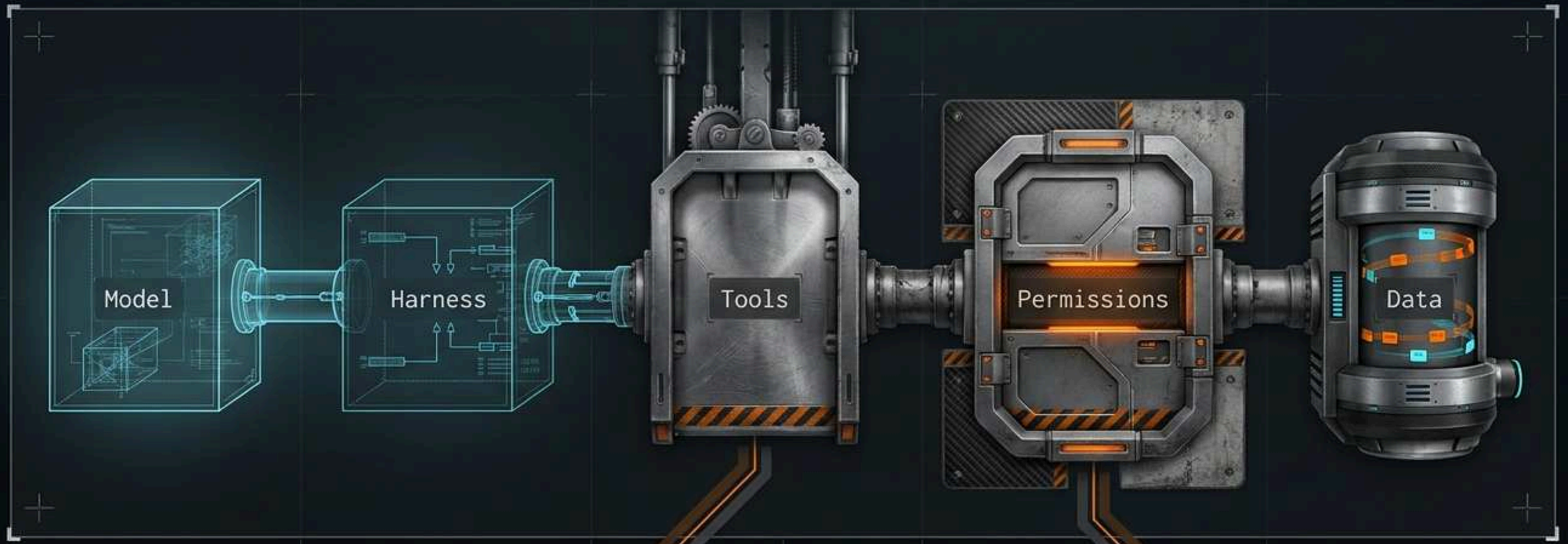
The Physical AI Frontier

Agent Paradigm Evolution and Implementation Mechanisms

Speaker: Michael Huo
Date: November 21, 2026
Session Time: 60 Minutes (Part 1: 30m)

