CIS 6280 · FALL 2026

WORLD MODELS

LECTURE 1

Jiatao Gu

University of Pennsylvania · August 25, 2026

Course staff

Jiatao Gu
Instructor

Mutian Tong
Teaching Assistant

Yong-Hyun Park
Teaching Assistant

Enxin Song
Teaching Assistant

Xinyue Ai
Teaching Assistant

Office hours and the preferred contact channel will be posted on the course website.

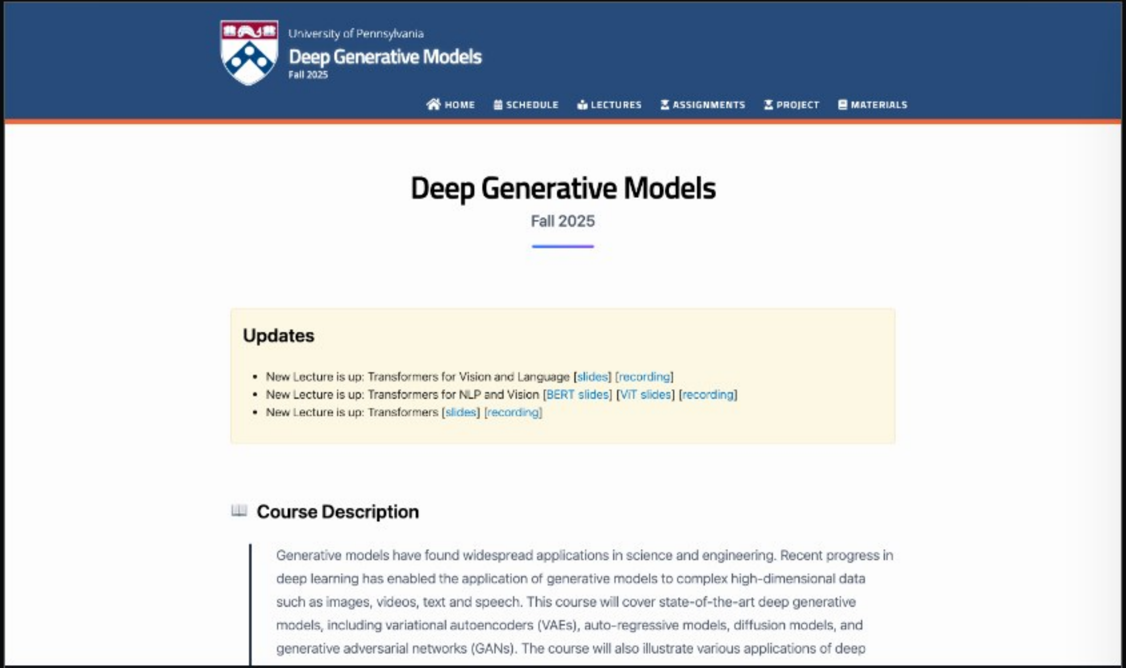
NOVEMBER 2025

I had an idea for a new course on generative models.

> I'm writing to ask about the process for proposing a new course for Fall 2026 on generative models. It would be a graduate-level course that builds on my current CIS 7000 special topics class, but with a stronger emphasis on both fundamentals and technical details.

FROM THE COURSE PROPOSAL EMAIL

Then I found two excellent courses.


University of Pennsylvania
Deep Generative Models
Fall 2025
HOME SCHEDULE LECTURES ASSIGNMENTS PROJECT MATERIALS

Deep Generative Models
Fall 2025

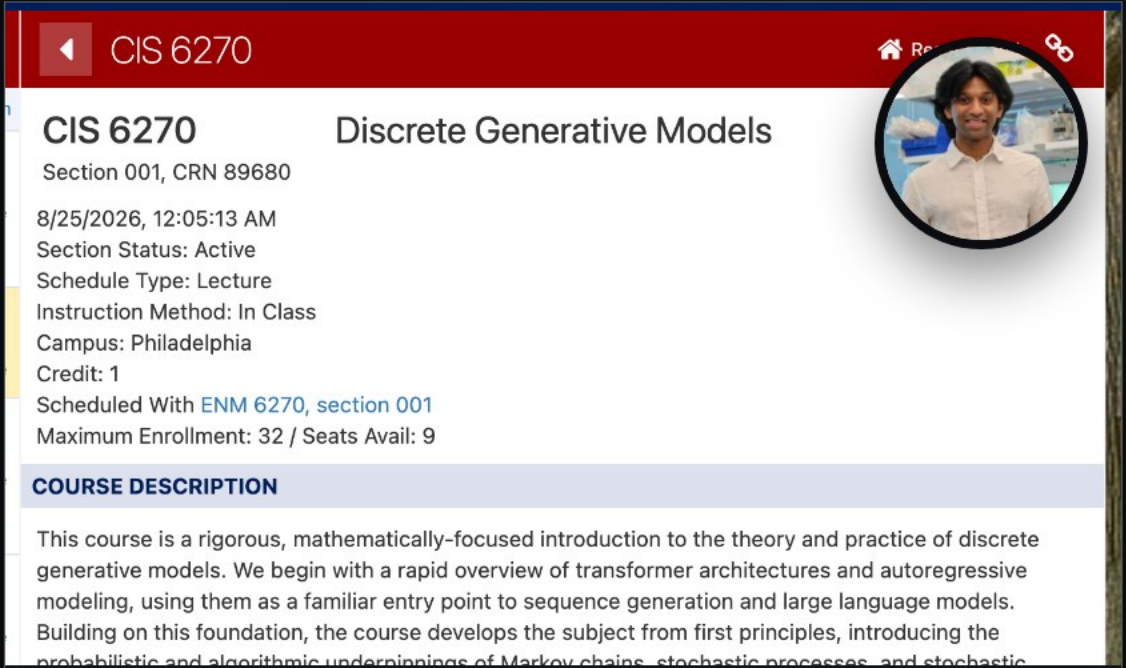
Updates

- New Lecture is up: Transformers for Vision and Language [slides] [recording]
- New Lecture is up: Transformers for NLP and Vision [BERT slides] [ViT slides] [recording]
- New Lecture is up: Transformers [slides] [recording]

Course Description

Generative models have found widespread applications in science and engineering. Recent progress in deep learning has enabled the application of generative models to complex high-dimensional data such as images, videos, text and speech. This course will cover state-of-the-art deep generative models, including variational autoencoders (VAEs), auto-regressive models, diffusion models, and generative adversarial networks (GANs). The course will also illustrate various applications of deep

René's Deep Generative Models Fall 2025


CIS 6270

CIS 6270 Discrete Generative Models
Section 001, CRN 89680

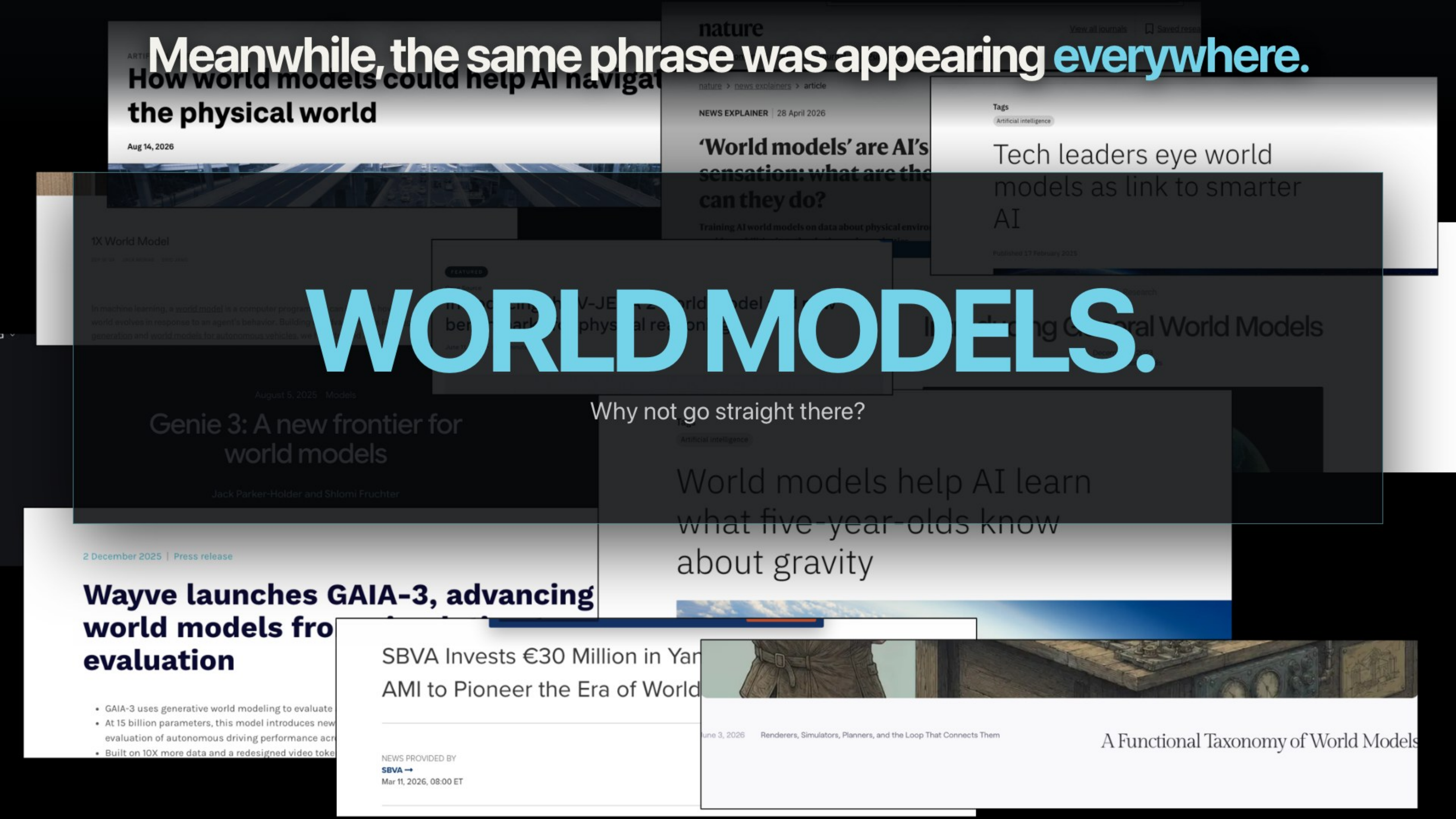
8/25/2026, 12:05:13 AM
Section Status: Active
Schedule Type: Lecture
Instruction Method: In Class
Campus: Philadelphia
Credit: 1
Scheduled With ENM 6270, section 001
Maximum Enrollment: 32 / Seats Avail: 9

COURSE DESCRIPTION

This course is a rigorous, mathematically-focused introduction to the theory and practice of discrete generative models. We begin with a rapid overview of transformer architectures and autoregressive modeling, using them as a familiar entry point to sequence generation and large language models. Building on this foundation, the course develops the subject from first principles, introducing the

Pranam's Discrete Generative Models Fall 2026

Meanwhile, the same phrase was appearing everywhere.



