



VISION MODELS, ON DEMAND

# Stop labeling data. Start describing it.

One sentence becomes a trained, deployable detection model that generates, labels and verifies its own data. **Zero human annotation.**

Built on a single AMD MI300X

FLUX.2 / SDXL

SAM 3 / YOLOE

YOLO • RT-DETR • RF-DETR

Gemma 4 via vLLM

PyTorch on ROCm

# Every vision model is built on a tax nobody can avoid.

## COLLECT

# 01

### Months before a model even exists

Thousands of on-site images before one can train — the cold-start problem: on day one there is no data, and the rarest cases may never occur on camera.

## LABEL

# 02

### \$0.02 to \$0.08 for every single box

Drawn by hand. A 100,000-image set runs \$50,000 and up.

## PAY AGAIN

# 03

### Change anything, and do it all over

New camera or site: collect, label and pay again. Data you rent, never own.

**A market worth billions, and it exists only because models cannot feed themselves.**

Sources: LabelYourData, BasicAI, GigaBPO annotation pricing guides (2025).

# One sentence in. A deployable model out.

## Five agents. Zero humans.

AGENT 01

### Prompt

Gemma 4 · vLLM

Turns your sentence into  
hundreds of varied scenes

AGENT 02

### Synthesis

FLUX.2 / SDXL

Renders each scene as a  
photorealistic image

AGENT 03

### Vision

SAM 3 / YOLOE

Labels every object, any noun  
works

AGENT 04

### Critic

Gemma 4 VLM + geometry

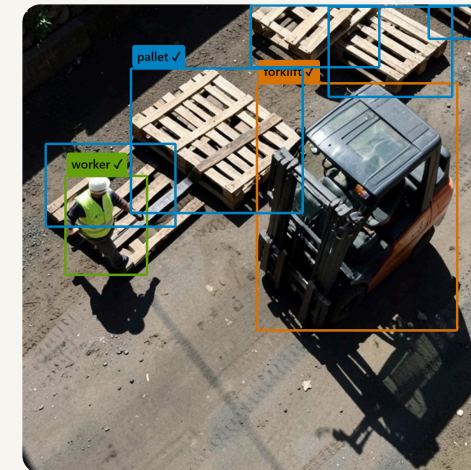
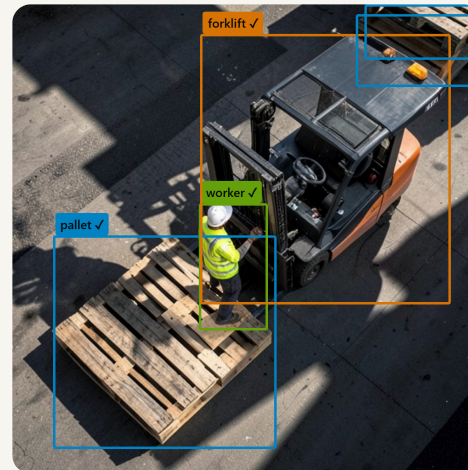
Checks and rejects bad labels.  
The trust layer

AGENT 05

### MLOps

YOLO on ROCm

Trains and exports a .pt / ONNX  
model



**Fully agentic, end to end:** on the MI300X the whole swarm stays resident and runs as **one parallel swarm on a single card**.

Measured live on the MI300X (run\_0044 → model\_0026): yolo26n, mAP50 0.96 · mAP50-95 0.95, 5,000 images, 6,027 verified labels, 40 epochs in 44 minutes.

# We give eyes to agents. **On demand.**

Besides the **UI** there is: One headless API, two callers: a person in the console, or an autonomous agent asking for a model for its exact task. Both get **trained, verified weights in minutes.**

Open-vocab models (SAM, OWL-ViT) see anything but are too slow for the edge. Fixed detectors are fast but blind to new tasks. **We generate the fast specialist, on demand.**

A field drone can ask the API for a blight detector and run it **onboard, offline**, minutes later. No data collection, no labeling, no connection home.

## \$52B

AI agents market by 2030

## \$48.5B

AI drones market by 2030

## \$2.9T

AI agent + robot value by 2030 (McKinsey)

Sources: MarketsandMarkets; McKinsey.