Gemini Notebook

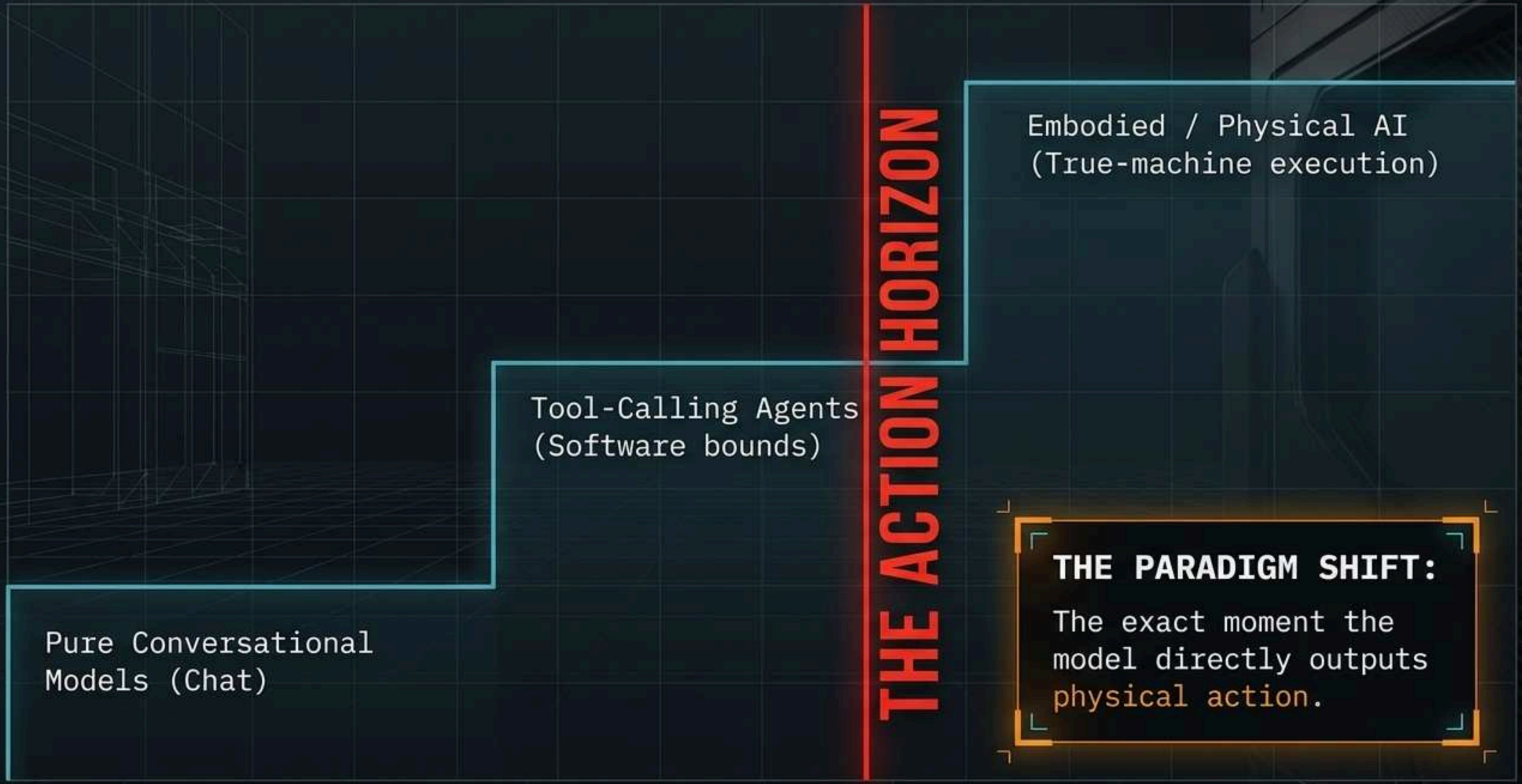
The Agent Formula Engine



Transition from software APIs to physical robotic arms & true-machine hardware.

Verifiable safety structures. The definitive bottleneck determining production readiness.