WORLD MODELS.

Why not go straight there?

Then, near the start of this semester, I started looking for another course on world models. for reference...

Are you able to find me a course in US university (not one lecture, not a workshop or tutorial) online exactly about "world models"? Exclude UPenn's CIS 6280...

Worked for 17s

Yes. I searched fairly aggressively across U.S. university course pages, syllabi, and special-topics offerings, with your criteria interpreted strictly:

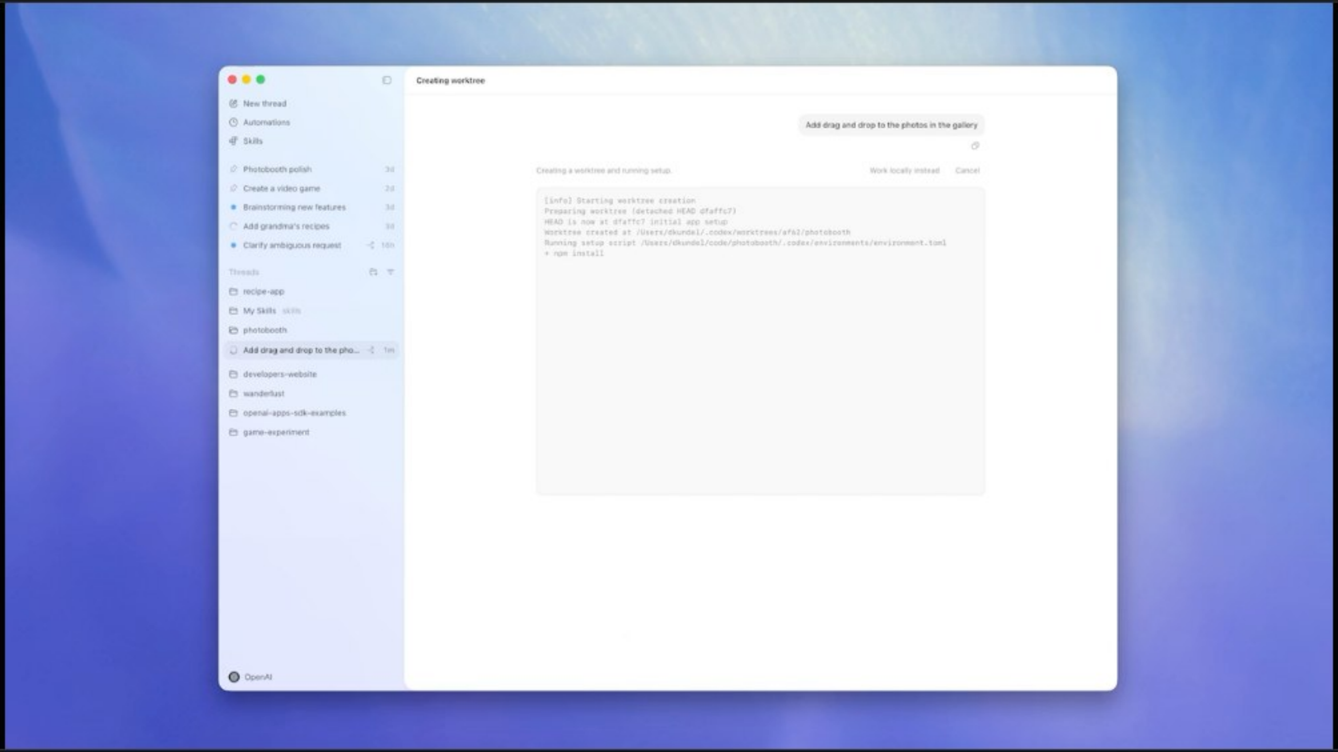
- a **real semester/quarter university course**
- at a **U.S. university**
- substantially/dedicatedly about **world models**
- not just one lecture/module inside a broader course
- not a workshop/tutorial/seminar event
- excluding **UPenn CIS 6280: World Models**

At the moment, I **cannot find another U.S. university course that clean**
UPenn CIS 6280 appears unusually specific: its official catalog descript
models, predictive environment dynamics, representations, planning/co
robotics. Courseleaf

WHAT IS A WORLD MODEL?

Digital agents already act in software.

> Watch for four things: what is visible, what can be changed, what gives feedback, and what can be undone?



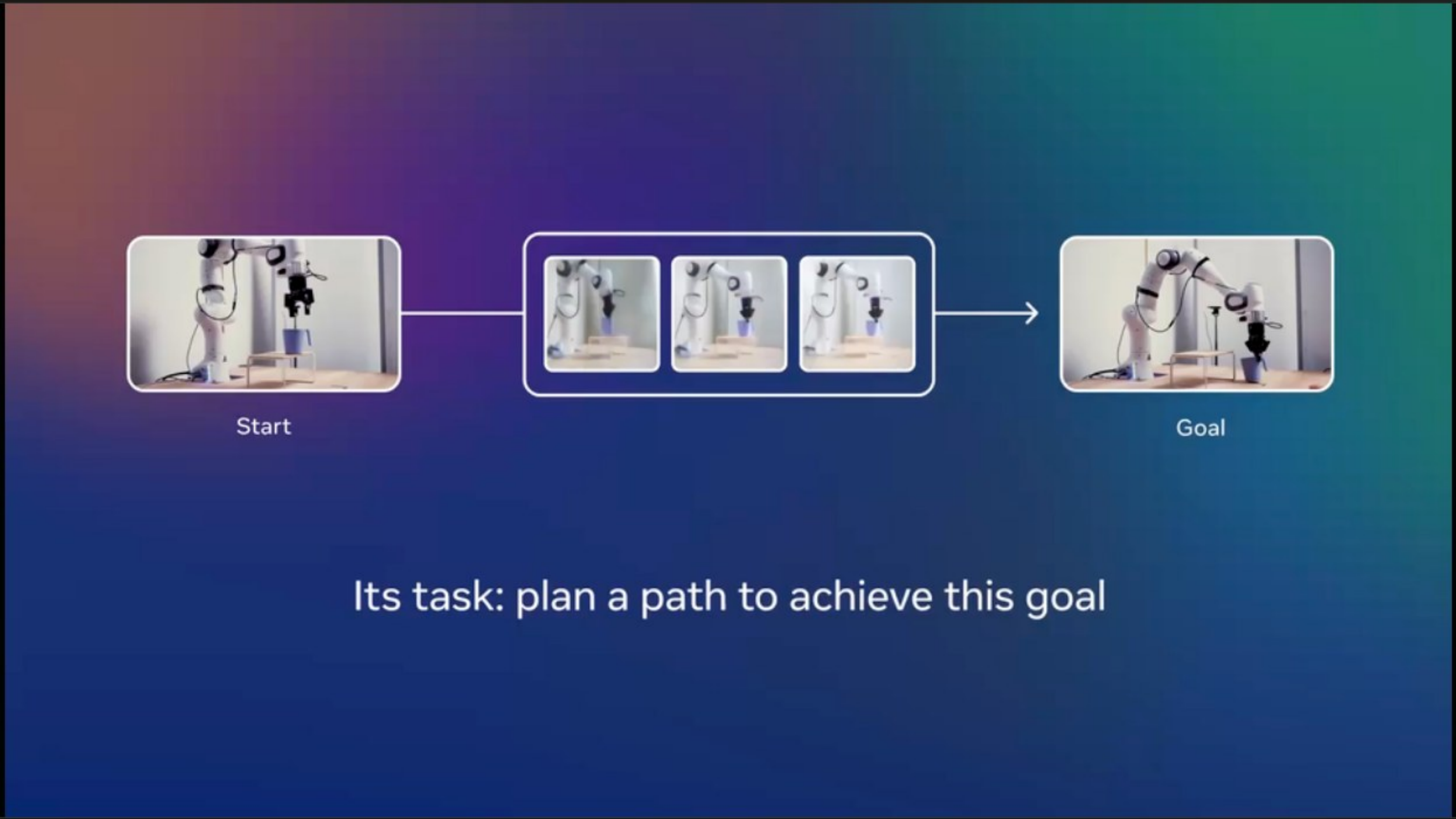
OpenAI Codex — an agent works inside a software project

Introducing the Codex app · OpenAI · 2026 · 00:40–00:58

legible state | explicit tools | fast tests | many reversible edits

Now shift to the physical world.

What must the robot anticipate before it moves?



A robot acts from a current scene toward a goal

V-JEPA 2 · Meta AI · 2025 · official demo

partial views

contact

delay

real consequences

What might prediction force a learner to discover?

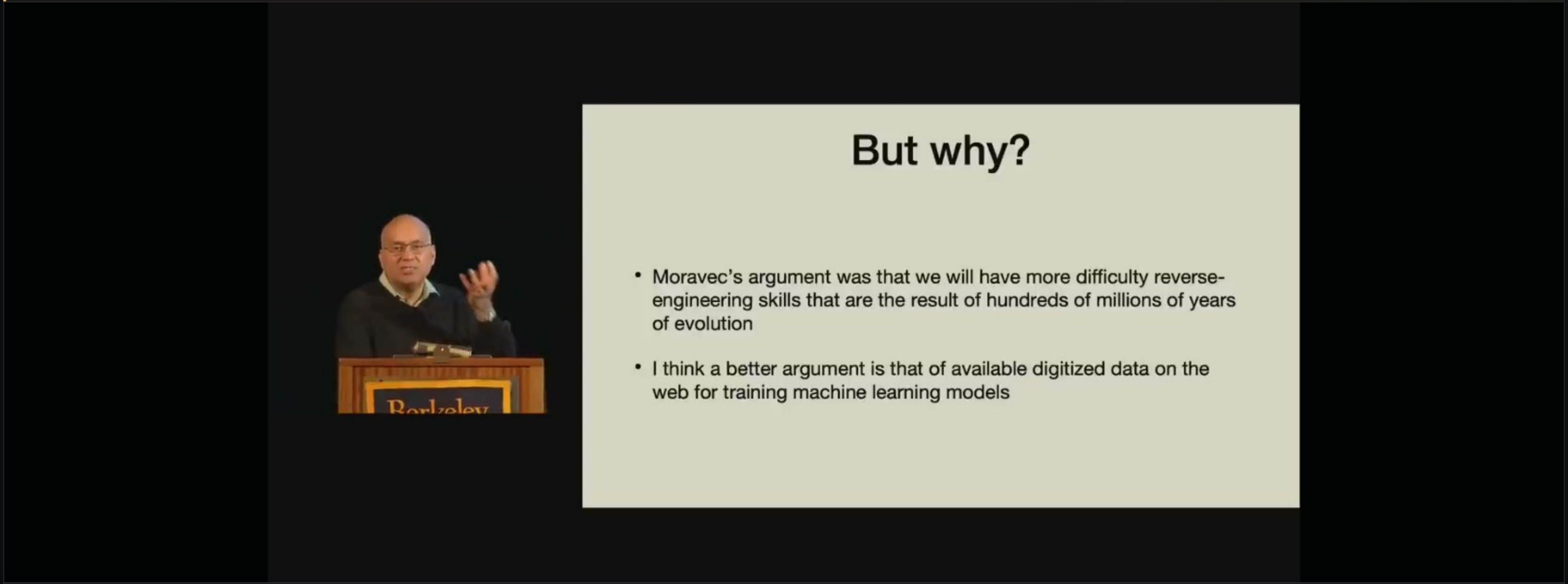
HYPOTHESIS 1 / 5 · What must be inferred beyond the next token?

