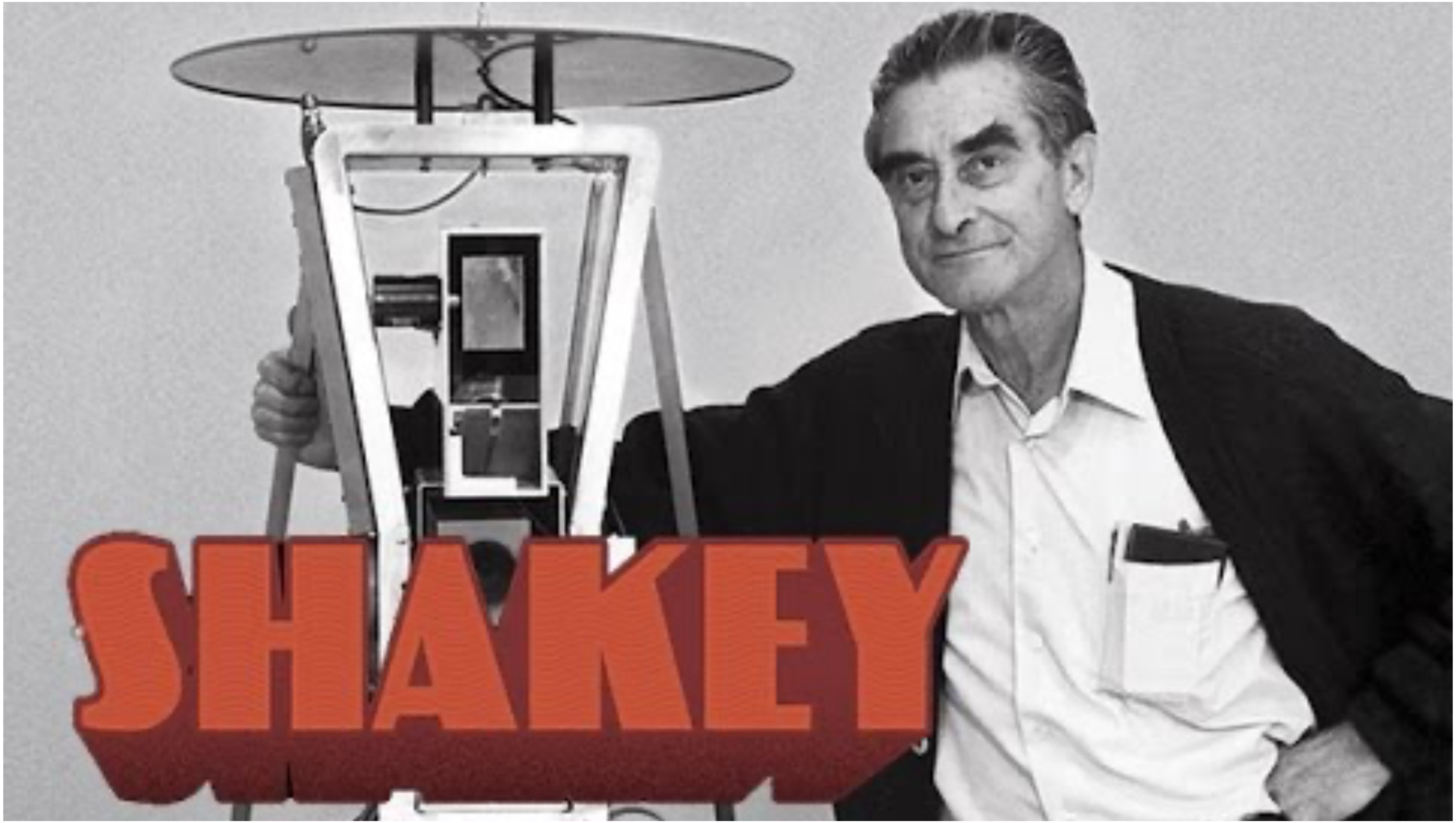
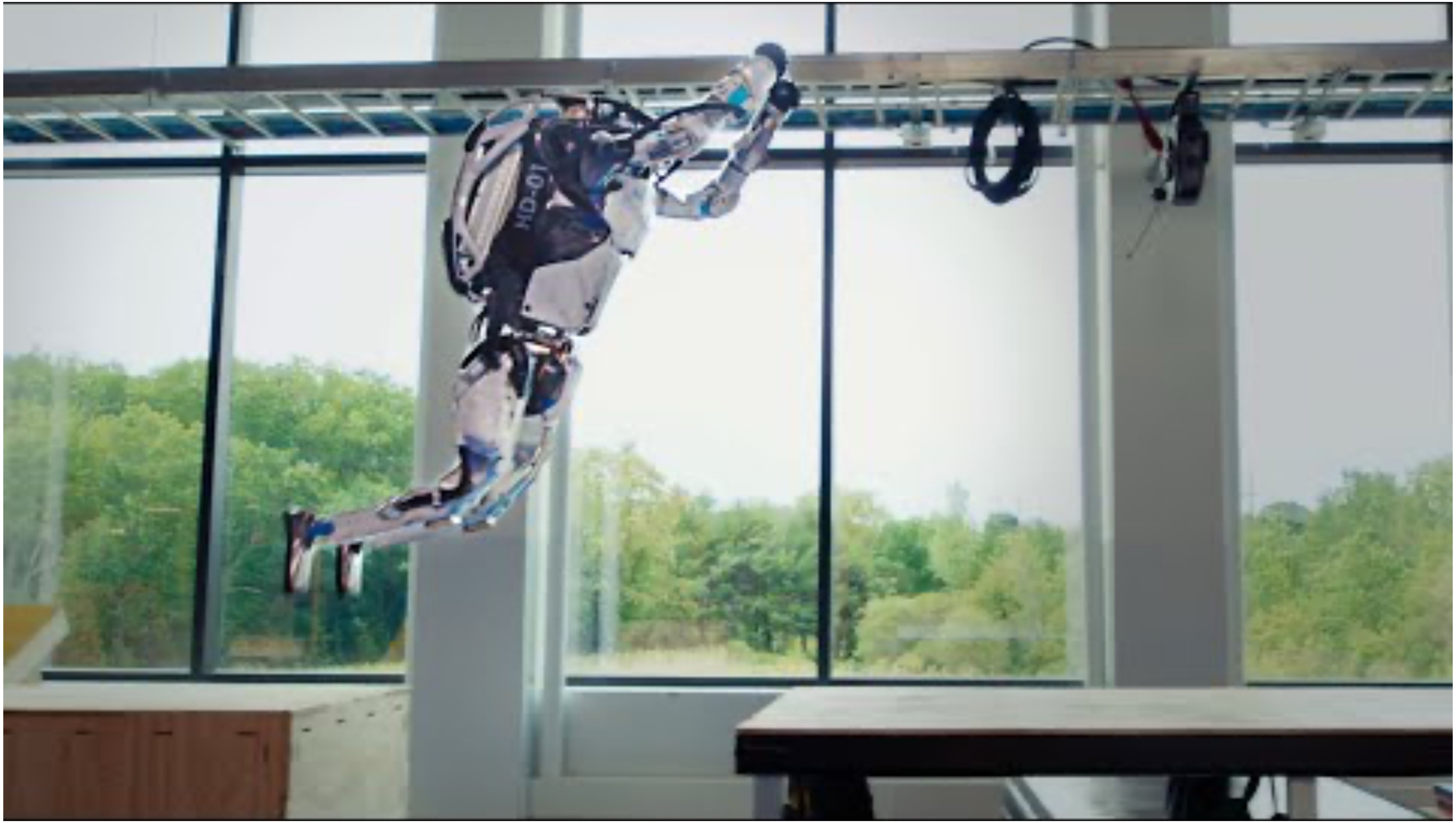

SHAKEY




Unitree G1
World's First Side-Flipping Humanoid Robot

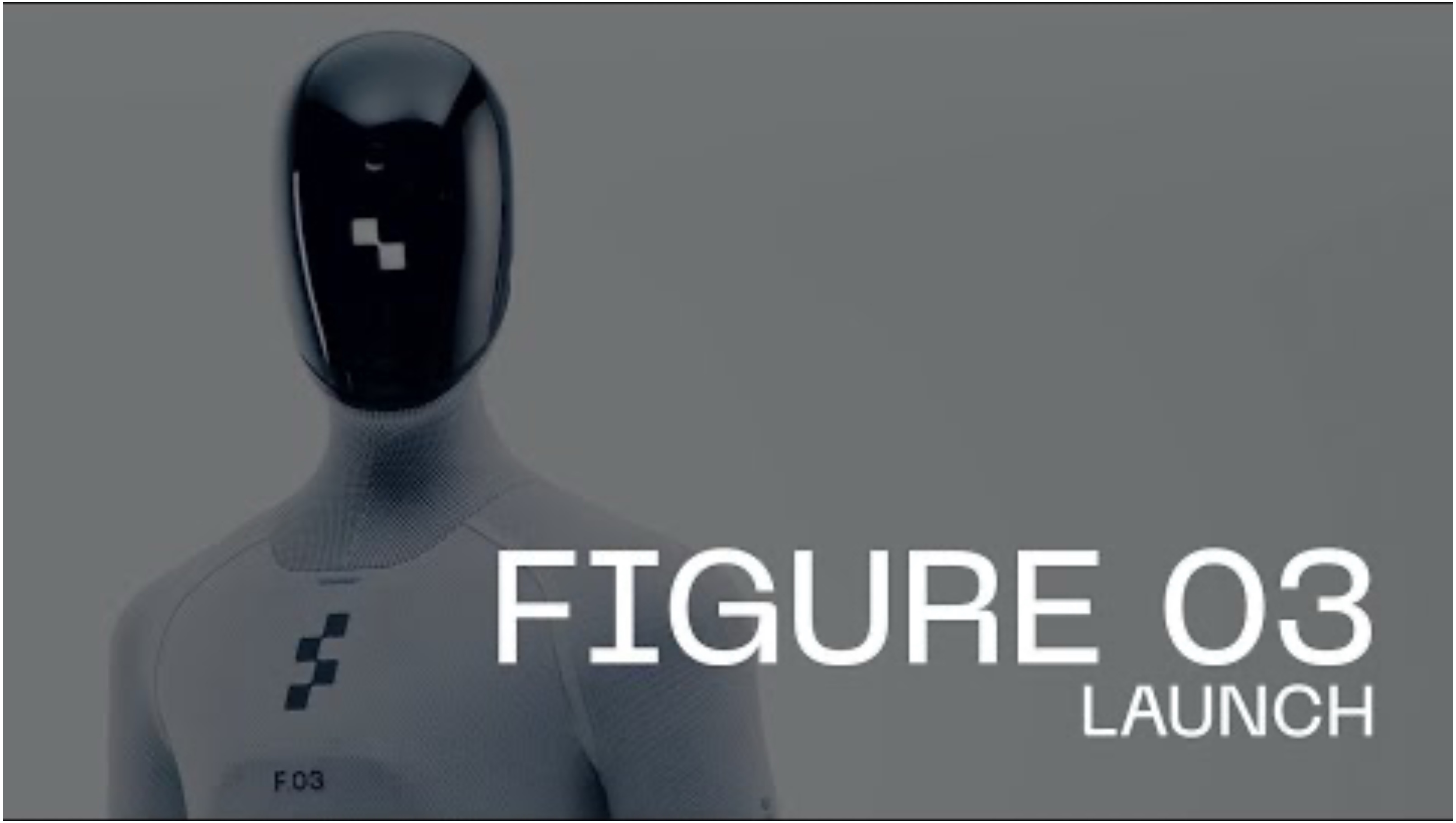


FIGURE 03
LAUNCH
F03

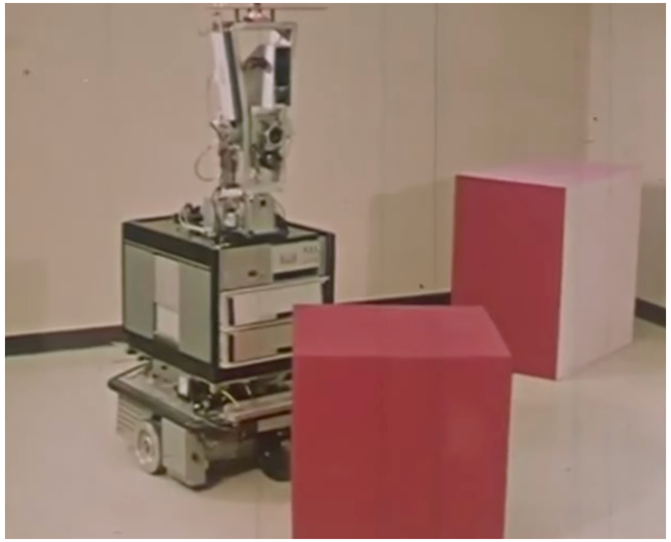
Embodied-AI:
Perception, Representation and Action

