Why Standardization Enables

Law of Scale — Physical AI

Yu Been Park — Managing Director, Self-Driving & Robotics Division

마음AI

GISC 2025

Company & Speaker Overview

Founded in 2014 and listed on KOSDAQ in 2021, MAUM.AI is widely regarded as a frontrunner in commercializing vision-language-action (VLA) models across brownfield industries.

Korea's Leading Physical AI Pioneer maum.ai

2025

- Founding member & inaugural chair of the Physical AI Association
- Officially registered as a member of Qualcomm's IoT Accelerator Program
- CES 2025: Co-exhibited in collaboration with Qualcomm
- WoRV (Physical AI Technology) accepted to ICRA; invited Oral Presentation
- 2025 AI Excellence Summit — AI Technology Innovation Award

2024

- CES 2024: Recognized as a leading company in the CTA Trend Report
- Delivered a large-scale AI docent for the Gangneung Winter Olympics commemorative metaverse exhibition
- 2024 CICON Awards: Winner, Integrated AI Business category
- AI Defense Innovation Forum: Grand Prize recipient
- Korea's Robot Company of the Year 2024 (selected)
- WoRV (Physical AI Technology): Outstanding Paper Award, NeurIPS 2024

Yu Been Park — Managing Director, Self-Driving & Robotics Division, MAUM.AI

- Leading transition from "AI System Integration" to a "AI Product" corporate structure
- Leading & Built Physical AI Team, "WoRV" from scratch
- NeurIPS'24 OPA, ICRA'25 invited Oral, 2 papers currently under review

Deployments:

- Autonomous VLA-based agency-certified speed sprayer(Korea & Japan)
- Special-purpose agriculture vehicles(Southeast Asia)
- Undisclosed autonomous vehicles in the Agriculture/Construction/Defense Domain

Co-founder, Manned Unmanned:

- Multimodal LLM "visual analyst" platform, focusing on visual insights & surveillance

Next GPT Moment is coming — Physical AI

The Physical AI market is vast—capturing even a small share can generate significant profits.
With earlier-than-expected revenue and greater industrial reach than ChatGPT, moving now secures share and asymmetric upside.

Morgan Stanley

“Robotics and Physical AI will likely have revenue impact earlier than expected”

Morgan Stanley via Investing.com
(post-GTC market analysis), Mar 2025

Investing.com

Why robotics and physical AI will likely have revenue impact earlier than expected

Investing.com | Author Navamya Acharya | Stock Markets

Despite skepticism stemming from past AI investments—such as autonomous driving in 2018, which has yet to yield widespread commercial success—historical patterns suggest semiconductor companies can generate revenue early in the investment phase.

“Most companies will make money when autonomous cars, or robots, are deployed, but we would argue that processor revenues can come earlier, as their hardware dominates the profits during that investment phase,” Morgan Stanley added

Physical AI Market Measured in Tens of Trillions of Dollars

Humanoid Robot Market

“10 Billion Humanoid Robots Will Appear in 25 Years”
— Elon Musk, CEO of Tesla

First $10T market in human history

**Morgan Stanley prediction*

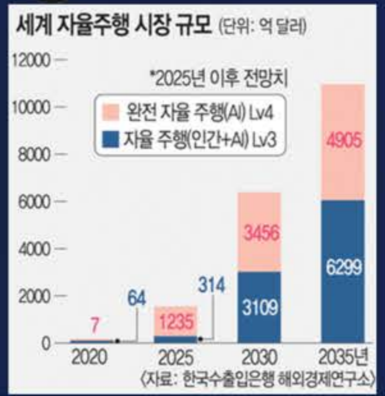
Manufacturing Robots Market

2035 Industrial Robot Market: $4T

**Roots Analysis prediction*

Autonomous Mobility Market

2035 Autonomous Driving Market: $1.1T ➡ 723% Growth



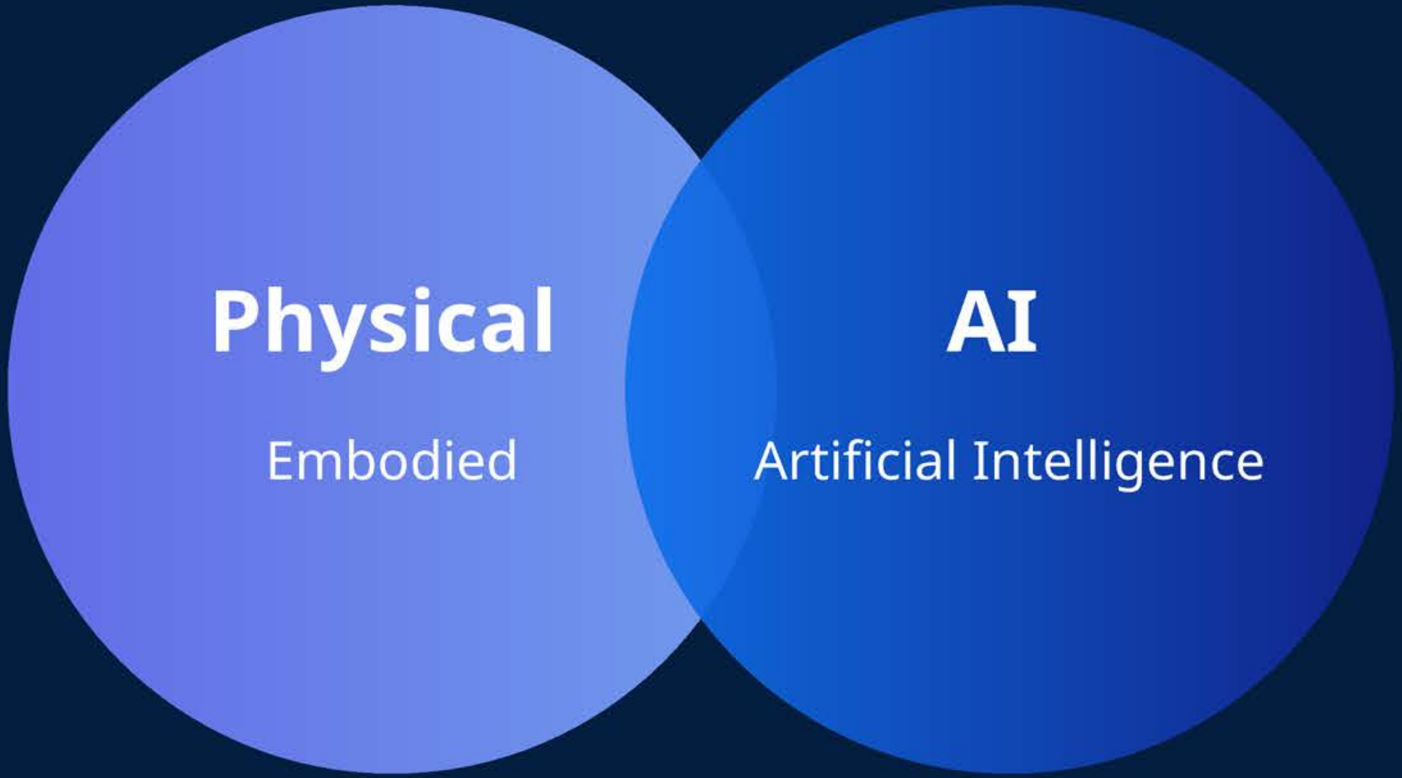
**Korea Eximbank prediction*

Even a small slice of the market can generate high profits

What is Physical AI?

