

# WEDA: WISE-Edge Developer Architecture

Driving Scalable Edge AI Solutions  
from Complexity to Speed

- ／ Overview
- ／ Edge AI Ecosystem
- ／ Strategic Partnership
- ／ Partner Collaborations

# WEDA



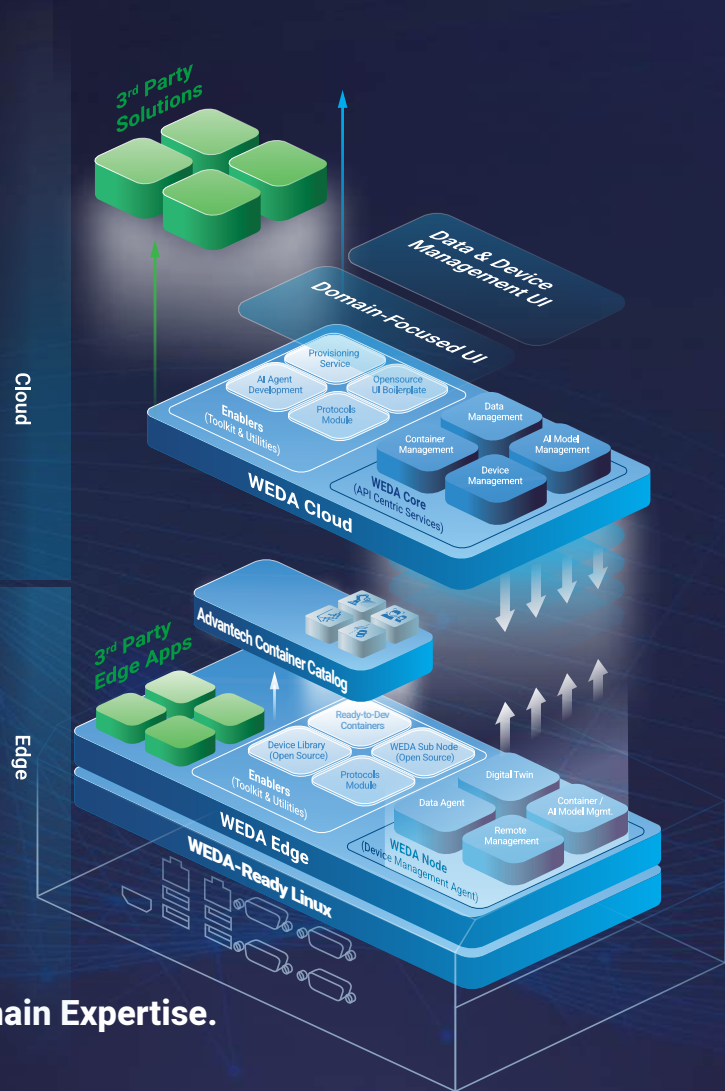
**ADVANTECH**

*Enabling an Intelligent Planet*

# WEDA: WISE-Edge Developer Architecture

## The Strategic Blueprint for Edge Intelligence

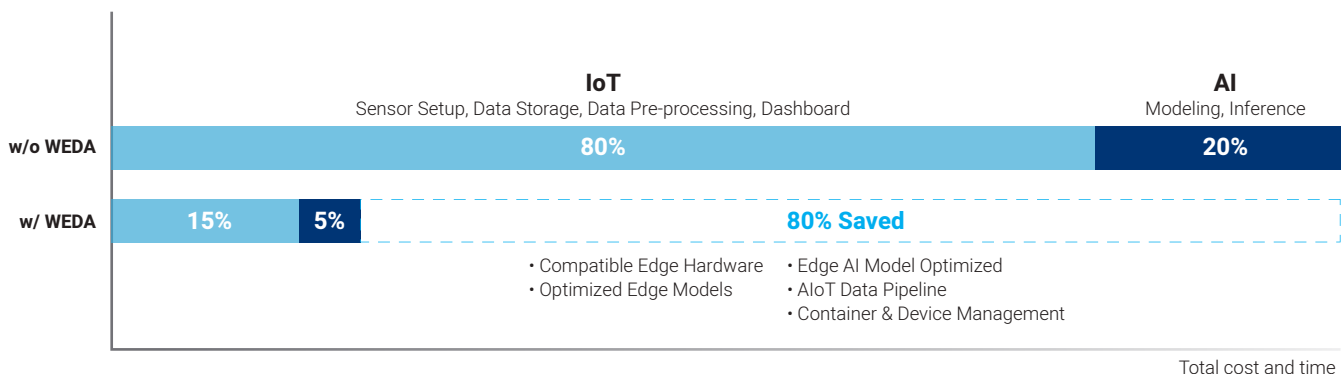
- From Connective Chaos to a Standardized Stack**  
 By integrating WEDA Edge and WEDA Cloud, we provide a complete, end-to-end architecture.
- From "Ready-to-Dev" to "Ready-to-Use"**  
 Leverage the Advantech Container Catalog (ACC) to bypass the complexity of infrastructure setup.
- Openness | API-Driven | Toolsets**  
 Openness of WEDA Edge, the APIs of WEDA Cloud, and the toolsets of ACC.
- AI-Agent Friendly Ecosystem**  
 Optimized for AI-Agent driven development. Standard interface with skills to automate full-stack logic.