- **DATA 8010 & ELEC 8111**
- **Instructor: Prof. Yanchao Yang**

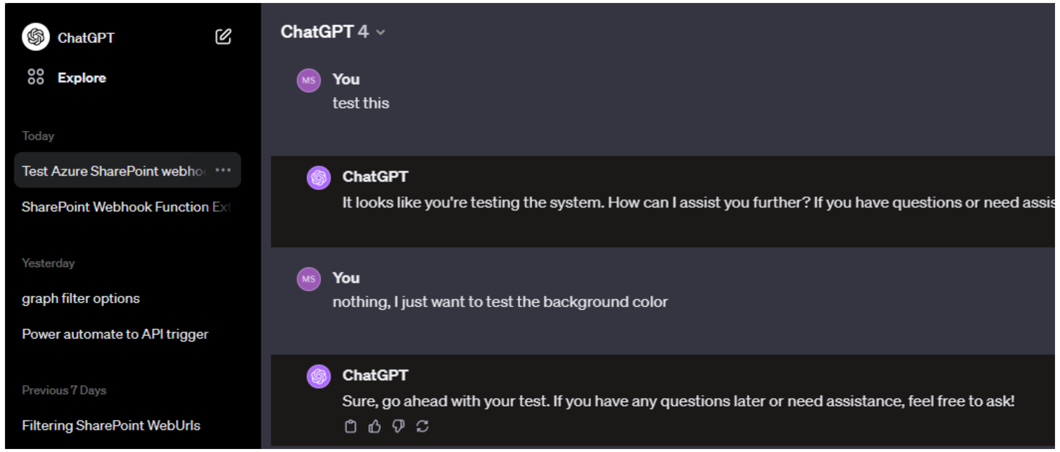
"What changed between 1969 and 2026?"

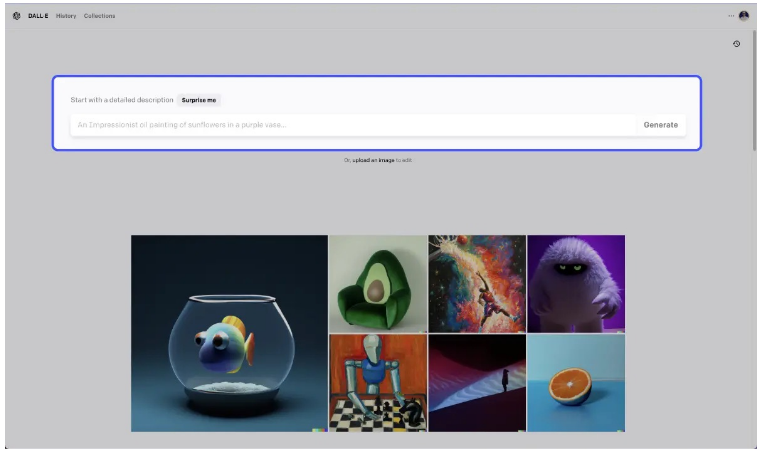
"And more importantly — what's still hard?"

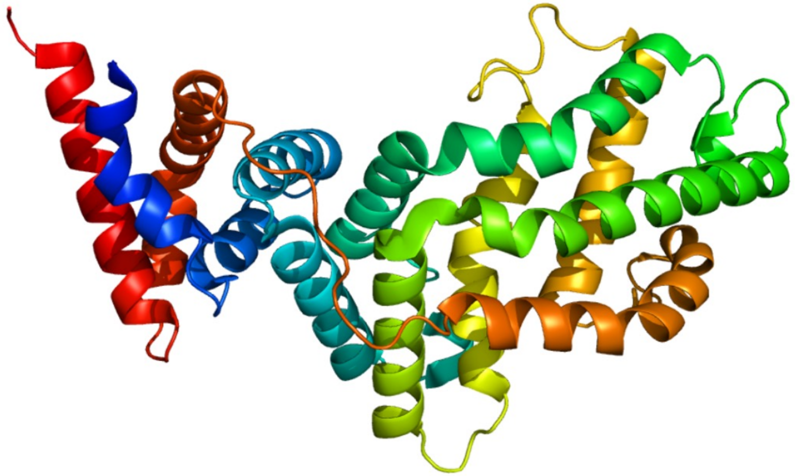
SHAKEY 1969 Humanoids 2025

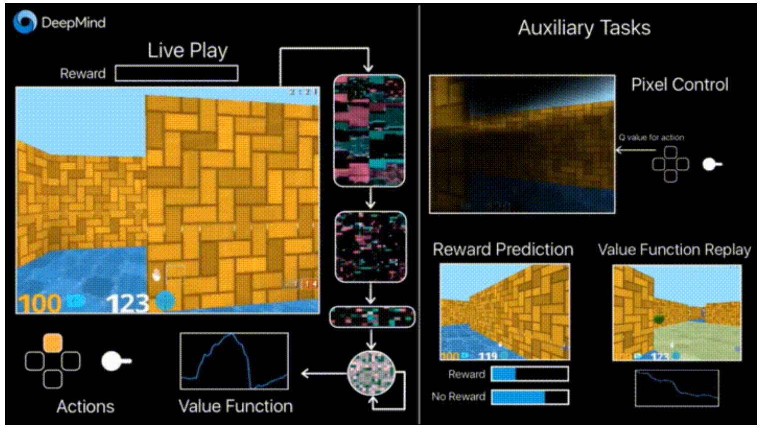


When we think of AI...



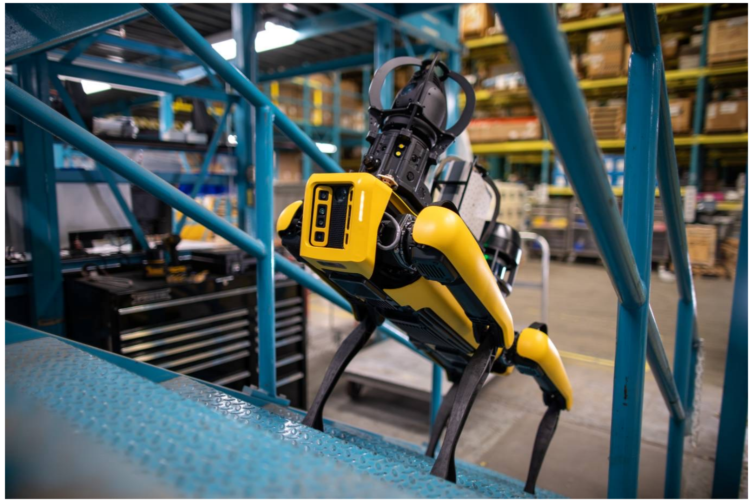


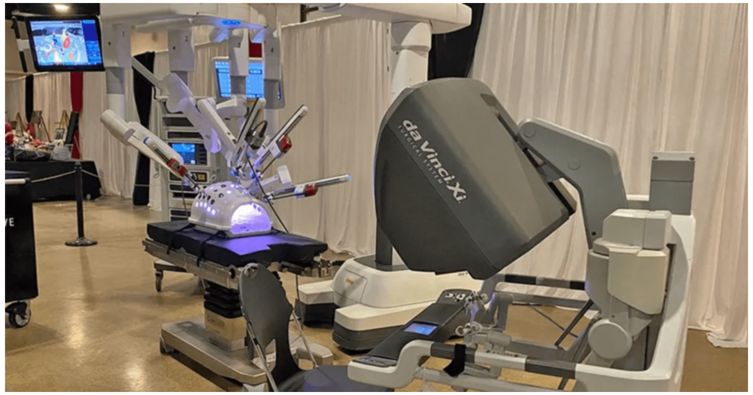




"These systems process information but don't act in the physical world."

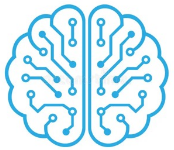
When we think of Embodied AI...

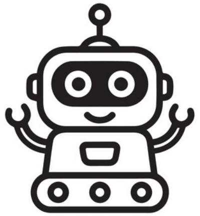




"These systems perceive, decide, AND act in the physical world."

The Fundamental Distinction





Disembodied AI

Flow: Input → Computation → Output

Processes: Tokens, pixels, symbols

Errors produce: Wrong answers

Environment: Digital sandbox

Retry: Can retry infinitely

Embodied AI

Flow: Perception → Computation → Action → World

Processes: Sensorimotor streams

Errors produce: Broken objects (or robots)

Environment: Physics constraints

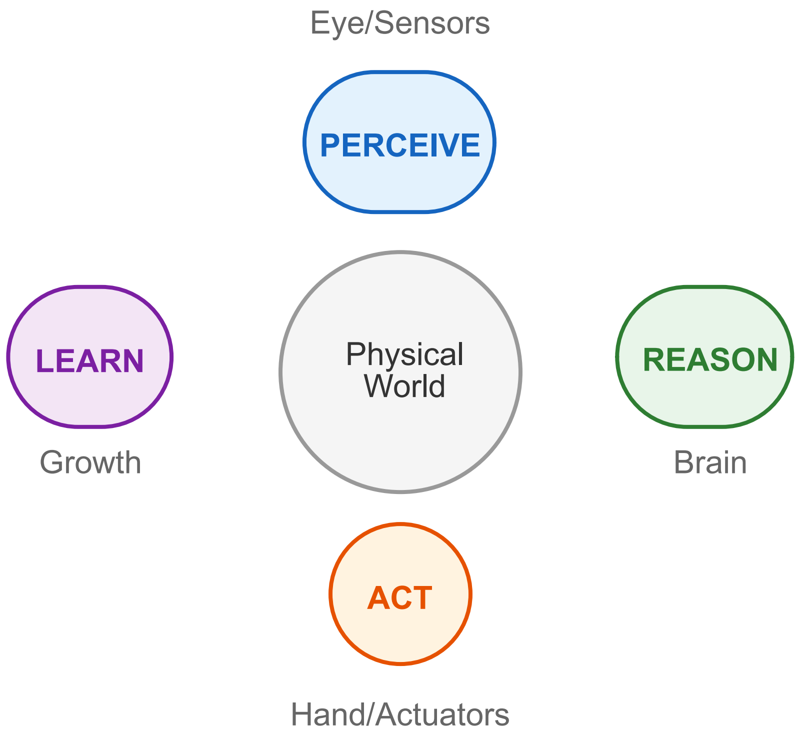
Retry: Actions have consequences

> *"The world is its own best model." — Rodney Brooks, 1990*

Formal Definition

EMBODIED AI studies intelligent systems that:

1. **PERCEIVE** the physical world through sensors
2. **REASON** about goals and constraints
3. **ACT** through actuators to change the world
4. **LEARN** from the consequences of their actions



Three Distinguishing Characteristics

Perception-Action Loops

Closed loop · Interleaved · Feedback

"The fundamental unit is not a forward pass but a closed loop where sensing and acting are interleaved."

Physical Grounding

Real-time · Safety · Consequences

"Subject to physics: gravity, friction, contact forces, inertia. Mistakes break things."

?

Uncertainty & Partial Observability

Noisy sensors · Partial observability · Dynamic world

"The world changes, often unpredictably. Other agents act independently."

Why Physical Grounding Matters: "Pick Up a Cup"

Disembodied AI (LLM)

Input: "Pick up the cup"

Output: "I would grasp the cup by its handle and lift it carefully."