New paradigm where AI systems have physical embodiment, interact with real-world environments, and perform physical human labor

Space-aware, real-time adaptive, environmentally coupled—AI built for the physical world



Physical AI Core Technology: VLA Model Architecture

MAUM.AI's proprietary core, WoRV (World model for Robotics & Vehicle control), is a VLA-based model that reproduces human motor control to enable autonomous driving and manipulation

VLA Model: Understand, Plan, Execute

Vision — Human-like scene understanding from cameras

Language — LLM-based common-sense and logical reasoning; understands & acts on natural-language instructions

Action — Executes precise navigation and tool/manipulator control from the reasoning results

Why "Language" Matters

VA Approach ➡ Requires Astronomical Amounts of Data (Tesla)

Behavior Cloning: learn a policy that reproduces expert actions

$$\theta^* = \underset{\theta}{\operatorname{argmin}} E_{(s,a)\sim D}\left[L(\pi_\theta(s), a)\right]$$

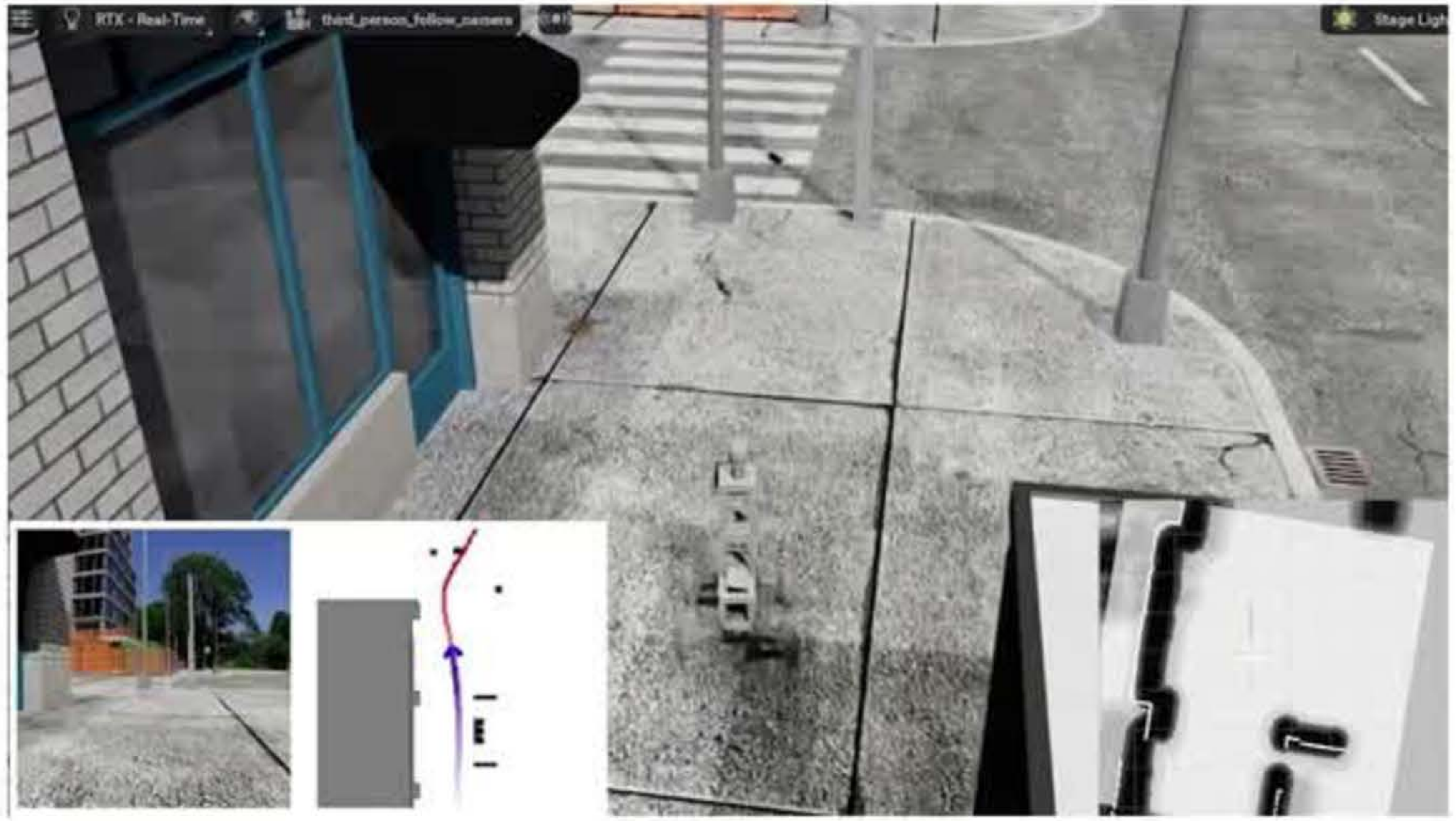
"Learn by observing and approximating human behavior (D)."

VLA Approach

"Perform common-sense reasoning and decision-making through language"

