





DATASHEET

ZEDEDA Edge Intelligence Platform

Intelligence Where It Matters Most



Edge AI deployments have challenges not found in the cloud:

-  Intermittent network connectivity
-  Hostile environments – cybersecurity threats compounded by physical security risks
-  No on-site IT staff, demanding more autonomous operations
-  Legacy systems that often feed telemetry to Edge AI models

The Edge Intelligence Imperative

To unlock AI's transformative value, enterprises must deploy intelligence where their business actually runs.

Manufacturers monitor production lines. Energy companies manage remote assets. Retailers track inventory in real time. Logistics operators run global fleets. The decisions that drive revenue, cut costs, and protect workers don't happen in the cloud. They happen at the edge.

That's where the data is generated. That's where it needs to be acted on. So that's where AI must live.

We call this Edge Intelligence. It's essential for companies with edge operations, because the alternative – routing data to the cloud for analysis and back again – is too slow and too costly to be competitive.

The moment for Edge Intelligence has arrived. Edge hardware is advancing rapidly, with dedicated accelerators such as GPUs and NPUs becoming commonplace. At the same time, AI models have been aggressively optimized for the edge, delivering robust performance while operating within strict compute, power, and thermal limitations.

ZEDEDA

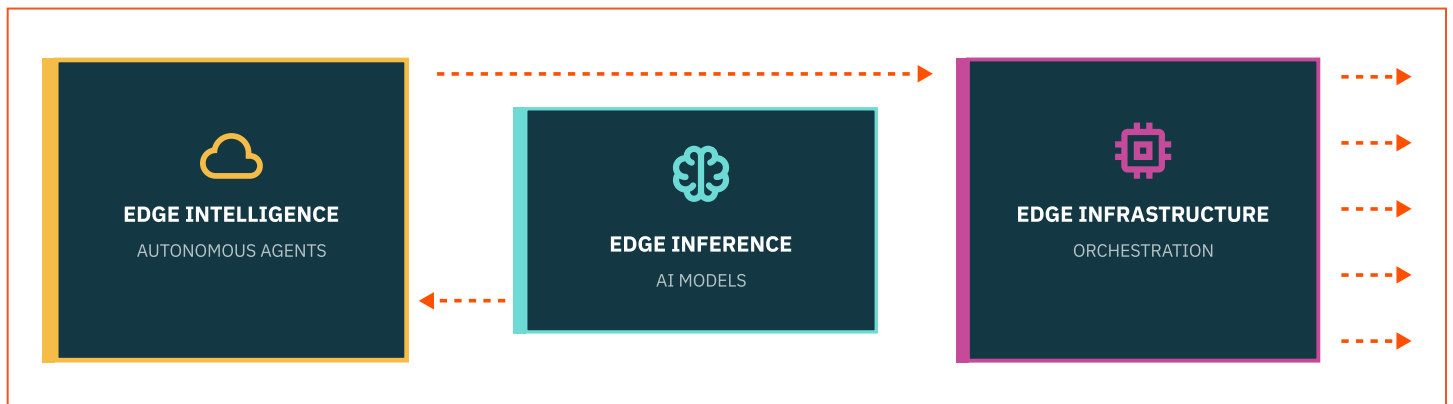


One Platform for Infrastructure, Inference, and Agents

To meet these challenges, ZEVEDA Edge Intelligence Platform delivers:

- ✓ Intelligence via autonomous agents
- ✓ Inference using curated AI models
- ✓ Infrastructure orchestration across your full AI stack

ZEVEDA Edge Intelligence Platform



Edge Intelligence and Inference Services

ZEVEDA Edge Inference Services combines a breakthrough approach to building and orchestrating Edge AI, comprising agents, models, applications, and infrastructure. Using it, Enterprise AI teams can rapidly create, test, and deploy autonomous edge agents with any AI model on any edge hardware platform, then operate them with the same edge orchestration already trusted by the world’s largest enterprises.

ZEVEDA Edge Inference Services includes prevalidated Agent Solutions that bundle everything you need to run an AI workload on the edge: model artifacts, Helm charts, configuration values, deployment metadata, and hardware – enabling you to deliver Edge AI faster with fewer integration surprises.