# The whole factory fits **in one GPU.**

AMD INSTINCT MI300X

## 192 GB

HBM3

Holds the whole swarm resident, streaming in parallel.

- Gemma 4, FLUX.2, SAM 3 and YOLOE **held warm together**
- Synthesis, vision and critic **stream in parallel**, no reloads

SINGLE NVIDIA H100

## ~~80 GB~~

HBM3

Cannot co-reside the stack. Swap per stage, or split across boxes.

- MI300X packs **2.4×** the memory of an H100
- Per-stage swapping means **no parallel swarm**

VRAM BUDGET · ONE MI300X

## 192 GB

HBM3

The resident swarm (~**125 GB, measured live**) already exceeds a full **H100's entire 80 GB** by half — with room to spare for training and heavier models as they land.

Sources: AMD specs; measured footprints.

# Gemma 4, the brain of the swarm.

One language model, resident on the MI300X beside FLUX.2 and SAM 3, doing three jobs at once.

## 01 · SCENE DESIGNER

### Designs the training set

The Prompt Agent hands Gemma your use case. It infers viewpoint and environment, then designs **hundreds of varied scene prompts**.

## 02 · SEMANTIC CRITIC

### Verifies every label

Geometry proves the box fits; **Gemma proves it is really a forklift**. Crops that fail never reach training.

## 03 · DOCUMENTATION

### Writes the model card

Intended use, limits and metrics for every trained model, **no human in the loop**.

Stack: google/gemma-4-26B-A4B-it served by vLLM 0.23 (ROCm) on the MI300X, held warm with the rest of the swarm.

# The industry stopped collecting data. It started generating it.

aionVIS rides the biggest shift in how AI gets built: from collecting real data to generating it. The missing piece is **trusting the labels**, which we ship.

NVIDIA, Tesla and Waymo already train on generated data; Gartner expects **synthetic to pass real by 2030**.

Usable human data runs dry between 2026 and 2032, and the rarest cases (defects, edge scenes, objects in orbit) never had a dataset at all.

## 2030

synthetic overtakes real (Gartner)

## ~35% / yr

synthetic data market growth

## \$17B

data-labeling spend by 2030, replaced

Sources: Gartner; Grand View Research; Epoch AI.

# One engine. Eight vision markets, each in the billions.

|                                                                                                    |                                                                                                               |                                                                                                           |                                                                                                           |
|----------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| <div>RETAIL</div> <div>\$18B</div> <div>Shelf analytics, loss prevention</div> <div>Amazon</div>   | <div>MANUFACTURING</div> <div>\$20B</div> <div>Defect inspection, rare-defect synthesis</div> <div>TSMC</div> | <div>AUTOMOTIVE</div> <div>\$19B</div> <div>Edge-case perception, rare road scenes</div> <div>Waymo</div> | <div>AGRICULTURE</div> <div>\$18B</div> <div>Crop, weed and disease detection</div> <div>John Deere</div> |
| <div>ROBOTICS</div> <div>\$4.4B</div> <div>On-demand perception per object</div> <div>Figure</div> | <div>LOGISTICS</div> <div>\$22B</div> <div>Warehouse safety, forklift detection</div> <div>Voxel AI</div>     | <div>AVIATION</div> <div>\$12B</div> <div>Runway debris, apron monitoring</div> <div>Xsight Systems</div> | <div>DEFENSE</div> <div>\$9.3B</div> <div>Air-gapped ISR, fully on-prem</div> <div>Anduril</div>          |

BEYOND COLLECTABLE

Debris in orbit, rare defects and rare targets have **no dataset at all**, so we generate both the images and the labels. Space-domain awareness alone: **\$1.9B**.

DETECTING AI ITSELF

The same engine that generates AI imagery trains **deepfake and AI-video detectors**, a trust-and-safety market heading to **\$8B by 2030**.

Sources: Mordor Intelligence, MarketsandMarkets, Grand View Research (2025 estimates).