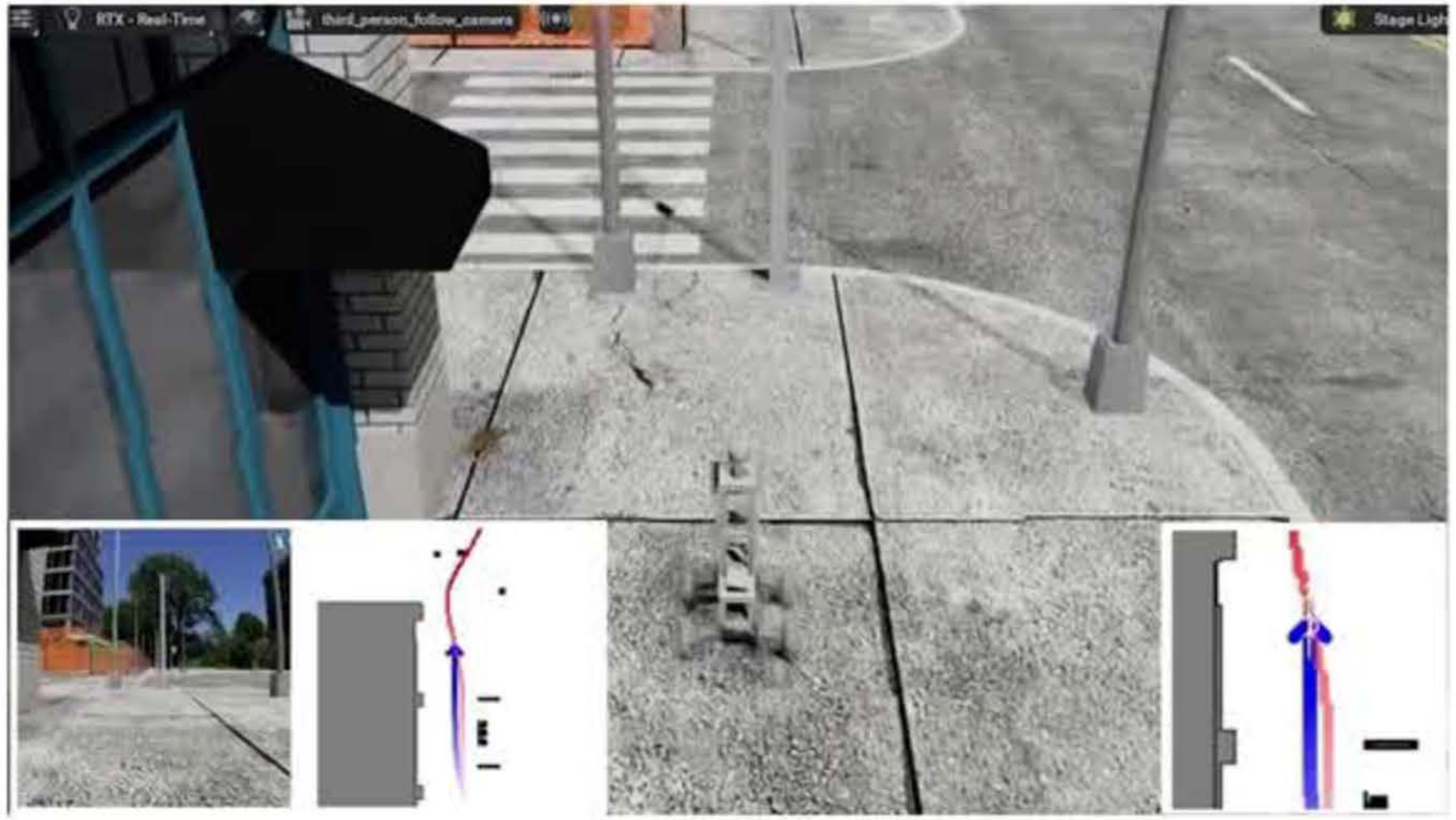
Comparison with Baseline

Language Instruction: You must use the crosswalk



Nav Stack

CANNOT Understand Language Instruction



CANVAS (ours)

2X

The Global Shift to Vision-Language-Action Models

The AI industry is moving beyond modular, rule-based planning to end-to-end reasoning models—single models that learn from sensor inputs and produce actions. This shift is converging on the VLA paradigm.

VLA-based End-to-End Adoptions by Leading Tech Companies

TESLA

E2E for FSD
trains on camera video and human driver control logs; advancing toward a VLA-style perception → reasoning → control stack

FIGURE

VLA for human-like learning
robots imitate and learn human behavior to perform new tasks without task-specific prior experience

Baidu 百度

E2E autonomous driving
pretrains large-scale driving data on a BEVFormer-based Transformer to unify perception through driving control

AGIBOT

VLA humanoid
integrates visual perception and natural-language commands to generate actions end-to-end

Data-driven:
How Physical AI Differs from Legacy Automation

Legacy Rule-Based Robots/Autonomy

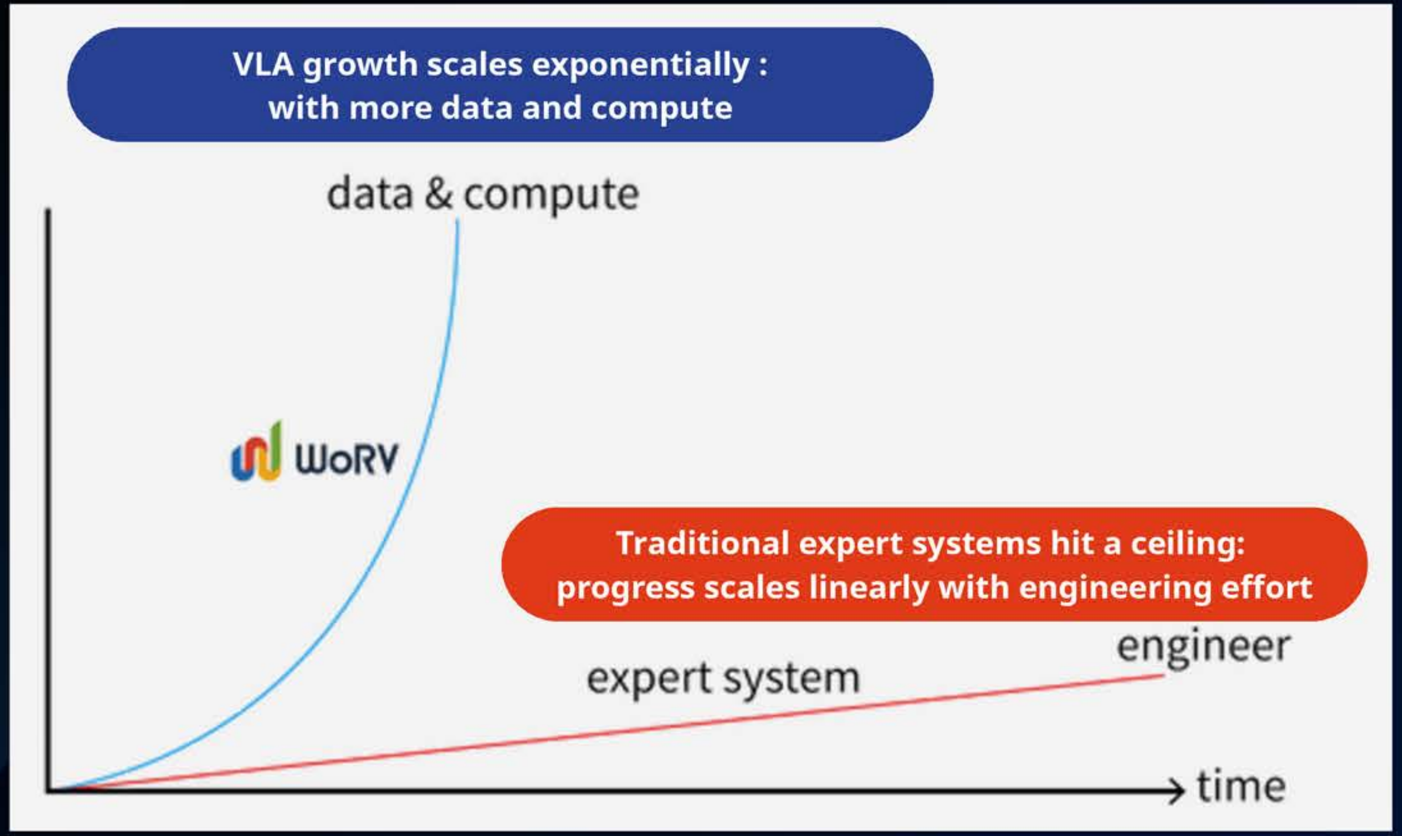
Hard-coded if–else rules

"If obstacle < 1m ahead and > 30cm high, turn right 90°, detour 2m, go straight."

Physical AI Robots/Autonomy

Data-learned policies & VLA-driven planning

"Assess slope, crop spacing, and surface; avoid narrow aisles; slow down in mud."



Rule-based

+ *Deterministic and predictable*
+ *Relatively low compute and latency*
+ *Strong performance in controlled environment*
- *Brittle to edge cases*
- *Scales poorly with engineering effort*
- *Slow/expensive to retarget for new tasks*

Data-driven

+ *Adapts to variability/unstructured environments*
+ *Performance scales with data and compute*
+ *Rich perception/planning; handles noise better*
- *Higher upfront complexity and cost*
- *Less transparent/stochastic policies*

The Broad Spectrum of Physical AI Across Industries

Physical AI attaches to existing industrial hardware to deliver human-level autonomy and immediate value across brownfield industries.

Physical AI is general-purpose—any machine that senses, decides, and acts can create new business value.

