Minsky™Edge

Model Governance & Deployment Orchestration for FPGA and Embedded Targets

Overview

Minsky™Edge is AI Labs' end-to-end MLOps platform purpose-built for embedded and FPGA deployment. It carries a model from raw data all the way to a signed, board-validated delivery bundle — driving vendor toolchains (Vitis-AI, TensorRT, ONNX Runtime, Vela) from a single governed interface and closing the loop between what a model does in training and what it actually does on the target board. Every number the platform reports carries a provenance stamp: measured on hardware, measured on a training host, or estimated. Nothing is quietly promoted from a simulation to a fact.

Each project lives inside its own isolated tenant workspace. Datasets are frozen to an immutable hash the moment a training run begins, so every model stays traceable to the exact data it learned from. Models are trained, quantized, compiled, and benchmarked against explicit power and latency budgets — and a model cannot reach Approved status without latency and power figures taken on a real target board. No exceptions, no workarounds, no simulated numbers standing in for hardware truth.

About AI Labs, Inc.

AI Labs, Inc. is an enterprise AI company delivering a full-stack, proprietary, sovereign, multi-vertical platform at global scale. Minsky™Edge is built on the same Minsky™ Platform Suite foundation that powers AI Labs' broader portfolio — engineered for organizations that need AI governance to extend all the way to the silicon, not stop at the model file.

11+	4	6	100%
Proprietary Platforms	Global Offices	Strategic Partnerships	Sovereign-Ready

Minsky™Edge Platform Modules

Dashboard	**Projects**	**Target Board Catalogue**	**Datasets**
Model Training & Distillation	**Compile & Benchmark**	**Acceptance Gate**	**Embedded Scaffolding**
Delivery & Bundles	**Self-Hosted Runners**	**Deployment Guide Generator**	**Any FPGA / Any Board** Fully customizable to your target hardware

Note: *The modules shown reflect a representative Minsky™Edge deployment. The platform itself is board-agnostic — it can be customized and extended to support any FPGA or embedded target, with new target board profiles, toolchains, and scaffolding added to the catalogue as needed.*

Key Capabilities

- **Immutable Dataset Manifests —** Every training run pins a frozen dataset hash, making any model traceable to the exact rows it learned from — a hard requirement for audit trails and regulated deployments.
- **Multi-Toolchain Compile —** Drive Vitis-AI, TensorRT, ONNX Runtime, and Vela from a single interface. Pre-flight checks catch activation mismatches and precision conflicts before a job is ever queued, saving hours of failed compile cycles.

- **Hardware-Only Approval Gate —** Latency (p95) and average power are measured on the attached target board. Simulated figures keep a model in Pending status until real evidence arrives — so nothing ships on an estimate.
- **Provenance-Stamped Metrics —** Every accuracy, latency, and power figure is tagged as hardware-measured, training-run, or estimated — and that tag travels through gate verdicts, reports, and delivery bundles end-to-end.
- **Self-Hosted Runner Network —** Register GPU boxes or board-wired agents inside your own network. Runners lease queued jobs, post results back automatically, and are always preferred over the built-in simulator.
- **Signed Delivery Bundles —** Assemble a versioned ZIP containing the artifact, compiled binary, benchmark report, scaffolding, and gate record — SHA-256 sealed and labeled Release or Draft.
- **Auto-Generated FPGA Deployment Guides —** Every completed project generates a detailed, board-specific deployment guide tailored to the exact FPGA target used — covering toolchain configuration, pin and interface mapping, compiled-binary integration steps, and board bring-up notes — so hardware teams move from a signed bundle to a running board without reverse-engineering the process.

Highlights

The Full Pipeline in One Tool

Five project stages — Datasets, Models, Compile & Benchmark, Scaffolding, Delivery — are navigable from a persistent header, not a chain of buttons you can only walk forwards.

Feedback Loop into Training

Benchmark results return as per-metric constraints and retrain advice, turning a missed latency budget into a concrete next step rather than a number reported once and forgotten.

Tenant-Isolated Workspaces

Every team gets its own private database; the dashboard shows the bound tenant database at all times so a silent fallback to a template is caught immediately.

By The Numbers

5
Governed pipeline stages — from frozen dataset to signed delivery bundle

4
Vendor toolchains supported: Vitis-AI, TensorRT, ONNX Runtime, Vela

10
Tracked capability areas per workspace

20k+
Lines of production code across a full Angular frontend and backend service

Why Choose Minsky™Edge

Hardware Truth, Not Simulation

Approval requires a real board reading. Teams stop shipping on optimistic estimates and start shipping on what the silicon actually measures.

Governance Built In, Not Bolted On

Frozen dataset hashes, provenance-stamped metrics, and signed SHA-256 bundles give compliance and audit teams a trail that exists by default — not as an afterthought.

One Interface, Every Toolchain

Vitis-AI, TensorRT, ONNX Runtime, and Vela are orchestrated from a single pane, with pre-flight checks that catch conflicts before a job wastes a compile cycle.

Sovereign, Tenant-Isolated Deployment

Every team's data lives in its own private, isolated workspace — deployable on-prem or in a sovereign cloud, aligned with AI Labs' sovereign-first architecture.

Who It's For

ML Engineers

Train, quantize, and gate model versions against frozen datasets without leaving a single interface — from first experiment to board-validated artifact.

Embedded Systems Teams

Compile for specific FPGA and MCU targets, measure real latency and power, and receive board-specific deployment guides automatically.

Program Managers & Leads

Track project status, gate verdicts, and delivery bundle provenance across multiple teams in isolated workspaces — one view, every project.

Compliance & Quality Teams

Every bundle carries a SHA-256 manifest, stamped metrics, and a gate record — the audit trail is built in, not bolted on.

Why Minsky™Edge Is a Game Changer

Embedded and FPGA AI has traditionally meant weeks of manual toolchain wrangling, hardware surprises discovered only after a board is already committed, and no reliable way to know if a simulated benchmark will hold up on real silicon. Minsky™Edge collapses that entire cycle into one governed pipeline — compressing the path from trained model to board-validated artifact from weeks of back-and-forth into days, eliminating the costly late-stage rework that comes from simulation-to-hardware mismatches, and replacing guesswork with a hardware-verified answer at every gate. The result is faster time-to-board, materially lower engineering cost per deployment, and a level of accuracy and reliability confidence that comes only from measuring on the target itself — not estimating from a training host.

Faster Time-to-Board

A single governed interface replaces manual toolchain switching and ad-hoc scripts, so teams move from trained model to signed deployment bundle in days, not weeks.

Lower Engineering Cost

Pre-flight checks and a hardware-only approval gate catch failures before a compile job or board cycle is wasted — cutting the rework that normally eats embedded AI budgets.

Higher Accuracy Confidence

Every metric that gates a release is measured on the real target board, not simulated — so what passes Acceptance is what actually performs in the field.

See Minsky™Edge in Action

Contact AI Labs to schedule a live walkthrough and attach your first target board.

www.ailabsinc.ai