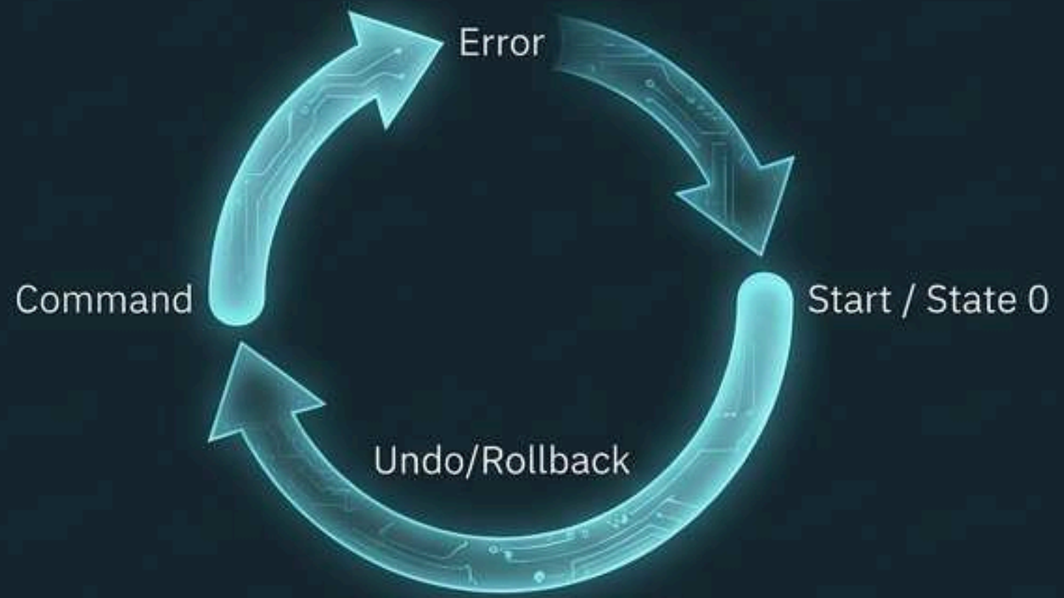
Gemini Notebook



Gemini Notebook

The Irreversibility Asymmetry

Software Agent



Physical AI



Because actions cannot be rolled back, Safety and Permissions must be forced to the absolute front of the execution flow. The Data and Permissions flywheels flywheels are an order of magnitude harder than software.

Gemini Notebook

	VLA (Vision-Language-Action)	**World / Action Models**
Core Mechanism	End-to-end direct translation from vision + language to physical action.	Internal prediction of world-state changes before action selection.
Key Advantages	Highly data-driven, fast deployment, proven in recent industrial tests.	Plannable, highly simulatable, superior interpretability.
Current Limitations	Weak long-horizon planning, highly opaque failure modes.	Persistent Sim-to-Real gap, data flywheels are still nascent.
Vanguard Models (Late 2026)	Gemini Robotics 2.5, Physical Intelligence π series.	NVIDIA Cosmos 3, EmbodiRSI.

Gemini Notebook

VLA Vanguard: Gemini Robotics 2.5

Deployment Context: Rolled out to robotics partners (e.g., Apptronik, Mercedes-Benz).



96%

Pick-and-pack success rate on Mercedes-Benz production lines using Apollo 2.

+30% to +100%

Task success rate improvement over previous generation.

+20% to +30%

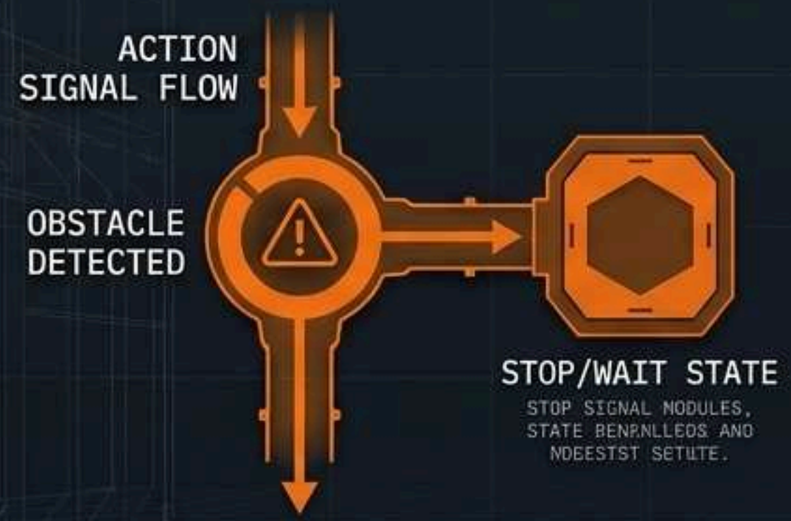
Execution speed increase.