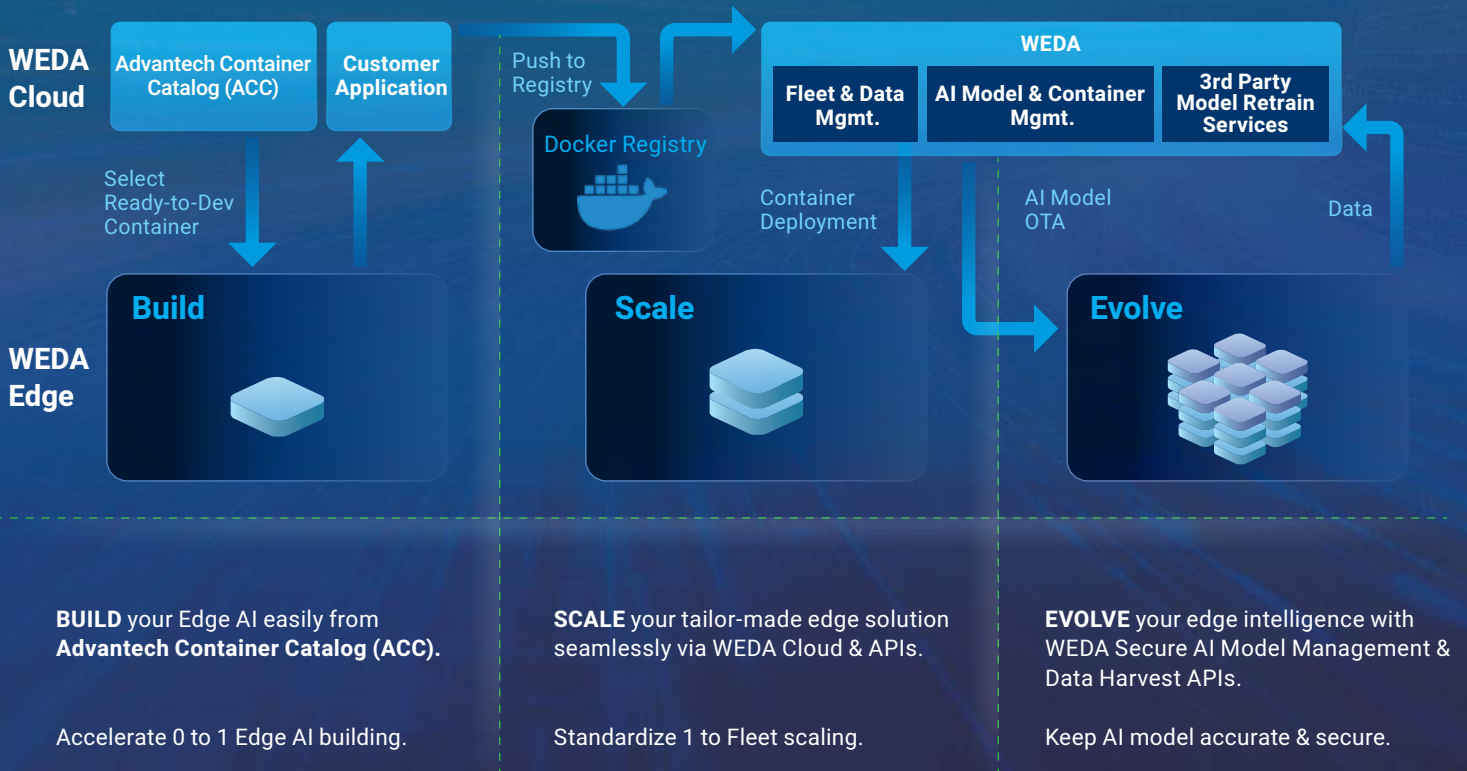
**Decoupling Complexity to Empower Your Domain Expertise.**

## Build Your Own Solutions – 80% Faster with WEDA

Accelerating Edge AI Development from Prototype to Production.



# WEDA User Journey: Your Highway to Production-Ready Edge AI Efficiency via Containerization & API Architecture



## Edge AI Ecosystem with WEDA Cross-Chipset Support

Cross-Chipset architecture across CPU, GPU, NPU, and DSP, enabling efficient, flexible, scalable and development.



### Advantech Container Catalog (ACC)

#### Explore Containers to Build Your Solutions

Advantech Container Catalog (ACC) provides ready-to-develop containers – from computer vision to Edge LLM, accelerating AI agent and vertical solution development.

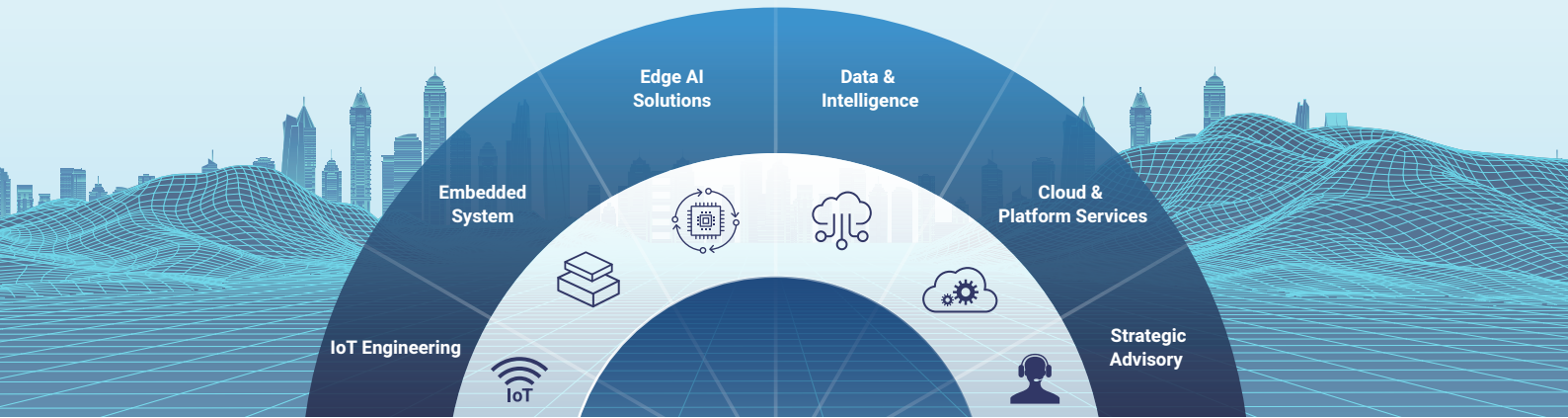
Scan NOW !



# A Strategic AI Partnership – Edge Intelligence Meets Digital Ingenuity

Advantech has formed a strategic AI partnership with Nagarro, a global digital engineering and consulting leader. Nagarro complements Advantech's hardware leadership with deep expertise in software development, including AI, embedded systems, and IoT-powered innovation. Together, we're building intelligent, secure, and adaptive ecosystems - where edge intelligence delivers real-time insights, smarter operations, and connected enterprise growth across industries.