Ilya Sutskever — prediction as compression of an underlying process

Fireside Chat with Ilya Sutskever and Jensen Huang · NVIDIA GTC · 2023 · 21:10–22:25

What is missing from data on the web?

HYPOTHESIS 2 / 5 · Which early experience is not digitized?



Jitendra Malik — web-scale language data misses embodied childhood experience

The Sensorimotor Road to Artificial Intelligence · UC Berkeley · 2023 · 15:18–16:22

What can interaction teach that text alone cannot?

HYPOTHESIS 3 / 5 · Where does evidence about action consequences come from?

Richard Sutton — learning from the consequences of action

Richard Sutton: Father of RL Thinks LLMs Are a Dead End · Dwarkesh Podcast · 2025 · 02:09–03:19

For physical interaction, what structure must be preserved?

HYPOTHESIS 4 / 5 · Which relations do you use before the glass falls?

Fei-Fei Li — geometry, 3D relations, prediction, and the urge to act

With Spatial Intelligence, AI Will Understand the Real World · TED · 2024 · 05:54–06:31

Before the bottle lands, what do you already predict?

HYPOTHESIS 5 / 5 · Choose one consequence. What did you not need to predict?

Yann LeCun — the bottle-drop test

Fireside Chat with Yann LeCun · RAISE Summit · 2026 · 03:42–04:22

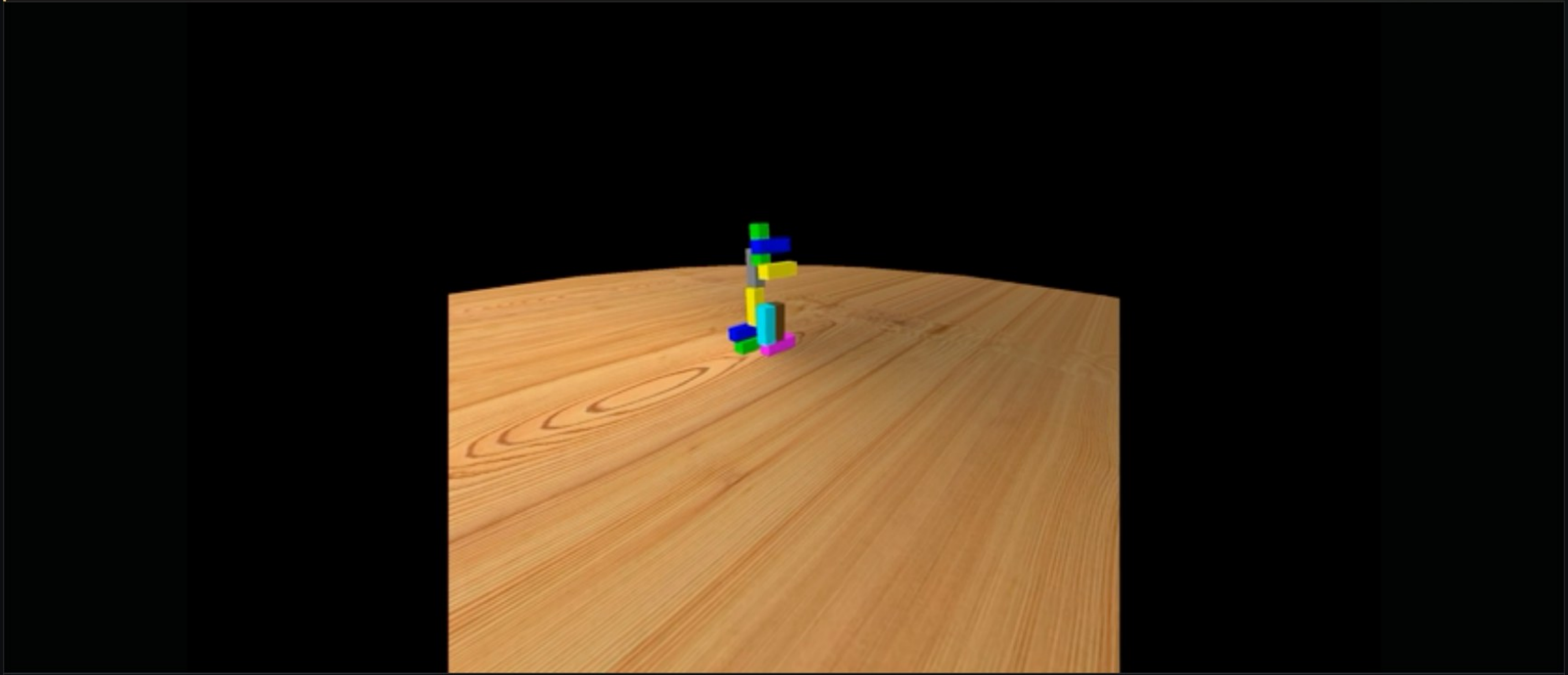
Five claims. One question.

Ilya	Prediction may reveal the process behind data.
Jitendra	Web data omits much embodied experience.
Richard	Interaction reveals what actions actually do.
Fei-Fei	Physical action depends on spatial structure.
Yann	Useful prediction need not reproduce every detail.

What lets an agent act beyond familiar situations?

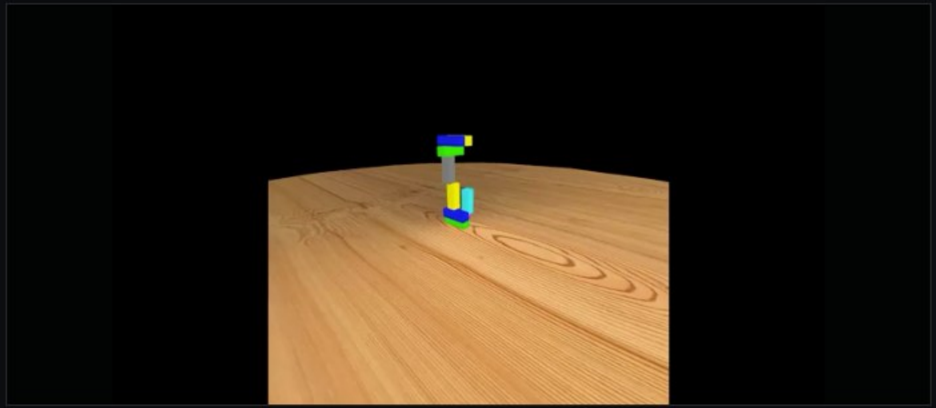
Will this tower fall?

Pause first. Predict one outcome and name the visual relation you used.

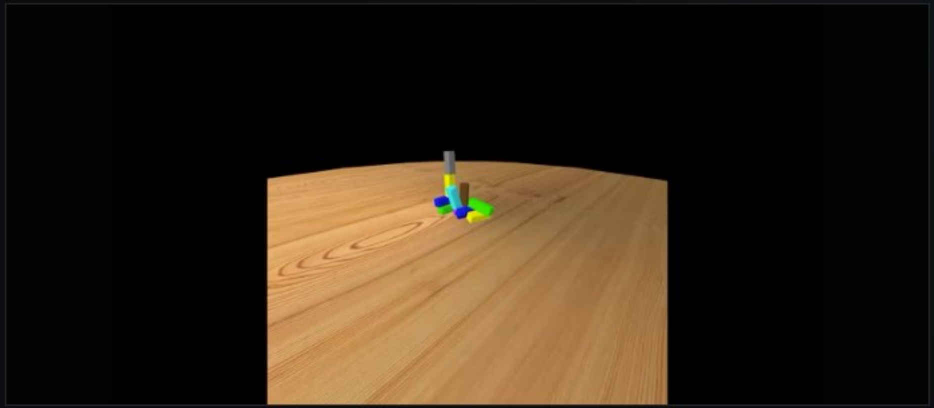


Block-stack stability experiment — predict first, then observe

A stochastic world model on gravity for stability inference · eLife · 2024



the world available now



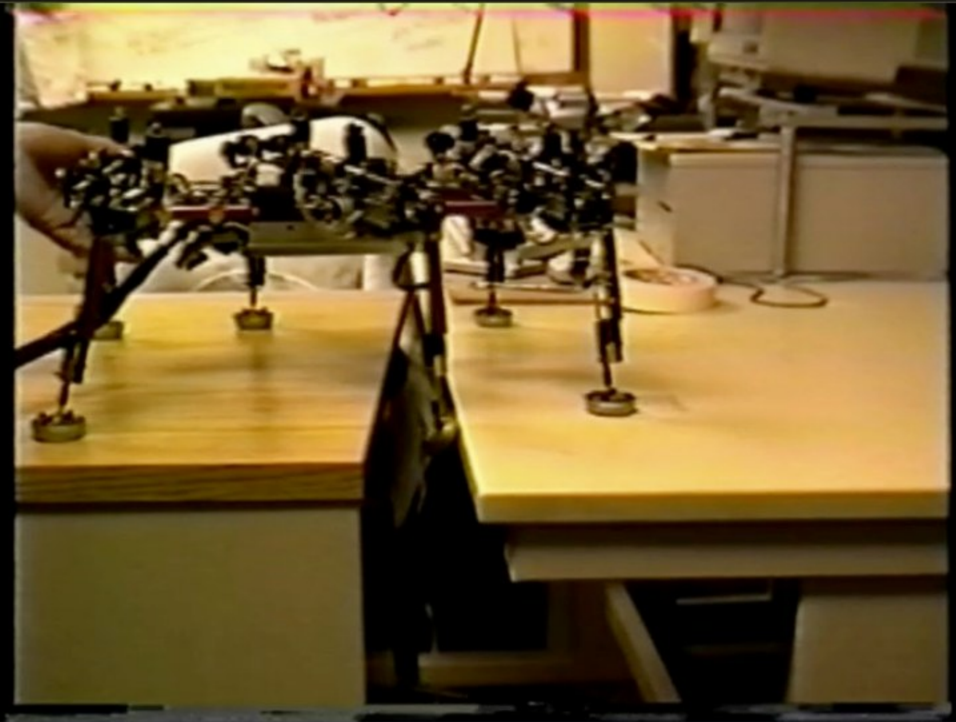
how it may change

World modeling is the ability to represent a world and predict how it may change—including the consequences of interaction.

This robot succeeds by reacting—not by first predicting what happens next.

Watch for behavior produced by fast feedback rather than a complete world replica.

MIT ARCHIVE



Rodney Brooks — robust behavior without a centralized world replica

Rod Brooks and Robotics · MIT CSAIL archive · excerpt 05:00–06:05

When the task stays the same, fast feedback can be enough.

observe → react → observe again

same objective | rapid feedback | cheap correction

The task: retrieve the cup without knocking over the tower.

