



Unleashing RAN Intelligence Powered By AI

Mobile Test Summit, 2025 AI in Action

Kuntal Chowdhury, 6G Technology Evangelist and Developer Relations

AI-Driven 6G Blueprints

Transforming Telco

Extreme Efficiency

- Spectral Efficiency
- Energy Efficiency
- Operational Efficiency with Agentic AI

Beyond Connectivity

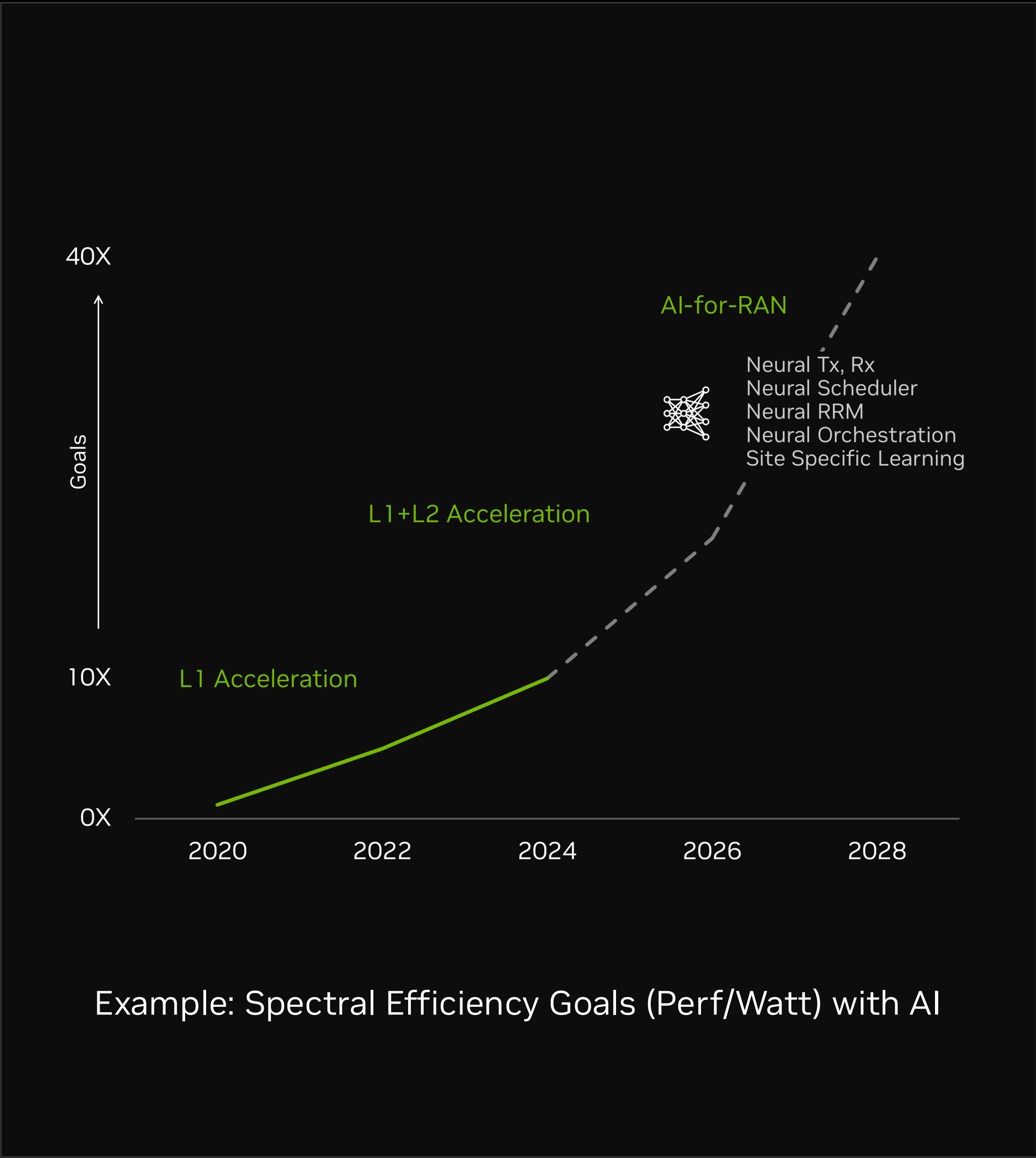
- Multi-modal sensing and comms
- Semantic Communications
- Applications and APIs

Network Fabric for AI

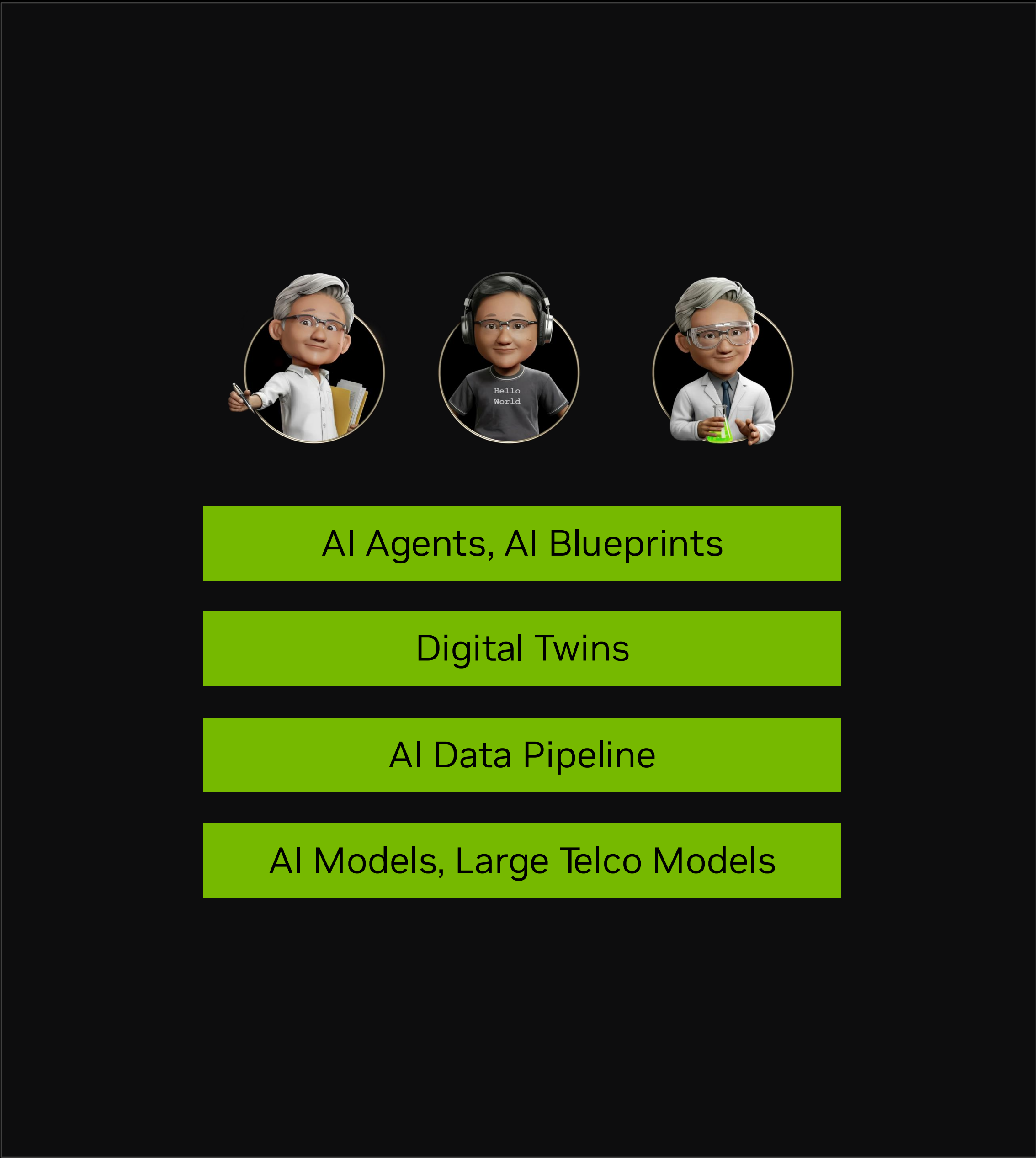
- AI Inferencing Everywhere
- Multi-purpose / AI-RAN Infrastructure
- AI-Grid / AlaaS

