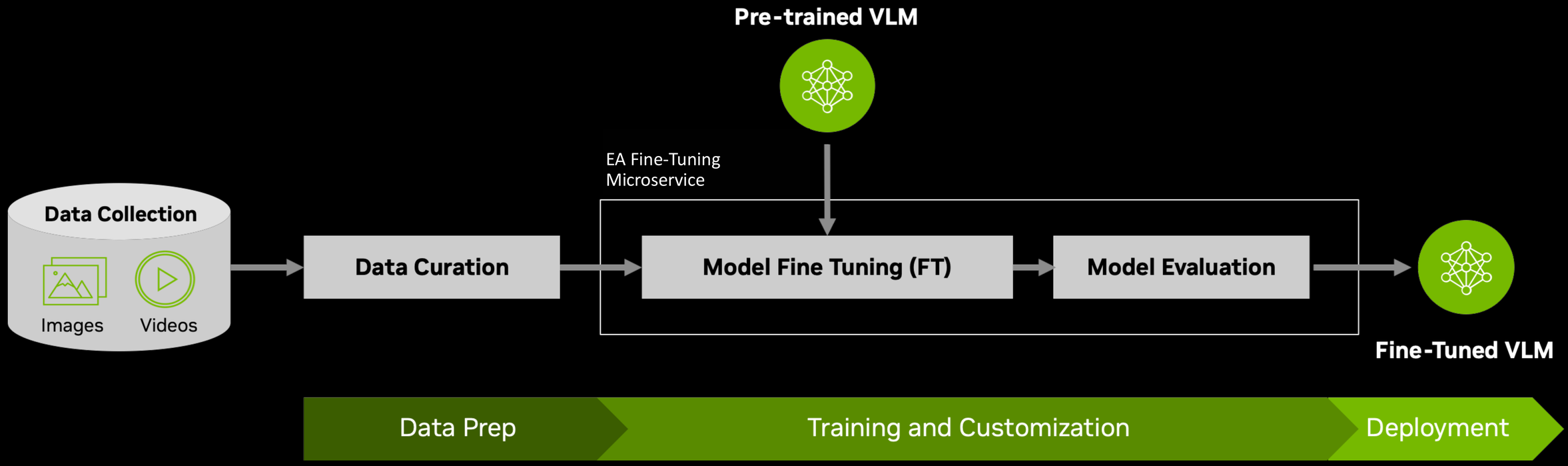
DATA PRIVACY

Secure on-prem solution

No need to access 3rd party services

Fine-Tuning Framework For Vision Language Models

VLM Finetuning Microservice — https://developer.nvidia.com/vlm-fine-tuning-microservice-early-access



DETECT SPECIFIC OBJECTS OR EVENTS

DOMAIN ADAPTION

MULTIMODAL UNDERSTANDING

Improving Caption Quality With Fine-Tuning

Video captioning fine-tuned with YouCookII dataset

Before Fine-Tuning:

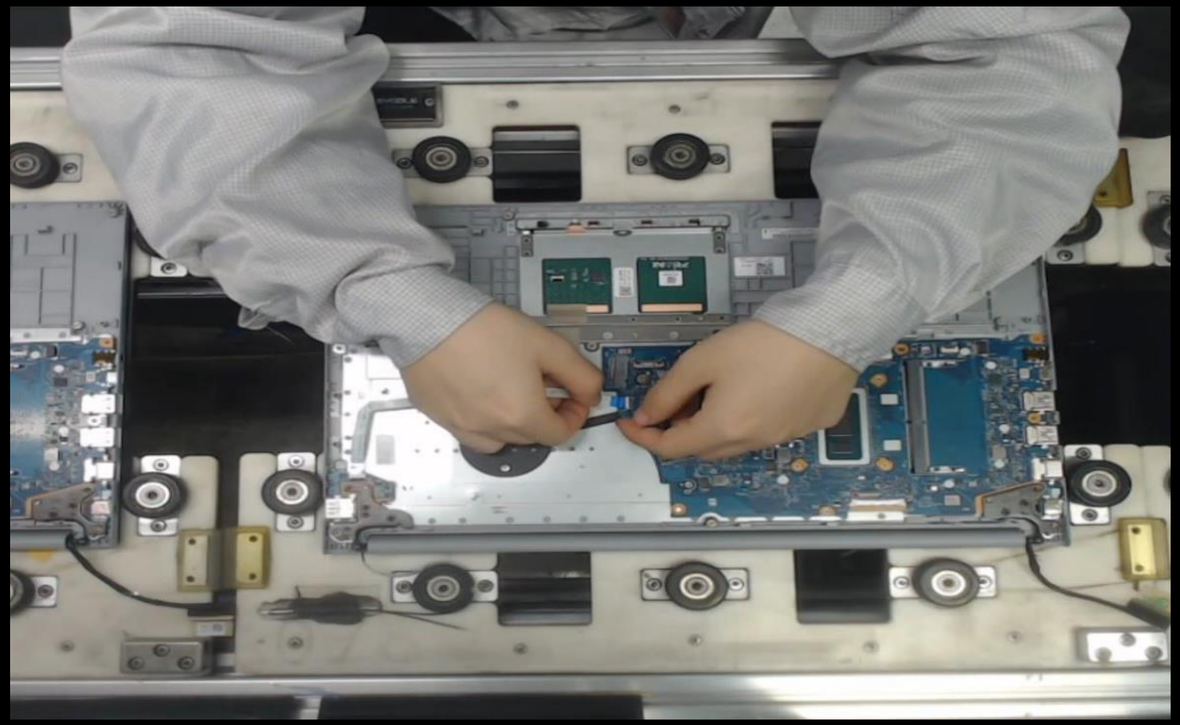
A man is cooking food in a pan and then covers it with a lid.