GOAL

retrieve the cup

the robot has retrieved cups before

CONSTRAINT

keep the tower standing

this exact arrangement is new

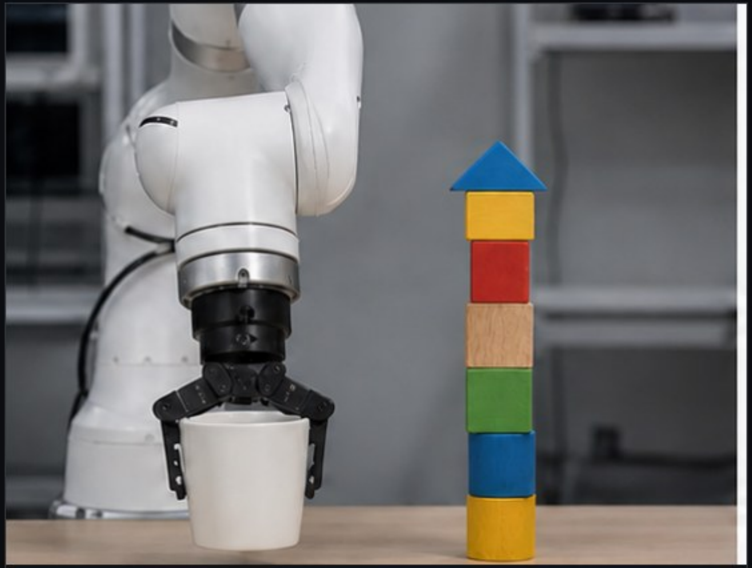
The objects are familiar. This combination of arrangement, goal, and constraint is new.

By the time the robot gets feedback, the tower is already falling.

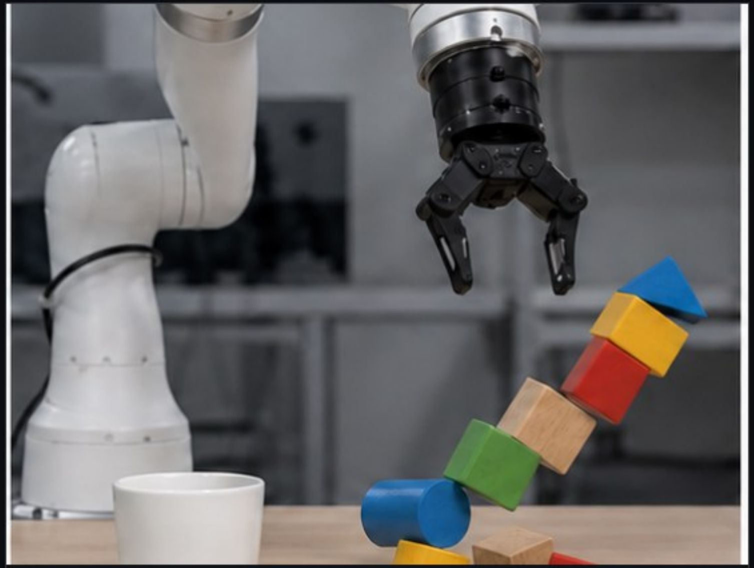
reach → contact → collapse → feedback

Feedback can improve the next attempt; it cannot undo this one.

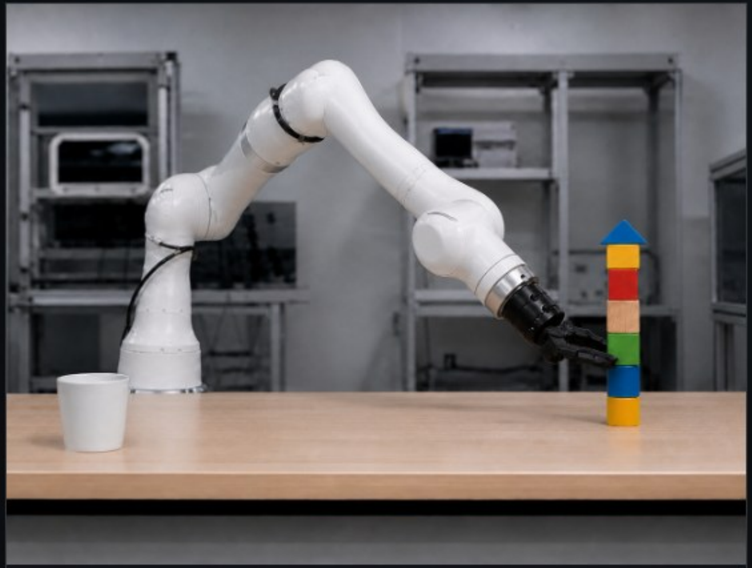
The future depends on what the robot does.



reach outside
cup moves · tower stays



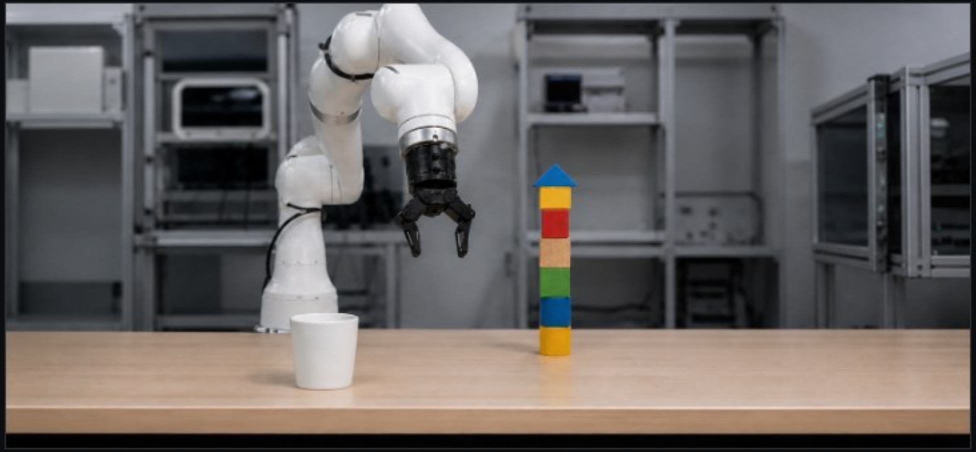
reach across
contact · tower falls



move tower
layout changes · path opens

The same scene can unfold differently under different actions.

Could the robot learn the right response directly?



current scene + task

LEARNED RESPONSE →

choose the safe reach

In principle, yes. With enough representative experience, a learned response may generalize to a situation it has not seen exactly before.

But training cannot densely cover every combination of scene, goal, and constraint.

A familiar response can solve a familiar task. A world model makes familiar consequences reusable.

FAMILIAR TASK

the same situation repeats

recognize the task → reuse the action

Reuse the action that worked before.

NEW GOAL OR CONSTRAINT

the useful action may change

reuse known consequences → choose for the new goal

The model predicts; the agent chooses again.

A world model supports reasoning about consequences.

The cup-and-tower task contains three recurring questions.

Representation	WHAT ABOUT THE SCENE MUST THE ROBOT KEEP TRACK OF?
Prediction	WHAT HAPPENS AFTER EACH POSSIBLE ACTION?
Interaction	HOW DOES ACTING REVEAL NEW EVIDENCE—AND CHANGE THE NEXT STEP?

A world model links all three; we separate them only to inspect each one.

THREAD 1 · RECURS THROUGHOUT

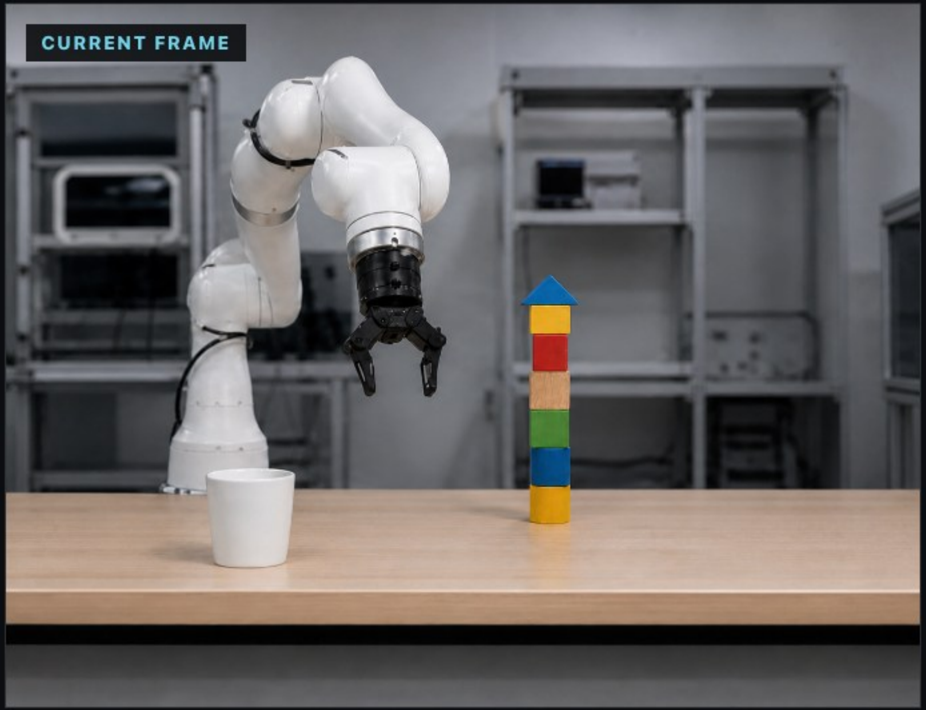
REPRESENTATION

What can experience teach the model to carry forward?

The robot receives this camera view—not "the whole world."

This image is one representation of the scene, shaped by its sensor and viewpoint.

One camera frame cannot reveal motion, contact history, or what lies outside the view.



\+

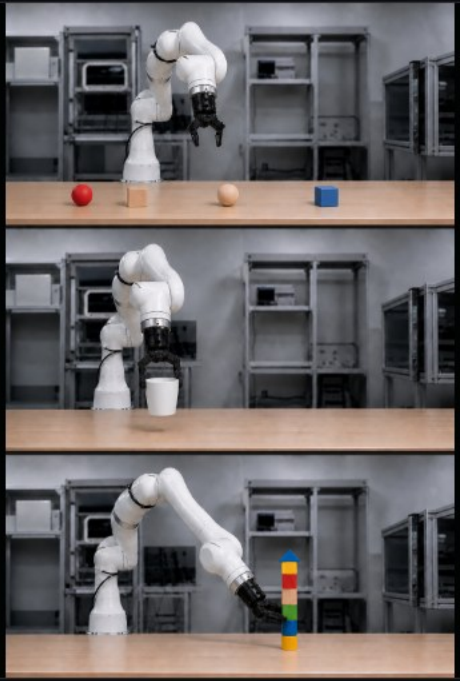
- where the arm was one moment ago
- whether contact just occurred
- what lies outside the image

→

working representation
supported by history

A world model can combine the current frame with history to represent what cannot be seen directly.

We cannot hand-design every variable.
The representation must be learned.



observations across experience

LEARNED REPRESENTATION

information that can be reused in a new scene

learned from many observations rather than specified by hand

Representation learning turns experience into information the model can reuse.