AI-Driven RAN Efficiency

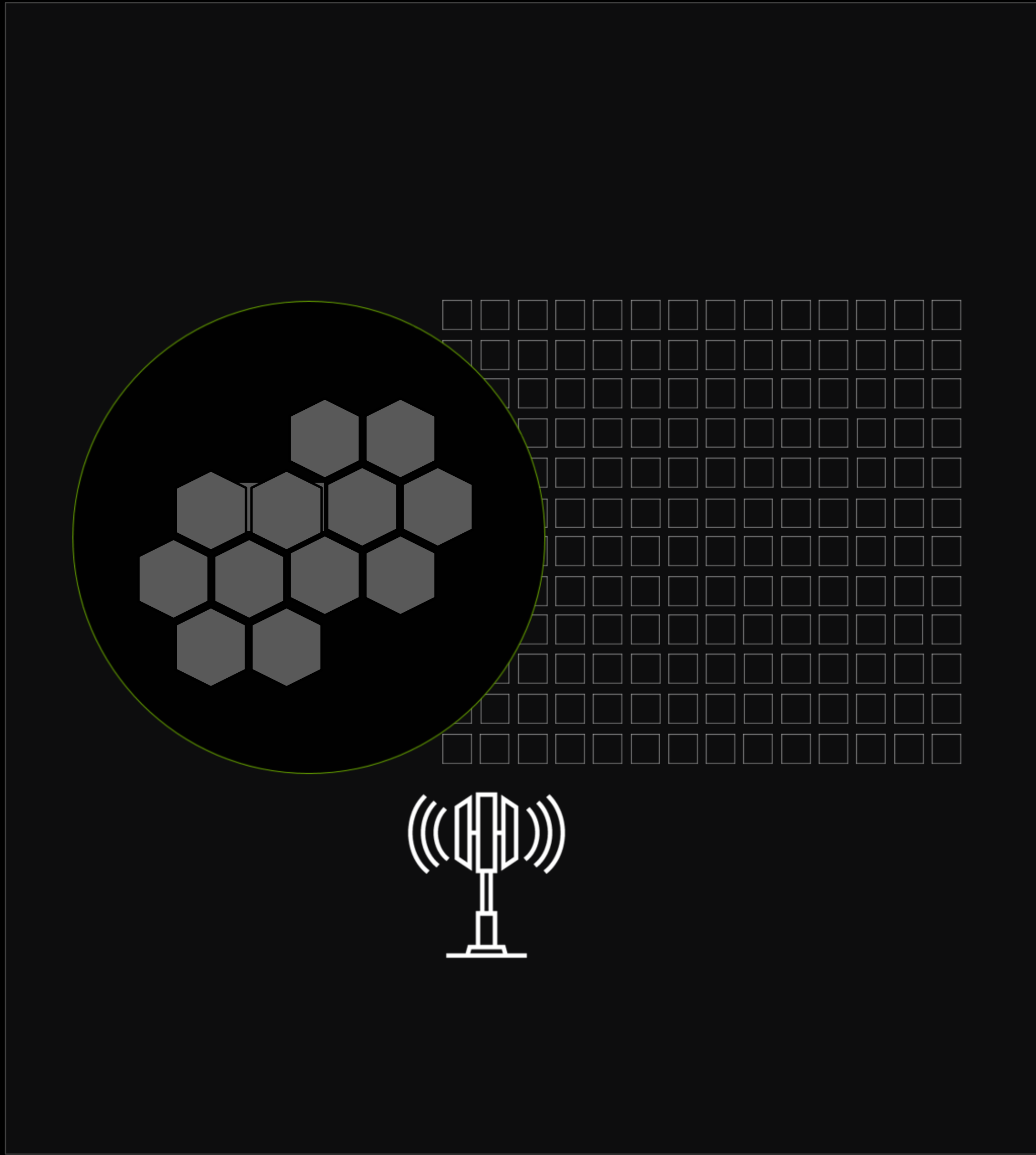
Key Capabilities



Neural RAN Algorithms (AI-for-RAN)
X-Factor Spectral Efficiency & Energy Efficiency



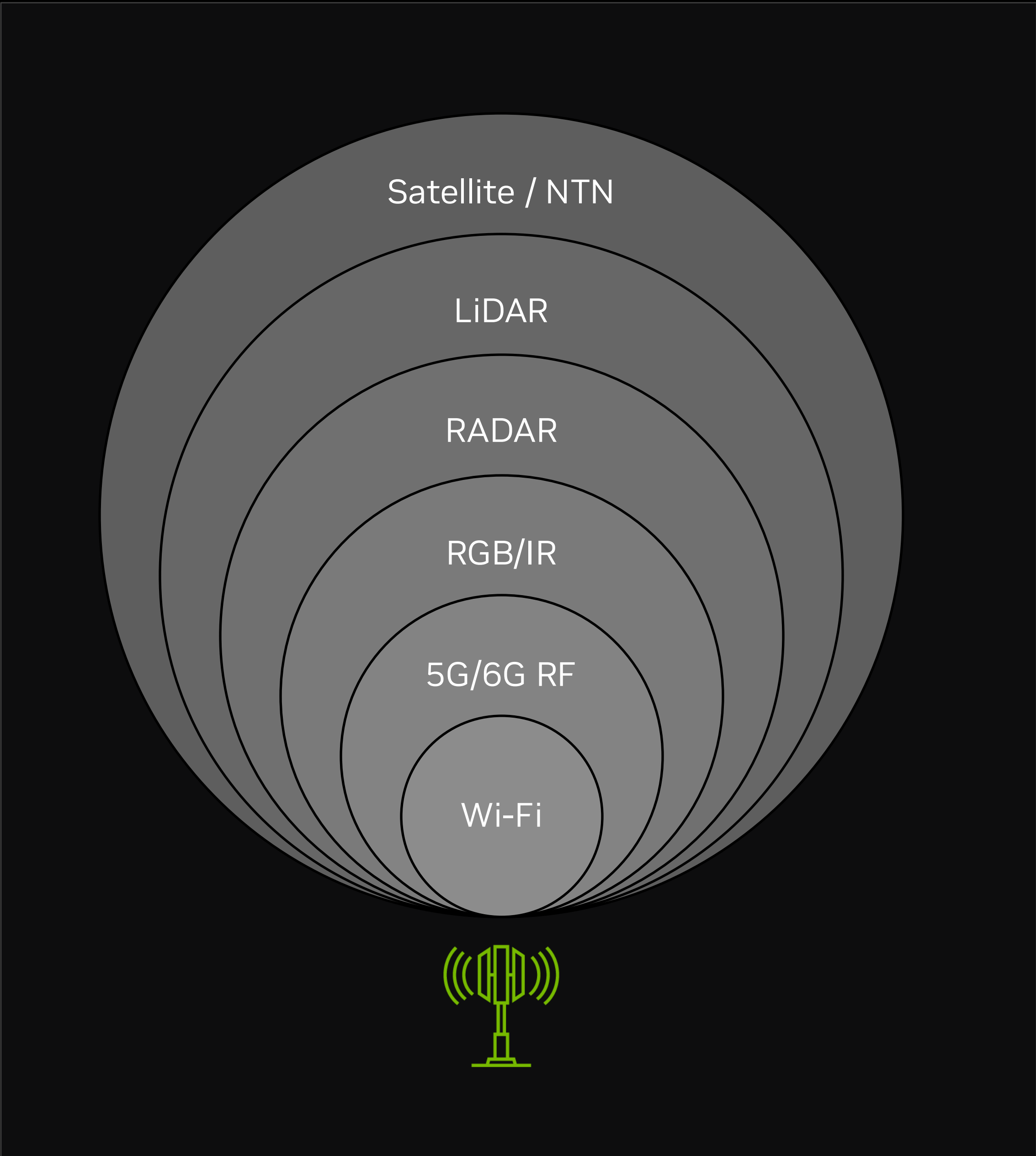
Agentic AI Core and AIOps
Agents to run 6G Network & Automate Operations