# Technologies **used.**

## THE MODELS

LANGUAGE · GEMMA 4

### The swarm's brain

Prompt Agent, semantic critic and model cards, served by vLLM.

SYNTHESIS · FLUX.2 / SDXL

### The world builder

Photoreal training scenes rendered from the designed prompts.

SEGMENTATION · SAM 3

### The eyes

Open-vocabulary labeling of any object by name; SAM 3 or YOLOE is a per-run user choice, honored verbatim.

DETECTION · YOLO / RT-DETR / RF-DETR

### The product

22 trainable architectures across detect, segment, OBB, pose and classify.

All open source or self-hosted: no third-party APIs; the hot path — Gemma 4, FLUX.2, RF-DETR — is Apache-2.0 end to end.

## THE PLATFORM

COMPUTE · AMD MI300X + ROCM

### One GPU, whole swarm

192 GB HBM3 keeps every model resident; PyTorch on ROCm.

ORCHESTRATION · AGENT SWARM

### Streaming pipeline

Producer/consumer stages, GPU slots, live VRAM telemetry.

BACKEND · FASTAPI + WEBSOCKETS

### Control plane

Run queue, REST + WS API, API-key auth, Mission Control feeds.

CONSOLE · NEXT.JS 16

### The cockpit

Wizard, Mission Control, dataset analytics, inference playground.

# Priced like infrastructure, not labor.

| PLAN       | PRICING                                       | WHO IT FITS                                                                                    |
|------------|-----------------------------------------------|------------------------------------------------------------------------------------------------|
| Developer  | Free, self-hosted. Export YOLO / COCO / ONNX. | Developers, researchers                                                                        |
| Team       | Per GPU-minute on managed MI300X.             | Most teams. The 500-image flagship run used ~\$1.25 of raw GPU time — quoted before you click. |
| Enterprise | Dedicated or on-prem. SSO, isolation, SLAs.   | Regulated, IP-sensitive, defense                                                               |

| BUYER         | ANNUAL       |
|---------------|--------------|
| SMB           | \$3k–\$25k   |
| Mid-market    | \$50k–\$250k |
| Enterprise    | \$300k–\$1M+ |
| Gov / defense | \$500k–\$5M+ |

The flagship run: 22,718 verified labels for ~\$1.25 of GPU time — \$450–\$1,800 at human labeling rates. Cloud-GPU margins (70–85%), not services (30–40%).

Sources: in-app pricing engine; spend figures are planning estimates.

# Everyone else needs your data, or your clicks.

|                        | SCALE AI · LABELBOX  | ROBOFLOW · ULTRALYTICS | AIONVIS               |
|------------------------|----------------------|------------------------|-----------------------|
| Who creates the data   | You collect it       | You collect it         | We generate it        |
| Who draws the boxes    | Humans, by hand      | Humans, assisted       | The agent swarm       |
| Humans in the loop     | Many                 | Fewer, still required  | Zero                  |
| Label trust layer      | Human QA             | Human review           | Self-verifying Critic |
| When the model misses  | New labeling project | Collect and relabel    | One-click closed loop |
| Time to a new detector | Weeks                | Days                   | Minutes               |

BUILT TODAY 22 architectures · 5 task types detect / segment / OBB / pose / classify · export to .pt, ONNX, TorchScript, OpenVINO

Meta just valued Scale AI at ~\$29B, a business built on human labor. Our self-correcting Critic is the moat: everyone can generate, few can trust it.

Sources: Scale AI / Meta, 2025; run history on the live node. Live now: aionvis.com — judges can sign in with their own AMD droplet.

# Startup roadmap.

From hackathon build to product. Four milestones, one direction.

Q4 2026



## The raise

Seed round on the hackathon build: a live product, a working MI300X deployment and \$1.25-per-run economics. Funds a founding team committed full-time to commercial development.

H1 2027



## Paid design-partner

Five to ten paying teams run aionVIS on their own GPU nodes. The trust suite, scene-level critic and the "why this label?" inspector, is built against their real workloads.

H2 2027



## Commercial launch

Managed MI300X capacity goes live: the metered per-GPU-minute Team tier, workspaces and quotas. Pilot teams convert into the first recurring revenue.