Edge Infrastructure Services

Your AI is only as reliable as its foundation. **ZEVEDA Edge Infrastructure Services** orchestrates and manages the full AI stack across thousands of edge nodes: agents, models, inference engines, applications, containers, virtual machines, and more.

Software, models, and agents deploy automatically. Operations teams gain full observability and manageability across every site, every device, and every workload. AI workloads run alongside legacy applications on the same hardware, increasing reliability while reducing CapEx.

The ZEVEDA Difference

Rapid Time-to-Value

Prove ROI quickly with curated AI models, validated hardware configurations, and guided deployment workflows that eliminate guesswork.

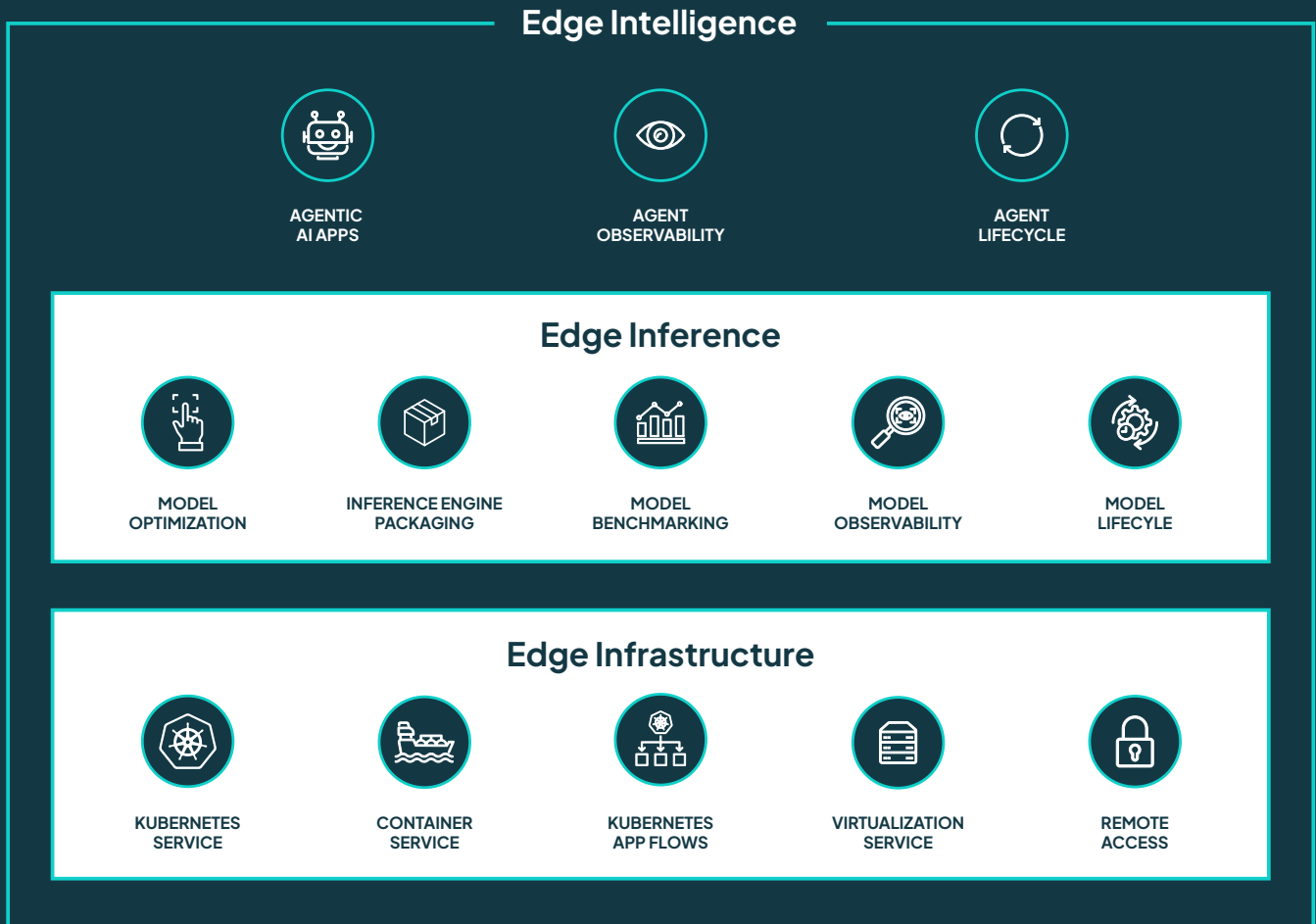
Zero-Compromise Security

Deploy with confidence using hardware-rooted security, encrypted agents and models, and remote attestation across your distributed infrastructure.