Extreme MIMO, Spectrum Aggregation
XL Radio Arrays, Unified Spectrum Giga band

Looking Beyond Connectivity

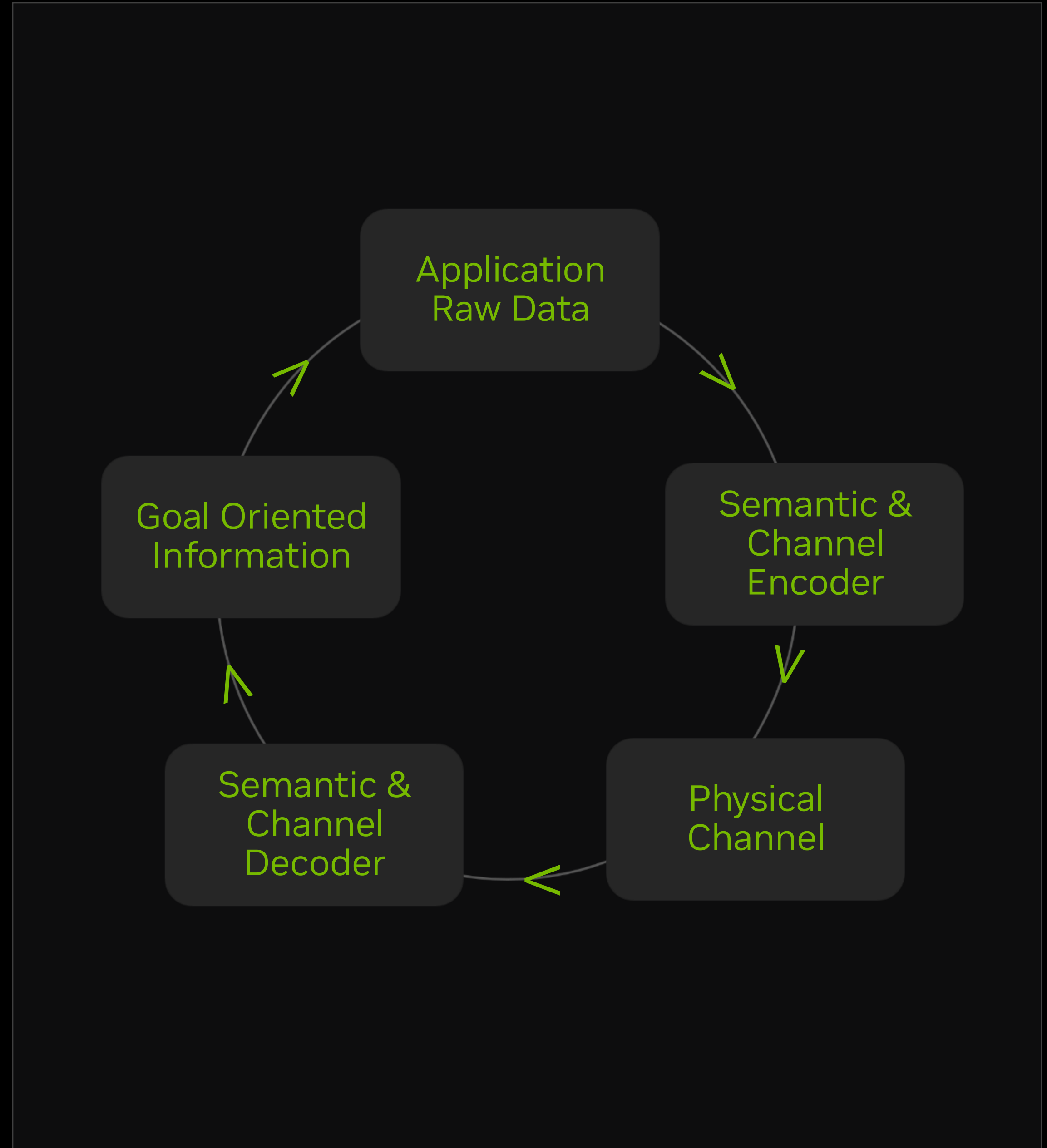
Key Capabilities



Multi-Modal Sensor Fusion
Integrated Sensing and Communications (ISAC)



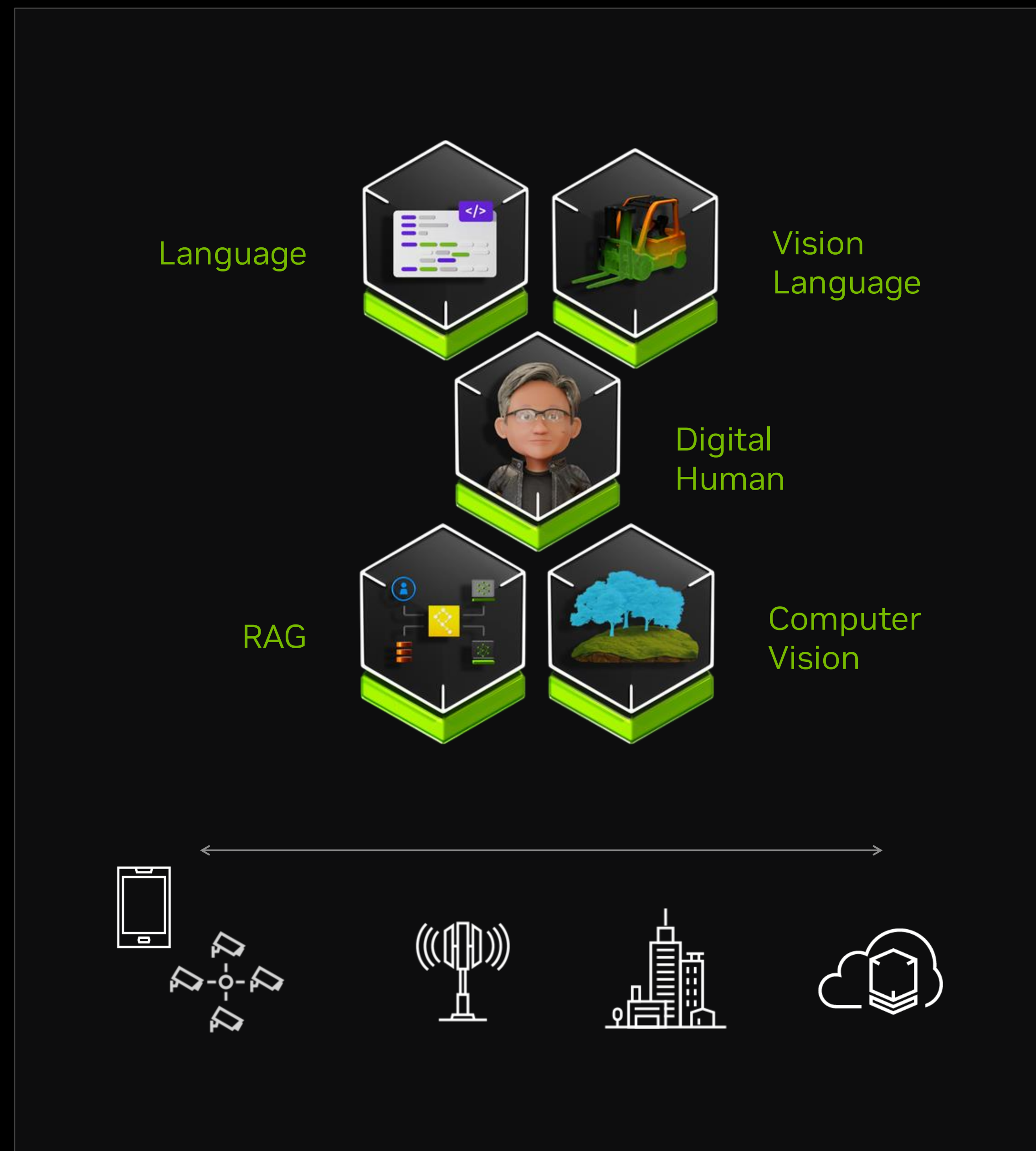
External Apps Integration
Network Exposure, Real-time L1 Integration, Agent Protocols