## Our Expertise Spans Across The Edge AI Value Chain



## Turning Edge Intelligence into Real Success



### Intelligent Vehicle Solution for an Automotive Leader

#### Automotive

Nagarro helped Hitachi Astemo monetize in-car data and improve driver experience with an Edge solution that captures real-time on-road information (weather, pothole, debris, etc.) and shares insights among subscribers and partners.



### Predictive Maintenance for a Manufacturing Leader

#### Manufacturing

Nagarro built a smart factory platform that processes sensor data locally to detect anomalies and predict mechanical deterioration before failures occur, reducing downtime by 15% and costs by 20%.



### AI For Safety: Privacy-First Monitoring of Passenger Safety in Railways

#### Transportation

Nagarro's privacy-first monitoring platform enhances passenger safety on platforms and in vehicles. It tracks anomalous behavior and crowd density without compromising privacy, delivering a safer and better public transport experience.

## Four Steps to Accelerate Edge AI Solution Delivery

Enabled by Advantech **WEDA**, Crafted by **Nagarro**

### 1 Data Capture

- Seamless integration with every industrial protocol
- Instant data, ready for AI

### 2 AI Training

- MLOps built-in, training made effortless
- AI shaped by real-world domain knowledge

### 3 Deploy & Scale

- OTA update any AI model & application
- Update unlimited devices with one-click

### 4 Inference & Self-Learn

- AI-driven device control
- Smarter AI models with continuous evolution



# Plug-and-Play Service Capability for Airports, Railways, Metro Stations & Large Public Venues



An offline, voice-only AI assistant deployed on kiosks that helps passengers navigate, discover services, and get real-time contextual guidance – without screens, typing, or internet connectivity. Built on edge AI, the solution ensures **zero dependency** on cloud, **near-instant responses**, and **complete privacy compliance**.



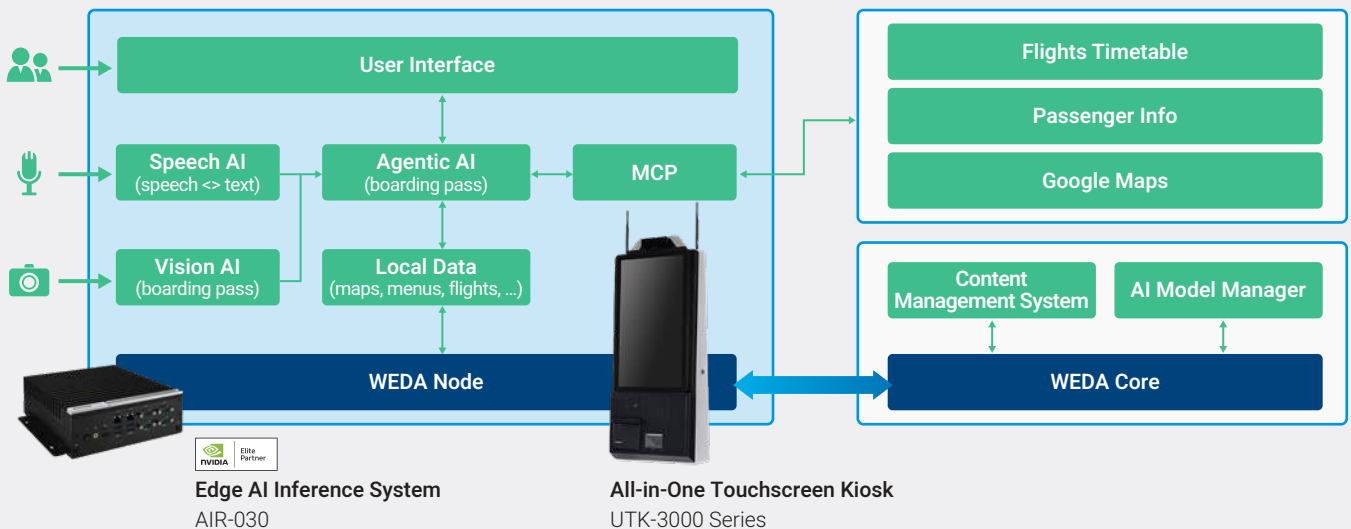
## Benefits

- Delivers fast, inclusive, assistance for all travelers.
- Reduces staff load and improves traveler service experience.
- Strict on-device processing and privacy-by-design to protect personal data.

## Nagarro x WEDA Architecture

- AI empowered service concierge tailored for hospitality in airports, hoteliers and more.
- Fast track solution development with WEDA Ready-to-Dev containers.
- Ensure service continuity and quality with remote updates of content and AI.

### AI-Powered Service Concierge



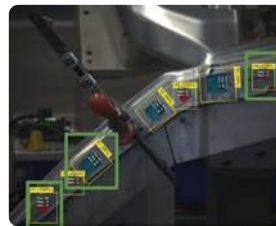
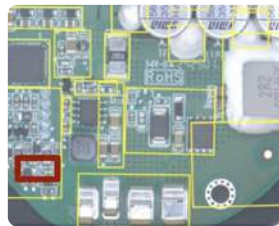
# Edge-Native Vision Built for Real Challenges in Modern Manufacturing



Overview.ai's **OV20i** and **OV80i** are ruggedized smart camera systems that bring advanced computer vision directly to industrial environments. Designed for quality engineers, automation specialists, and manufacturing teams, these devices eliminate the complexity of traditional vision systems while providing enterprise-grade AI capabilities.