Prediction can train a representation without manual labels.

ordinary experience unfolding over time

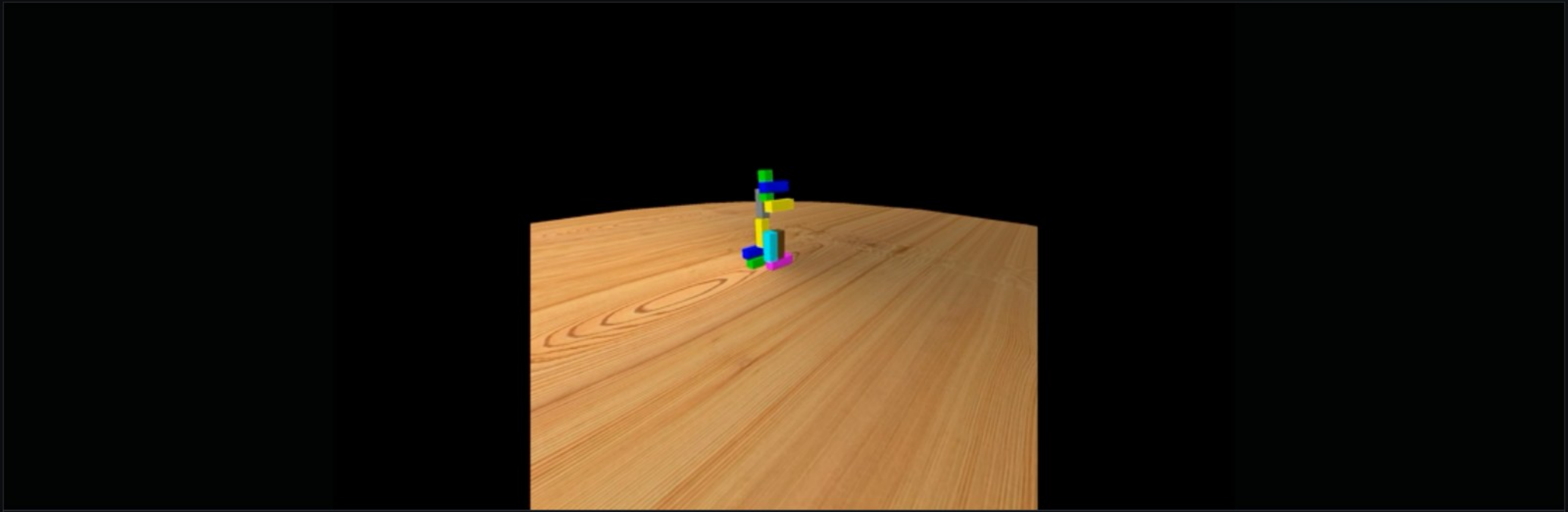
TO PREDICT THIS CROWD

track who is where and how each person is moving

the next frame reveals what actually happened

The future supplies the answer: what actually happened.

Whether it predicts pixels or features, prediction teaches the model what matters.



Different support → different outcome

Block-tower stimulus · synthetic illustration

For this tower, support is what separates “fall” from “stay.”

What part of the future should the model predict?

NEXT OBSERVATION

Predict the future camera frame.

Compare it with what the camera actually saw.

FUTURE REPRESENTATION

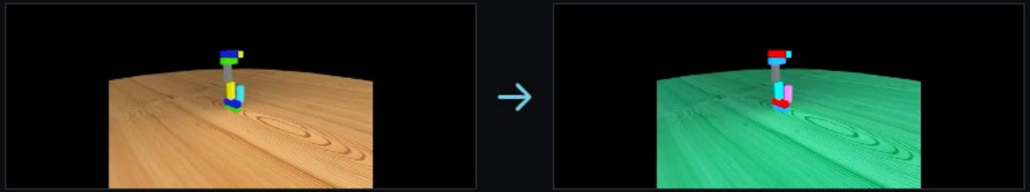
Predict features computed from that future frame.

Compare them in the learned feature space.

It can predict the next image—or a learned representation of that image.

Color can change without changing the future.
Support predicts whether the tower falls.

COLOR CHANGES



geometry unchanged

color need not change the stability prediction

SUPPORT PREDICTS FAILURE

this support pattern → collapse

the representation must preserve it

A useful representation keeps the differences that change predicted consequences.

THREAD 2 · RECURS THROUGHOUT

PREDICTION

Prediction asks about futures. Generative modeling learns how to model them.

Prediction can ask about change —with or without an action.

THE WORLD CHANGES

What may happen next?

weather Where will the storm move?

video What will the person do next?

object Will the motion settle or continue?

The system watches; no action from our agent is required.

AN AGENT INTERVENES

What may happen next if it does this?

robot What if it pushes?

vehicle What if it accelerates?

experiment What if we apply a force?

The action becomes part of the information used to predict.

Both are prediction. Only the second depends on an action.

Knowing the action still does not determine the future.

same visible history

same light push

The action changes the possibilities; hidden factors can still leave several outcomes plausible.

When several futures are possible, prediction becomes a generative modeling problem.

SAME QUERY

same scene + same push

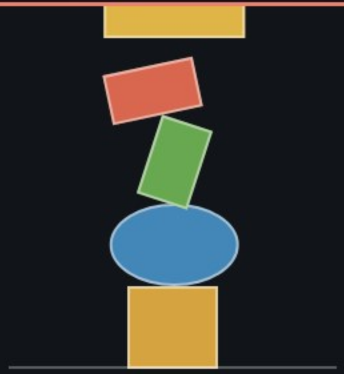
Run the same model more than once.

→

RUN 1
stays upright

RUN 2
falls left

RUN 3
falls right

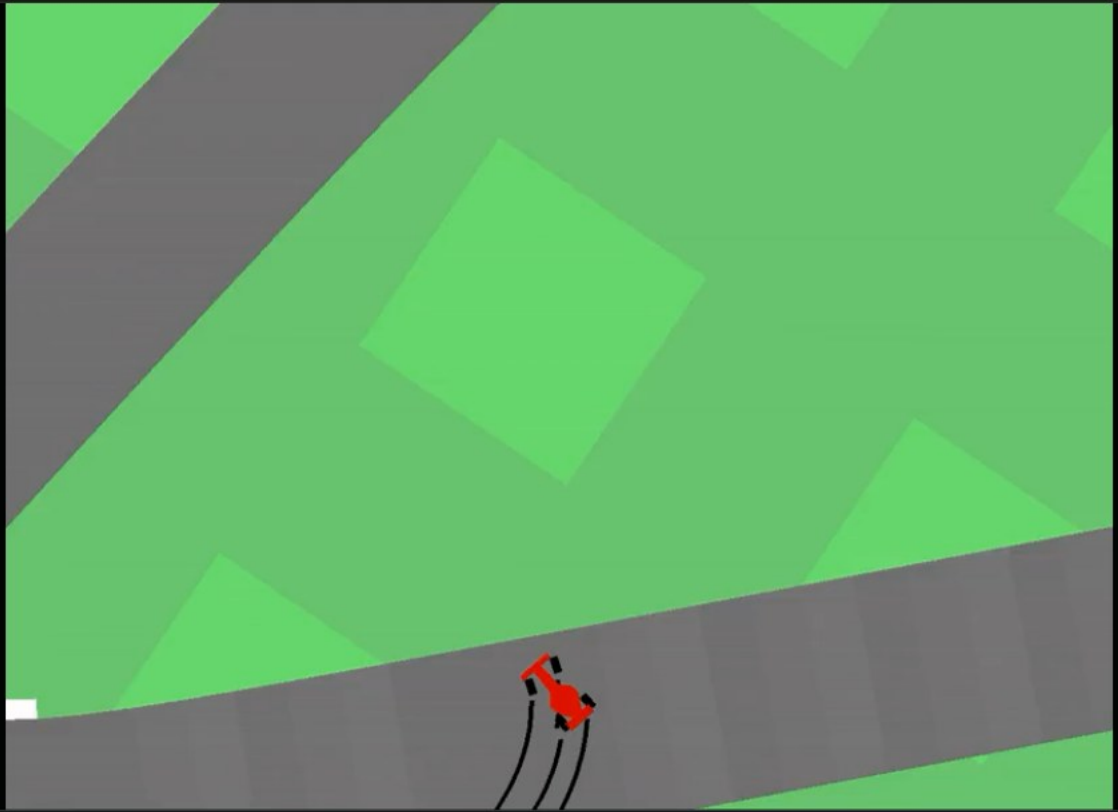


FAILURE
blocks change shape ✕

One rollout = one sample. Repeat the same query to reveal the plausible futures the model has learned—and the impossible futures it invents.

During a long rollout, a mistake becomes the next input.

Find the first error. Then watch it become part of the generated history.



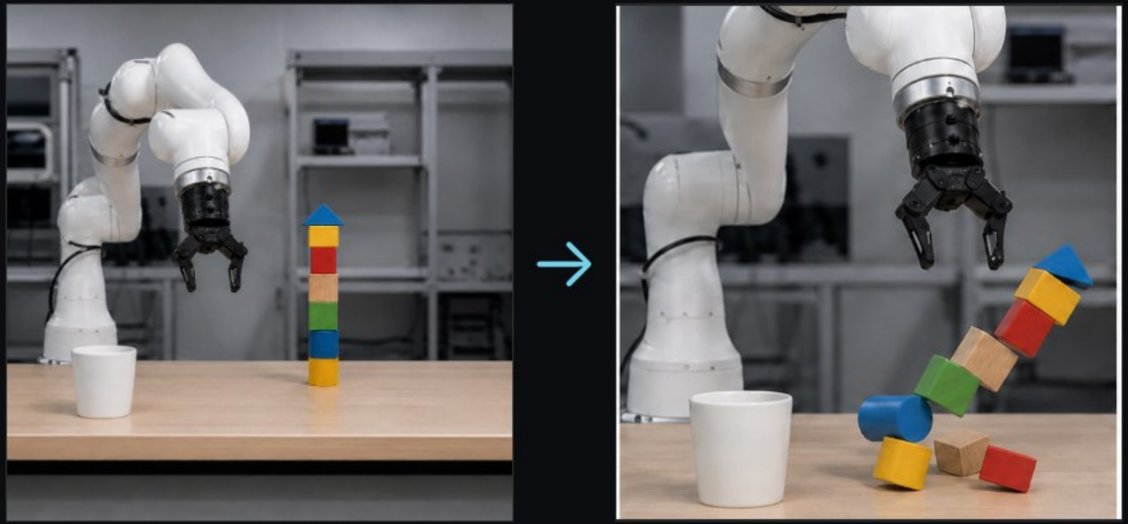
World Models — generated rollout leaves the road

Ha & Schmidhuber · 2018

One-step accuracy is not enough when future predictions depend on earlier predicted frames.