Evaluation Breadth: Validated across 200+ distinct tasks, encompassing complex operations like flexible object manipulation (tying bows) and advanced tool usage.

Gemini Notebook

Brute-Forcing Safety in VLA Architectures

Inference-Time Safety Routing



+33%
safe completion rate in obstacle scenarios.

+43%
safety improvement in human-proximity scenarios.

ASIMOV-Jailbreak Benchmark

(Open Sourced on Hugging Face)



99.5%
Resistance rate achieved by Gemini 4 Argon (Simulation/Inference model), up from Gemini Robotics ER 2's 82%.

CRITICAL CLARIFICATION: Official DeepMind data applies to the Argon simulation model. The Gemini Robotics 2.5 action model itself remains unrated via public API and is restricted to closed partnerships.

Gemini Notebook

VLA Ecosystem Acceleration

Gemini 4 Hydrogen & Antigravity

Model: A faster, highly efficient general model with advanced spatial/embodied reasoning (Artificial Analysis Robotics 0.1 Score: 58).

Tooling: Open-source Google Antigravity Robotics Plugin.

Impact: Compresses robotic arm configuration (driver installation, camera discovery, calibration, MuJoCo simulation, closed-loop control) from WEEKS to HOURS.

Physical Intelligence (π Series)

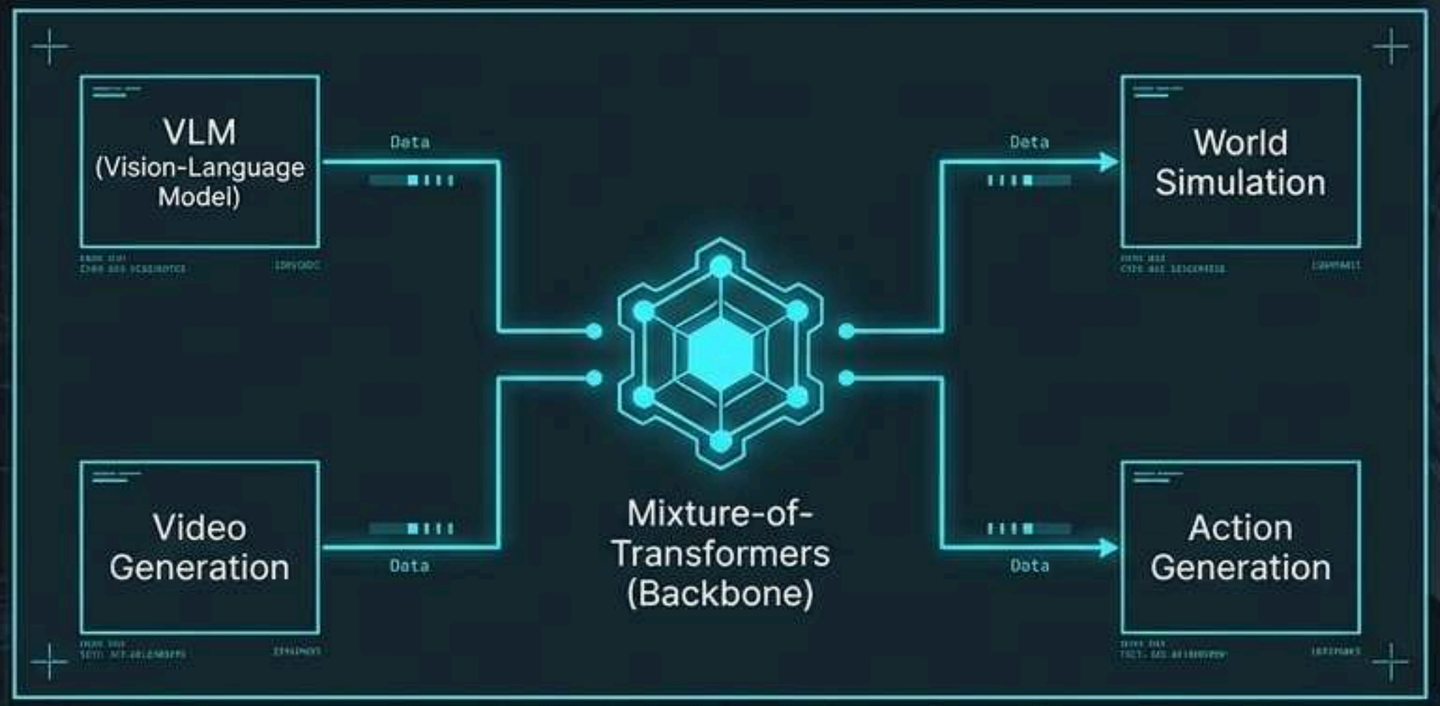
Models: π0, π0.5 (VLA architecture).