## Benefits

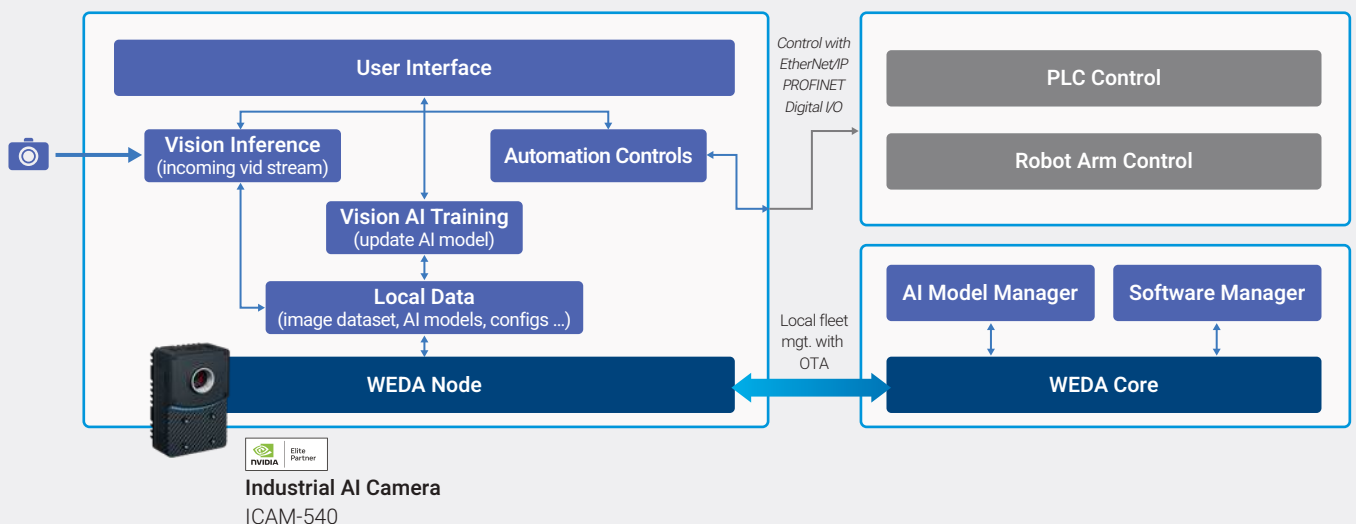
- **Zero Cloud Dependency:** All AI processing happens locally for maximum reliability.
- **Industrial-Grade Connectivity:** Native Ethernet/IP and PROFINET integration.
- **Self-Training AI:** Continuous learning from production data.
- **Plug-and-Play Setup:** Recipe-based configuration, no programming required.



## Overview x WEDA Architecture

- Secure tunnels for managed servicing & OTA.
- Device fleet management for live system health checks.
- Data infrastructure for IoT data, enabling continuous Edge AI accuracy.

## All-In-One Vision Analytics

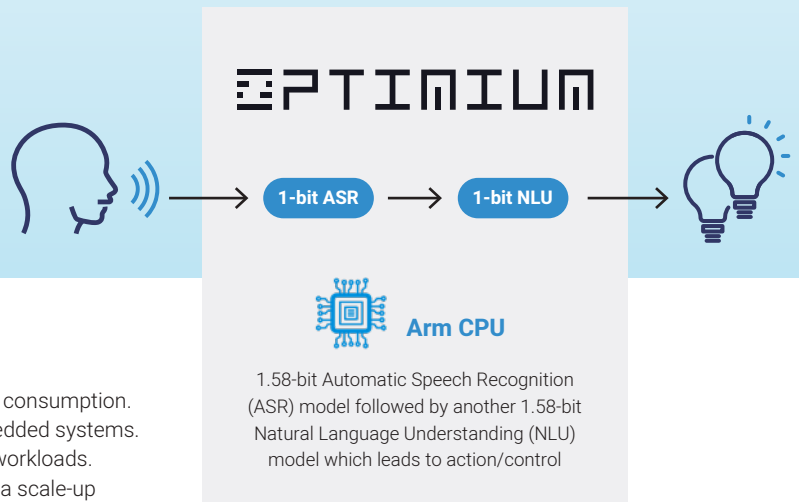


# Smart Home Service Provider Enables High-performance Voice & Language Models With Minimum Memory & Power Usage

Deactivate the age limit function



Mass-deployed our voice control pipeline which uses **<100MB** of memory on approximately **1.2 million set-top boxes** in Korea. As such, ENERZAI achieves Pareto-optimal trade-offs between efficiency and accuracy – deploying speech, language, and vision models optimized for Arm-based SoCs.

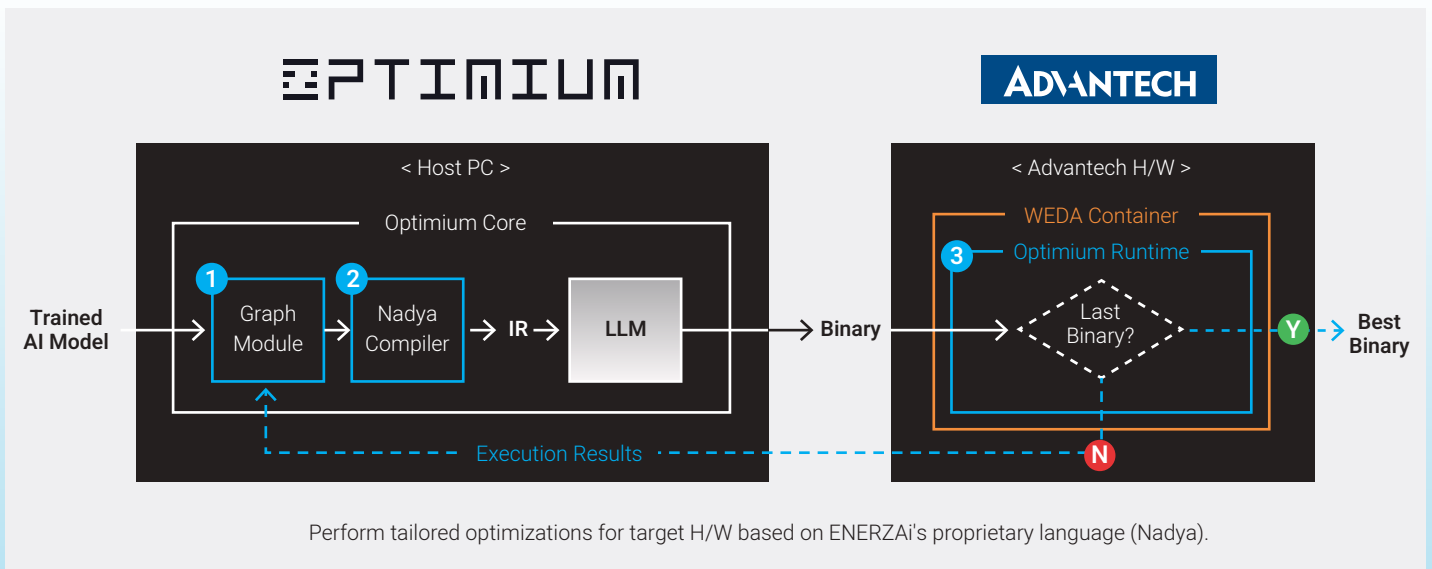


## Benefits

- High-performance voice control with minimal memory & power consumption.
- Extreme low-bit (1-bit) inference for resource-constrained embedded systems.
- Run entirely on CPU, enabling NPU to be allocated for other AI workloads.
- Integrated with WEDA Node, featuring seamless device and data scale-up management.