Semantic Communications
Delivering Meaning, Application specific Optimizations

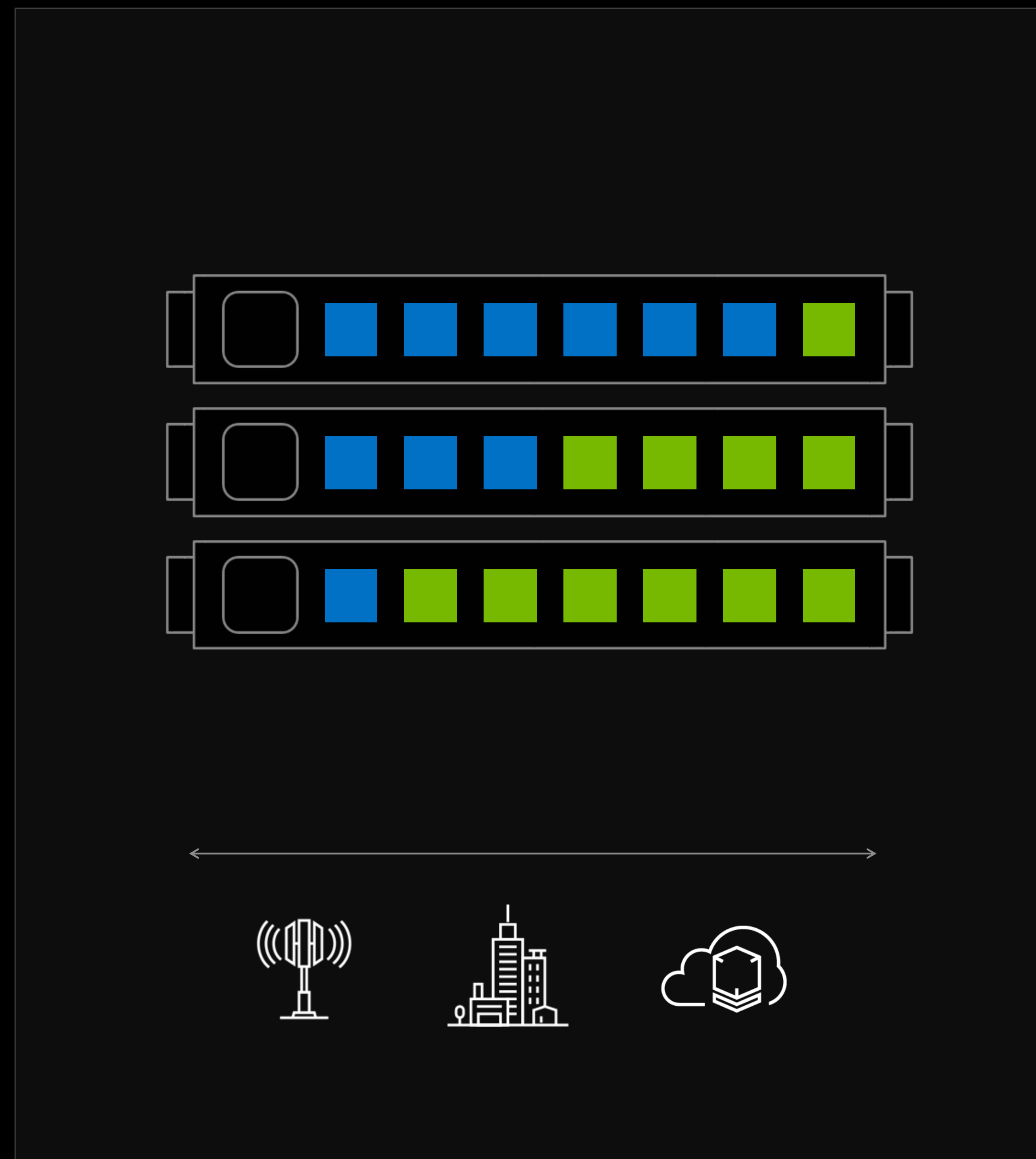
RAN As A Distributed Global AI Inference Engine