Status: Pi API beta opened October 13, 2026.

Capabilities: Cloud-hosted data upload, fine-tuning on managed compute, and inference via managed gateways. Early weights/code available via openpi repo.

Gemini Notebook

World Model Vanguard: NVIDIA Cosmos 3



Key Specifications:

Processing:
Unified processing of language, image, video, audio, and physical action.

Scale Versions:
Super (648),
Nano (168),
Edge (48).

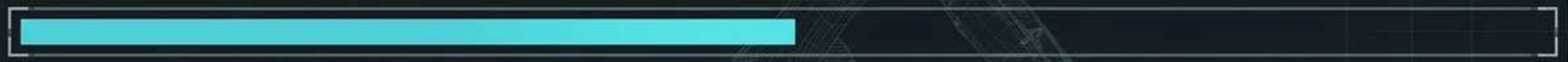
Updates:
Released May 2026.
Oct 2026 update introduced interactive world simulation checkpoints (multi-view driving, dual-arm robotic sim).

License:
OpenMDW-1.1
(Open Source).

Gemini Notebook

THE SIM-TO-REAL DELTA (EMBODIRSI V4)

SIMULATION LEAP



BASE SIMULATION SUCCESS: 50.4%



ENHANCED SIMULATION SUCCESS (TWO-ROUND UPDATE): 83.5%

THE REAL-WORLD TRANSFER (THE CRITICAL DELTA)

OLD PARADIGM

REQUIRED 200 REAL-MACHINE DEMONSTRATIONS TO ACHIEVE 75.0% SUCCESS.

EMBODIRSI PARADIGM

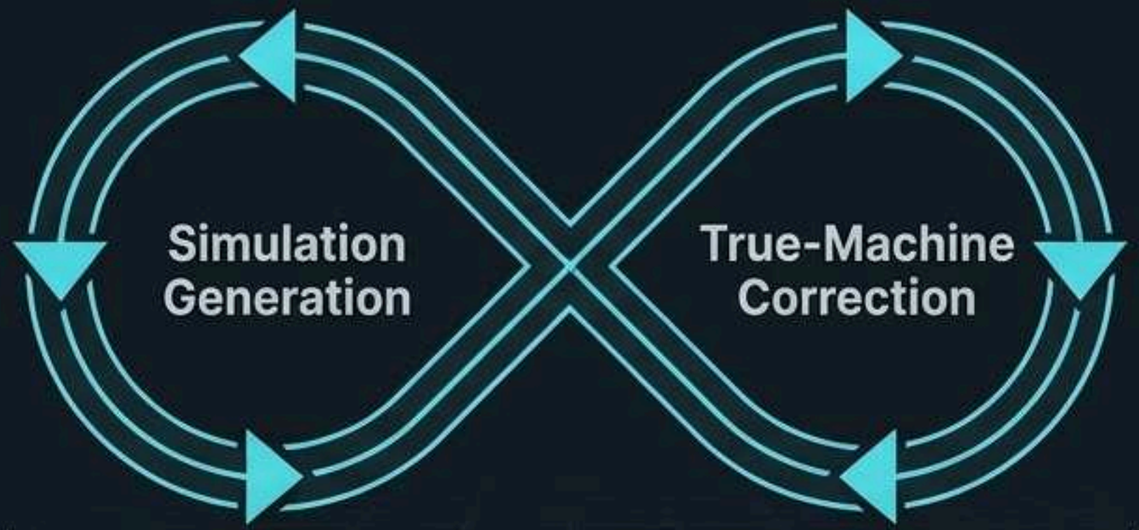
REQUIRES ONLY ~10 TRUE-MACHINE CORRECTION TRAJECTORIES PER SUB-TASK TO ACHIEVE 83.1% REAL-WORLD SUCCESS.