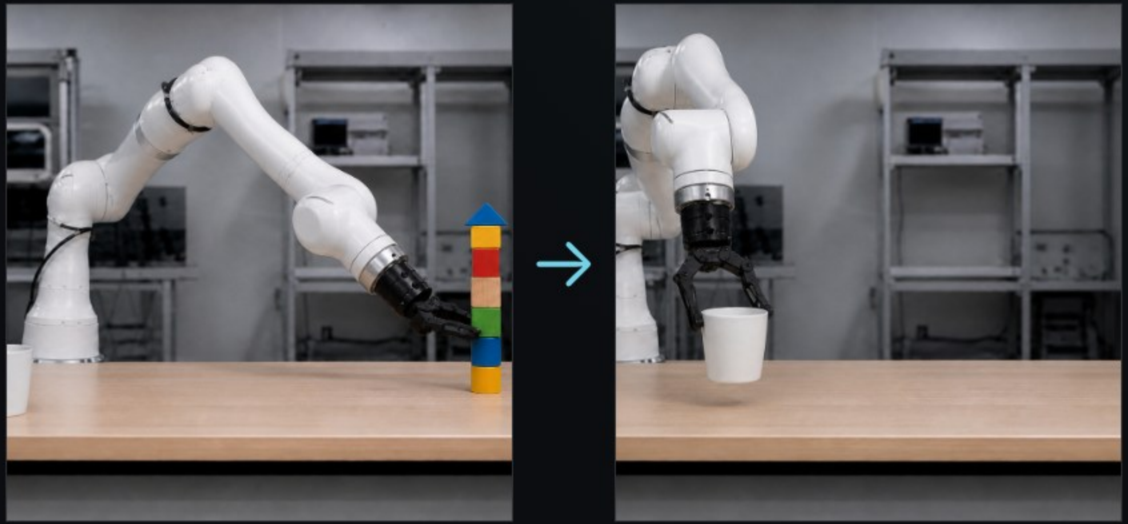
Prediction must extend far enough for the goal and constraint to matter.

grab now



early progress · task failure

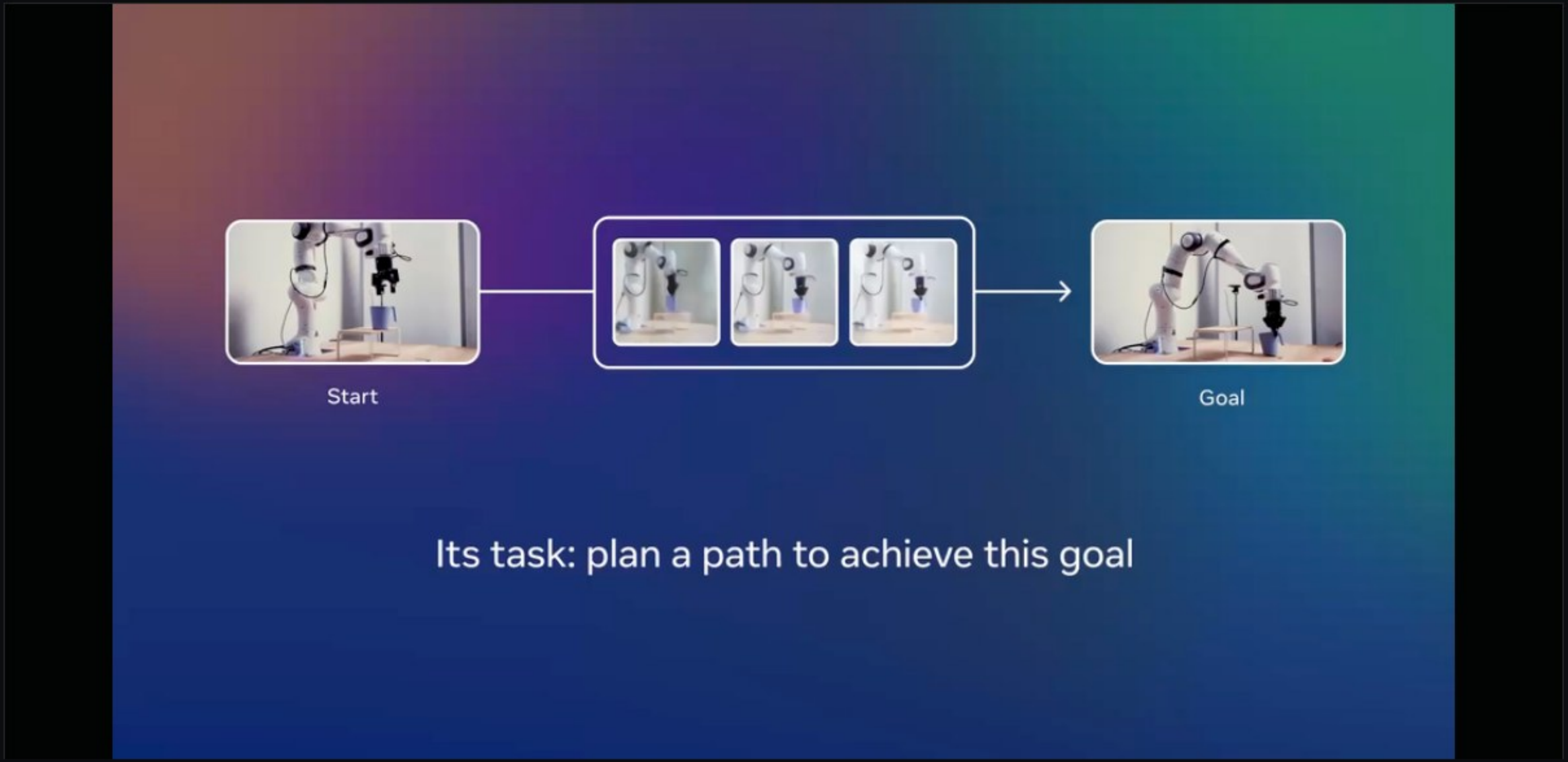
move obstacle first



slower first step · goal and constraint satisfied

The relevant horizon is long enough to reveal whether the task actually succeeds.

Prediction proposes a result.
The world returns evidence.



BEFORE PLAY

What should change?
What should stay the same?

AFTER PLAY

What actually changed?
Where did expectation fail?

Interaction compares an expected consequence with an observed one.

Once the agent can act, the current scene is not enough.

SAME CURRENT SCENE

WATCHING

What may happen next?

The world may change without an action from our agent.

INTERVENING

What may happen next if the robot moves the tower?

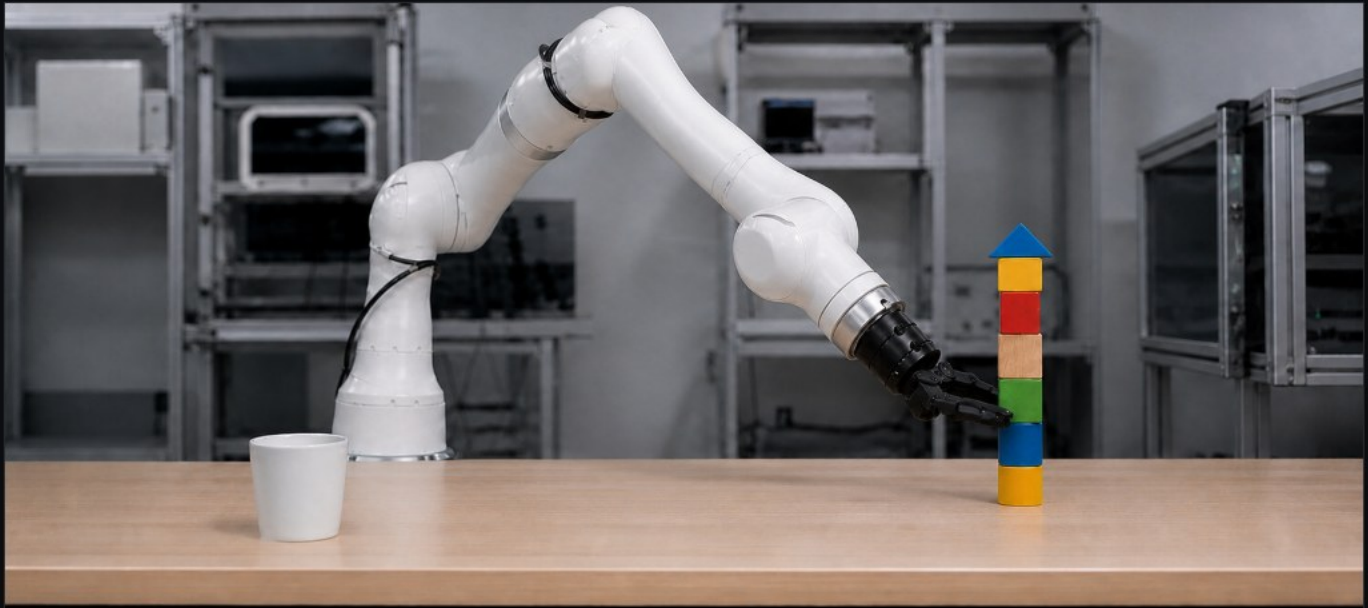
The action changes which consequences are relevant.

WORLD MODEL
predicts consequences

AGENT + GOAL
chooses an action

The action changes the prediction question; it does not guarantee one outcome.

A predicted consequence is still only a proposal.



PREDICTED CONSEQUENCE

Move the tower → the path should open.

NOT YET OBSERVED

→ EXECUTE

NEW OBSERVATION

?

confirm · contradict · surprise

Prediction offers possibilities. Interaction supplies evidence.

THREAD 3 · RECURS THROUGHOUT

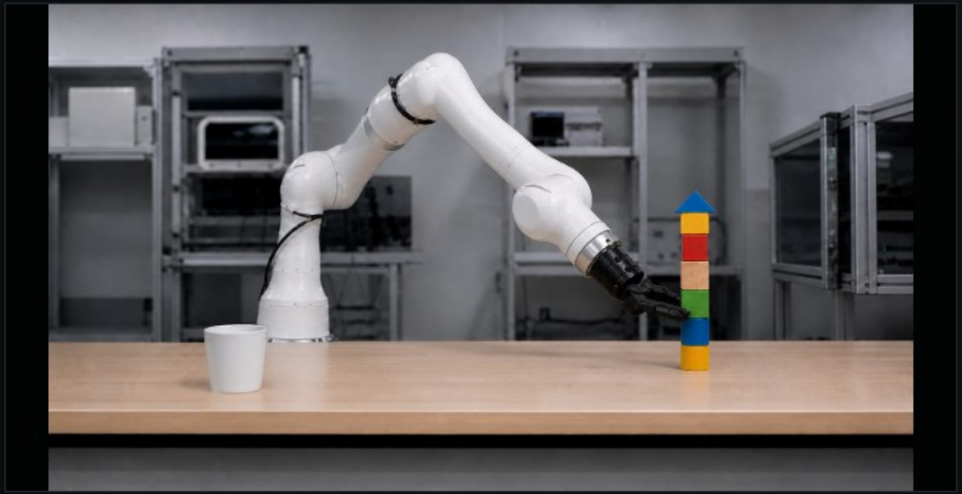
INTERACTION

ACT · OBSERVE · CHOOSE AGAIN

Feedback changes what the agent does next.

predict: moving the tower should open the path

ACT · OBSERVE



new evidence: where is the tower now?