Agriculture
Autonomous Tractor
Speed Sprayer Robot

Public Services
Autonomous Cleaner
Last Mile Delivery Robot

Defense
Reconnaissance Robot
Tactical Support UGV

Construction
Autonomous Heavy Equipment
AGV/AMR Material Transport

Manufacturing
AMR
Welding Robot

Physical AI — Build the Data, Build the Lead

As open source narrows model gaps, the edge shifts to data.
Early movers who secure the domain specific data flywheel will set the pace.

Adaptive Fidelity Digital Twin
Cost-Effective Data Generation

+Data
+Data
+Data

Data Assets Appreciate **VS** Model Assets Depreciate as technology advances

Data ⬆ Generalization ⬆

Domain action data compounds → Better in-domain generalization

More runs, smarter robots

Model Scarcity Due to Open Sourcing ⬇

Open-source accelerates → Standalone model IP depreciates

Models are shared; data isn't

In the Physical AI era, domain-specific action data is a company's exclusive, defensible Moat & Asset

WoRV: Korea's First **Commercialized** Physical AI

Maum AI's VLA-based Physical AI technology, "WoRV," has been successfully applied to autonomous orchard Speed Sprayers.
The technology is already at the commercialization stage while also achieving academic recognition,
including the Outstanding Paper Award at NeurIPS 2024, establishing Maum AI as a globally acknowledged leader in Physical AI.

WoRV

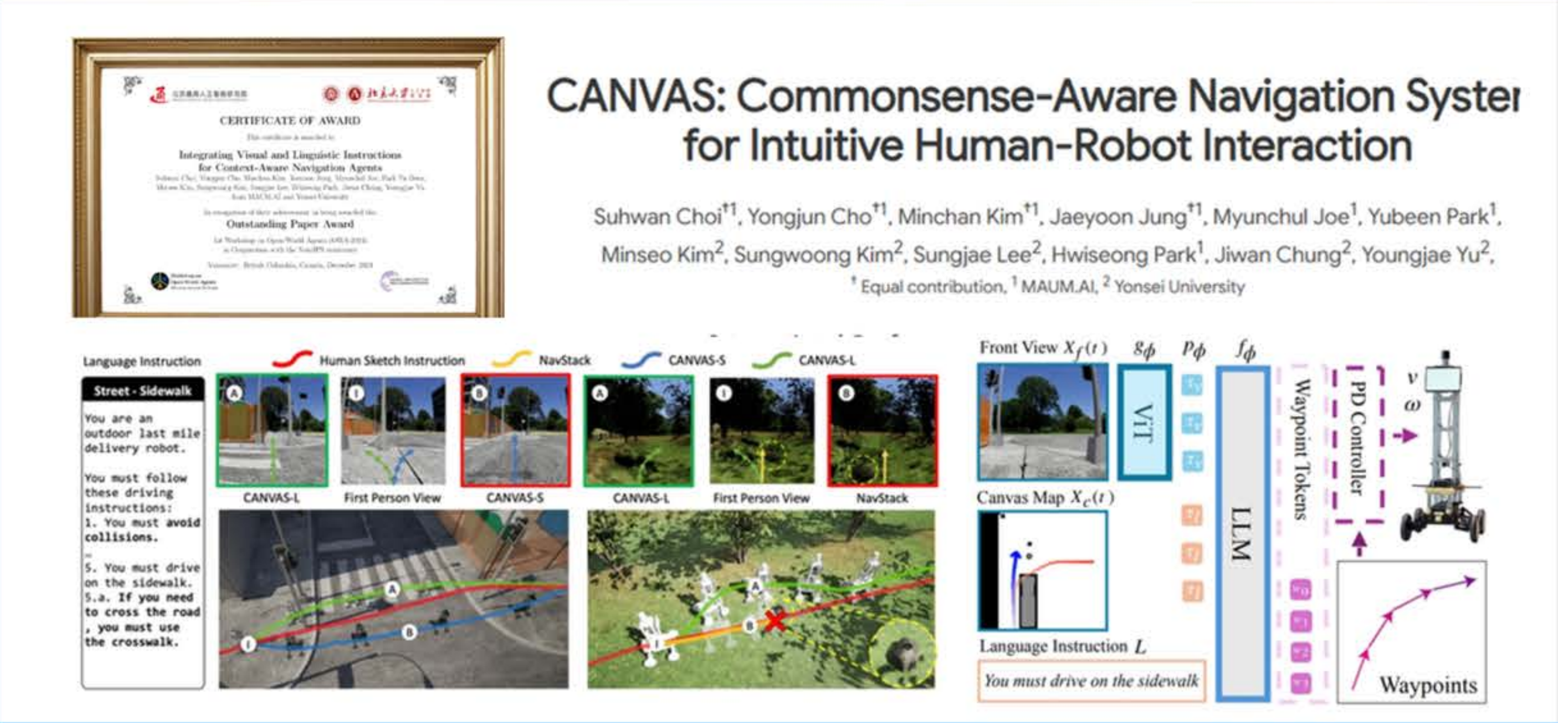
World model for Robotics and Vehicle control

World model for Robotics and Vehicle control

First to pass the national autonomous agricultural-machinery certification via Physical AI

Commercial deployment contract (100 units) with agricultural machinery company GINT for orchard speed sprayers

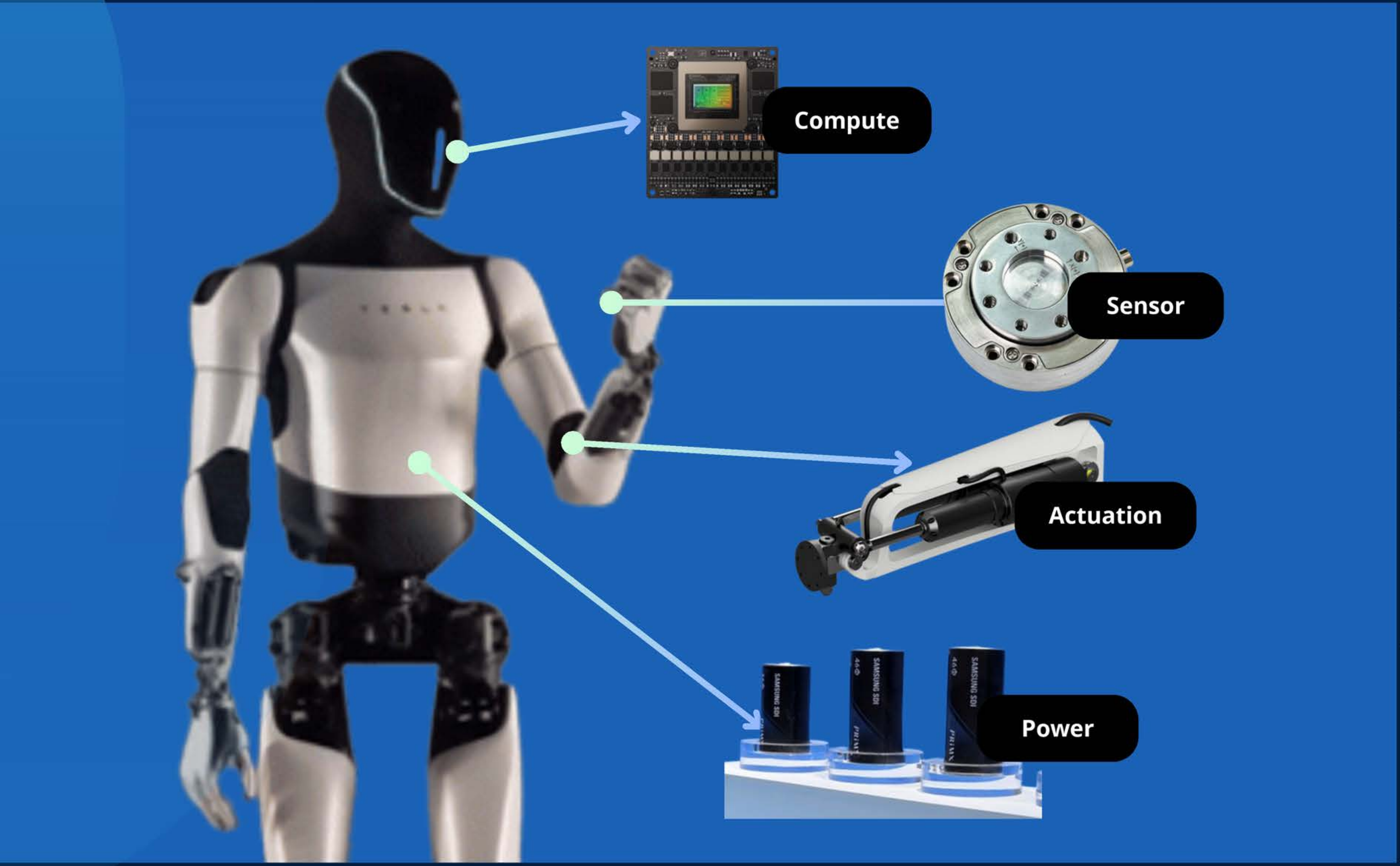
Globally recognized innovative technology