TAKEAWAY: TASK-SPECIFIC SIMULATION FLYWHEELS EFFECTIVELY COLLAPSE REAL-WORLD DATA REQUIREMENTS.

Gemini Notebook

The Data Flywheel Engine

EmbodiedSmith Prototype: Scene generation and task generation mutually improve one another in simulation, enabling long-horizon task training. (Note: Tasks and weights still defined by humans).



EmbodiedRSI v2 (Nov 12, 2026):
Hypothesis-graph-guided continuous learning.

STATUS ALERT: Currently limited to policy and code-level improvements.

Telemetry Data in the Loop

RoboCasa365 Overall Success: **77.0%**
(vs best baseline 40.1%)

LIBERO-Pro Success: **86.8%**

True-Machine Zero-Shot Transfer (via harness): **~71.3%**

Gemini Notebook

The Current Boundary of RSI (Recursive Self-Improvement)



What is Solved Today

Sim-to-Real flywheel is operational (Sim Iteration → Minimal True-Machine Correction → Sim Update).

AI autonomously updates data sets and policy executions.

The Hard Frontier

AI cannot yet modify its own underlying neural architecture.

Humans remain strictly in the loop for: Task Definition, Weight Selection, and Safety Boundary Definition.



Conclusion: Full Autonomous RSI—where a physical system redesigns its own model structure—remains firmly over the horizon.

Gemini Notebook

The Permissions Bottleneck



SYNTHESIS STATEMENT: A 96% production-line success rate is an illusion without pre-filtered safety.

Capability is irrelevant without verifiability.

Before every physical action, a bounded decision must occur: Allow, Reject, or Handoff to Human.

The true bottleneck to deploying Physical AI is no longer the model's intelligence; it is the verifiable, industrial-grade Permission structures (like ASIMOV-Jailbreak routing) that allow the model to legally and safely exist on a factory floor.

Gemini Notebook

Short-Term Horizon (Late 2026)



What to Expect

SURGE IN PRODUCTION-LINE, INDUSTRIAL-GRADE EMPIRICAL TESTING.

DRASTIC IMPROVEMENTS IN SIMULATION-TO-REAL DATA EFFICIENCY (COLLAPSING THE 10-CORRECTION BARRIER FURTHER).

WIDE-SPREAD INDUSTRY ADOPTION OF OPEN-SOURCE SAFETY BENCHMARKS AND TOOLS (ANTIGRAVITY, ASIMOV).

What Not to Expect

COMPLETE ARCHITECTURAL SELF-EDITING (FULL RSI). HUMANS WILL STILL DEFINE TASKS, GUARDRAILS, AND WEIGHTS.

Gemini Notebook

PART 1 CONCLUSION

The paradigm has irrevocably shifted from conversational text to irreversible true-machine execution.

The data flywheels are spinning, but the human remains the critical node at the permissions gate.

PART 2 PREVIEW

UP NEXT: ZHANG LIJIA

Topic: The Mathematical Foundations of World Models.

Focus: Moving from empirical deployment back to the theoretical mathematics driving simulation architectures.

Gemini Notebook