PATH OPEN	reach for the cup
PATH STILL BLOCKED	move again—or try another side

The observation becomes the starting point for the next prediction.

One interaction updates the situation.
Many interactions can improve the model.

THIS EPISODE

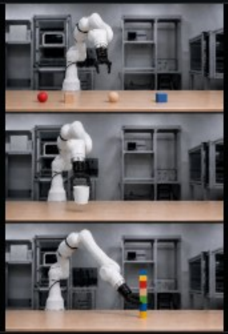
Update what is happening now.

- **observe** the path is partly open
- **update** the cup is still blocked
- **choose** move again or try another side

No retraining is required for this immediate update.

ACROSS EXPERIENCE

Learn better consequences.



soft push → shifts

hard push → topples

support first → stays upright

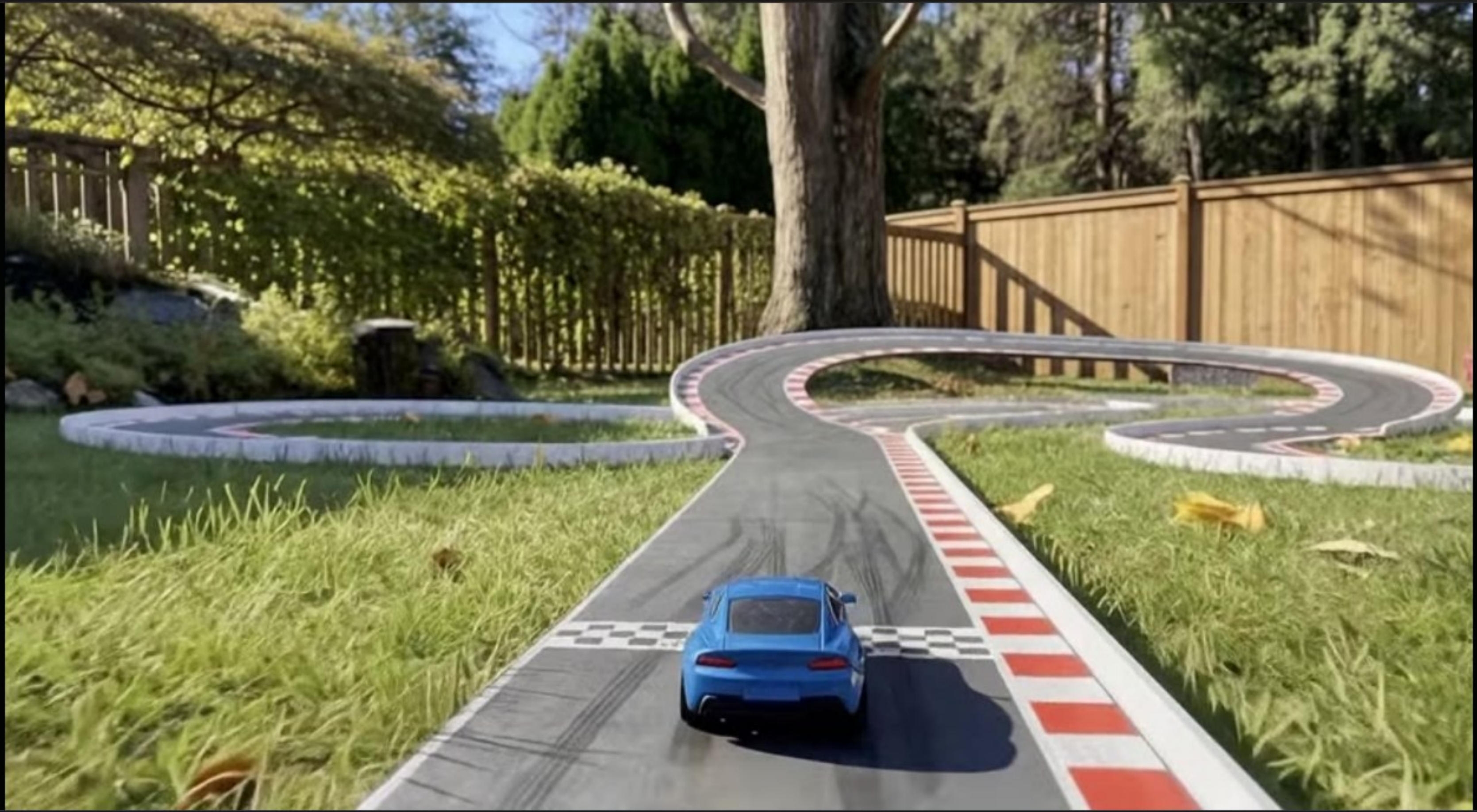
Accumulated trajectories can improve future predictions.

Immediate feedback changes this decision; accumulated experience can improve reusable predictions.

WHAT CAN WORLD MODELS DO?

Build an interactive world—and step inside.

As the user moves, what changes—and what stays coherent?



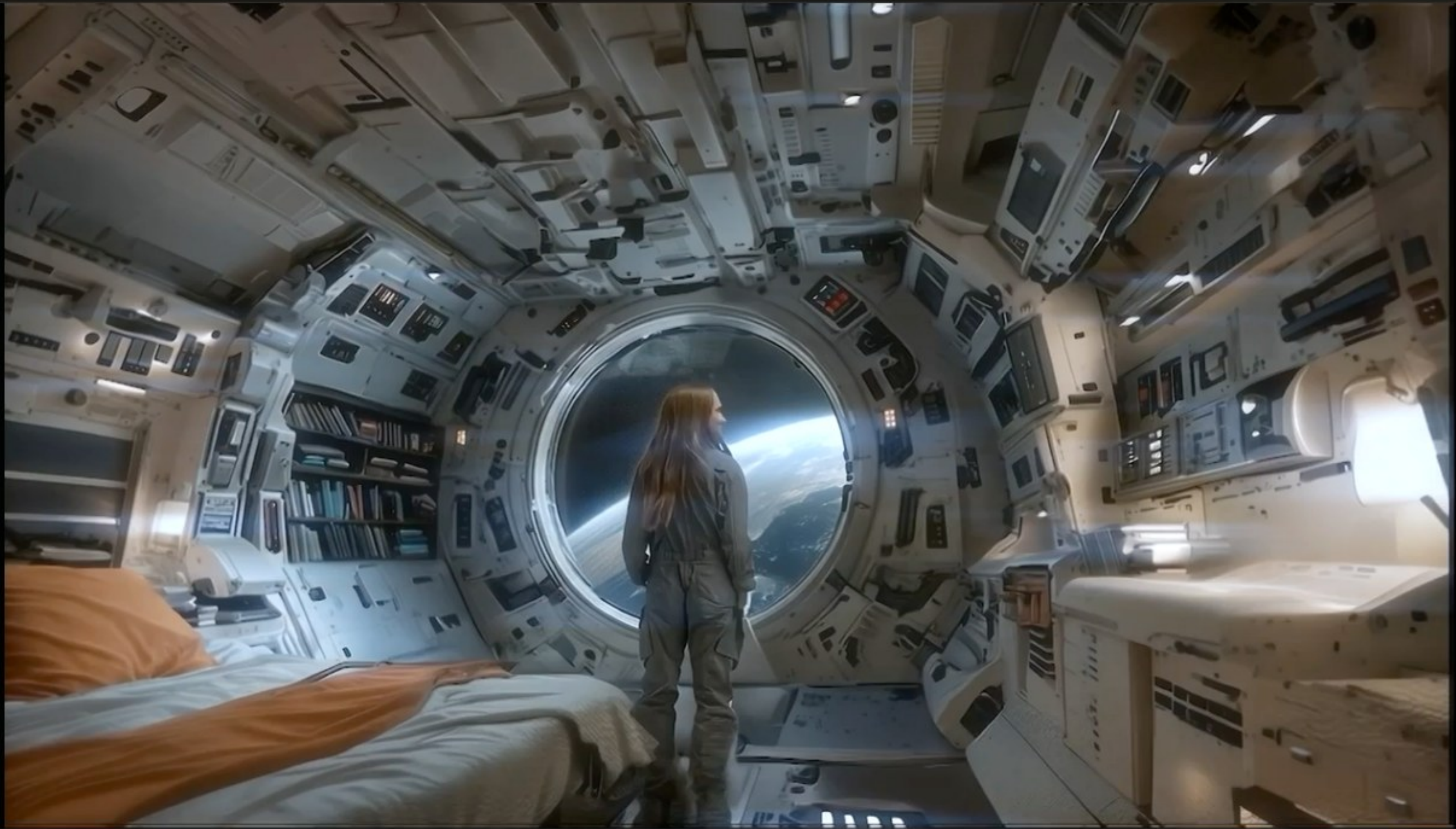
Genie 3 — a generated continuation responds as the user explores

Genie 3 · Google DeepMind · official demo

The application is not one video: it is a world that continues in response to the user.

Turn an image into a persistent 3D world.

Turn around: does the same room remain? Edit an object: does the change persist?



Marble — persistent and editable spatial generation

Marble · World Labs · official demo

A spatial world model can support navigation, editing, design, and simulation.

Test a driving decision before putting the car on the road.

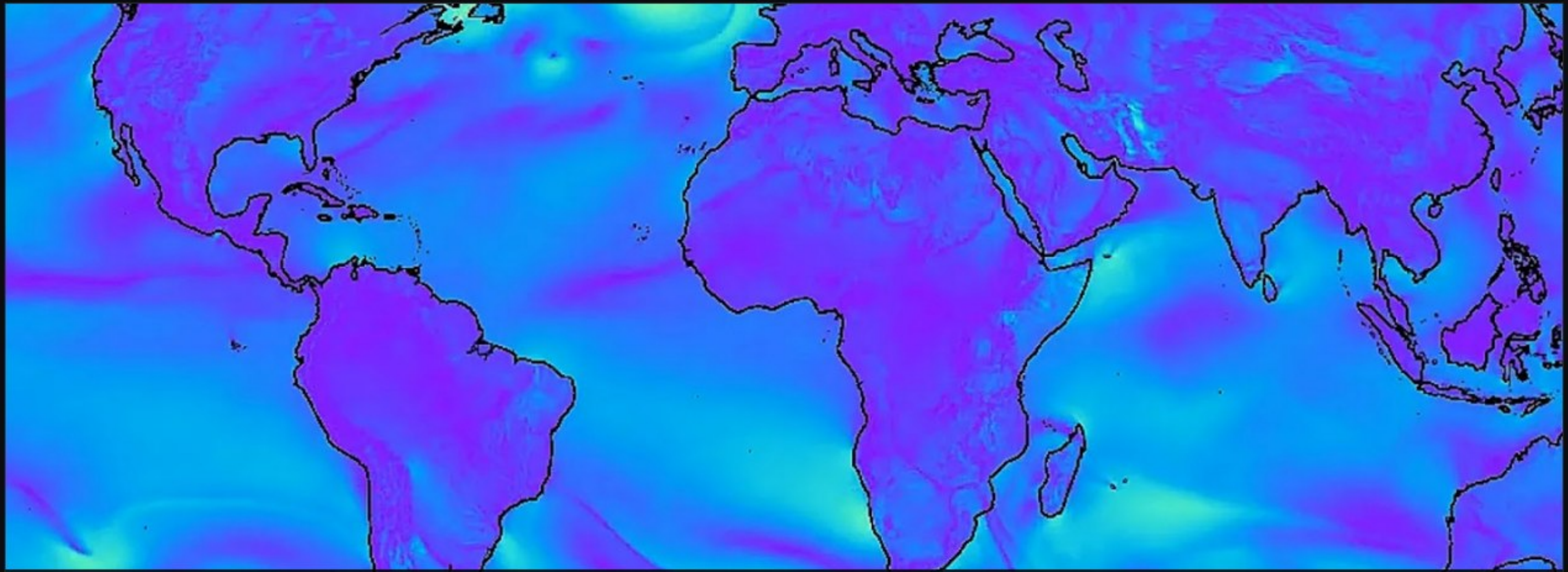
Same recorded context. Watch where “slow down” and “speed up” begin to diverge.