After Fine-Tuning:

This video begins with a close-up of a person's hand lifting a stainless steel placed on a gas stove. The pot contains a mixture of what seems to be pasta and possibly some vegetable or meat. The person's hand is partially visible, wearing a blue and white striped shirt. The stove has a black base in a kitchen setting, and the lighting is warm, suggesting an indoor environment.

Pegatron Reaches 98% Understanding Of SOP On An Assembly Line

Detect missing actions or unusual actions in manufacturing process



- **Use case:** Pegatron analyzes 7-step I/O cable insertion process on assembly line. Alert workers when differences are spotted compared to standard operating procedure (SOP)

- **Solution:**
  - AI Blueprint for Video Search and Summarization
  - Model: VILA 3B (research)
  - Fine-tuned with 8 videos (12 seconds each)
  - Training time / GPU used:
    1*A100, ~3 hours;
    8*A100, ~1.5 hours

- **Results:**
  - 98% recognition rate of actions after VLM finetuning

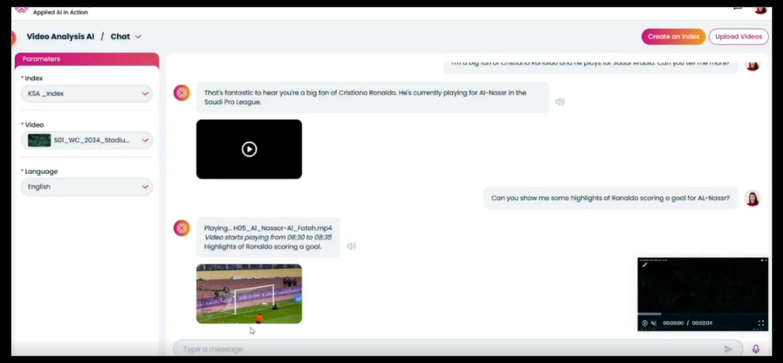
AI Agents Offer Valuable Insights and Streamline Processes

Siemens transforms shopfloor operations with Industrial Copilot

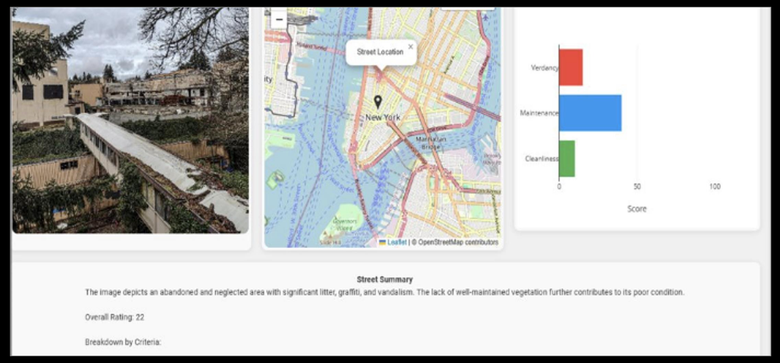
Stanley Black & Decker halves onboarding costs with DeepHow