Describes the action in words

GAP

Embodied AI

① Estimate cup pose from noisy depth
② Plan collision-free trajectory
③ Apply appropriate grasp force
④ Handle unexpected perturbations
⑤ Verify success via feedback

Actually performs the action

> *"The gap between knowing and doing is where embodied AI lives."*

Quick show of hands:

1. Who has programmed a physical robot before?
2. Who has worked primarily with simulations?
3. Who is coming from a pure ML/AI background?

Intelligence is not disembodied computation.

Intelligence emerges from the dynamic interaction

between an agent’s

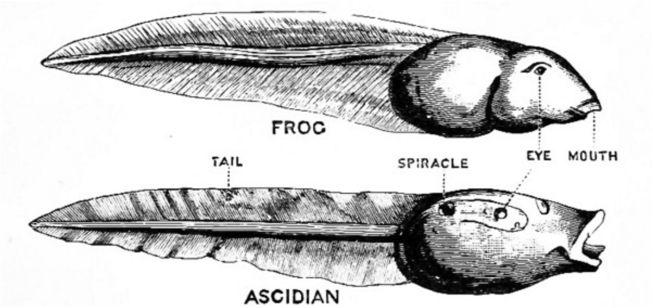
body, **brain**, and **environment**.

Embodied Cognition tradition

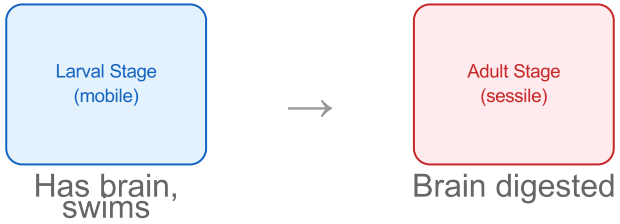
(Varela, Thompson & Rosch, 1991; Brooks, 1991; Pfeifer & Bongard, 2006)

Evidence from Biology: The Sea Squirt

Larva stage



Adult stage



Sea Squirt Lifecycle

Larval Stage

Has a notochord, neural tube, and cerebral ganglion (primitive brain). Swims actively to find a suitable location to settle.

After Settling

Once attached to a surface, the sea squirt **digests its own brain**, absorbing it for nutrients. It no longer needs to move.

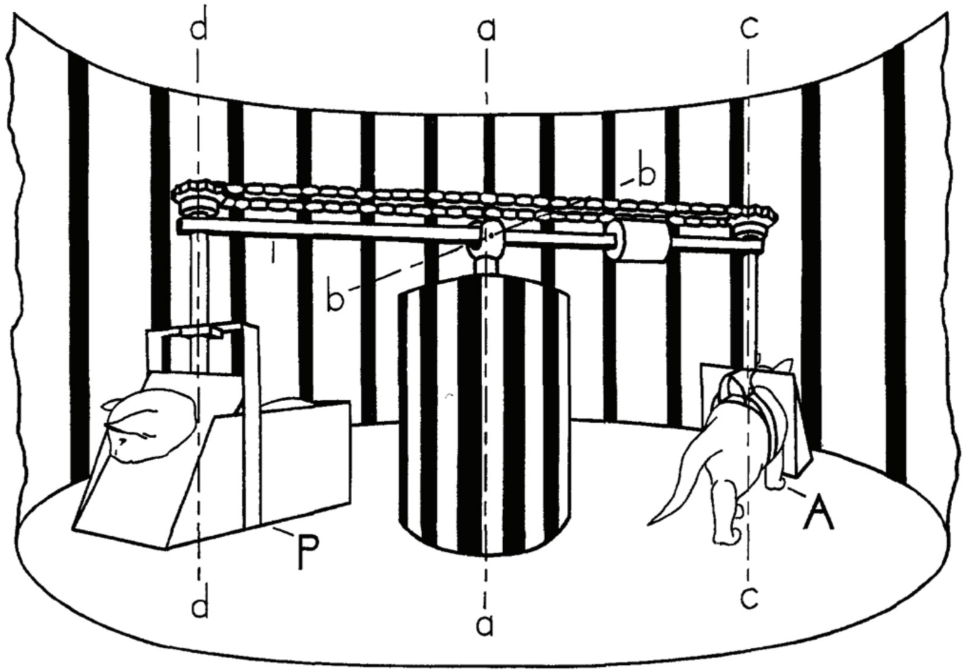
Implication

"If you don't need to move, you don't need a brain."

"The brain exists to produce adaptable, complex movements." — *Daniel Wolpert*

The Kitten Carousel Experiment (Held & Hein, 1963)

Held & Hein (1963)



Setup

Two kittens in identical visual environments. One walks (active), one is carried (passive). Both see the same things.

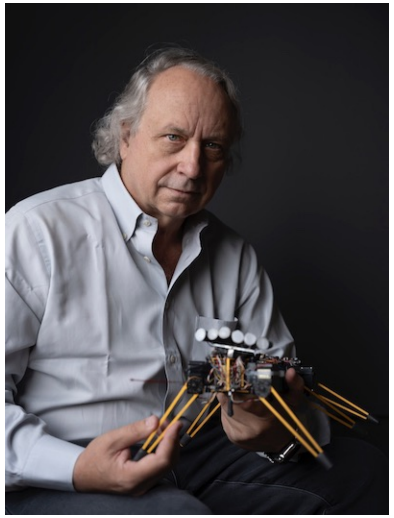
Results

Only active kittens developed normal depth perception and visual cliff avoidance.

Implication

"Perception requires action to develop properly."

Rodney Brooks: Intelligence Without Representation



Key Papers

"Intelligence Without Representation" (1991)

"Elephants Don't Play Chess" (1990)

1. The Representation Bottleneck

Building accurate world models is intractable. The world is too complex, changes too fast, and has too many relevant details to capture in symbolic form.

2. Behavior Emerges from Interaction

Intelligent behavior doesn't require a central executive or explicit planning. It emerges from the interaction of simple behaviors with the environment.

3. Situatedness and Embodiment

Robots must be situated in the real world, not in simulated or abstract environments. The physical body and its interaction with the world are essential.

> *"The world is its own best model."*

The Subsumption Architecture

← SENSORS

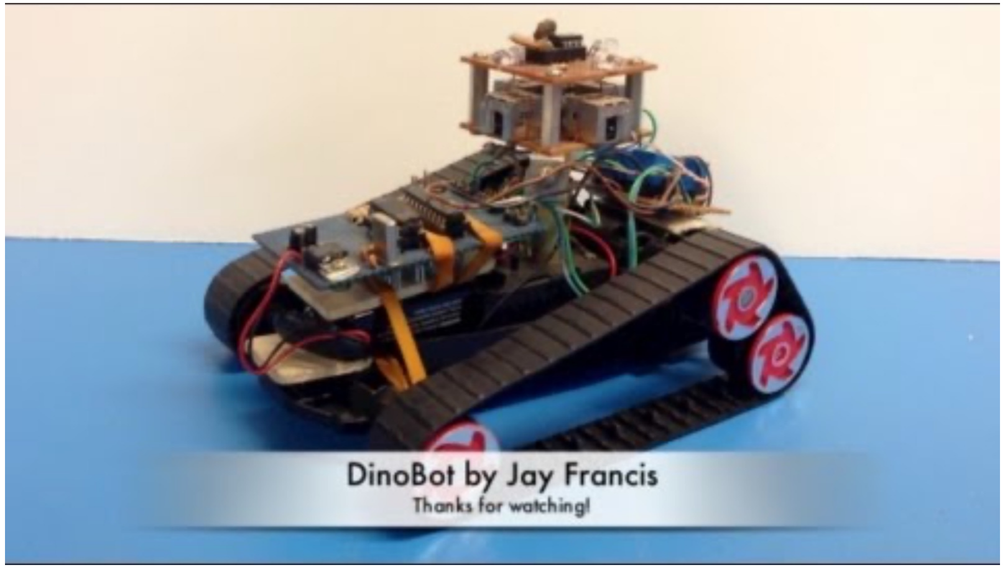
Level 3: EXPLORE
Explore distant places
↓ inhibits

Level 2: WANDER
Move around aimlessly
↓ inhibits

Level 1: AVOID OBSTACLES
Don't hit things
↓ inhibits

Level 0: BASIC LOCOMOTION
Move legs

ACTUATORS →



Key Insight

Each layer is a complete behavior system. Higher layers modulate (not replace) lower layers.

How It Works

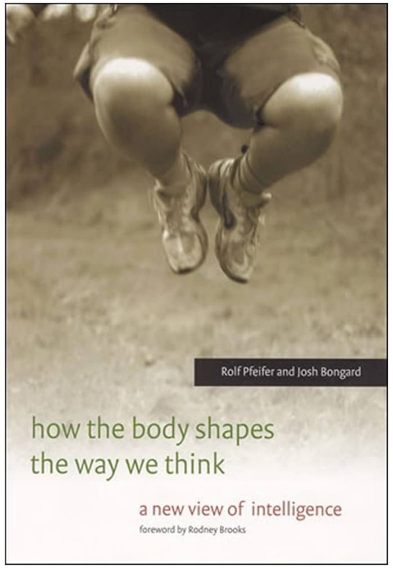
Lower layers run continuously. Higher layers can suppress outputs when appropriate, but lower layers remain autonomous.

Example Robots

Allen (1985), Herbert (1988), Genghis (1989)

How the Body Shapes the Way We Think

Pfeifer & Bongard (2006)



1. Morphological Computation

The body does part of the "thinking." Physical structure constrains and guides behavior without requiring explicit computation.

Example: A well-designed hand naturally conforms to objects during grasping.

2. Ecological Balance

Match complexity of body, brain, and environment. "Cheap design" — let physics do the work instead of computation.

Example: Passive dynamics in walking reduce motor control complexity.

3. Embodied Development

Intelligence develops through physical interaction with the world. It cannot be programmed from scratch — it must be learned through experience.

Example: Infants learn sensorimotor skills through months of exploration.

Morphological Computation in Action: Passive Dynamic Walkers

Mechanical Schematic

No motors | No sensors | No computer

Gravity = energy | Geometry = control

Passive Dynamic Walker

What It Is

A mechanical device that walks down a gentle slope using only gravity. No motors, no sensors, no computer. The geometry of the legs creates a stable walking gait.

Why It Matters

Demonstrates that intelligent-looking behavior can emerge from physical design alone. The "control" is embedded in the morphology, not computed by a brain.

Lesson for Robot Design

"If you design the body right, you need less brain."

Source: Cornell Passive Dynamic Walking Project; MIT Leg Lab

Connecting Philosophy to Practice

Challenge	Embodied Cognition Insight	Modern Application
LLMs hallucinate	Lack of physical grounding — no feedback from reality to correct false beliefs	VLAs ground language in physical action and perception
Policies don't generalize	No sensorimotor development — passive learning from fixed datasets is insufficient	Sim-to-real transfer, domain randomization, active learning
Robots are brittle	Over-reliance on planning — world models fail when reality diverges	Reactive control + learned policies that adapt online
Data efficiency	Passive learning insufficient — agents must actively explore	Curiosity-driven exploration, intrinsic motivation

Key Message: Foundation models are powerful but lack grounding. Embodiment provides the reality check.

"Those who cannot remember the past are condemned to repeat it."

— George Santayana

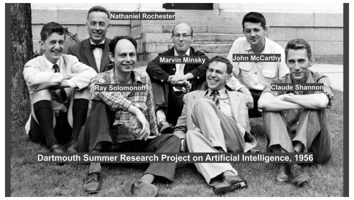
Reframed:

To understand why learning-based methods are transforming robotics, we must understand what classical methods achieved —

and where they fell short.