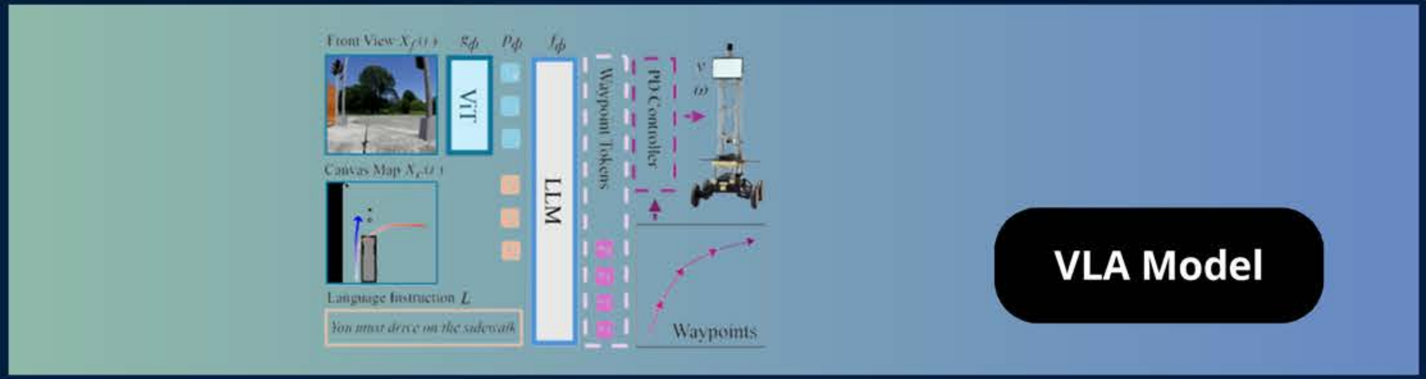


Outstanding Paper Award at NeurIPS 2024, one of the world's most prestigious AI conferences

Physical AI: A Sophisticated, Interconnected Value Chain

The value chain of Physical AI is extensive and intricately interconnected.
Due to the variety of sensors and heterogeneity of data formats, collecting and reutilizing data is highly challenging.

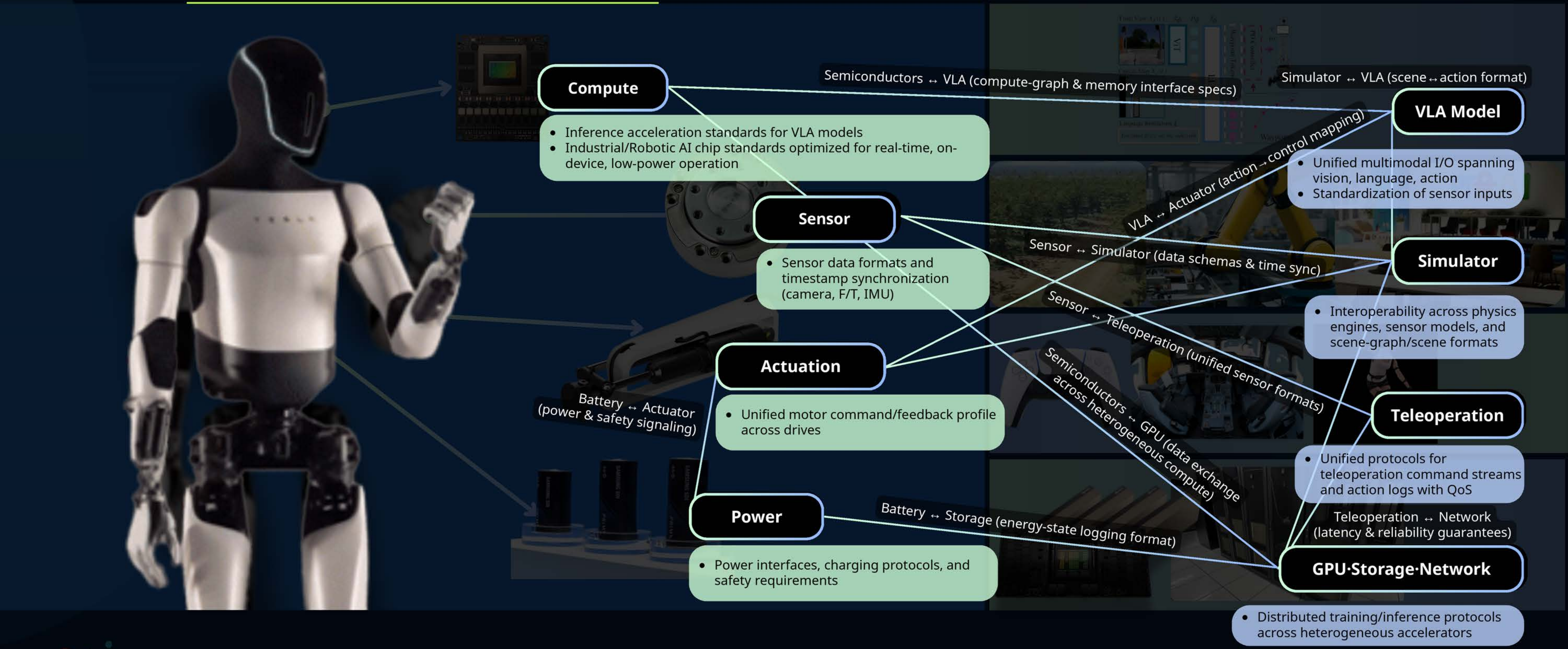




Physical AI: A Sophisticated, Interconnected Value Chain

The value chain of Physical AI is extensive and intricately interconnected.
Due to th… a formats, collecting and reutilizing data is highly challenging.