Key Capabilities



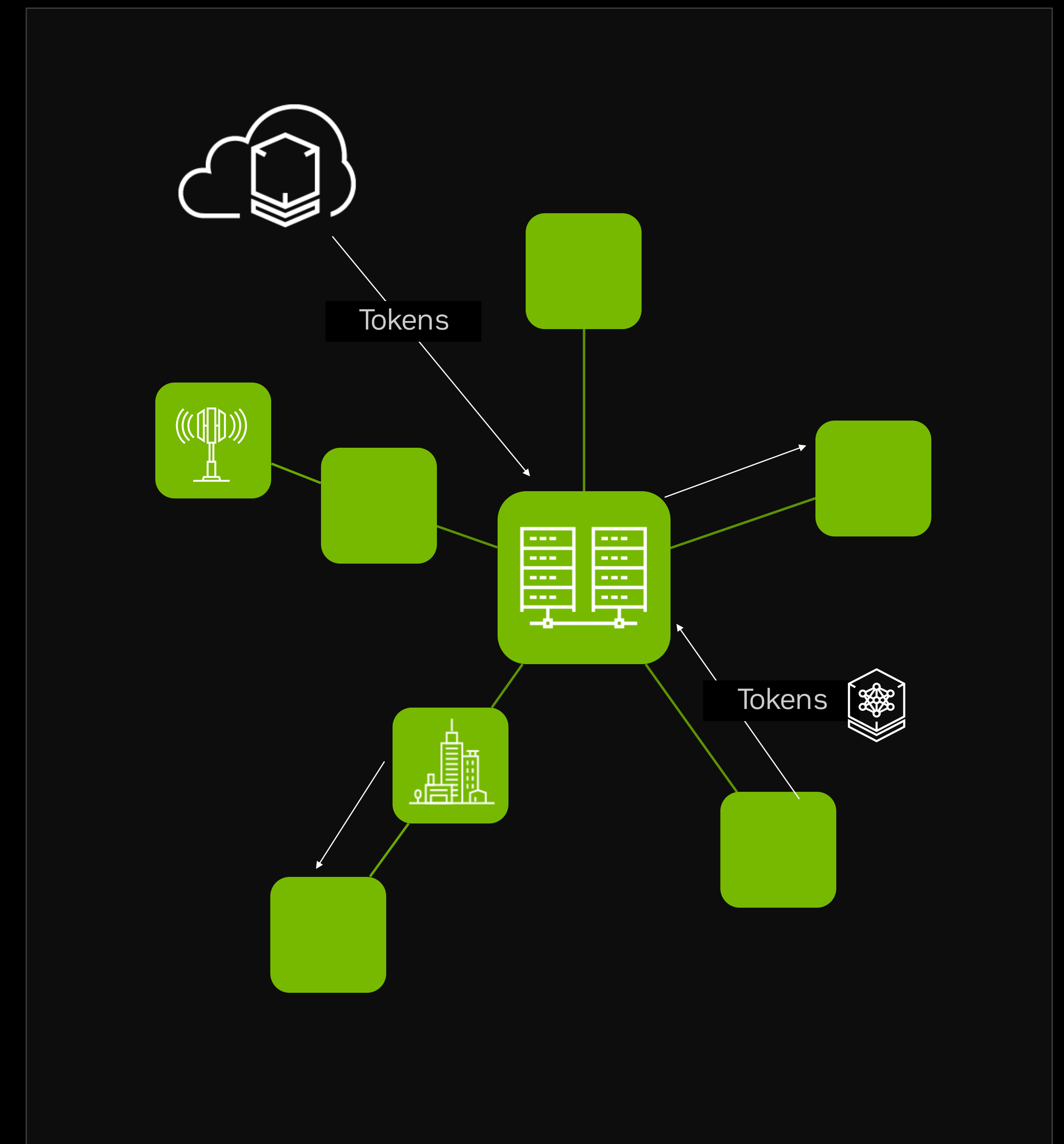
AI Inferencing Everywhere

Stateless Microservices, Optimized for any compute



Multi-purpose AI and RAN Infrastructure

Software-Defined, Any Workload, Any Deployment



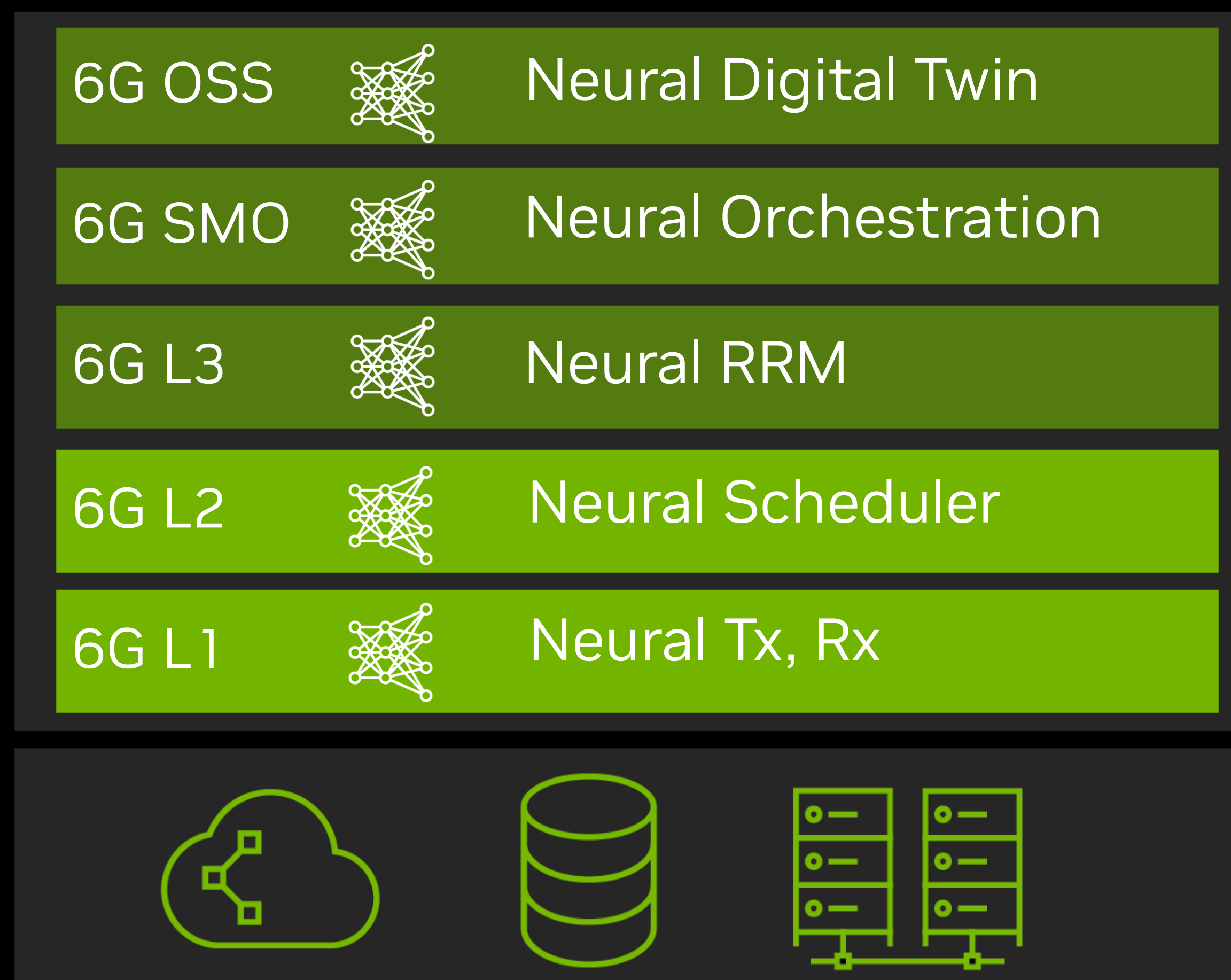
AI Grid Enabling AI-aaS

Intra/Inter-Network Orchestration of AI workloads

NVIDIA's Intelligent RAN Vision

Intelligence in every layer

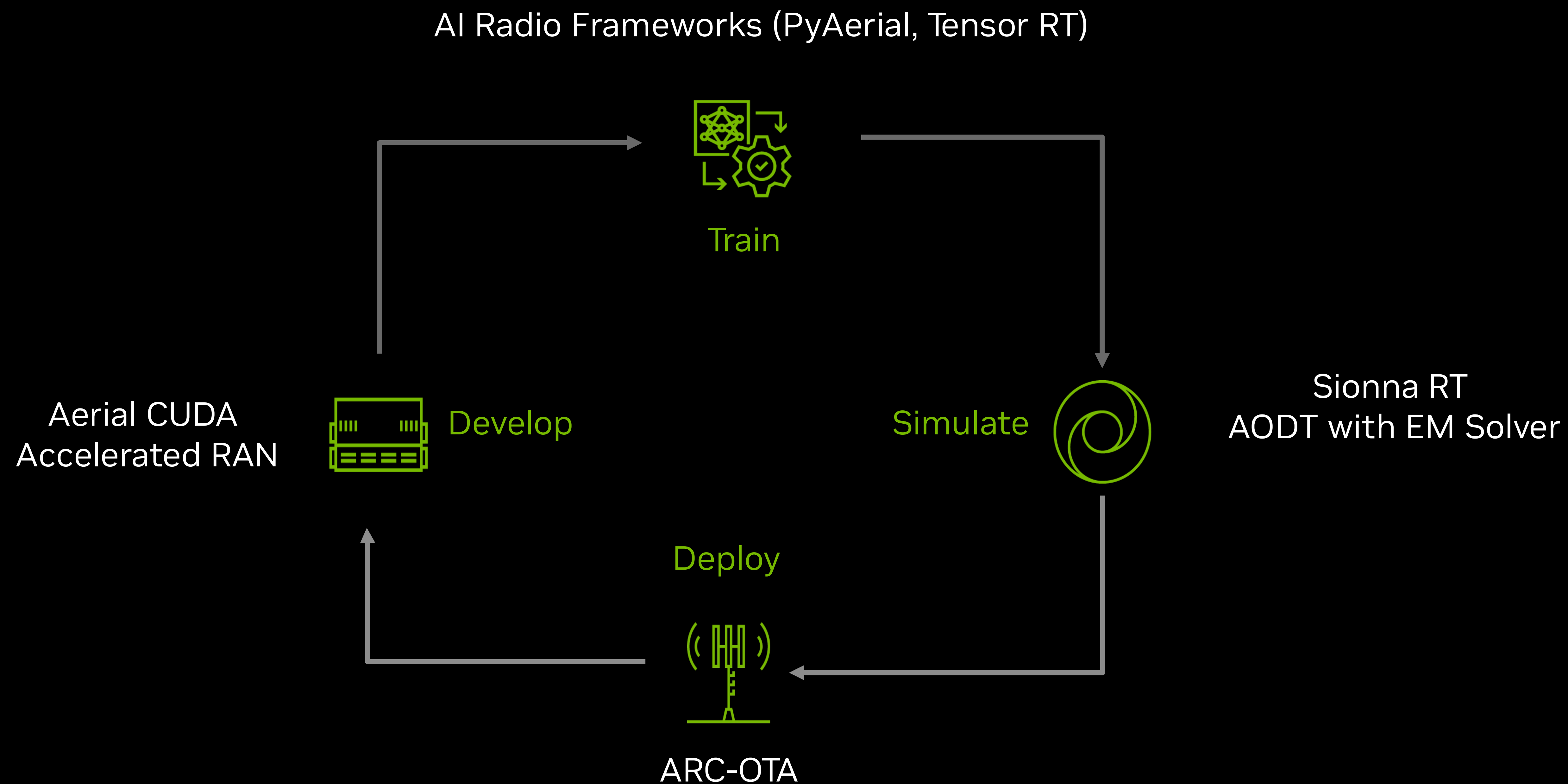
Aerial CUDA-accelerated RAN Stack



- **AI-Native** RAN
- **Site-specific** AI/ML, for:
 - Higher spectral efficiency
 - Lower energy consumption
 - Higher signaling efficiency to reduce latency
- **Accelerated computing** platform:
 - Capable of efficiently running AI/ML workloads
 - Accelerated computing + comms **continuum**
 - Provide multi-tenancy for GenAI and XR apps
 - Enable fast product development velocity
- **Simulation Tools** :
 - SLS and LLS
 - AI/ML
 - Digital Twins

AI-RAN Research Platform

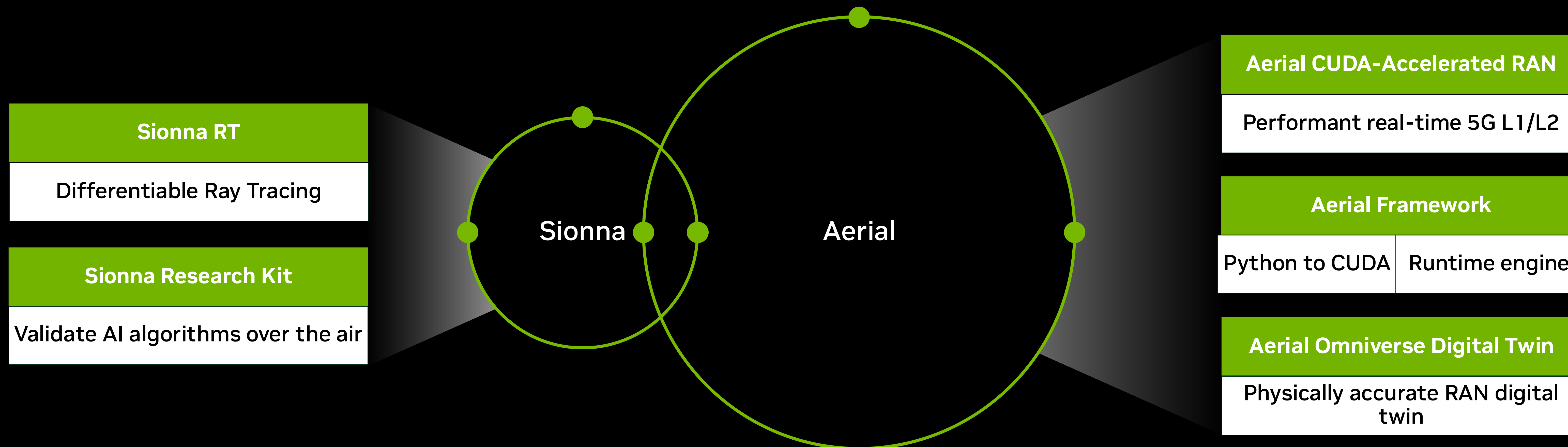
Develop + Train + Simulate + Deploy -> Iterate



