World of World Modeling

Shane Gu 顾世(World)翔

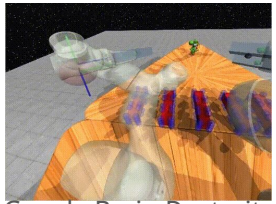
Senior Staff RS, Google DeepMind
Feb 25, 2026
Stanford CS234

Profile

- Japan-born Chinese Canadian, with PhD (2018) from UK and Germany
- Employments: Google Brain (2015-2022) → OpenAI (2023) → Google DeepMind (2023-present)
- Philosophy: Move on to a new field if **(impact remaining) / (top talents)** ratio becomes low

$$p(y_1, .., y_k) = \Gamma(k) \left(\prod_{i=1}^{k} \exp(x_i) \frac{y_k^\tau}{y_i^\tau} \right) \left(\sum_{i=1}^{k} \exp(x_i) \frac{y_k^\tau}{y_i^\tau} \right)^{-k} \tau^{k-1} \prod_{i=1}^{k} y_i^{-1}$$
$$= \Gamma(k) \tau^{k-1} \left(\sum_{i=1}^{k} \exp(x_i) / y_i^\tau \right)^{-k} \prod_{i=1}^{k} (\exp(x_i) / y_i^{\tau+1})$$

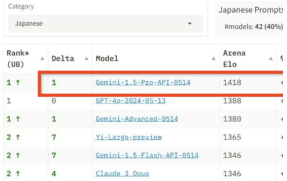
Gumbel-Softmax (2016)



Google Brain Dexterity Moonshot Co-Lead (2019)

OpenAI Japan entry (2023/02)

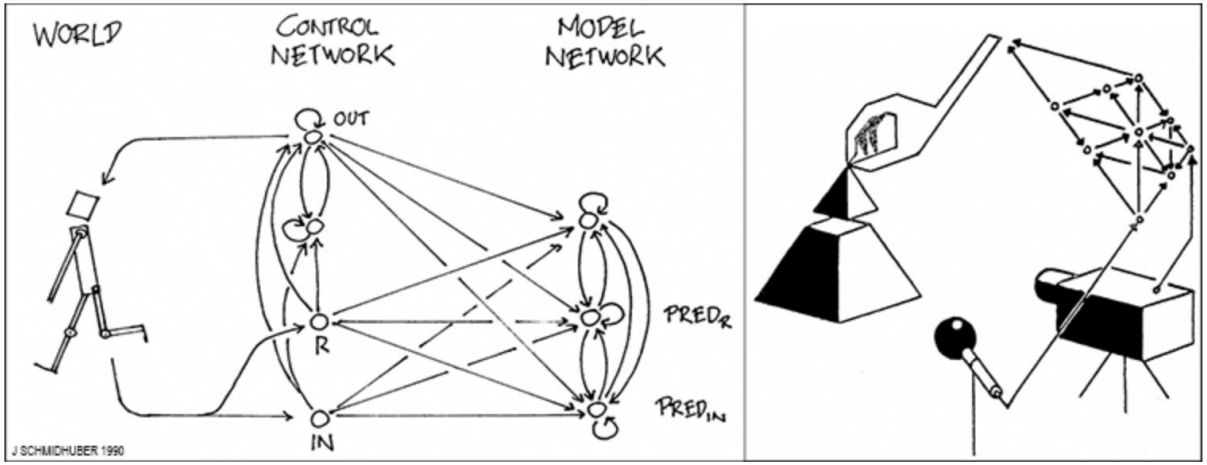
ChatGPT Post-Training (2023)



Gemini Post-Training
Multilinguality Lead (2024-25)

Gemini Thinking / RL (2025-)
Veo / Genie (2026-)

What is a world model?