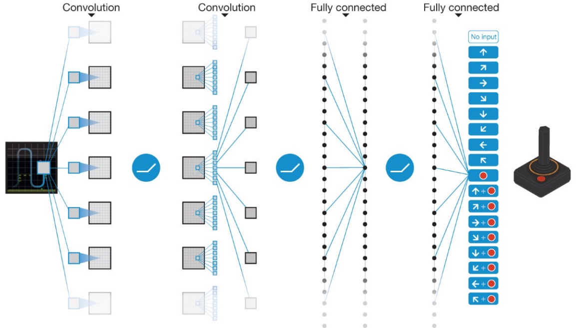
The Evolution of Embodied AI

Dartmouth
1956

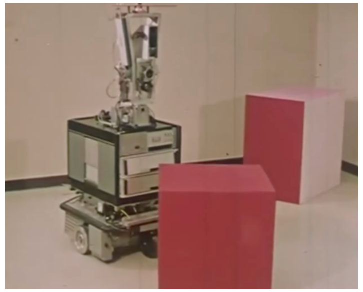


Subsumption
1986

AlexNet/DQN
2012-15



Classical/Symbolic	Behavior-Based	Probabilistic	Deep Learning	Foundation Models
1956 – 1970	– 1985	– 1995	– 2012	2022 – NOW

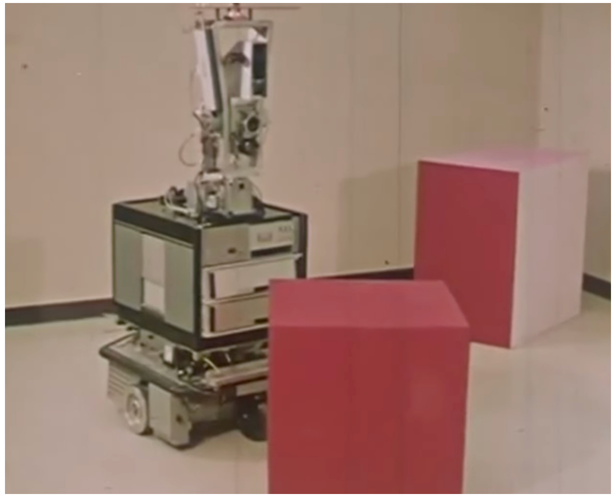


SHAKEY/STRIPS
1969-71

Stanley (DARPA)
2005

RT-1/RT-2/OpenVLA
2022-24

SHAKEY: The First Mobile Robot to Reason (1966-1972)



Built at SRI International (Stanford Research Institute)

What SHAKEY Was

First robot to integrate **perception**, **planning**, and **action** in a unified system.

Could navigate rooms, push boxes, climb ramps — all while reasoning about its goals and the state of the world.

Technical Innovations

A* Search Algorithm (Hart, Nilsson, Raphael, 1968) — Optimal pathfinding, still used today

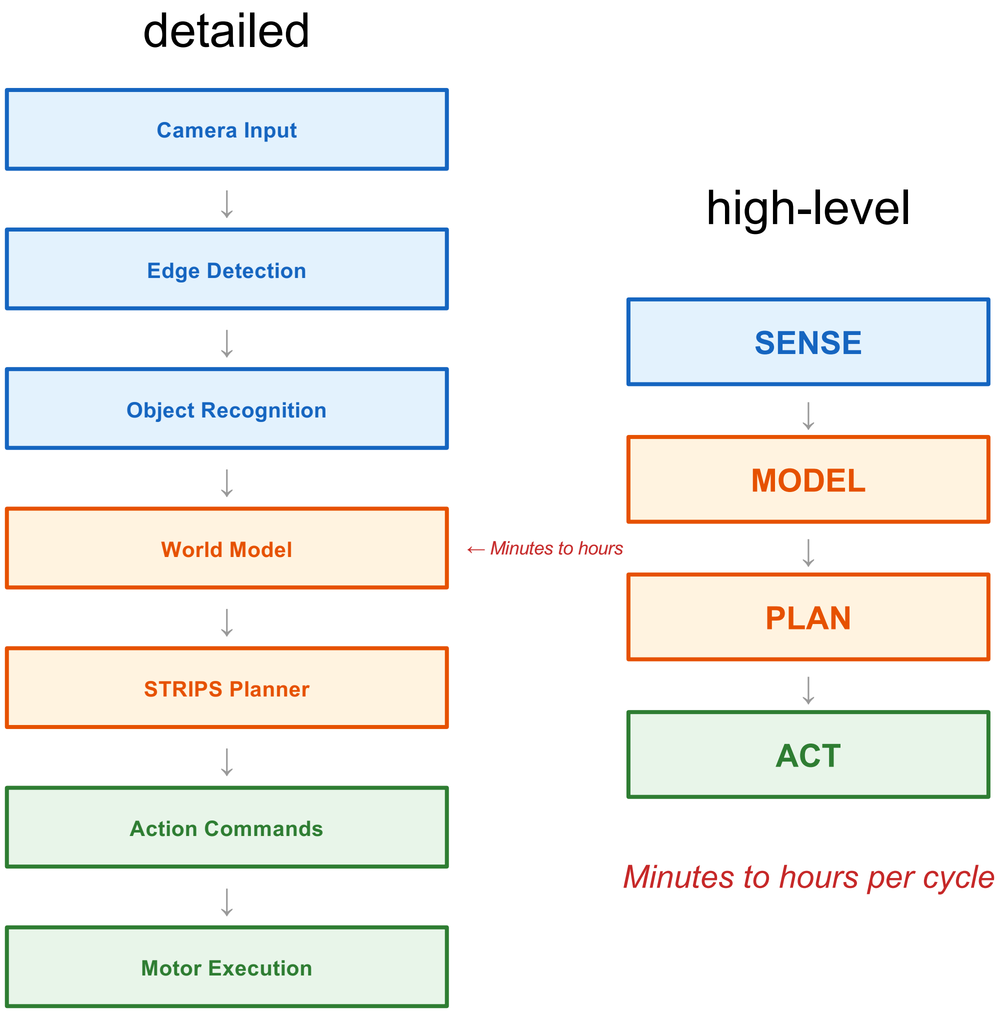
STRIPS Planner (Fikes & Nilsson, 1971) — Foundation of automated planning

Hough Transform — For SHAKEY's vision

Historical Significance

SHAKEY established the Sense-Model-Plan-Act paradigm that dominated robotics for decades.

SHAKEY's Architecture: Sense-Model-Plan-Act



Key Characteristics

Sequential processing. Build complete world model before acting. Each step takes significant time.

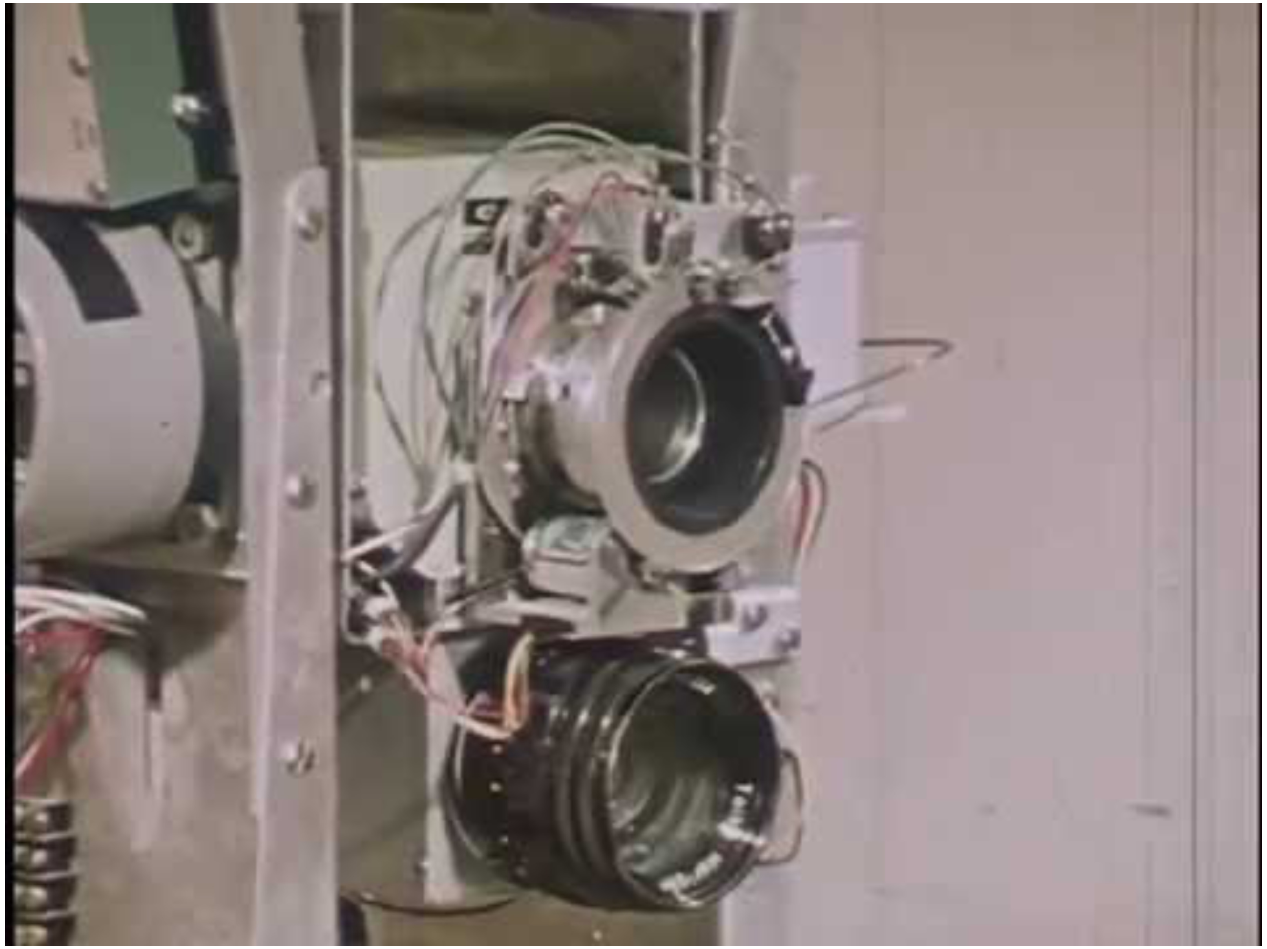
The Paradigm

SMPA dominated robotics for decades. Assumes world can be fully modeled.

Critical Limitation

"Works in controlled labs. Struggles in the real world where environments change faster than the robot can plan."

SHAKEY in Action (1969)



> "Notice the deliberate pace — SHAKEY stops to think between actions."

Industrial Manipulators (1970s-1980s)

Unimate (1961)
GM Assembly Line

PUMA (1978)
Programmable Arm

The Success Story

Precise, repeatable, tireless. Transformed manufacturing worldwide.

The Limitation

Structured environments only

Pre-programmed trajectories

No perception of novel situations

"Blind and rigid"

"Industrial robots succeeded by eliminating uncertainty, not by handling it."

Why Didn't Robots Take Over?

Key Events

1973: Lighthill Report criticizes AI progress

1980s: Expert systems boom, then bust

Robots didn't "take over" as predicted

1973	1980s	1990s
Lighthill	AI Winter	Recovery

Why Robotics Stalled

1. Sensing was hard

Cameras expensive, low-resolution

2. Computation was slow

Couldn't process in real-time

3. Generalization failed

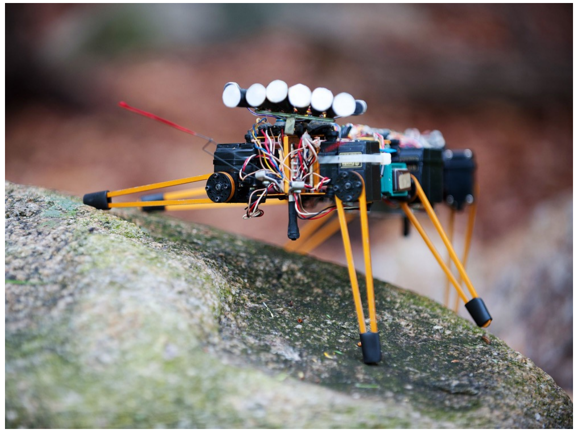
Every new task required reprogramming

4. The world is messy

Factories controlled, homes not

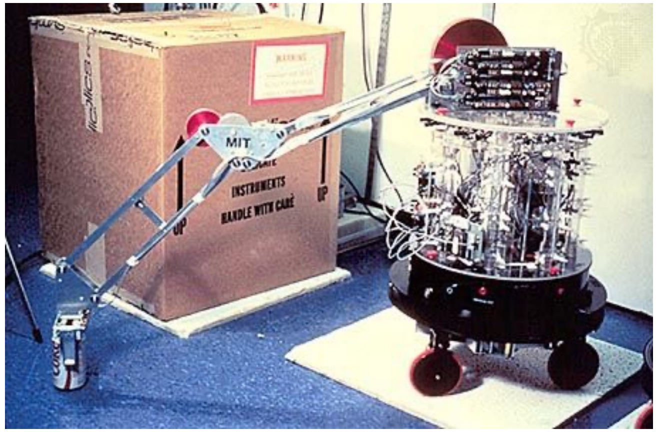
Behavior-Based Robotics (1985-1995)

Sense-Model-Plan-Act	Behavior-Based
Serial processing	Parallel behaviors
Build complete model	React to immediate stimuli
Plan then execute	Act while perceiving
Centralized control	Distributed, emergent



Genghis

6-legged walker



Herbert

Soda can collector

Commercial Success

Roomba (2002) — built on behavior-based principles. Over 40 million sold.