Build Video Analytics AI Agents At Scale

ASOCS
Situational awareness of forklifts, AGVs and drones in factories



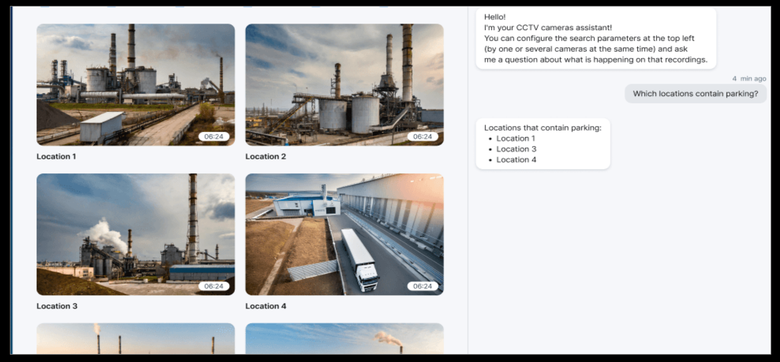
Centific
AI Agent for live events to provide a better fan experience



Deloitte
Extract insights from thousands of city cameras

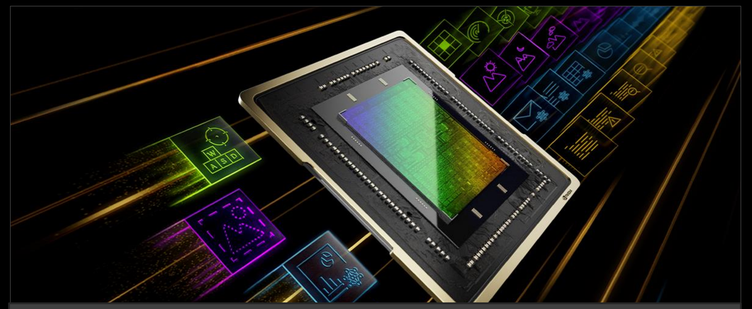
ITMAX
Generate deeper insights for traffic events to quicken first/second responders' response time

Linker Vision
Provide real-time incident alerts and reduce response time



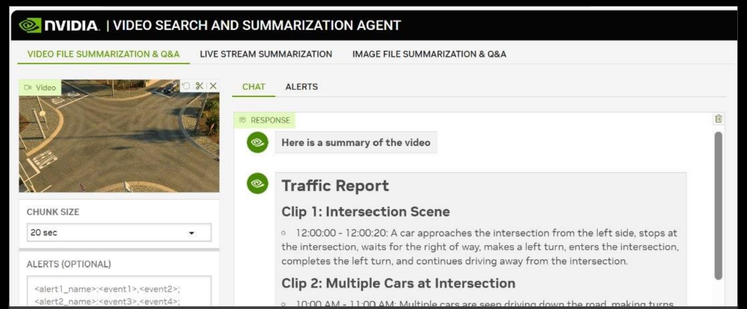
SoftServe
AI assistant to respond to user queries about specific events

Key Features For Upcoming VSS GA Release



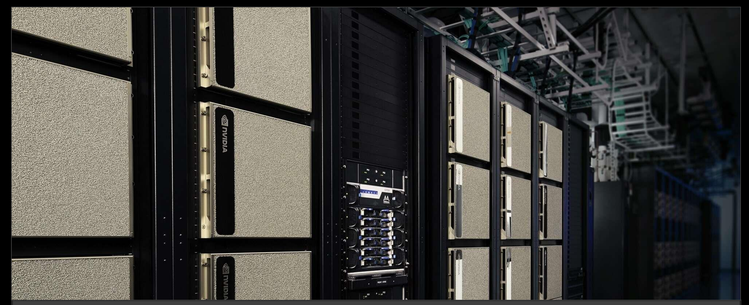
Single GPU deployment

RTSP short- and live multi-streaming



Burst mode video ingestion

Audio transcription



One-click deployments through
CSPs and Launchables

Customizable video ingestion pipeline
with CV & multi-view 3D tracker

Get Started With VSS Blueprint

1

EXPLORE

Learn about VSS blueprint

Free

Try VSS blueprint preview on build.nvidia.com

Apply for Early Access

2

TRY

Try on cloud with Launchables

Starting at $14/hour

- Try VSS Launchable
- Register for VSS early access
- Click “Deploy Launchable”
- Sign in to Brev.dev
- Get access to cloud instances
- Upload videos to test VSS

3

BUILD

Customize and finetune for your business needs

Flexible development

- Public CSPs
- NVIDIA DGX Cloud
- On-prem hardware
- GPU workstations*
- NVIDIA DIGITS*

Access VLM fine-tuning microservice

4

DEPLOY

Deploy services and applications at scale

Optimized deployment with NVIDIA NIMs

For NVIDIA AI Enterprise licensing, contact your account manager.