Proven at Scale

Scale AI across your edge fleet with a platform battle-tested in the world's most demanding environments. Built to create, deploy, and operate enterprise Edge AI end-to-end. From software to silicon, brownfield apps to agentic models, and everything in between.

Platform Capabilities



Capability	What It Does	Business Value
Agentic Solution Templates	Declarative Helm charts that bundle model, runtime, preprocessing, and business logic.	Accelerate Edge AI deployments with validated solution blueprints.
Comprehensive Model Lifecycle Management	<p>Model versioning, import, and promotion. Ingest models via Jupyter SDK, external providers, or manual upload.</p> <p>Asynchronous import from NVIDIA NGC, Hugging Face, AWS S3, SageMaker, Azure ML, Azure Blob Storage, and MLflow, with integrated credential management.</p> <p>Deploy directly to ZEVEDA Edge Kubernetes clusters via GitOps.</p>	Guarantees traceability, reproducibility, and rapid iteration across the model lifecycle
Inference Engine Packaging	Inference engine packaging for OpenVINO, NVIDIA Triton, vLLM, and Ollama, delivered as Helm charts.	Reduces engineering effort, promotes reuse, and enforces best-practice configurations.
Model Optimization	Hardware-specific model optimization for edge silicon.	Train on high-end cloud silicon, then optimize and run on cost-effective edge devices.
Model Benchmarking	Model benchmarking on NVIDIA and Intel silicon Inference-speed, latency, throughput, and stress testing on a curated device pool including NVIDIA Jetson and Intel NUC.	Delivers data-driven sizing, cost optimization, and SLA assurance before production deployment.
Cloud-Managed Edge Node and Cluster Orchestration	<p>GitOps integration for governance of agents, models, apps, and other workloads.</p> <p>Lifecycle management of edge nodes.</p>	<p>Zero-Touch provisioning of full stack, from agents to operating system, without onsite IT staff.</p> <p>Automated, reliable rollouts and rollbacks across edge fleets.</p>
Virtualization and Container Services	Hypervisor and container runtime management on diverse hardware profiles.	Run legacy brownfield workloads and new, containerized applications, models, and agents on the same device, reducing hardware capital expenditure.

Dashboards, Observability, and APIs	Central UI shows model inventory, organization health, benchmark results, deployment status, and user activity. Alerting, events, resource utilization, and analytics Comprehensive REST APIs	Visibility, reports, status, and performance of hardware, apps, models, and agents. Gives leadership immediate visibility into Edge AI adoption, operational health, and performance across the fleet.
Multi-tenant Organization Model	Isolates models, users, and permissions per organization, team, project, or environment.	Enables secure collaboration, clear ownership, and a clean staging-to-production handoff.
Zero-Trust Security and Access Control	Secure by design. Hardware root of trust via Trusted Platform Module (TPM). Measured boot and remote attestation. Cryptographic device identification. Data encryption at rest and in transit. Distributed firewall for every app. Physical security via port deactivation and isolation. Role-based access control (RBAC).	Protects edge assets, including PII from video feeds, model weights, and application code, against both cyber and physical threats.
App and Model Marketplace and Catalog	Partner-provided and ZEDEDATA-curated solutions available in a unified catalog.	Simplifies discovery and deployment of validated Edge AI solutions.

To learn more about how we can help you deliver intelligence to the edge, [contact us](#) today.



About ZEDEDA

ZEDEDA unlocks the value of AI where it matters most, enabling enterprises to create, secure and operate edge AI at scale. ZEDEDA's Edge Intelligence products and solutions are used by global distributed enterprises to rapidly realize and deploy autonomous intelligence wherever they operate, turning real-time data into real and tangible business outcomes and decisions. Trusted by the world's largest organizations, ZEDEDA is backed by world-class investors, with teams in the United States, Germany, India, and the United Arab Emirates. For more information, visit www.ZEDEDATA.ai.



CONTACT@ZEDEDATA.COM