Lessons from the Classical Era

1. Representation is Hard

Building accurate world models is intractable.

The world changes while you model it.

2. Real-Time Matters

Deliberation taking minutes is useless for dynamic tasks.

Reactive systems handle unexpected changes.

3. The Body Helps

Mechanical design simplifies control.

Physical compliance handles contact.

4. Generalization Remains Elusive

Every approach worked in its niche.

Transferring to new domains required starting over.

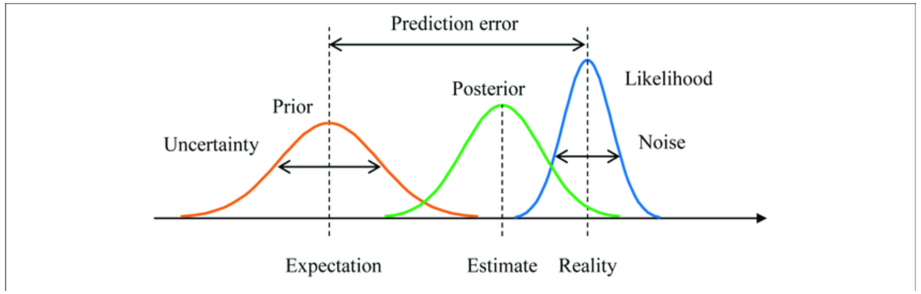
The Probabilistic Revolution: Embrace Uncertainty

Classical Approach

"Build a perfect model, plan the optimal action"

Probabilistic Approach

"Maintain beliefs about the world, act to reduce uncertainty"



Prior → Observe → Posterior

Mathematical Foundation

$$P(\text{state} \mid \text{observations}) \propto P(\text{observations} \mid \text{state}) \cdot P(\text{state})$$

Bayes' rule applied to robotics: Update beliefs based on sensor observations. The robot maintains a probability distribution over possible states rather than a single estimate.

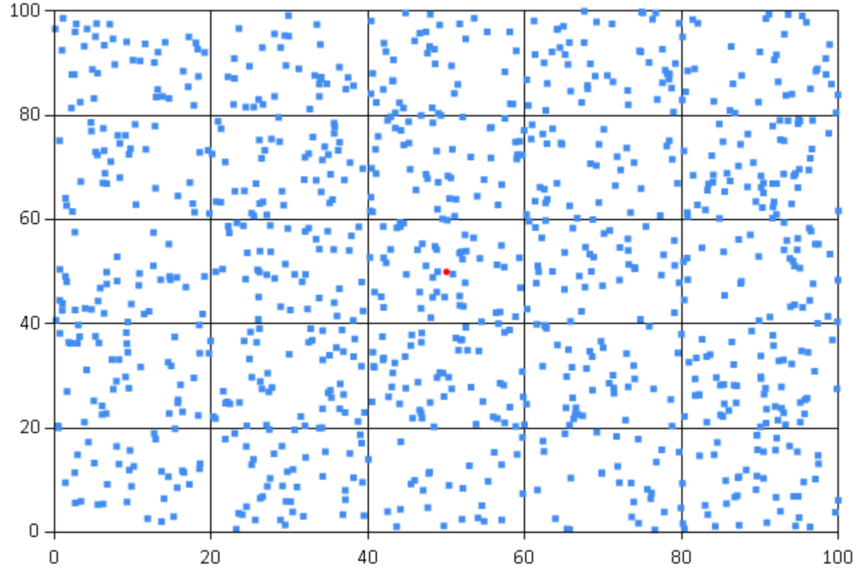
Key Algorithms of the Probabilistic Era

Algorithm	Problem Solved	Key Idea
Kalman Filter (1960)	State estimation	Optimal fusion of prediction + observation (Gaussian)
Particle Filter (1993)	Non-Gaussian estimation	Sample-based belief representation
SLAM (1990s)	Localization + Mapping	Build map while localizing within it

Dispersed particles → Clustering → Converged

t = 0 | t = 5 | t = 10

Particle Filter Convergence



The DARPA Grand Challenges

2004 Grand Challenge

Course: 142-mile desert route

Prize: $1 million

Result:

No vehicle finished

Best performance: 7.3 miles

2005 Grand Challenge

Course: 132-mile desert route

Prize: $2 million

Result:

5 vehicles finished

Winner: Stanley (Stanford, Sebastian Thrun)

Key Insight: In just one year, the field went from complete failure to five finishers. The challenge catalyzed rapid progress by creating a concrete, measurable goal.

2007 Urban Challenge

The Challenge

Navigate 60 miles of urban environment

Obey traffic laws, handle intersections

Interact with other (robotic) vehicles

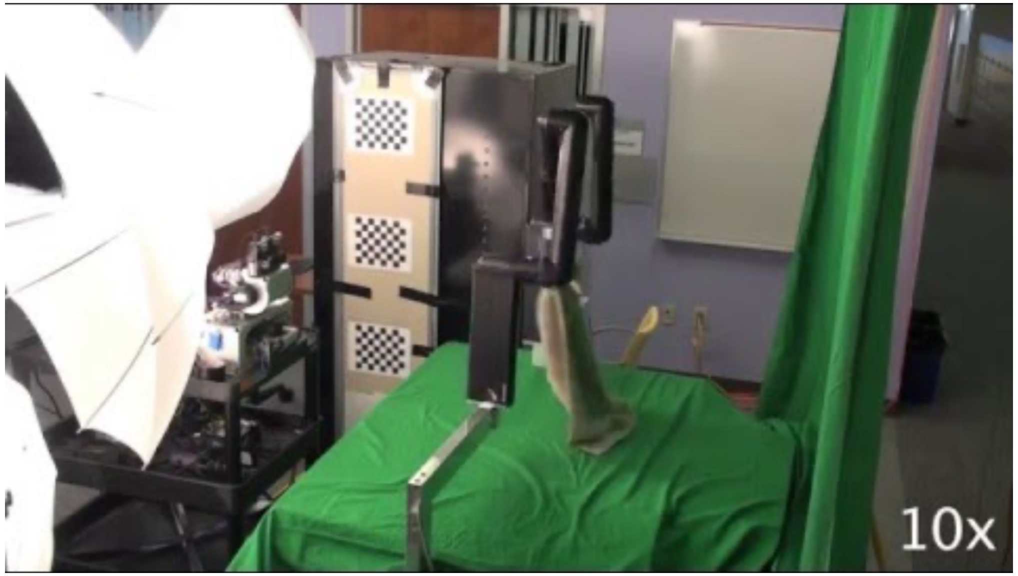
Winner

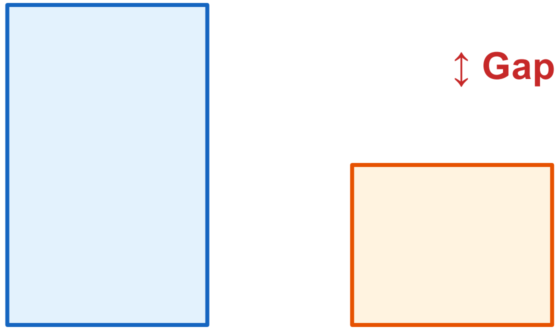
Boss (Carnegie Mellon University)

Industry Impact

"Launched the autonomous vehicle industry. Waymo, Cruise, Aurora — all trace their lineage to these challenge teams."

Meanwhile, in Manipulation...





Navigation
Rapid progress

Manipulation
Slower progress

1995 → 2020

Navigation

SLAM matured

Autonomous vehicles emerged

Indoor navigation reliable

Manipulation

Contact physics harder

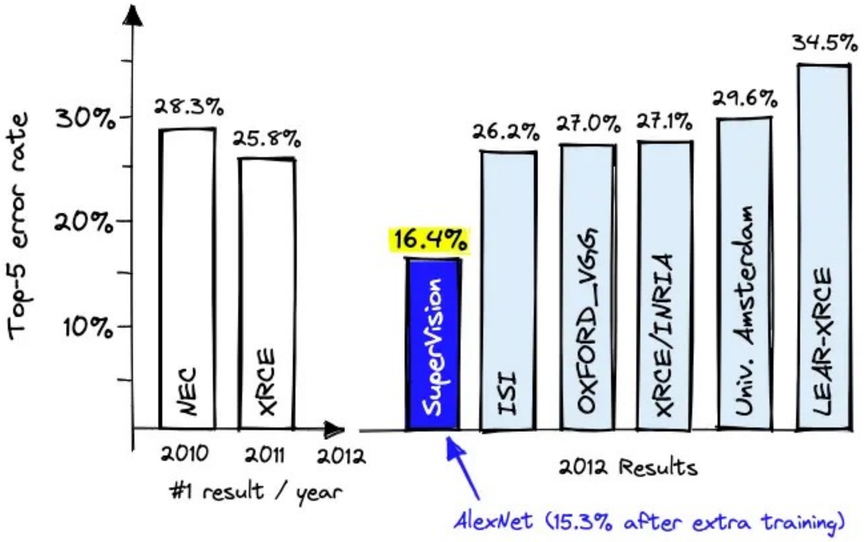
Object variety enormous

Dexterity challenges remain

Why the Difference?

Navigation can plan in 2D space. Manipulation involves 6-DOF contact dynamics, deformable objects, and tool use.

The ImageNet Moment (2012)



Top-5 Error Rate on ImageNet

26.2%
2011

15.3%
AlexNet 2012

AlexNet (Krizhevsky, Sutskever, Hinton, 2012)

Deep convolutional neural network

GPU-accelerated training

Won ImageNet by large margin

Why It Mattered for Robotics

Demonstrated that learning from data beats hand-designed features.

Suggested the same might apply to control.

Launched the modern deep learning era.

Deep Reinforcement Learning (2013-2015)

DQN Playing Atari Breakout

Source: Mnih et al., Nature 2015

DQN (Mnih et al.)

49 Atari games from raw pixels

No game-specific engineering

Human-level performance

Key Equation

$$Q(s, a; \theta) \approx r + \gamma \max Q(s', a'; \theta^-)$$

The Insight

"Raw sensory input → Neural network → Action. End-to-end learning."

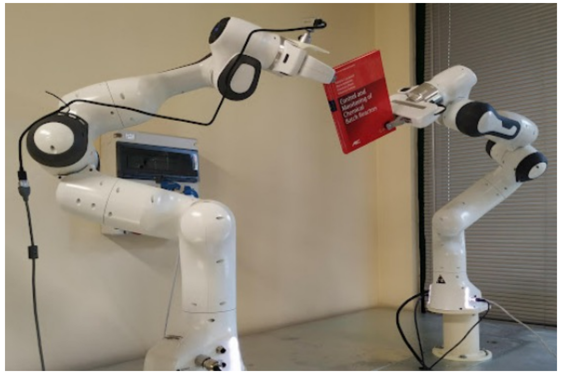
From Games to Robots — The Challenge

Atari Games	Real Robots
Perfect observation	Noisy, partial observation
Deterministic physics	Complex, continuous physics
Fast (1000s games/hr)	Slow (~10 episodes/hr)
Safe failure	Failure breaks things
Unlimited retries	Expensive mistakes

Key Question

"How do we get millions of training episodes without destroying millions of robots?"