<https://developer.nvidia.com/aerial>

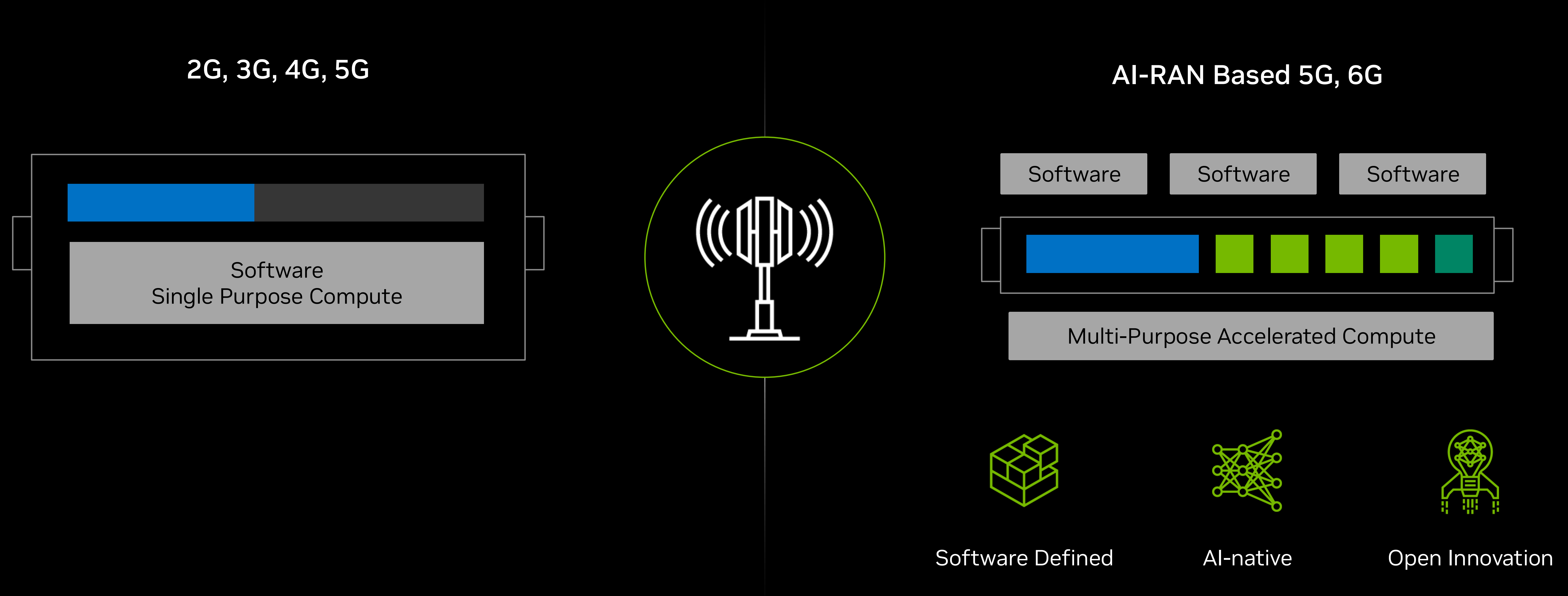
Simulation | Development | Validation

Key Continuous Learning R&D Components



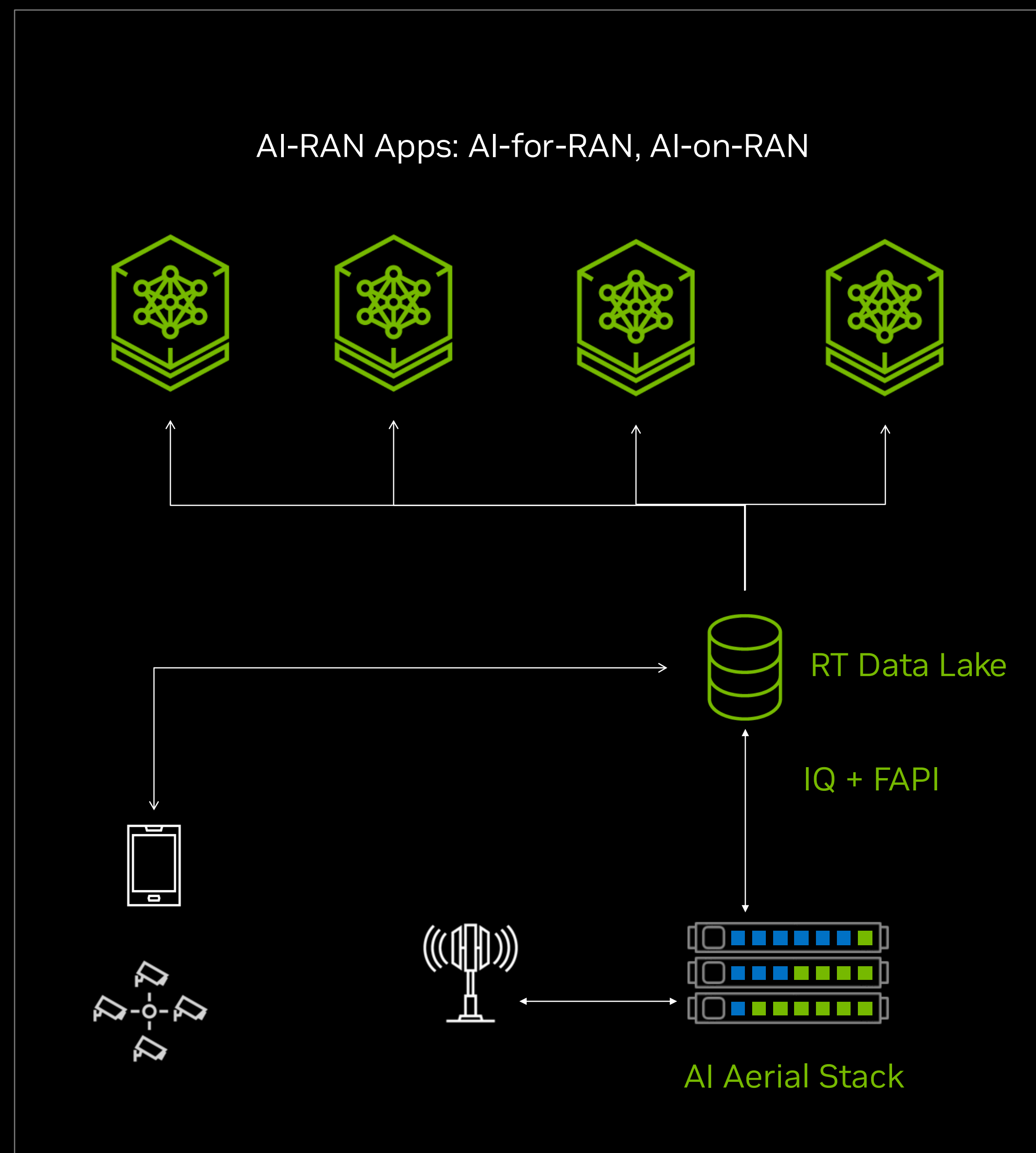
AI-RAN | The New Architecture for 6G and the AI Era

Single Purpose to Multi-Purpose RAN



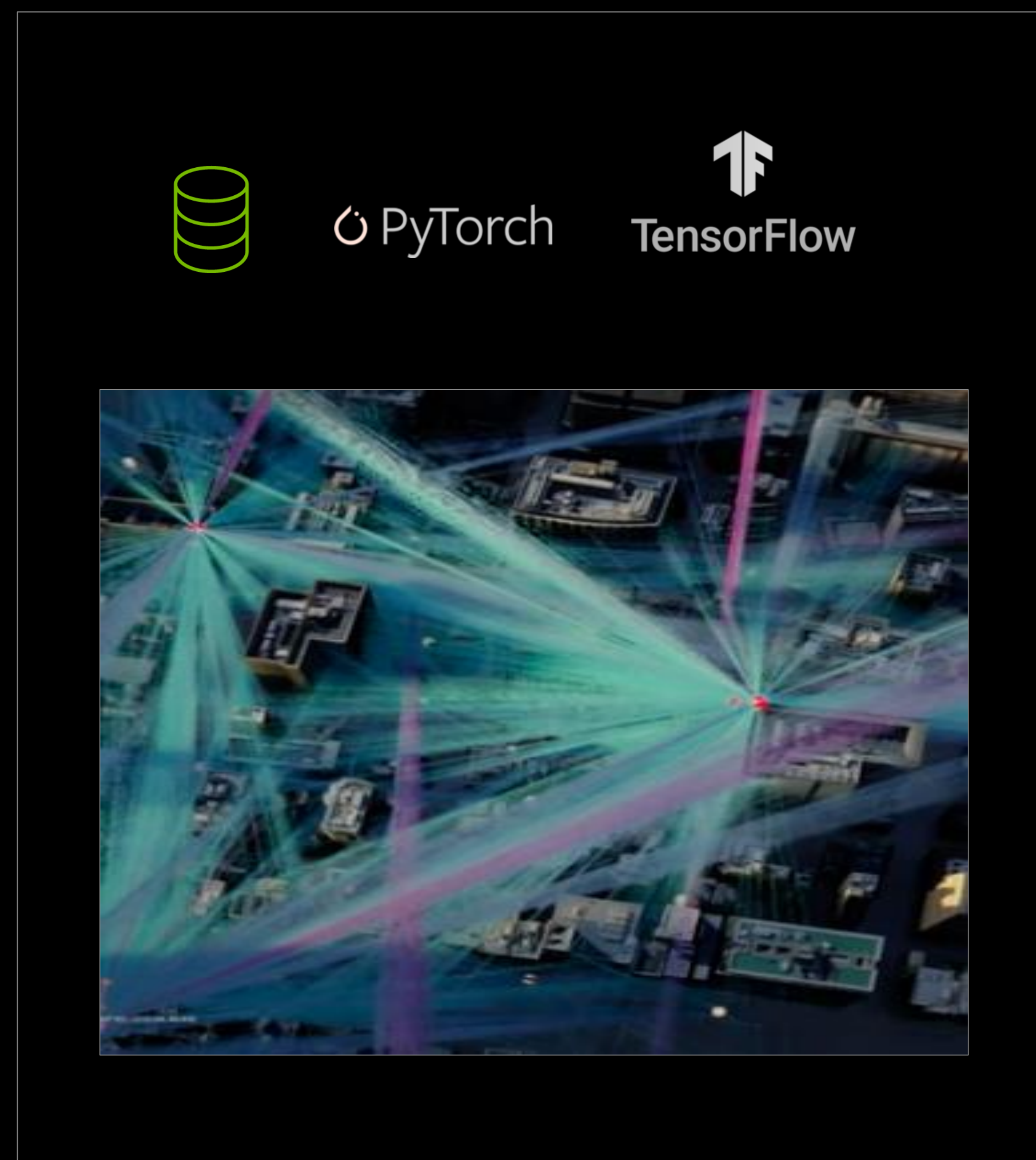
Key Platform and Tools Innovations For The 6G Developers