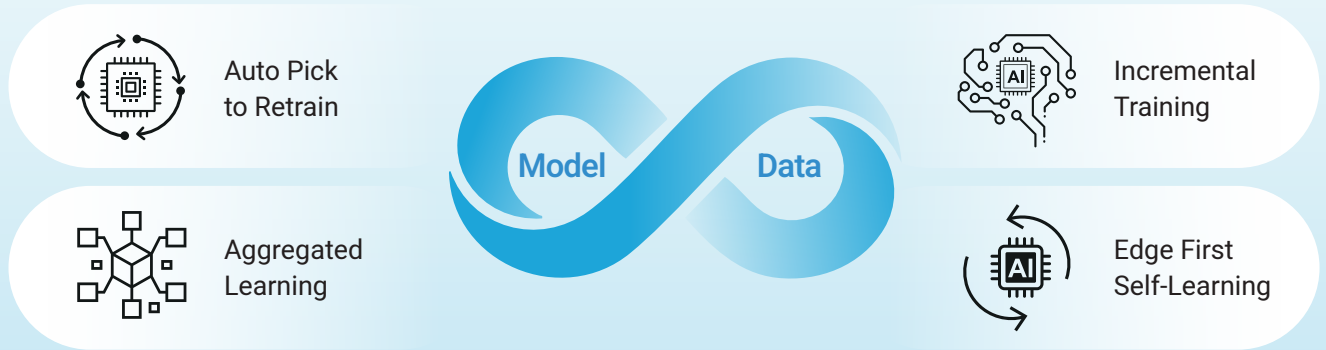
## Optimum x WEDA Workflow

- Above 1-bit models are powered by ENERZAI proprietary AI compiler, Optimum.
- Optimum plays key role in enabling extreme low-bit inference as other inference engines rarely support sub-4-bit inference.
- Deploy ENERZAI low-bit models alongside WEDA Ready-to-Dev LLM containers on Advantech Edge hardware for fast Voice AI Agent integration.



# Vision AI on Edge, Perpetual Business Intelligence

## eAI-Craft Edge Slim-fit Toolchains



### X10 AI Model Regressions, 9.5X Boosted

10 Regressions	Man-made AI Mode Builds (Man made & QC)	eSL-Craft Automated (AI-made & MAN QC)
<b>Dataset Processed</b>	<b>50K</b>	<b>650K</b>
Labeling	12	22
Filter corner case	6	6
Data augmentation	3	3
Model training	6	6
<b>Total Man-month</b>	<b>27</b>	<b>37</b>
Man-month per Dataset	0.5	0.06
<b>Acceleration</b>	<b>9.5x, productivity boosted</b>	

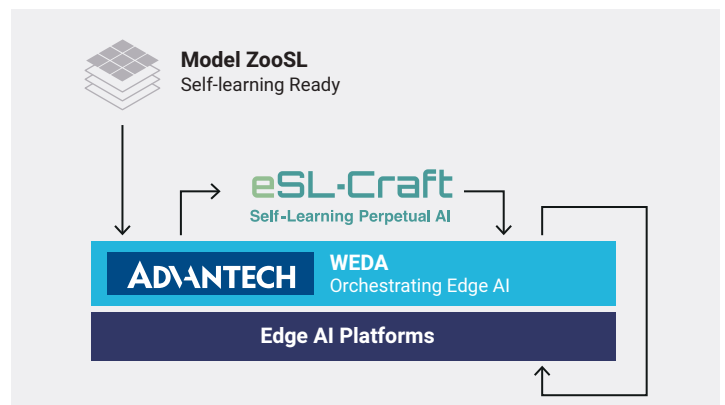
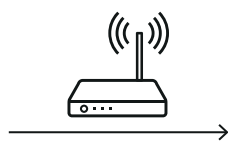
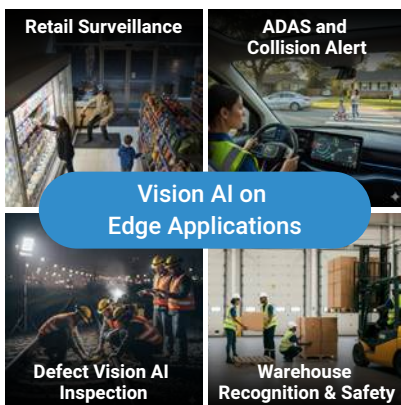
Automated annotation/ filter processes by eSL-Craft pipeline.

Pruning and Quantization tools to enhance model optimization for Edge.

10 manpower saving, up to \$80K per project. (blended \$8K avg. man-month cost for US)

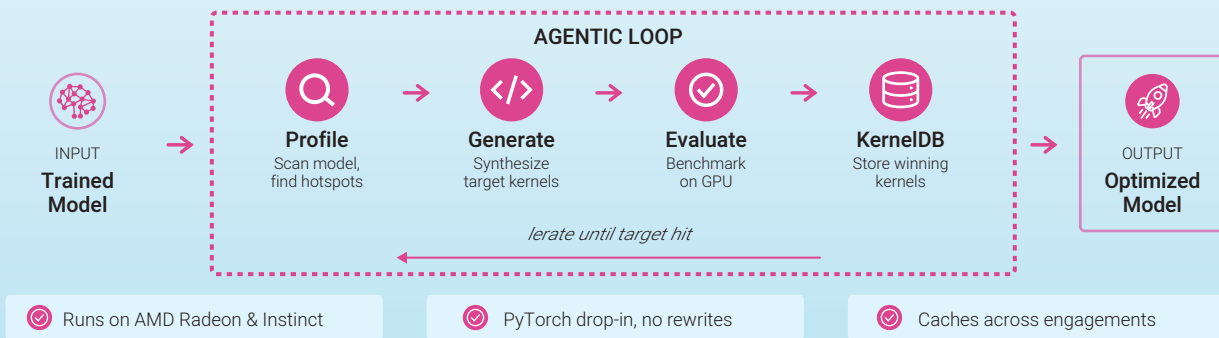
### eSL-Craft x WEDA Workflow

- eSL-Craft integrated with WEDA flow, providing easy access to IoT data.
- Supplemented with Model ZooSL to enable "one click plug-and-play".
- Self-Learning with GPU on edge AI devices for continuous model accuracy without major re-training effort.