The Answer: Simulation



Physical Robot

↔

Digital Twin

Core Idea

1. Build simulated version of robot and environment

2. Train policies in simulation (fast, safe, parallel)

3. Transfer learned policies to real robot

Challenge: The Sim-to-Real Gap

Physics approximations

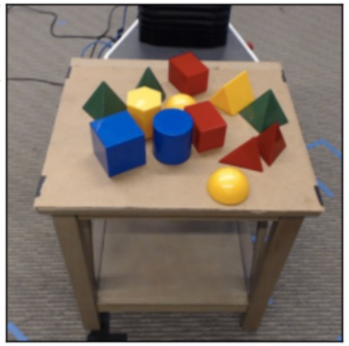
Visual differences

Sensor/actuator differences

Domain Randomization (Tobin et al., 2017)

Training

Test



Key Idea

"Train on such a diverse distribution of simulated environments that reality looks like 'just another sample.'"

Randomized Parameters

Textures and colors

Lighting conditions

Camera positions

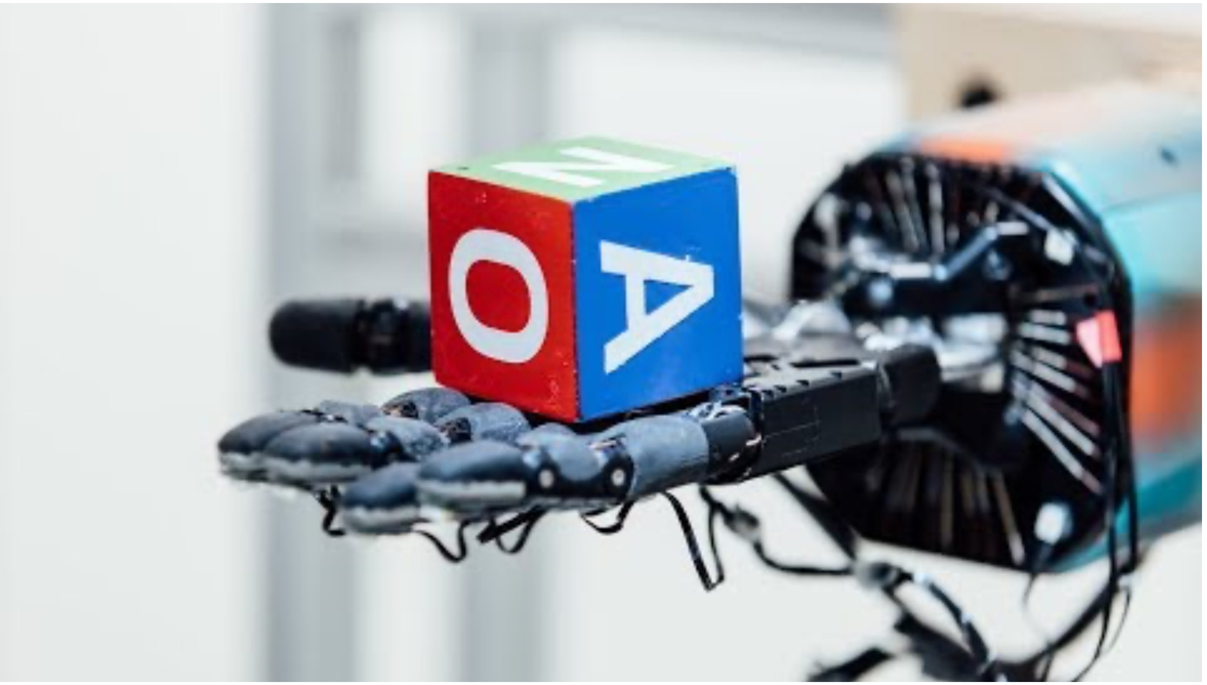
Physics parameters

Result

Object detection with **1.5cm accuracy** using zero real training data.

OpenAI Rubik's Cube (2019)

Shadow Hand Solving Rubik's Cube



Achievement

Robot hand solves Rubik's Cube

Policy learned entirely in simulation

60% success rate on real robot

Automatic Domain Randomization (ADR)

100+ randomized parameters

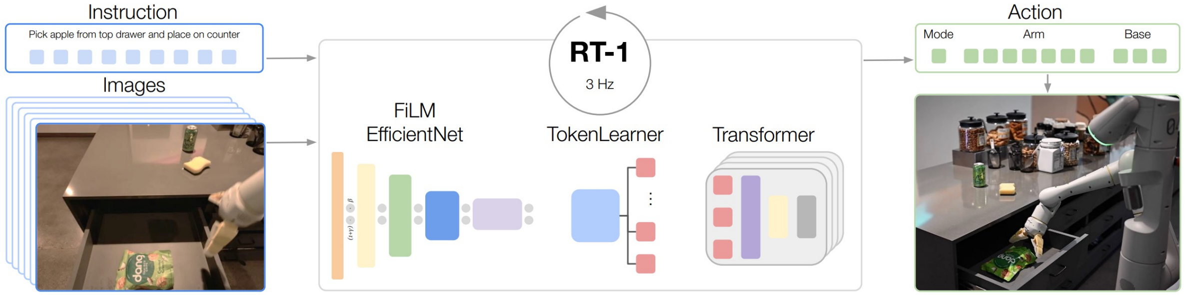
Equivalent to **13,000 years** of experience

Insight

"Massive simulation + aggressive randomization can bridge the sim-to-real gap."

RT-1: Robotics Transformer (2022)

Image + Language → Tokenizer → Transformer → Action Tokens



RT-1 Architecture

Key Facts

130,000+ demonstrations

13 robots collecting data over **17 months**

Transformer architecture

700+ distinct tasks

Performance

97% success on seen tasks

76% success on unseen tasks

Innovation

"Tokenize images and actions. Apply transformer architecture. Benefit from scale."

RT-2: Vision-Language-Action Models (2023)

RT-2 Following Novel Instruction

The Leap

Co-fine-tune vision-language model on robotics data

Actions represented as text tokens

Inherits reasoning from web-scale pretraining

Emergent Capabilities

"Pick up the extinct animal"

→ **Picks up dinosaur toy**

"Move the Coke can to Taylor Swift"

→ **Associates face with name**

Reasoning never seen in robot training data.

The Current Moment — Open-Source VLAs

2022	2023	2023	2024	2024
RT-1	**RT-2**	**RT-X/OXE**	**Octo**	**OpenVLA**
130k demos 700 tasks	VLM + Actions Emergent skills	22 robots 1M+ trajectories	Open source Diffusion policy	7B params Beats RT-2-X

Proprietary | **Open Source**

Key Message: Foundation models for robotics are becoming accessible. The field is democratizing rapidly.

Where We Are Now

Era	Core Question	Approach
Classical	How do we model the world?	Build representations, plan
Behavioral	How do we react to the world?	Design behaviors, let emerge
Probabilistic	How do we handle uncertainty?	Maintain beliefs, reduce uncertainty
Learning-Based	How do we acquire skills?	Collect demos, train policies
Foundation Models	How do we build generalists?	Pre-train at scale, fine-tune

The Question Shift

From **"Can robots learn?"** to **"How do we scale robot learning?"**

The Platforms of Embodied AI

Manipulator	**Mobile Robot**	**Humanoid**
Precision & Dexterity	Navigation & Exploration	Whole-Body Intelligence

"Three platforms, distinct challenges, shared foundations"

Manipulators: The Workhorse of Robot Learning

Why Franka Became the Standard

7 degrees of freedom (redundant)

Torque sensing at every joint

1 kHz control rate

Safe for human collaboration

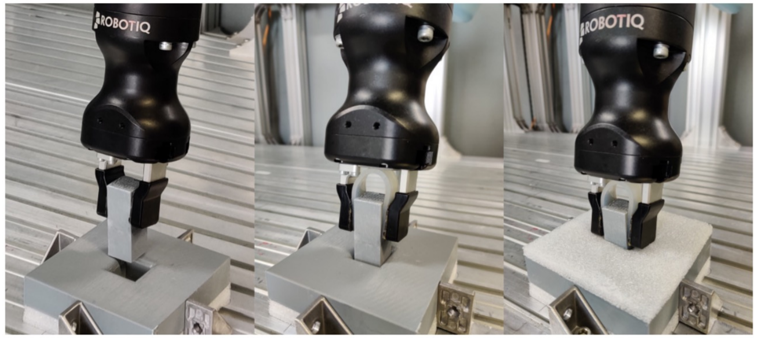
Open programming interface

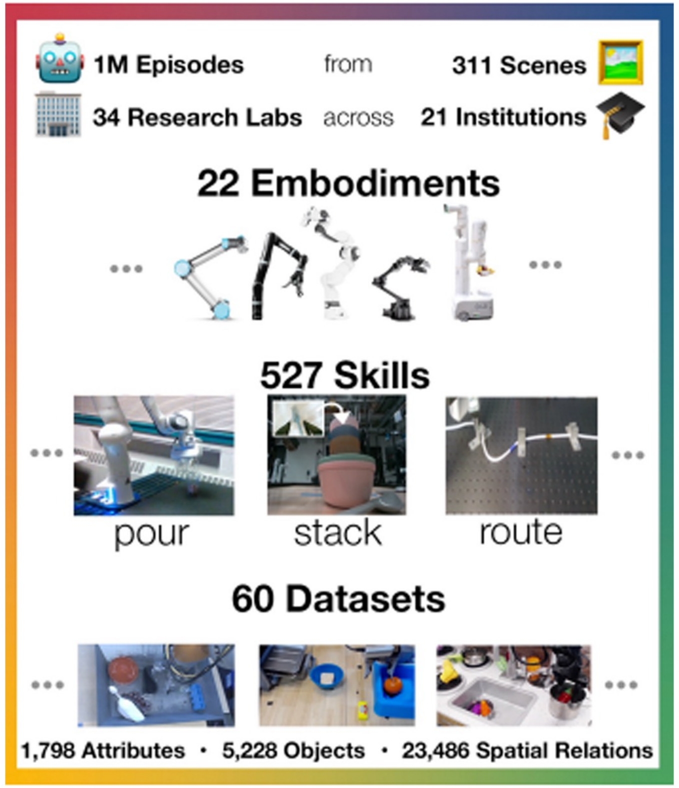
~$30,000 (research accessible)

Impact

Used in 100+ research labs worldwide. Featured in RT-1, RT-2, Diffusion Policy, and most major manipulation research.

Manipulation Challenges





1. Contact-Rich Interaction

Insertion, assembly, tool use

Challenge: Contact physics is discontinuous

2. Dexterous Grasping

Novel objects, clutter, deformables

Challenge: Infinite object variety

3. Generalization

Same skill, different contexts

Challenge: Policies fail in new settings

State of the Art in Manipulation

Key Methods

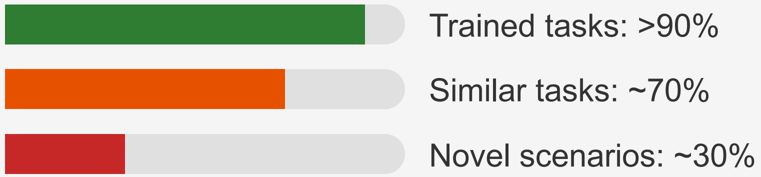
Diffusion Policy
Chi et al., 2023

ACT (Action Chunking)
Zhao et al., 2023

RT-1 / RT-2 / RT-X
Google DeepMind, 2022-2023

OpenVLA / Octo
Open-source VLAs, 2024

Current Capabilities



> *"High performance on trained tasks, but generalization remains the frontier."*

Mobile Robots: Navigation and Beyond

Wheeled

TurtleBot, Clearpath

✓ Simple, efficient

✗ Terrain limited

Legged

ANYmal, Spot, A1

✓ Terrain versatile

✗ Complex control

Aerial

DJI, quadrotors

✓ 3D movement

✗ Limited payload

Hybrid

Wheeled-legged

✓ Combines benefits

✗ Complex design

Navigation Challenges

SLAM

Where am I? What's around me?