* Coming Soon

Launchables Are Links To Deploy Pre-Configured Sandbox Instances

Launchables takes the guesswork away so developers can use your solution in 1-click

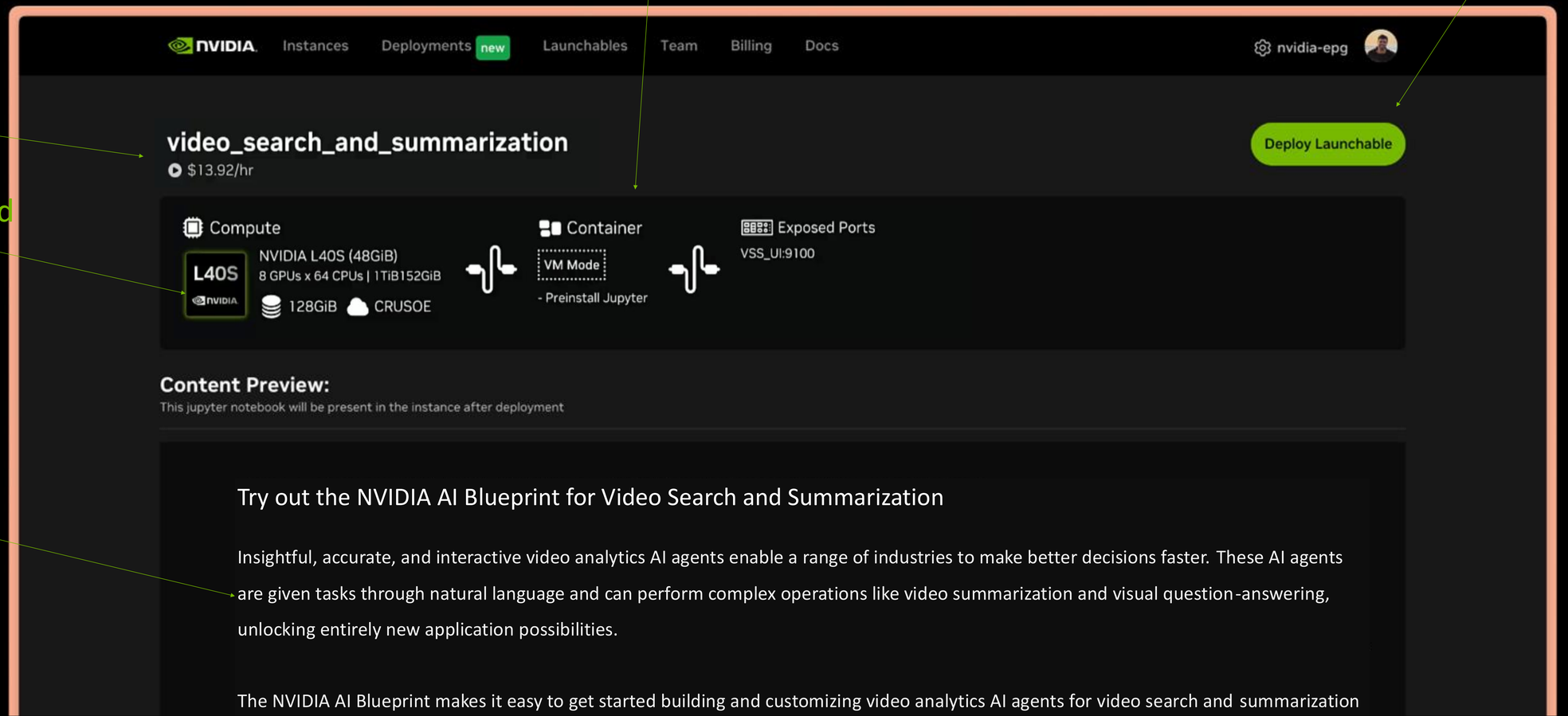
Software is pre-configured

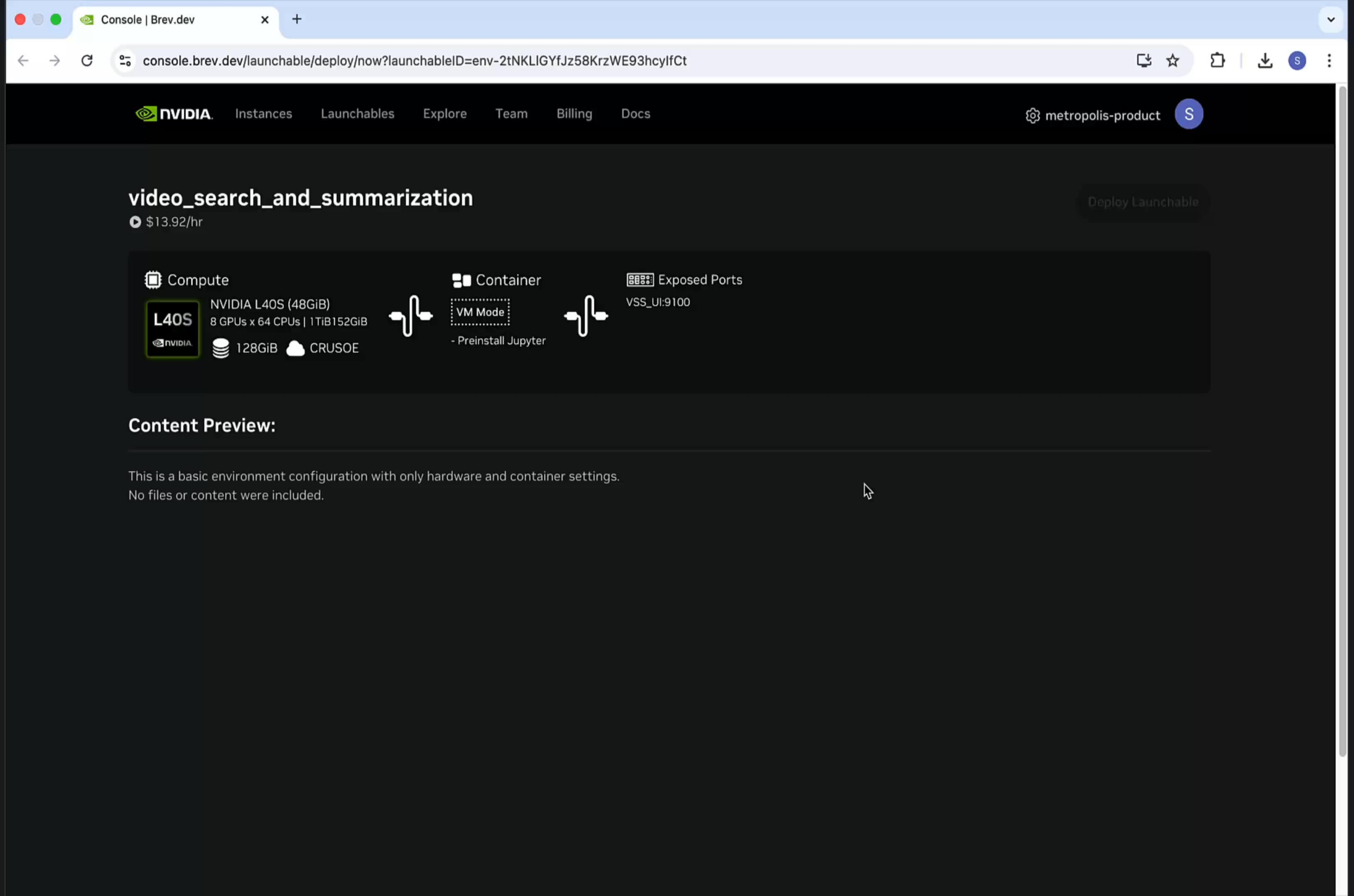
1-click deploy

Price

Compute is defined & provided

Preview of the notebook or content




Console | Brev.dev
console.brev.dev/launchable/deploy/now?launchableID=env-2tNKLIGYfJz58KrzWE93hcyIfCt
NVIDIA
Instances
Launchables
Explore
Team
Billing
Docs
metropolis-product
S
video_search_and_summarization
$13.92/hr
Deploy Launchable
Compute
L40S
NVIDIA L40S (48GiB)
8 GPUs x 64 CPUs | 1TiB152GiB
128GiB
CRUSOE
Container
VM Mode
- Preinstall Jupyter
Exposed Ports
VSS_UI:9100
Content Preview:
This is a basic environment configuration with only hardware and container settings.
No files or content were included.