GAIA-4 — action-conditioned counterfactual sensor futures

GAIA-4 · Wayve · 2025 · official demo

Action-conditioned futures can expose consequences before a real vehicle commits.

Forecast the planet—not just the next video frame.



global weather variables over time

10 days

forecast horizon in the reported system

< 1 minute

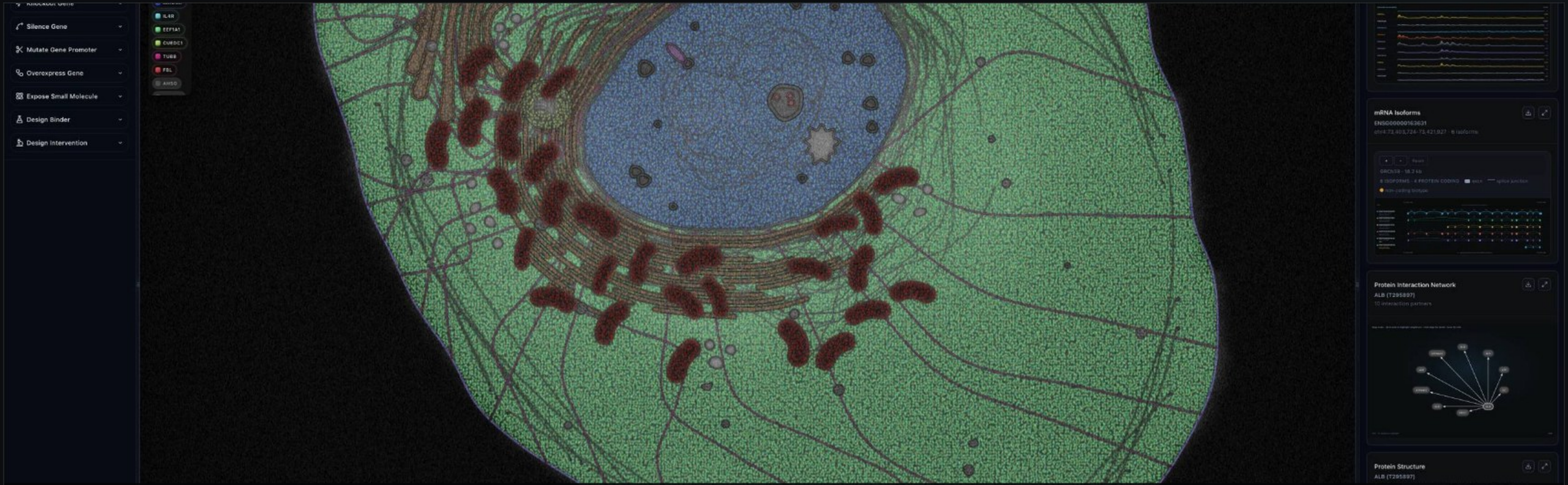
reported runtime for one forecast

The “world” can be represented by physical variables rather than pixels.

GraphCast — learned global weather forecasting

Google DeepMind · Science 2023

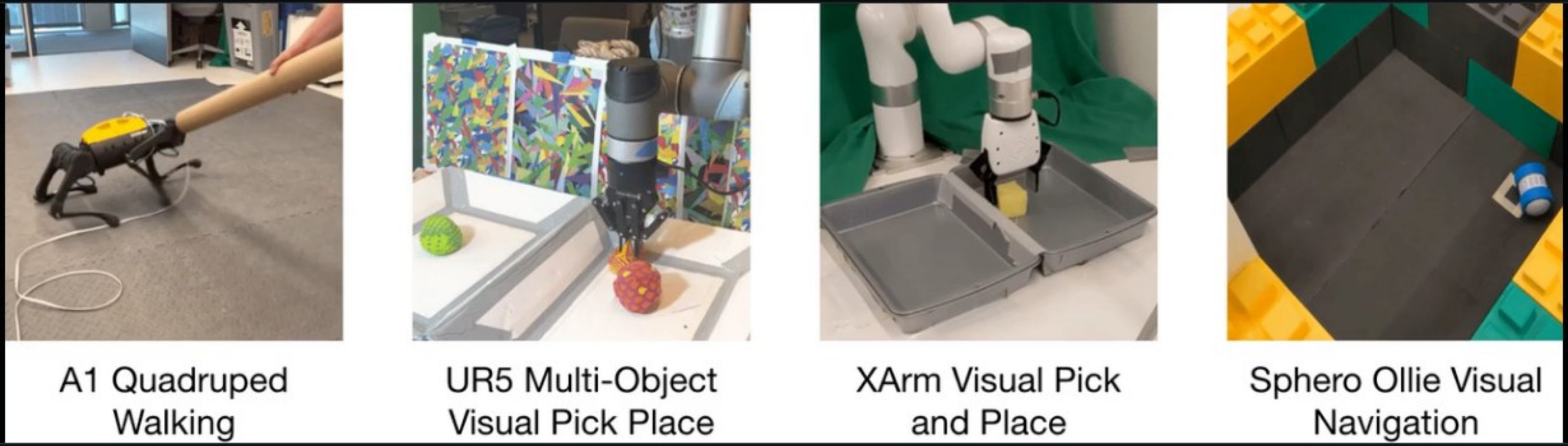
Test a gene edit or drug before the wet lab.



A virtual cell can be perturbed, observed, cloned, and perturbed again.

The ambition is a persistent cell state whose response can be read across biological scales.

Let a robot practice inside a learned world.



DayDreamer — world-model learning and imagined behavior learning on real robots

DayDreamer · Hafner et al. · CoRL 2022

Imagined practice can reduce hardware trials—but only if the learned model is accurate enough.

What should this robot remember, predict, and check?

REPRESENTATION

Where are the cup and tower—even when the arm blocks the view?

PREDICTION

What happens if the arm reaches over—or moves the tower first?

INTERACTION

After the first move, what should the robot inspect before choosing again?

Pairs · 60 seconds · then one answer per question.

Different worlds, the same three questions.

REPRESENTATION	What does the model observe?	What must it remember?
PREDICTION	What could happen next?	How does an action change the future?
INTERACTION	What can the agent change?	What feedback returns?

We will ask them about videos, robots, weather, cells, and software agents.

Tentative schedule · 24 lectures

Schedule

24 lectures and one final project presentation, plus four confirmed no-class dates and one tentative cancellation, following the Penn academic calendar. Assignment and project deadlines are collected under Coursework. Slides and readings appear here as the semester goes on.

Representation, prediction, and interaction are recurring themes, not sequential phases. Later application units—including 3D and robotics—bring these questions back together.

Download course calendar (.ics) ↓

Aug 25	01	**World Models: An Overview** Observation, state, transition, memory, action, prediction, simulation, planning, and reasoning.	Lecture 1
Aug 27	02	**World Modeling: Probabilistic Formulation** Trajectory distributions, latent state, partial observability, and one-step versus rollout objectives.	Lecture 2
Sep 01	03	**Environments, Simulators, and Rollouts** Environment interfaces, known transition rules, reset and step, observations, actions, feedback, and trajectory collection.	Lecture 3
Sep 03	04	**State-Space Models** Transition and emission models, filtering, smoothing, observability, and belief-state inference.	Lecture 4

The course website is the source of truth · topics and order may change.

Tentative schedule · Lectures 01–06

01 World Models: An Overview

02 World Modeling: Probabilistic Formulation

03 Environments, Simulators, and Rollouts

04 State-Space Models

05 Self-Supervised Representation Learning I: Foundations

06 Self-Supervised Representation Learning II: JEPA

Topics and order may change.

Tentative schedule · Lectures 07–12

07 Generative Model Foundations I: Latent-Variable and Adversarial Models

08 Generative Model Foundations II: Autoregressive Models

09 Generative Model Foundations III: Diffusion and Flow Matching

10 Generative Model Foundations IV: Normalizing and Autoregressive Flows

11 Sequential Generative World Models

12 Model-Based Reinforcement Learning, Planning, and Control

Topics and order may change.

Tentative schedule · Lectures 13–18

13 Video World Models I: Generative Architectures and Rollout	**14** Video World Models II: Interaction, Memory, and Efficiency
15 Spatial World Models I: Geometry and 3D Representations	**16** Spatial World Models II: 4D Dynamics and Interaction
17 Neural Physics and Learned Physical Dynamics	**18** World Models for Robot Learning I: Simulation, Control, and Sim-to-Real

Topics and order may change.

Tentative schedule · Lectures 19–24

19	World Models for Robot Learning II: Vision-Language-Action and World-Action Models
20	LLMs as World Models: Simulation, State Tracking, and Grounding
21	Reasoning Models: Deliberation, Recurrence, and Test-Time Computation
22	World Models for Digital Agents
23	Evaluating World Models: Utility, Robustness, and Open Problems
24	Course Synthesis and Open Problems

Topics and order may change.

Your grade has four components.

ATTENDANCE
In-class attendance

3
Coding assignments

1
Major project
proposal · checkpoint · presentation · report + code

FINAL
Written exam
one-page cheat sheet allowed

Exact grading weights will be announced separately.

Three coding assignments: environment, world model, policy.

01

Environment

Design your own environment: define its rules, observations, actions, and feedback.

→

02

World model

Learn a representation and predict how the world changes.

→

03

Policy

Use the model or feedback to choose actions and complete a task.

> Each assignment isolates a part of the same agent–world loop.

Major project: teams of 1–3 design a world model.

PROVIDED

Environment

A shared world and evaluation setting.

→

YOUR TEAM · 1–3 PEOPLE

World model

Choose the representation, prediction target, and learning method.

→

SHARED RESULT

Leaderboard

Compare what works under the same evaluation.

> Proposal → checkpoint → presentation → report + code

Three assignments lead into one research project.

SEP 09 → 23	SEP 30 → OCT 19	OCT 14	OCT 21 → NOV 18	NOV 16	DEC 03	DEC 14
Assignment 1	Assignment 2	Proposal	Assignment 3	Checkpoint	Presentation	Report + code