Open & Source Code Access | Simulation Tools | Python-CUDA Compilers



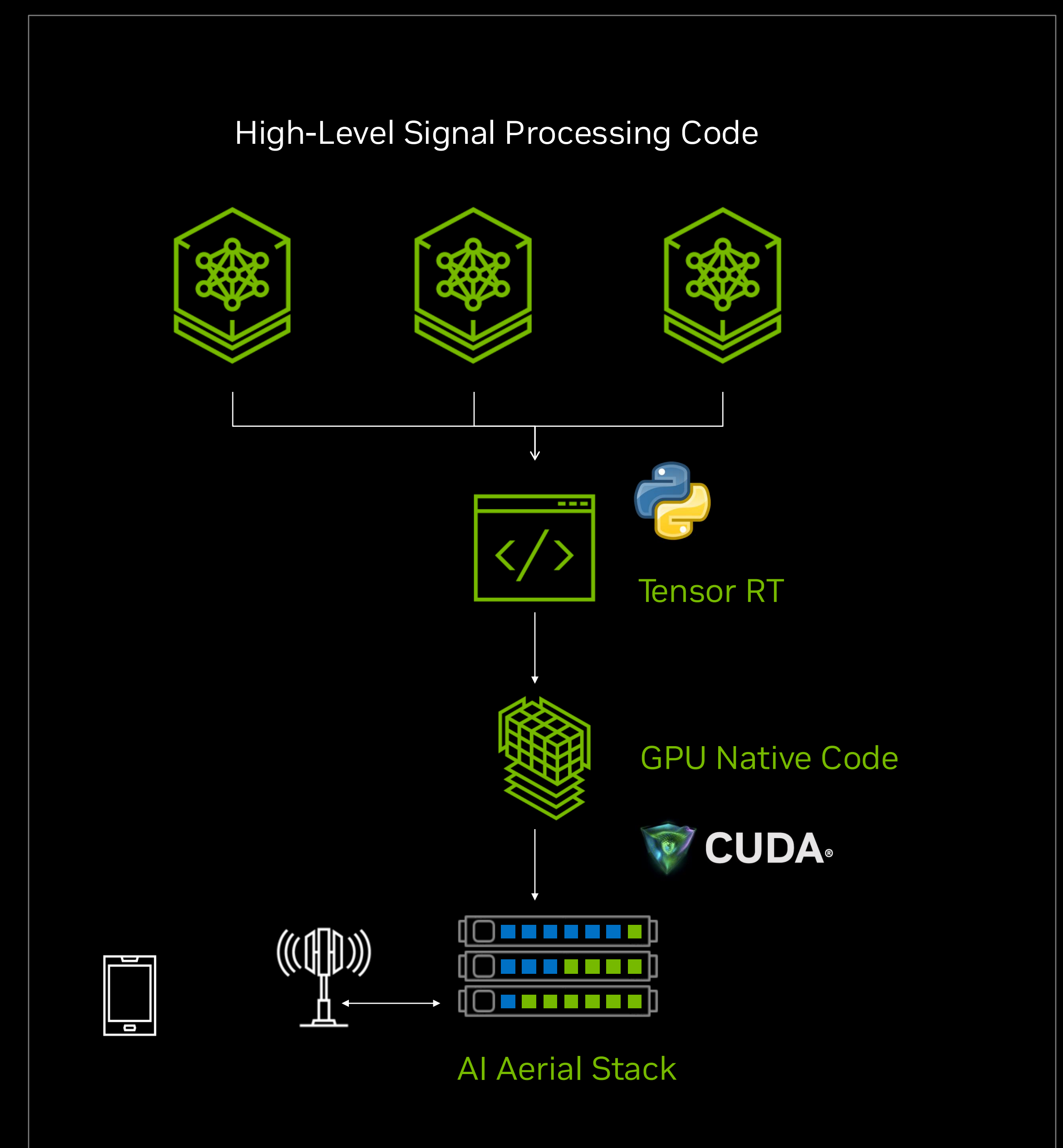
Low Level & Real-time Data Collection

Enables AI/ML algorithm development at any RAN layer



RF Simulation Integrated with AI Tools

Synthetic data generation and AI model validation



Real-Time OTA Validation

Rapid high-level code (Python) to real-time code compilation

NVIDIA AI-RAN Platforms from Research to Commercialization

Continuous Software-Defined Innovation with NVIDIA



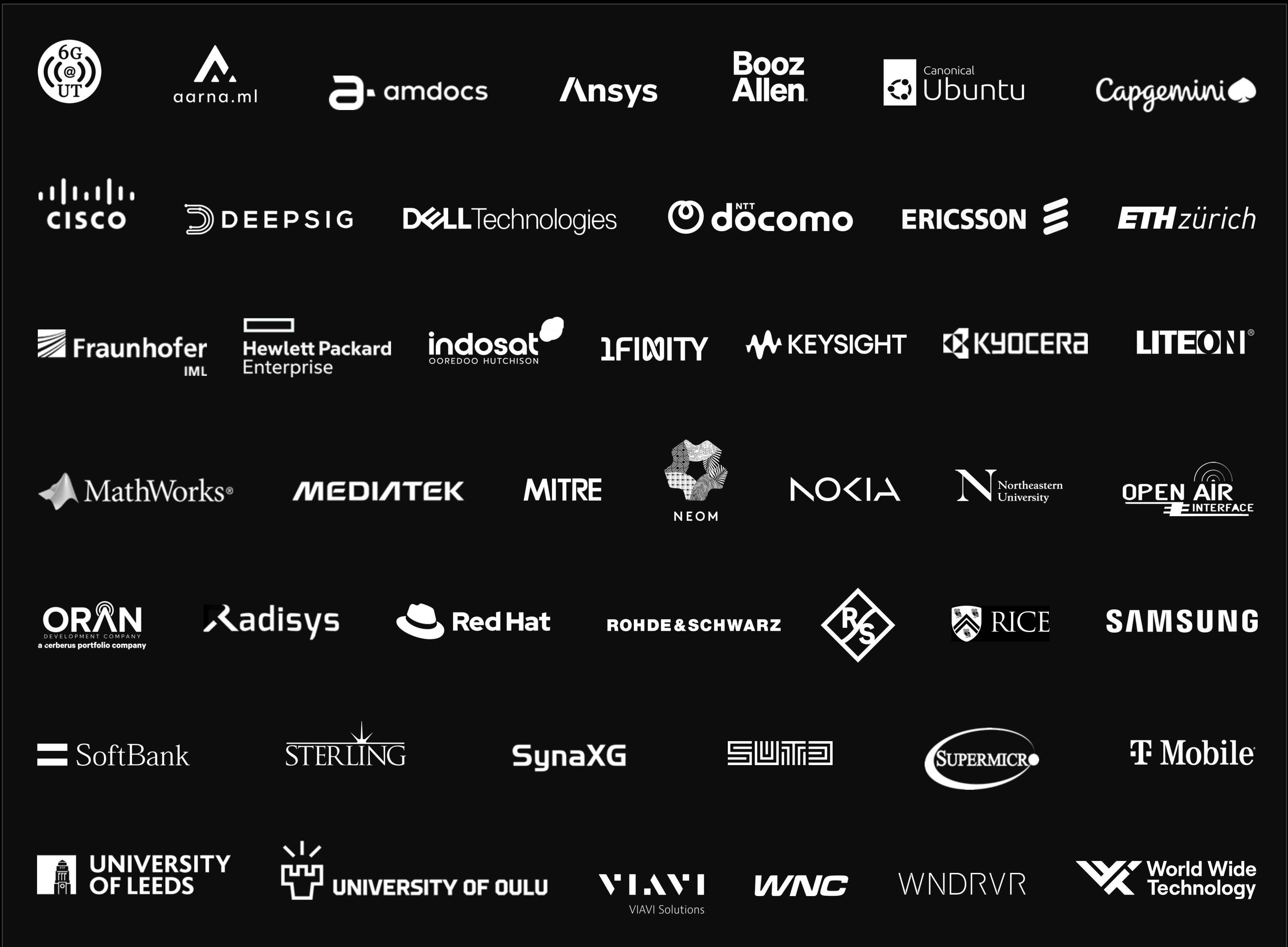
Research Platforms

Commercial Platforms - Aerial RAN Computer (ARC)

Multiple Platform Options for C-RAN/D-RAN through 2030			
CPU	Grace → Vera x86 Options	DPU	BF3 → BF4
GPU	L4 → Hopper → Blackwell → Rubin	NIC	CX7 → CX8 → CX9

Enabling 6G Researchers Everywhere

NVIDIA AI Aerial Research Ecosystem



Research and Industry Collaborations



Industry Bodies