Get Started With VSS Blueprint

1

EXPLORE

Learn about VSS blueprint

Free

Try VSS blueprint preview on build.nvidia.com

Apply for Early Access

2

TRY

Try on cloud with Launchables

Starting at $14/hour

- Try VSS Launchable
- Register for VSS early access
- Click “Deploy Launchable”
- Sign in to Brev.dev
- Get access to cloud instances
- Upload videos to test VSS

3

BUILD

Customize and finetune for your business needs

Flexible development

- Public CSPs
- NVIDIA DGX Cloud
- On-prem hardware
- GPU workstations*
- NVIDIA DIGITS*

Access VLM fine-tuning microservice

4

DEPLOY

Deploy services and applications at scale

Optimized deployment with NVIDIA NIMs

For NVIDIA AI Enterprise licensing, contact your account manager.

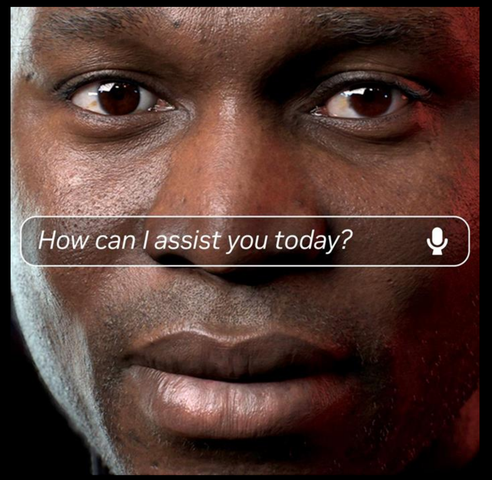
* Coming Soon

NVIDIA GTC, March 17-21
What's Next in AI Starts Here

GTC is a unique developer conference that brings together thousands of innovators, developers, engineers, researchers, creators, and business leaders to explore how accelerated computing and AI are helping humanity solve our largest, most complex challenges.

From NVIDIA CEO Jensen Huang keynote to hundreds of inspiring sessions, exhibits, technical hands-on training workshops, and unique networking events, GTC is the place to see real-world examples of applied AI in action.

Registration is required to join in person. You can also attend virtually at www.nvidia.com/gtc/keynote on March 18, 10:00 a.m. PDT (no registration required).



Cesium

March 17–21, 2025 | www.nvidia.com/gtc

Computer Vision and Video Analytics

Featured Talks

S72784 AI Agents for Real-Time Video Understanding and Summarization

S72902 Build your next Vision AI application for Physical AI on a Digital Twin

S72758 Agentic AI for Physical Operations

S71858 Level Up Retail Store Security and Customer Insights with Video Analytics AI Agents

S73304 Upskill Manufacturing Workforces with AI Agents

Workshops and Trainings

DLIT71406 Build Next-Gen Agents With Large Vision Language Models

DLIT72753 Build a Co-Pilot for Industrial Process Monitoring and Quality Control

DLIT72055 Build Visual AI Agents with RAG Using NVIDIA Morpheus, RIVA, and AI Blueprints for Video Search and Summarization

Partner Talks

S71858 Infosys - Level Up Retail Store Security and Customer Insights with Video Analytics AI Agents

S72673 Pegatron - Harmonizing Digital Twins and AI Factory: Unlocking Industrial Autonomy

S73304 DeepHow - Upskill Manufacturing Workforces with AI Agents

S72394 Centific - Connecting Smart Cities to the tourist experience with Agentic AI

S72944 ITMAX - Revolutionizing Traffic Management with Behavioral Simulation and Multi-Image Reasoning

S71410 K2K - Redefining Urban Evolution with Gen AI in Palermo, Italy

S71824 Linker Vision - City-Scale AI with Digital Twins

S73076 Accenture - Revolutionizing Warehouse and Factory Management With NVIDIA "Mega" Blueprint for Robot Facilities

S72896 Siemens - Challenges and Breakthroughs in RAG Implementation for On-Prem Industrial Copilot Assistants

March 17–21, 2025 | www.nvidia.com/gtc


NVIDIA