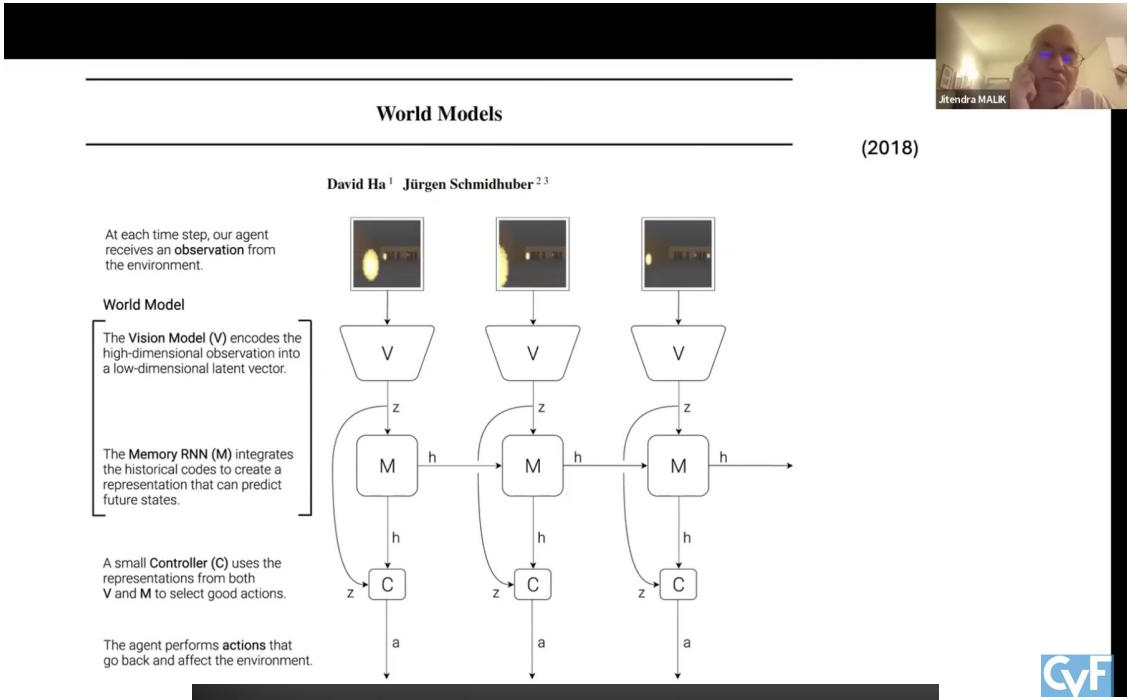
Jurgen Schmidhuber:
1990: Introduced RNN-based world models for planning.
1990: Proposed high-dimensional, multi-vector reward signals.
1990: Formulated deterministic policy gradients for RNNs.
1990: Invented adversarial artificial curiosity (foundation of GANs).
1991: Introduced neural network distillation and computational consciousness.
2004-2005: Applied world models to physical AI and self-healing robots.

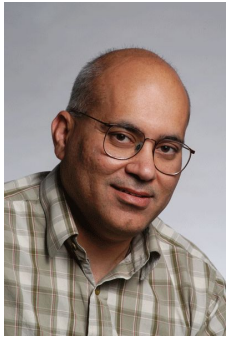
Jürgen Schmidhuber (Dec 2020, updated 2025)
Pronounce: You_again Shmidhoobuh

AI Blog
Twitter: @SchmidhuberAI

1990: Planning & Reinforcement Learning with Recurrent World Models and Artificial Curiosity

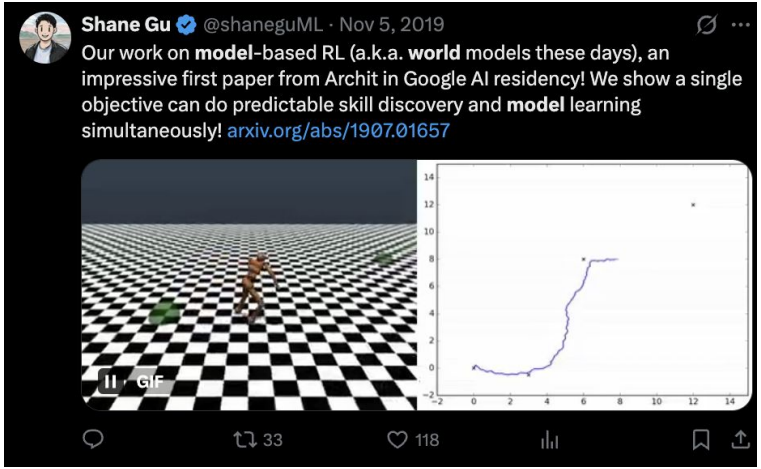
What is a world model?