VISION



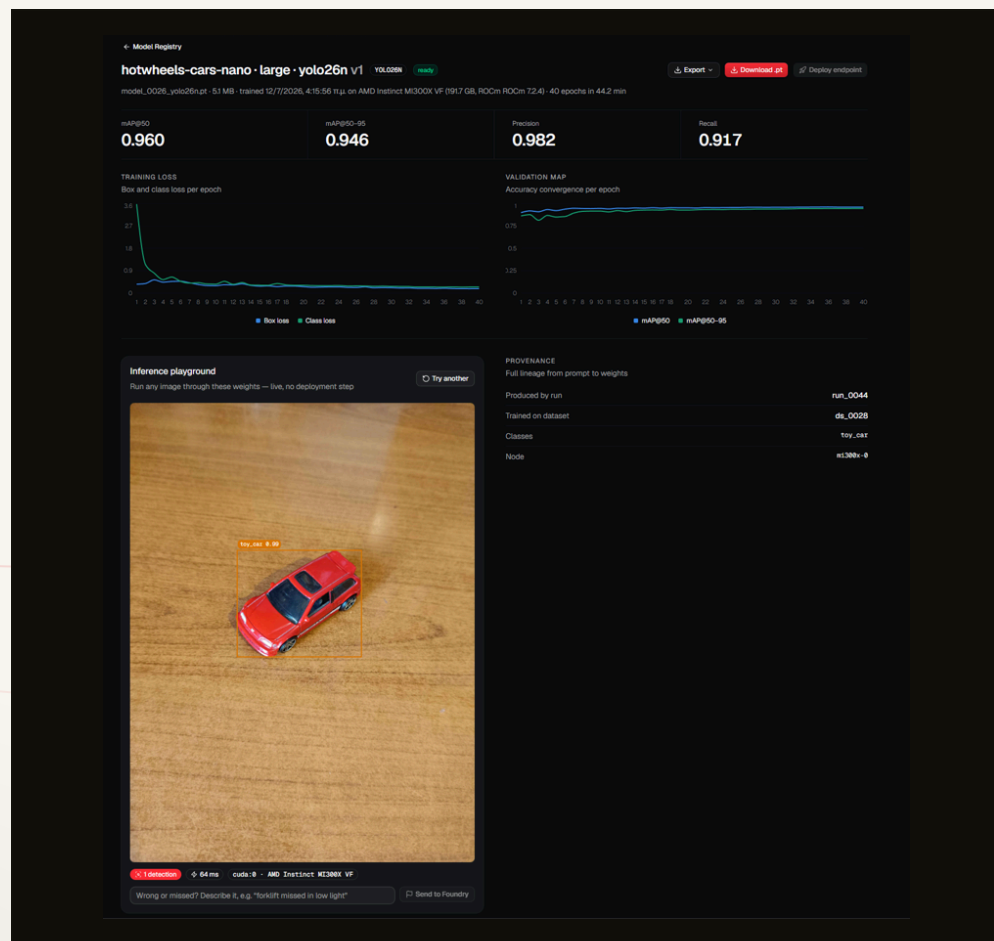
## Eyes on demand

The perception layer agents call when they need to see: a trained, verified model for any task, minutes after asking.

**LIVE TODAY** **aionvis.com**: demo in one click, or sign in with your own AMD Developer Cloud droplet; per-person API keys, revocable individually.

# From a sentence **to this.**

Trained in **44 minutes** on AMD, then shown a **real phone photo it had never seen** — a toy car it nailed at **99%**.



**44** min

**Training** · 40 epochs on one MI300X

**\$1.47**

**Total cost** · AMD Developer Cloud

**5,000**

**Training images** · fully synthetic, zero human labels

**99** %

**Confidence** on the real phone photo we took

# Not another annotation tool. We give machines their eyes.

One sentence becomes a trained, verified detection model on a single AMD MI300X. **Stop labeling. Start describing.**

 [github.com/0xAthkot/aionvis](https://github.com/0xAthkot/aionvis)

 [aionvis.com](https://aionvis.com)