"Standardization"

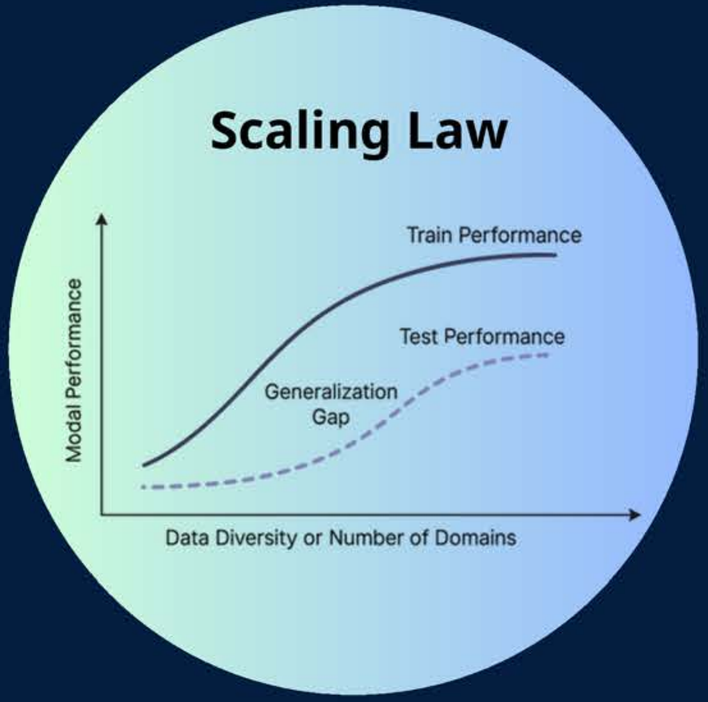


Why Data Standardization is Essential for Scale

LLM scaled because of Data Standardization —Unicode/UTF-8 text to subword tokens— resulting reproducible Scaling Law and compute-optimal training; **VLA still lacks a comparable Data-Time-Action IR**, with only pieces like MCAP logs and PTP/802.1AS timing emerging.

LLM

TEXT DATA



SUCCESS

ChatGPT Gemini Claude

VLA

VIDEO/ LANGUAGE/ ACTION DATA

Standard

?

FAIL

Current Issues with Physical AI “Data Standardization”

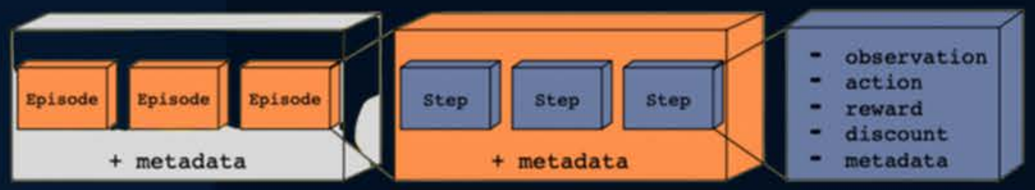
Today’s Physical AI data formats have two problems: **(1) Weak Large-scale Video/Media Support**
(2) Single-format Pipelines preclude Simultaneous Optimization of Recording and Training

Problems with the two main data formats currently considered standard

RLDs, Google

Data format originally defined for RL tasks

Vulnerable to large-scale media(video) processing



Explosive data volume and high I/O load

Storing 20 TiB of data without H.265 encoding can expand to 2,000 TiB (100× increase)

LeRobotDataset, Hugging Face

LeRobot

Single-format to optimize both (1) data collection and (2) model training is inefficient

Characteristics required for each task differs significantly, making it impossible to fully optimize both simultaneously.

Formats storing tens of thousands to hundreds of thousands of small files are extremely inefficient for large-scale distributed file systems

Performance degradation and I/O bottlenecks manifest across all processes, encompassing file downloading and preprocessing/training.

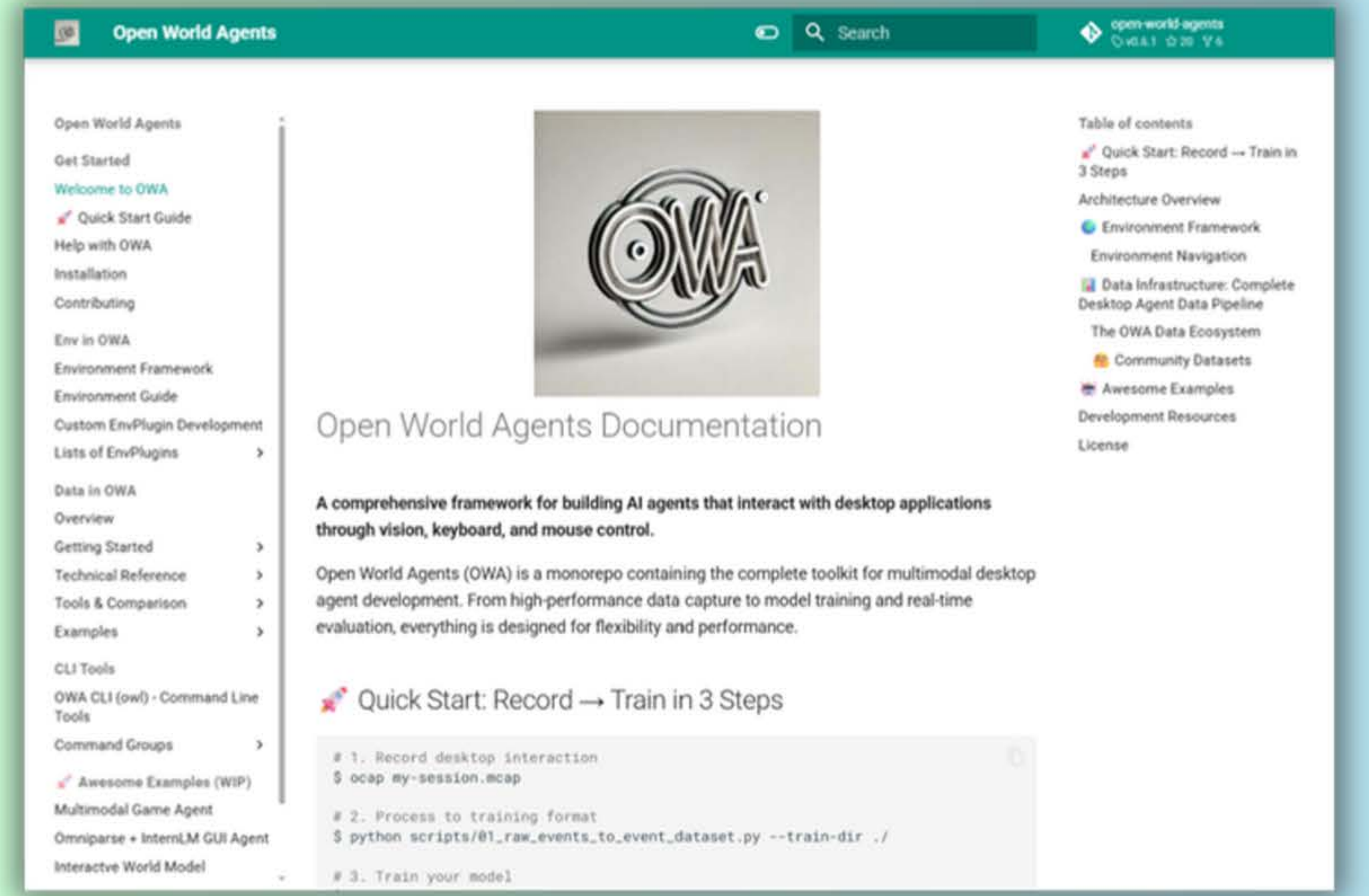
Leads to performance degradation, slower speed, and I/O bottlenecks

Downloading even less than 20 GB from the HuggingFace Hub can take over an hour on a 1 Gbps network

Our Proposed Solution: OWAMcap & Data Preprocessing Pipeline

Maum.AI has identified the optimal data pipeline through its research and commercialization process.