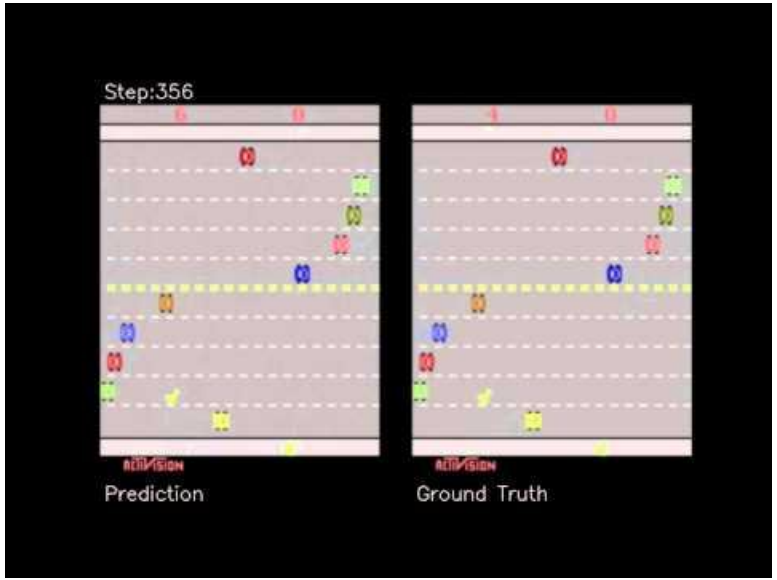
Jitendra Malik: “I think the main contribution of this paper is that it introduced the term 'world models.' This term caught on, but I'm not a particular fan of it because there was already a good term in use: 'dynamics model,' which control theorists started using in 1960. The consequence of using the term 'world models' is that we now have so much confusion about which definition of a world model we are discussing.”

What is a world model?



Shane Gu:
“World model is the ‘model’ in model-based RL.”

What is a world model?

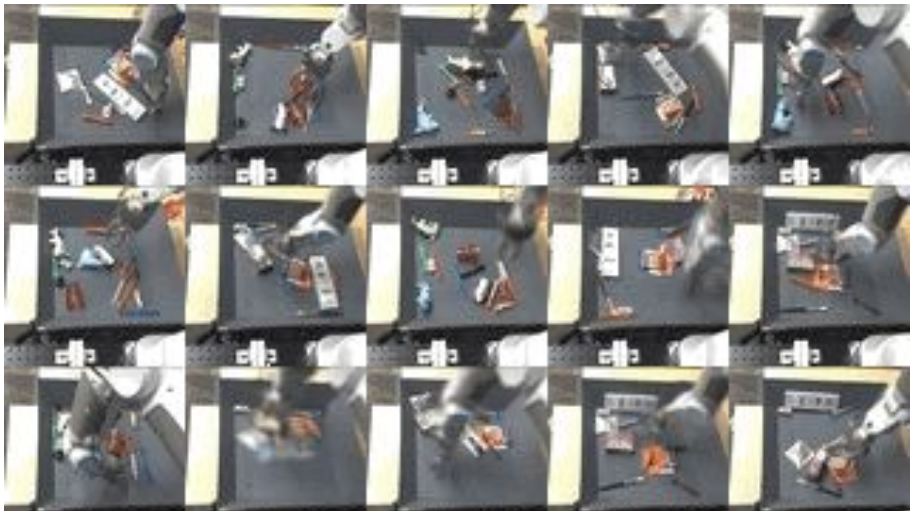


Action-Conditional Video Prediction using Deep Networks in Atari Games

Junhyuk Oh Xiaoxiao Guo Honglak Lee Richard Lewis Satinder Singh
University of Michigan, Ann Arbor, MI 48109, USA
{junhyuk,guoxiao,honglak,rickl,baveja}@umich.edu

Abstract

2015



Unsupervised Learning for Physical Interaction through Video Prediction

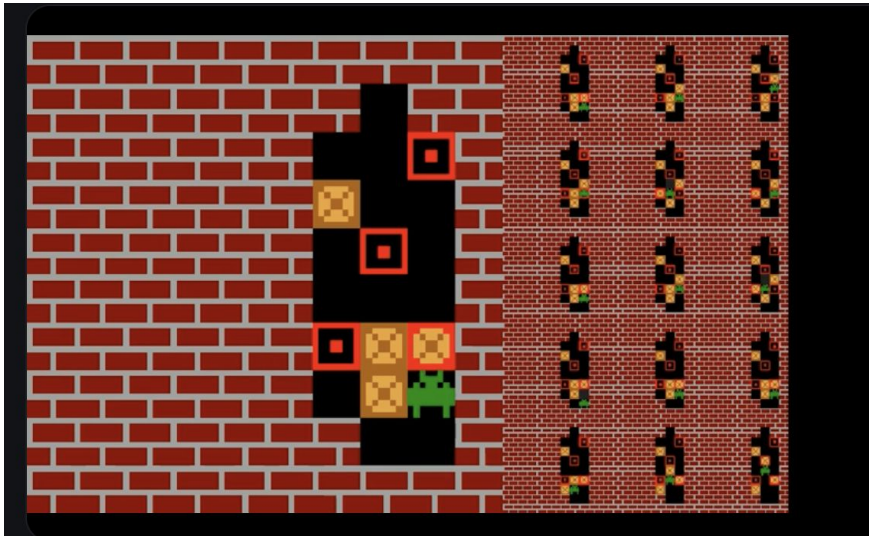
Chelsea Finn*
UC Berkeley
cbfinn@eecs.berkeley.edu