↔

Integrated Navigation

All three challenges interconnected

↔

Long-Horizon Planning

A to B, dynamic obstacles, replanning

↕

Semantic Understanding

"Go to kitchen" • "Find red chair" • Natural language commands

Learning-Based Navigation

Approaches

End-to-End RL

Raw visual input → actions directly

Modular Pipeline

Perception → Planning → Control

Hybrid

Learning for perception, classical for planning

Key Benchmarks

Habitat Challenge (Meta AI)

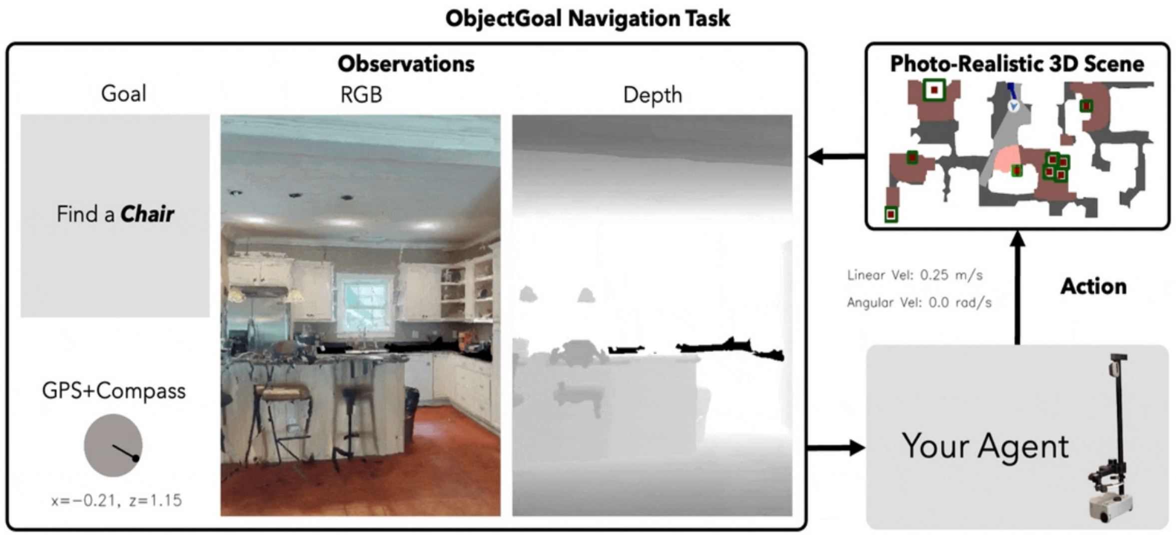
iGibson (Stanford)

Habitat Navigation Challenge 2023

Overview

In 2023, we are hosting the ObjectNav and ImageNav challenge in the Habitat simulator [1].

Task #1: ObjectNav focuses on egocentric object/scene recognition and a commonsense understanding of object semantics (where is a bed typically located in a house?).



Current State

Near-human performance in simulation. Real-world transfer improving rapidly with sim-to-real techniques.

Humanoids: The Grand Challenge

Atlas

Boston Dynamics

H1

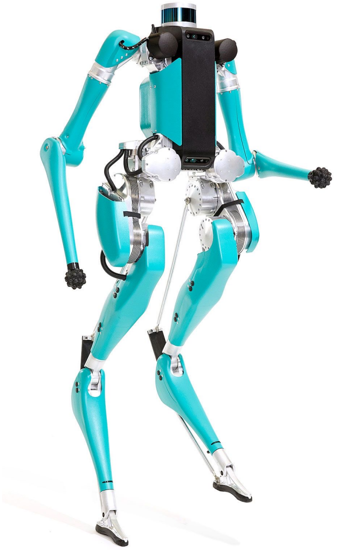
Unitree

Figure 02

Figure AI

Optimus

Tesla



Digit

Agility

Why Humanoids?

Human environments, human tools, human collaboration — the ultimate test of embodied AI.

Current Humanoid Platforms

Platform	Specs	Approach	Key Achievement
Atlas	Hydraulic→Electric	Model + Learning	Parkour, gymnastics
H1	Electric, 180cm, 47kg	Learning-based	3.3 m/s running, backflip
Figure 02	Electric, commercial	VLA-controlled	Factory deployment
Optimus	Electric, low-cost target	Learning-based	Autonomous operation
Digit	Electric, warehouse	Hybrid control	Commercial deployment

> *"2023-2025 saw more humanoid progress than the previous two decades."*

Humanoid Challenges

1. Whole-Body Coordination

30+ DOF, walking while reaching

Can't separate locomotion and manipulation

2. Balance and Recovery

Bipeds inherently unstable

Must handle pushes, slips, unexpected contacts

3. Energy Efficiency

Batteries limit runtime (~2 hours)

Humans operate all day on modest calories

4. Generalization

Humanoid form should enable broad capabilities

Current systems still task-specific

Learning-Based Humanoid Control

Recent Breakthroughs

Radosavovic et al. (2024)

Transformer policy with in-context adaptation. Single policy adapts to different terrains and perturbations without retraining.

Cheng et al. (2024)

RL from motion capture data. H1 humanoid performs diverse athletic motions learned from human demonstrations.

Helix (Figure AI, 2025)

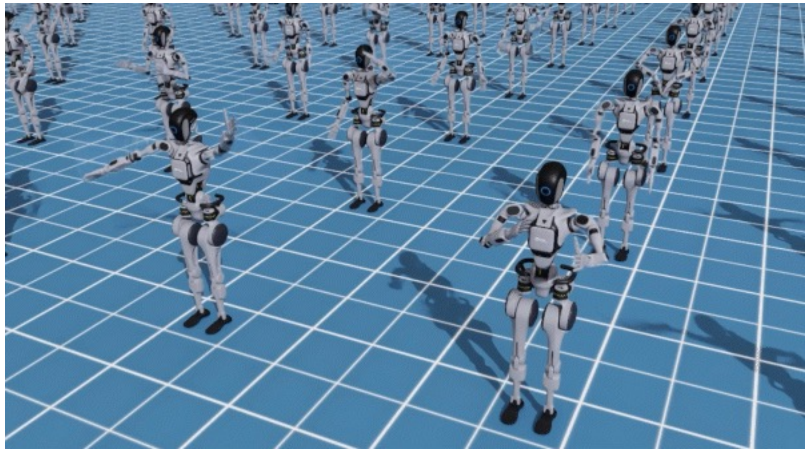
VLA for 35 DOF humanoid at 200 Hz control. Multi-robot coordination, deployed in production factories.

The Three Platforms — Summary

Aspect	Manipulators	Mobile Robots	Humanoids
Primary Challenge	Contact, dexterity	Navigation, SLAM	Whole-body, balance
Typical DOF	6-7 + hand	2-12	20-40+
Control Rate	100-1000 Hz	10-100 Hz	100-1000 Hz
Maturity	Most mature	Mature	Emerging
Data Availability	Good	Moderate	Limited

"Each platform pushes different aspects of embodied AI."

Why Simulation Has Become Central



🤖

1 Real Robot
10 eps/hr

VS

1000s Simulated
10,000 eps/hr

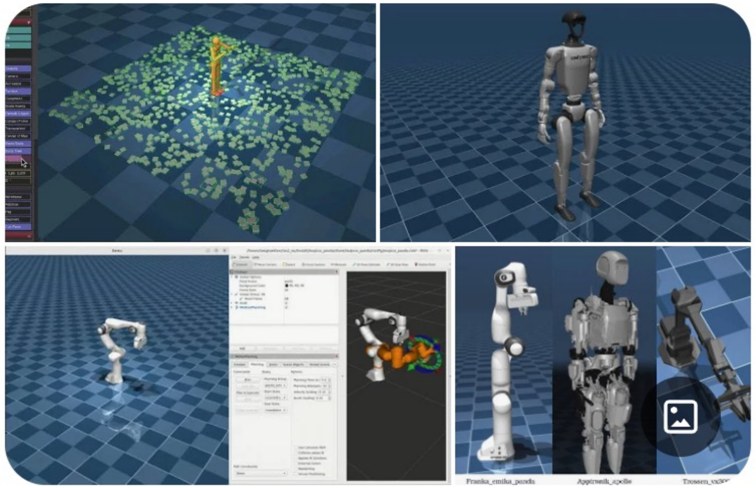
1. Scale
10 eps/hr (real) vs. 10,000 eps/hr (sim)

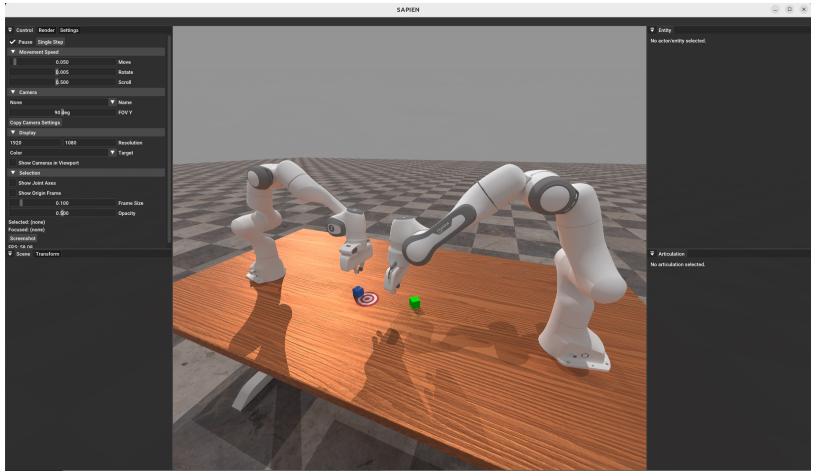
2. Safety
Simulated failures are free

3. Parallelization
Thousands of environments on one GPU

4. Reproducibility
Same sim, same seed → same results

The Simulation Stack for This Course





MuJoCo

Primary Use: Foundations

Key Features: Gold standard, open source

→ **HW1, HW2**

Isaac Lab

Primary Use: GPU RL

Key Features: 4096+ parallel envs

→ **HW3**

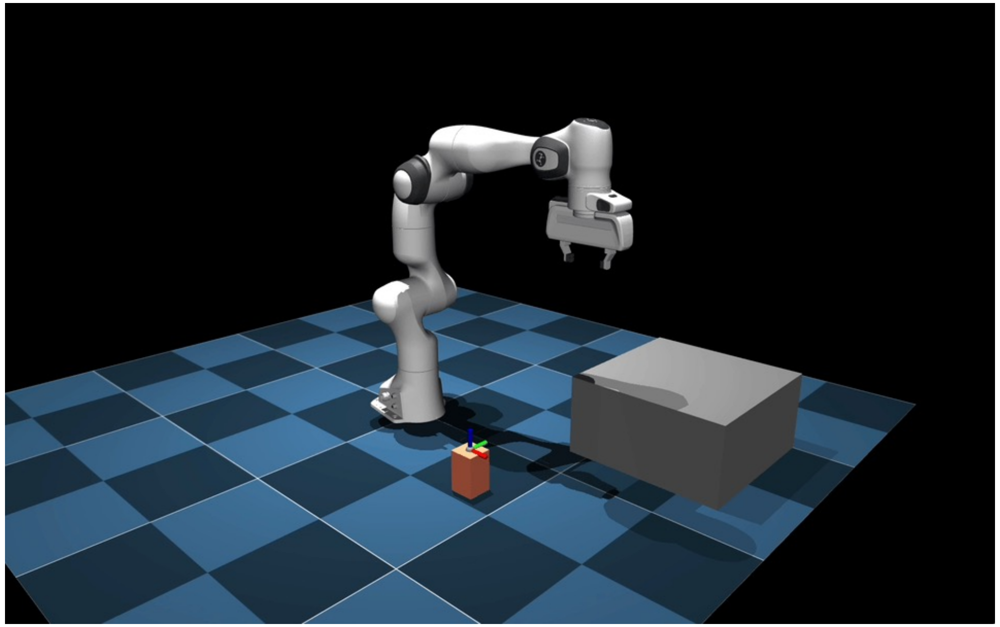
ManiSkill3

Primary Use: Benchmarks

Key Features: 30k+ FPS, comprehensive

→ **Optional**

MuJoCo: The Foundation



History

Created by Emanuel Todorov (2012)

Acquired by DeepMind (2021)

Open-sourced under Apache 2.0 (2022)

Why MuJoCo

Accurate contact dynamics

Fast: 10M+ steps/second

Excellent documentation

Dynamics Equation

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + g(q) = \tau + J^T f$$

You'll implement this in HW1.