# Closing the Gap Between Hardware and AI

YASP automatically compiles any model into hardware-tuned kernels, while AI agents profile, optimize, and validate performance on real silicon — eliminating CUDA lock-in and the need for code rewrites, while delivering accelerated execution efficiency at 67% lower cost on high-compute platforms.



## Benefits

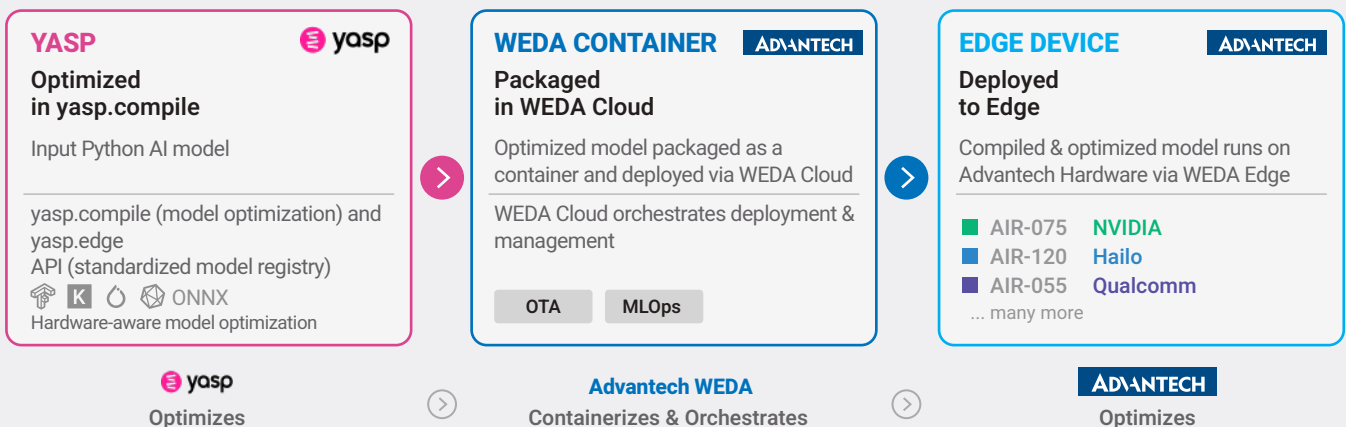
- **Run anywhere, rewrite nothing:** Same model, different hardware flag. No framework changes, no new runtime. Minutes, not months.
- **Real performance, not generic code:** YASP generates kernels tuned to each chip's architecture. Hardware-specific, verified on real silicon, compounding over time.
- **No compiler team required:** AI agents replace the scarce compiler engineers. Every new target, every model revision; handled automatically.

## YASP x WEDA Workflow

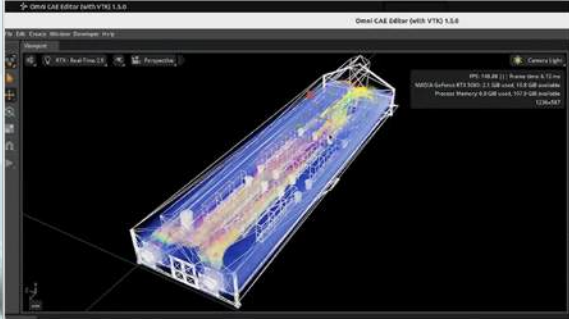
- **Optimize:** Feed any PyTorch model into `yasp.compile`. AI agents generate hardware-tuned kernels for the target Advantech device and register the optimized model via `yasp.edge` API.
- **Package & Deploy:** Containerize the optimized model and WEDA Cloud pushes it to edge devices over-the-air. No manual setup, full MLOps orchestration.
- **Run on Edge:** The compiled model runs natively on Advantech hardware - Jetson, Hailo-8, Qualcomm, and more — managed and updated through WEDA.



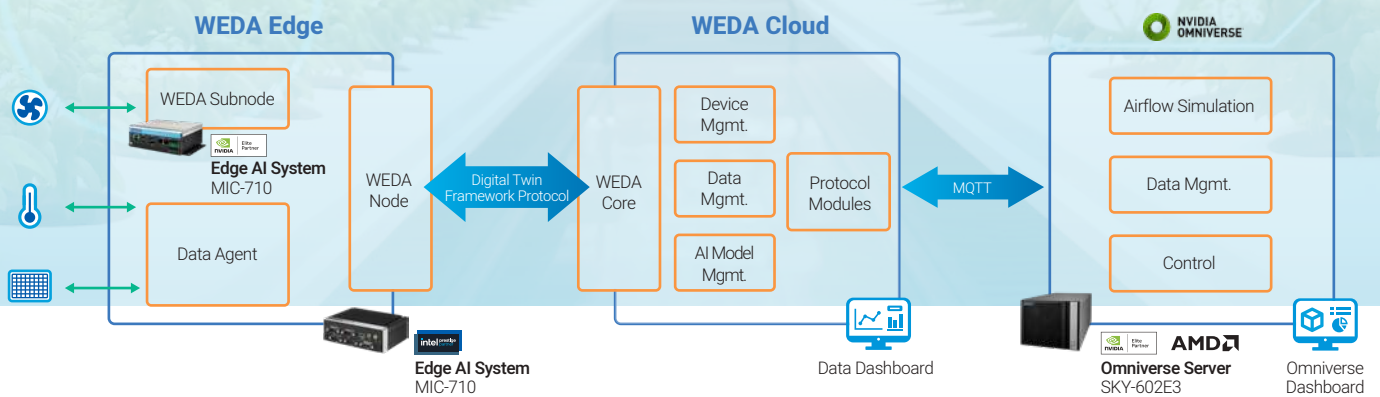
## From Model to Edge: One Pipeline



# WEDA Digital Twin for Smart Greenhouse



Through the integration of **WEDA and NVIDIA Omniverse™**, we have transformed greenhouse management from reactive monitoring to predictive 3D simulation. By optimizing airflow and device orchestration with AI, we can now prevent crop loss within minutes and **significantly reduce daily energy costs** by tens of thousands, achieving a truly sustainable 'Greenhouse X-Ray' for precision farming.