Project **<Open-world-agent>**



1 OWAMcap format

Media data are stored externally via MediaRef,
with only references retained inside the .mcap file

Dataset Compression Enabled

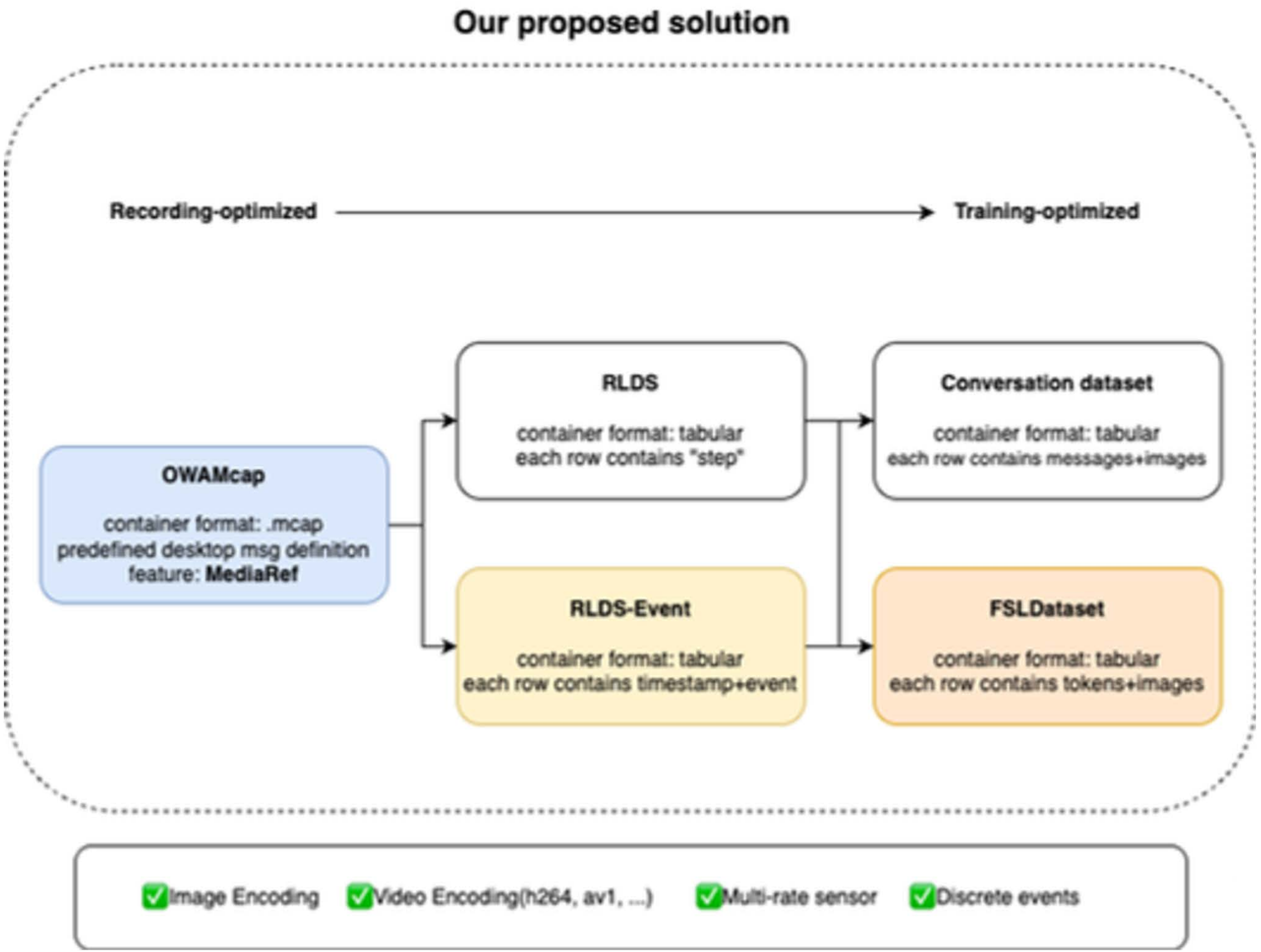
When converting the CS:GO dataset into the OWAMcap format, compression achieved a ratio of 34.45× (689 GiB → 20 GiB)

2 Data Preprocessing Pipeline

Converting recording-optimized OWAMcap format into training-optimized formats within the preprocessing pipeline

Optimized Training Efficiency

Consume media 10.2 times more rapidly while utilizing 41 times less bandwidth.



Our proposed solution
Recording-optimized
Training-optimized
OWAMcap
container format: .mcap
predefined desktop msg definition
feature: MediaRef
RLDS
container format: tabular
each row contains "step"
RLDS-Event
container format: tabular
each row contains timestamp+event
Conversation dataset
container format: tabular
each row contains messages+images
FSLDataset
container format: tabular
each row contains tokens+images
✅Image Encoding
✅Video Encoding(h264, av1, ...)
✅Multi-rate sensor
✅Discrete events

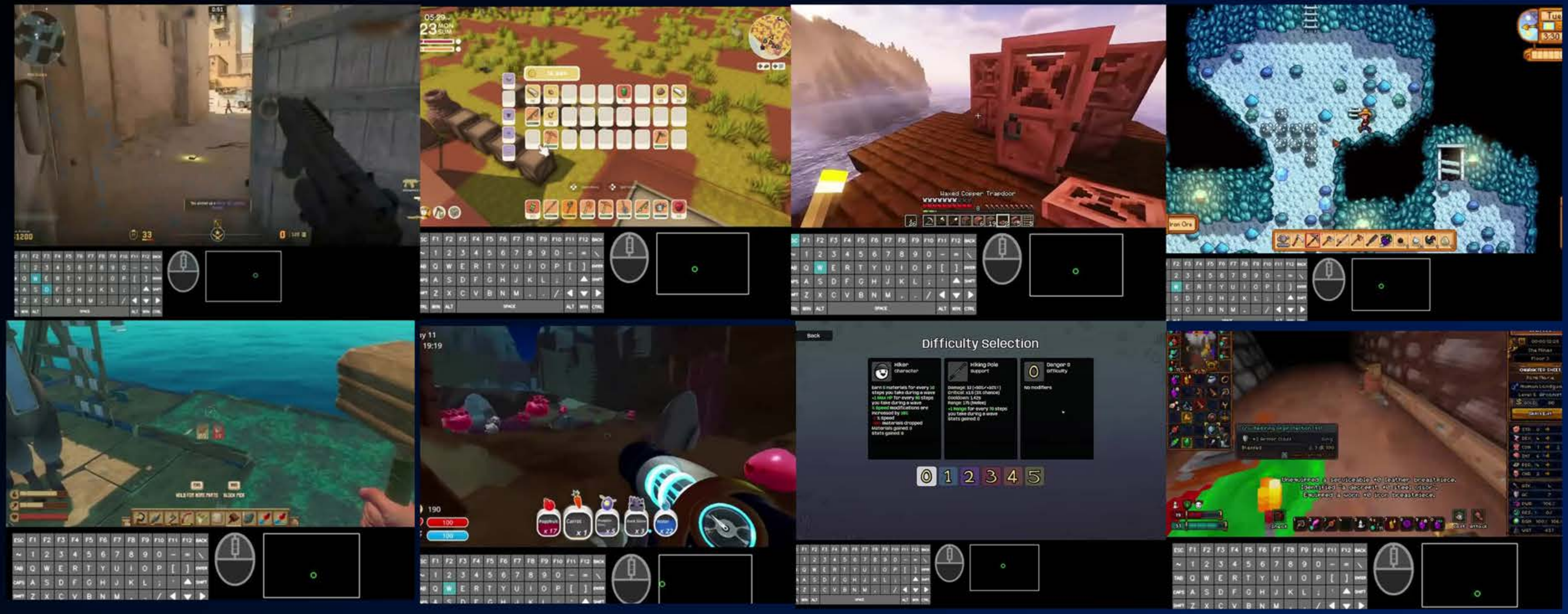
MAUM.AI's Solution – Desktop-to-Embodied AI Transfer