Isaac Lab: GPU-Accelerated Learning

Capabilities

Thousands of parallel environments on single GPU

End-to-end GPU pipeline

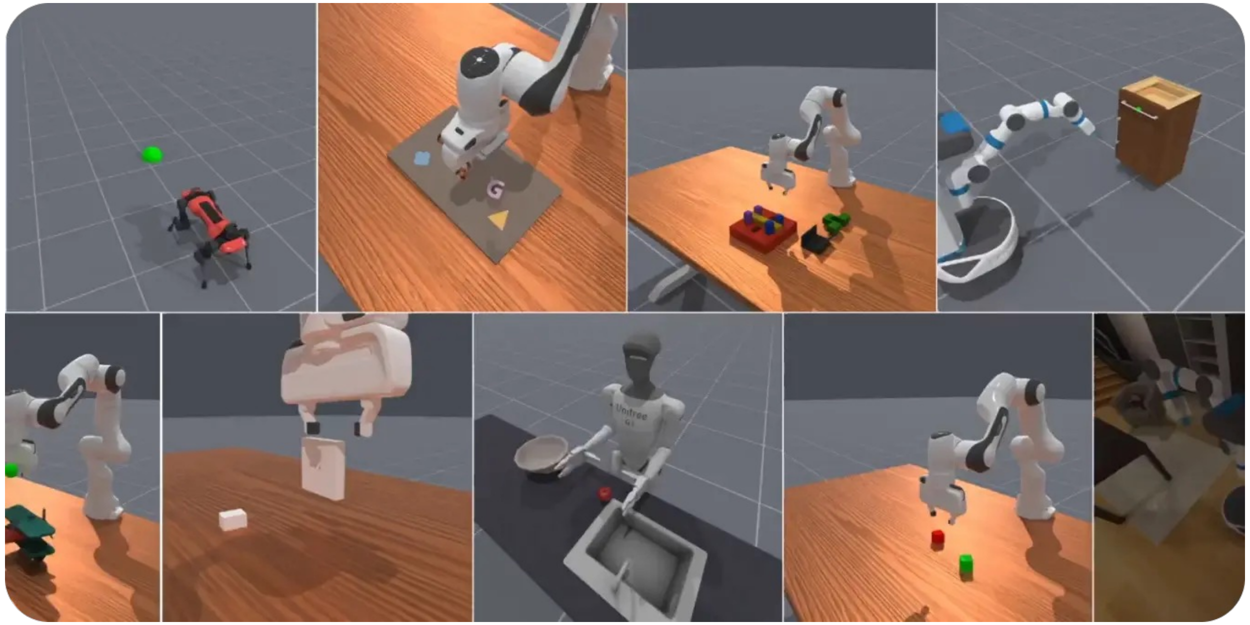
Photorealistic rendering

ROS 2 integration

Use Case

"When you need to train RL policies fast, Isaac Lab is the tool."

ManiSkill3: Benchmarks and Baselines



Key Features

SAPIEN-based simulation

Diverse manipulation tasks

Pre-implemented baselines

30,000+ RGBD FPS

Why It Matters

Standardized evaluation across methods. Easy comparison with published baselines. Quick start for research projects.

The Sim-to-Real Gap

Simulation

Perfect physics model
Clean observations
Ideal actuators

THE GAP

?

Physics approximations
Contact model differences
Sensor noise patterns
Actuator dynamics
Visual appearance

Reality

Complex physics
Noisy sensors
Imperfect motors

Bridging Strategies

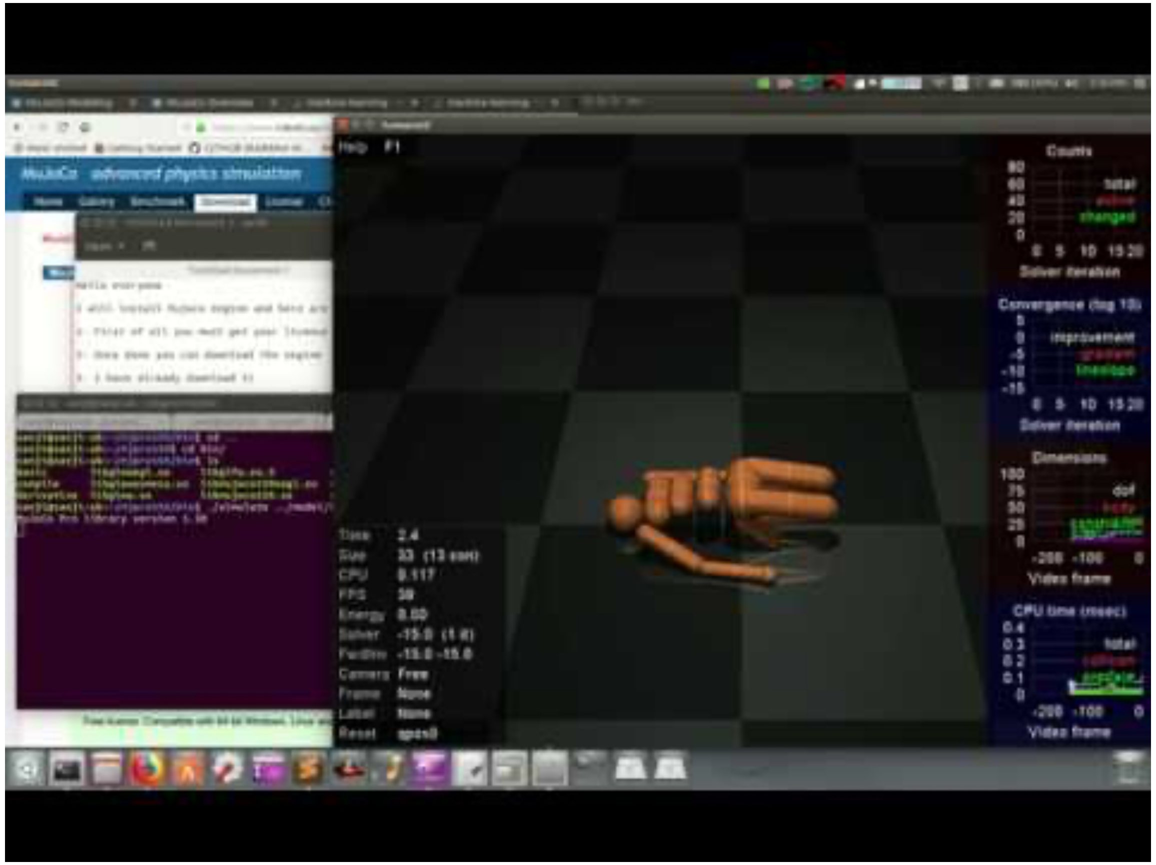
Domain Randomization • System Identification • Domain Adaptation • Real-World Fine-Tuning

Live Demo: MuJoCo Setup

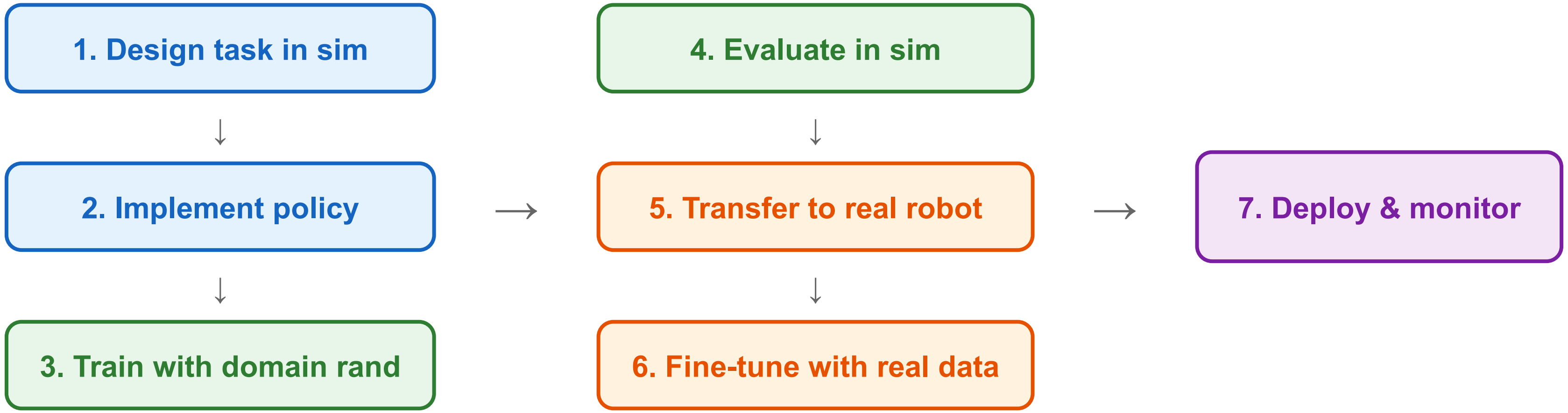
```
import
mujoco
import
mujoco.viewer

# Load model
model = mujoco.MjModel.from_xml_path("franka_panda.xml")
data = mujoco.MjData(model)

# Launch viewer
with
mujoco.viewer.launch_passive(model, data) as viewer:
while
viewer.is_running():
mujoco.mj_step(model, data)
viewer.sync()
```



Simulation-First Development



"Simulation is not a shortcut — it's the foundation."

Your Journey Starts Now

Course Structure: Four Modules

Module 1: Foundations

L2: Kinematics, Dynamics, Sim

L3: Control

L4: Planning

Module 2: Perception

L5: Visual Perception

L6: Tactile Sensing

Module 3: Learning

L7: RL

L8: Imitation Learning

L9: Generative Models

Module 4: Foundation

L10: VLAs

L11: World Models

L12: Humanoids & Future

Foundations → Perception → Learning → Foundation Models

Grading Structure

25% Presentation	25% Paper
30% Homework	20% Participation

Presentation (25%)

45 min presentation + 15 min Q&A

Paper (25%)

15-20 pages, 30-50 papers reviewed

Homework (30%)

Three assignments (10% each)

Participation (20%)

Questions + Discussion

Four Takeaways from Today

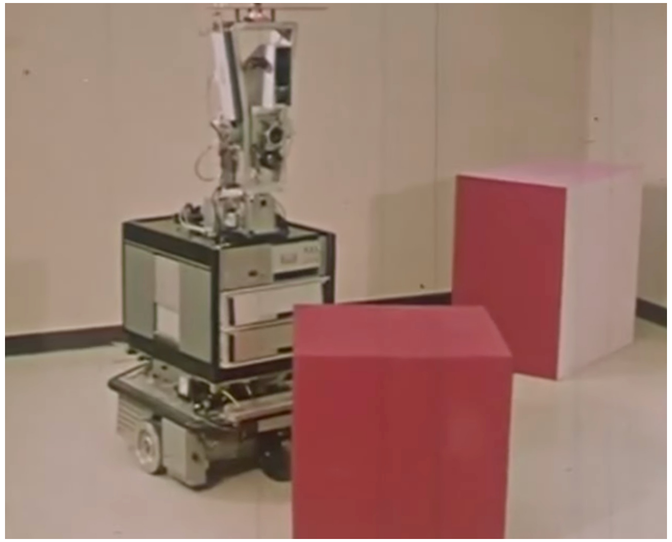
1. Embodied AI = Intelligence through physical perception-action loops

2. Historical Evolution: SHAKEY → Behaviors → Probabilistic → Learning → Foundation Models

3. Three Platforms: Manipulators, Mobile Robots, Humanoids — distinct challenges, shared foundations

4. Simulation-First: MuJoCo, Isaac Lab are your laboratory for this course

Final Thought



SHAKEY

1969

→

Modern Humanoid

2025

"The robots in those opening videos aren't magic — they're the product of ideas you'll understand by the end of this course."

"Let's build that understanding together."

"Want to understand the essence of intelligence?"

Study Embodied AI.

Q & A