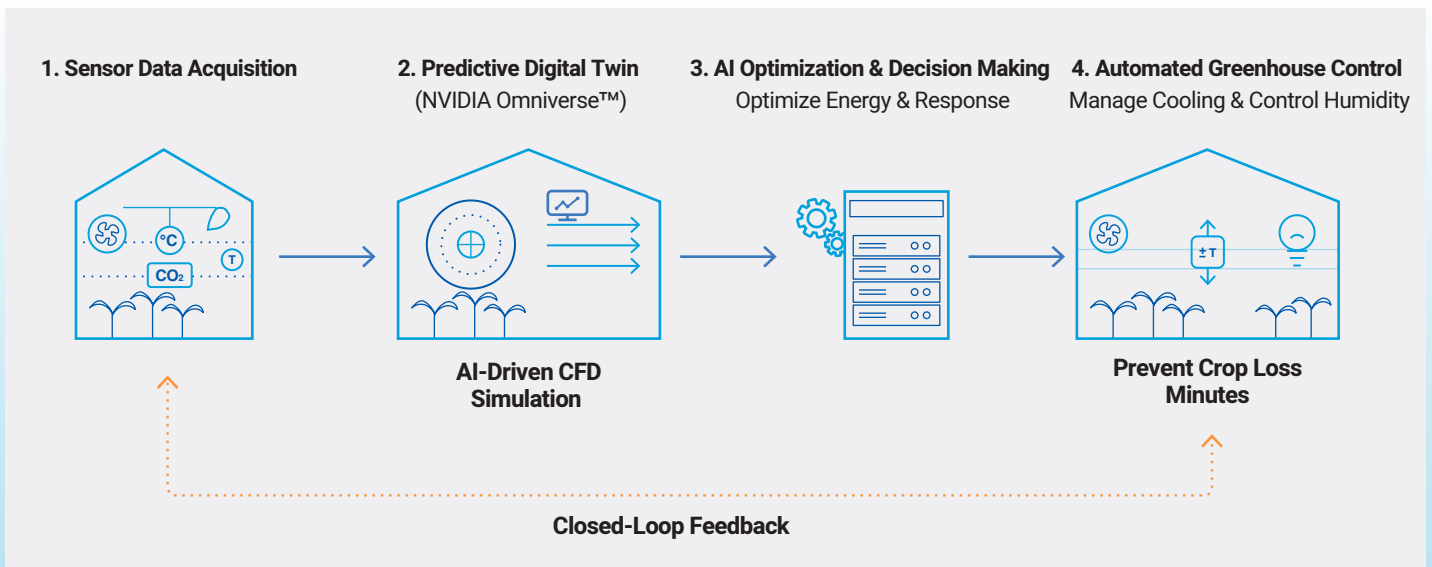


## Benefits

- **Immersive 3D Management:** Transition from traditional 2D monitoring to high-fidelity 3D digital twins via NVIDIA Omniverse™ integration.
- **Predictive Efficiency:** Identify optimal cooling paths using CFD airflow Simulation prior to physical installation, reducing hardware trial and error.
- **Substantial Cost Savings:** AI-driven orchestration of high-energy devices results in daily electricity savings of tens of thousands.
- **Rapid Risk Mitigation:** Prevents catastrophic crop loss within minutes of heat spikes, ensuring stability in intense light environments.

## Workflow

- **Sensor Data Acquisition:** Provides a seamless edge-to-cloud architecture, linking WEDA Sub-nodes to Core servers.
- **Predictive Digital Twin (NVIDIA Omniverse™):** Leverages predictive CFD simulations to enable growers to visualize hidden climate patterns.
- **AI Optimization & Decision Making:** Utilizes AI algorithms to intelligently manage environmental factors and energy-heavy equipment.
- **Automated Greenhouse Control:** Ensures optimal crop yield and operational scalability by responding to environmental threats in real-time.



ADVANTECH



## 3D AI Virtual Human Kiosk with Edge LLM



A next-generation Human-Computer Interaction solution leveraging high-fidelity BearVFX 3D AI Virtual Human with Advantech's **localized Edge LLM** and **RAG system** for 24/7 immersive, **low-latency**, and secure interactive experiences.