Scaling Action Data up to Web-Scale, essential for building Robotics Foundation Models

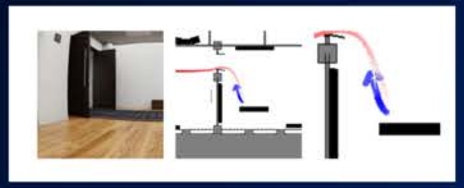
Solution: Pseudo-Labeling on Diverse Youtube Videos that embody motion

Pseudo-Labeled Desktop Game Videos can help "Real World Robotics"

Pseudo-Labeled Web-Scale Youtube Game Videos



Robotic Manipulation & Navigation



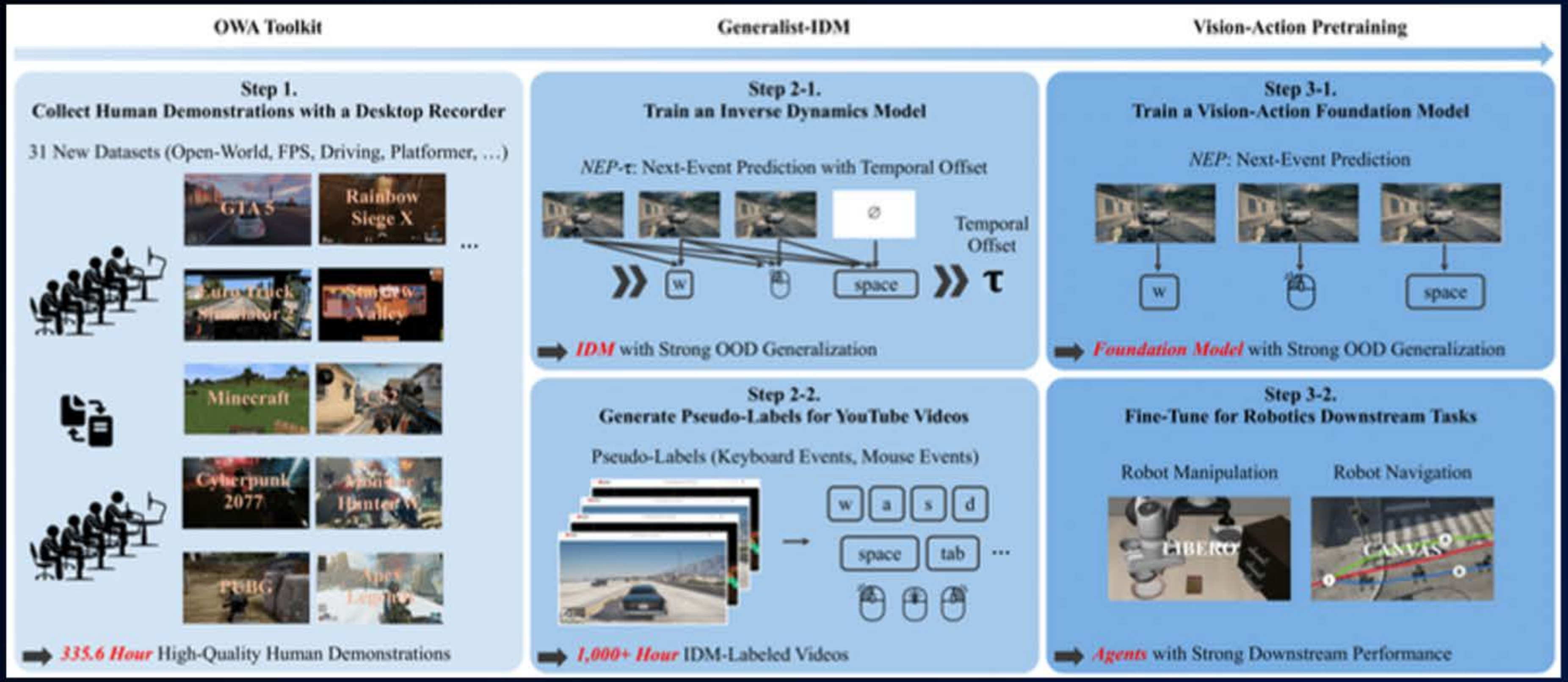
MAUM.AI's Solution – Desktop-to-Embodied AI Transfer

Scaling Action Data up to Web-Scale, essential for building Robotics Foundation Models

Solution: Pseudo-Labeling on Diverse Youtube Videos that embody motion

We show that with 335 hours of labeled action data, we can expand the data to web-scale 1,000k+ youtube videos with pseudo-labeled actions, making it contribute to real-world robotics tasks via Vision-Action Pre-Training.

We also open-source OWA (open-world-agents), the toolkit used for the whole pipeline, including data collection, training and evaluation : https://github.com/open-world-agents/open-world-agents



Lead Principle Investigators

Haebin Seong

Core Contributor of Open-World-Agents Repo
Critical Role in Performance Enhancement regarding Inference Pipeline and Optimization
Key Manager of Full Data Collection Pipeline, serving **indispensable roles of managing individuals** on data collection, data processing, visualization, contract management, and marketing.

Jaeyoon Jung

Critical Role in Training Pipeline, Downstream Navigation Task, allocation of computing resources

Suhwan Choi

Core Contributor of Open-World-Agents Repo
Critical Role in Training Pipeline, Downstream Manipulation Task

Ongoing RFM (Robotics Foundation Model) Research by Maum AI

As Korea's leading AI company, Maum AI continues its research on Physical AI Robotics Foundation Models.

LIBERO-long(10)

Ours

Maum AI's RFM (Robotics Foundation Model), despite being only 1B, demonstrates **higher performance than most existing 7B open-source VLA models**.

Benchmark	spatial	object	goal	10 (long)	total
DiT Policy (fine-tuned) (from OFT)	84.2	96.3	85.4	63.8	82.4
Octo (fine-tuned) (from OFT paper)	78.9	85.7	84.6	51.1	75.1
OpenVLA(from OFT paper)	84.7	88.4	79.2	53.7	76.5
PI-KI (from PI-KI)	98	97.8	95.6	85.8	94.3
OpenVLA-OFT(from OFT, simple-input)	96.2	98.3	96.2	90.7	95.3
baseline(seminit), full-input, causal	90.2	94	86.2	41.4	77.94
baseline, full-input, causal	76	96.2	95.4	38.4	76.5
ours(0910), full-input, causal	96	98.4	89.8	89.4	**93.4**

LET US COME TOGETHER TO BUILD A UNIFIED ECOSYSTEM FOR PHYSICAL AI

— ONE THAT DEFINES THE GLOBAL STANDARD FOR THE NEXT ERA OF INTELLIGENCE