User's query



Local Audio, 3D & animation solutions



Edge computed LLM

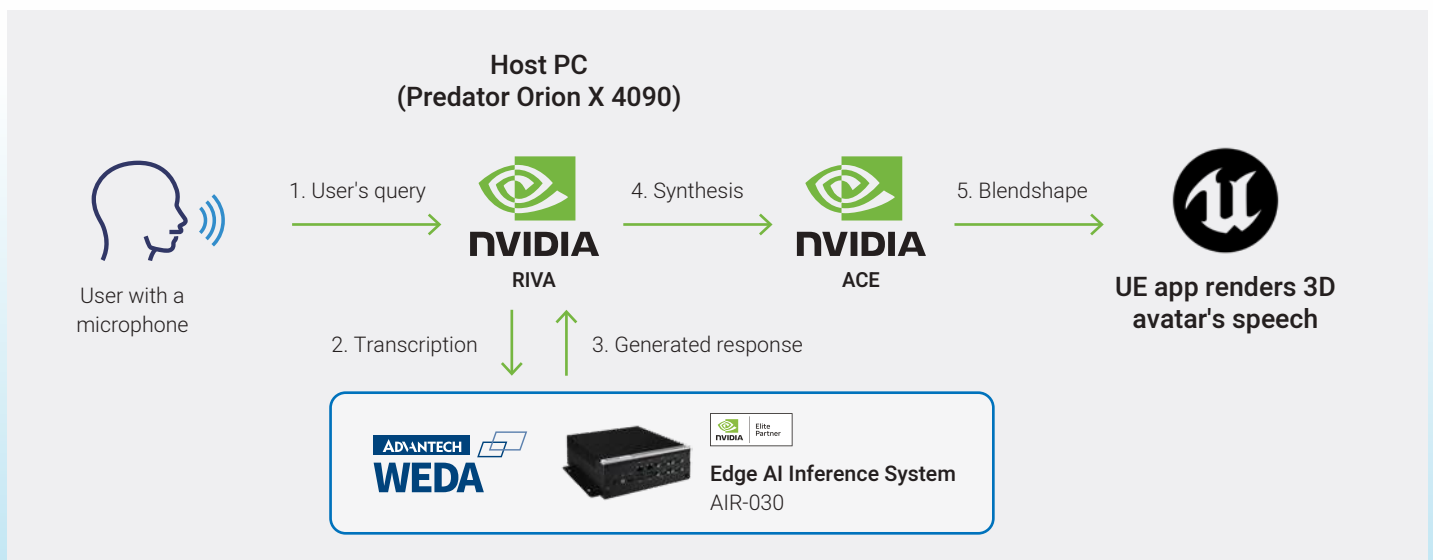
User talks with virtual avatar



### Benefits

- **Immersive Multimodal Interaction:** Combines vision, speech, and AI reasoning to deliver lifelike, human-like engagement beyond traditional chatbots.
- **24/7 Autonomous Service & Cost Efficiency:** Automates frontline operations, reduces labor dependency, and ensures consistent, round-the-clock service.
- **Real-World Data Intelligence:** Captures offline customer behavior, demographics, and interaction patterns to generate actionable business insights.
- **Precision Marketing & Personalization:** Transforms interaction data into targeted recommendations, enabling smarter marketing strategies and higher conversion.
- **Elevated Brand Experience:** Creates distinctive, high-end digital avatars that enhance brand perception and foster deeper emotional connections with customers.

### BEARVFX x WEDA Workflow

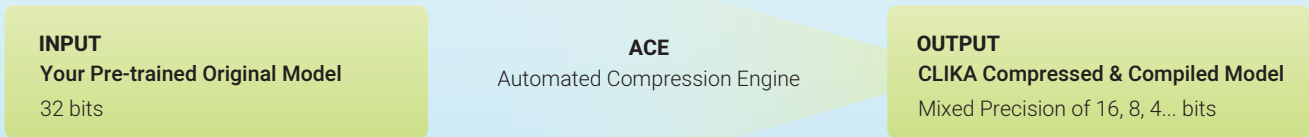


# Compress Once, Deploy Everywhere

The infra layer between your AI models and device fleet.

## Precision-Aware Compression

CLIKA's Automated Compression Engine (ACE) doesn't just shrink your model. It analyzes every layer, finds where precision can drop and where it can't, and builds a custom mixed-precision recipe (down to 1-bit where possible) tailored to your specific model and target hardware. The result: the smallest possible model that still performs.



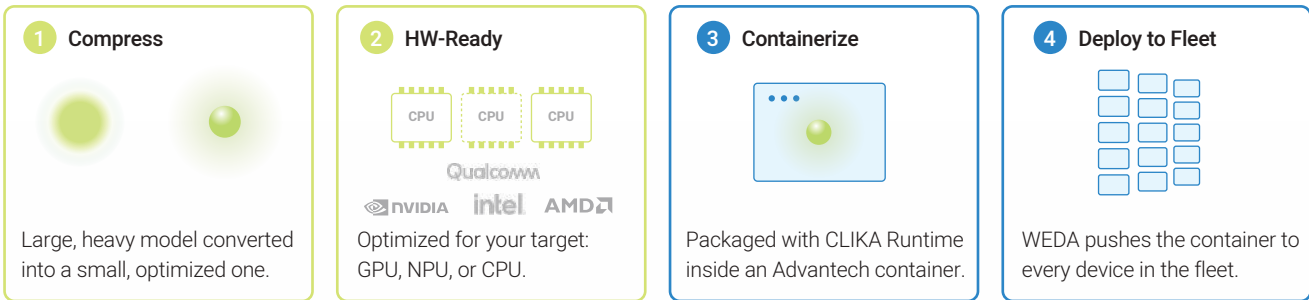
## Benefits

One compression pass gets you a model that runs faster, fits in less memory, and costs less to deploy at scale. The key difference: you don't sacrifice accuracy to get there. CLIKA finds the compression ceiling, not just a compression shortcut.

- up to **18x** Faster
- up to **90%** Smaller
- up to **90%** Saving
- up to **100%** Accuracy

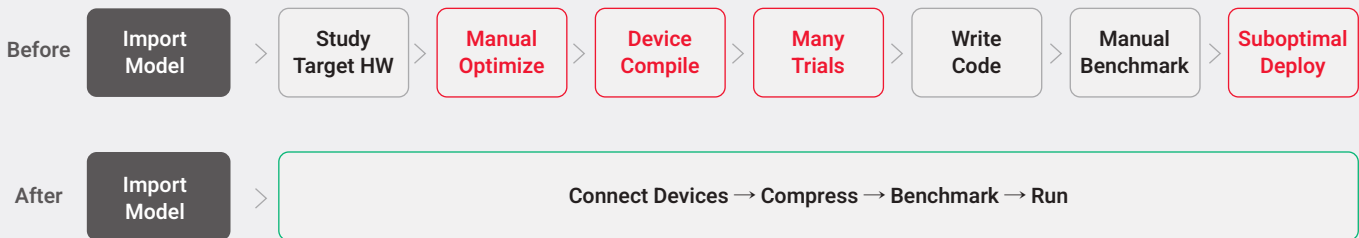
## CLIKA x WEDA Workflow

CLIKA prepares the model. WEDA delivers it everywhere. Together, we close the gap between "model ready" and "running in production across every device."



## Before & After

What used to take seven fragile steps full of errors and a team of engineers is now a single automated pipeline. No more guessing which optimizations work, no more rewriting code per chip, no more praying the benchmark holds in production.



Know more about WEDA



Request WEDA Free Account