Ian Goodfellow
OpenAI
ian@openai.com

Sergey Levine
Google Brain
UC Berkeley
slevine@google.com

2016

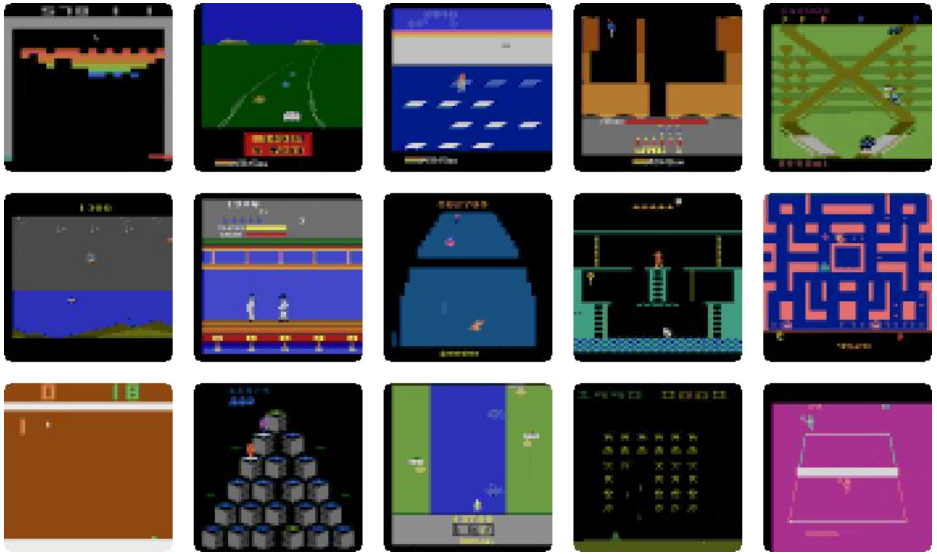
What is a world model?



Imagination-Augmented Agents for Deep Reinforcement Learning

2017

Théophane Weber* Sébastien Racanière* David P. Reichert* Lars Buesing
Arthur Guez Danilo Rezende Adria Puigdomènech Badia Oriol Vinyals
Nicolas Heess Yujia Li Razvan Pascanu Peter Battaglia
Demis Hassabis David Silver Daan Wierstra
DeepMind



Mastering Atari with Discrete World Models

Danijar Hafner Timothy Lillicrap Mohammad Norouzi Jimmy Ba

ICLR 2021

2021

Talk Outline

- What is prediction?
  - Solomonoff induction and universal intelligence
  - Causality and OOD generalization
  - Empowerment and 3 levels of predictability maximization
- Forward and inverse world models
  - Shooting and collocation methods
  - Goal-conditioned value functions and (generalized) decision transformers
- Physical and symbolic world models
  - Video models as reasoners
  - Futures of world modeling
- (Forgot) Control as inference: e.g. particle smoothing as optimal control, MCTS vs beam search

I won’t talk about any Gemini / Veo / Genie. Please check out new models! Or ask me questions after the class.

What is prediction?

What's the "best" world model?

the "induction" machine: given data, infer the **rules / causes / programs**

program: P

```c
#include <stdio.h>

int main()
{
    printf("Hello World");

    return 0;
}
```

Generation

Induction

data: $D_{t-1}, D_t, D_{t+1}, \ldots$

e.g.

$D_{t+1} = D_t + 2$

$\{1, 3, 5, 7, \ldots\}$

Solomonoff Induction

- Formalization of Bayes theorem in computation theory

$$p(P|D) \propto p(D|P)p(P)$$

- Occam's razor (related: Kolmogorov Complexity, Minimum Description Length)

$$p(P) \propto 2^{-|P|}$$

- Inspired:

C-Test [Hernandez-Orallo et al1998]

Sequence Prediction Test

Complexity	Sequence	Answer
9	a, d, g, j, _ , . . .	m
12	a, a, z, c, y, e, x, _ , . . .	g
14	c, a, b, d, b, c, c, e, c, d, _ , . . .	d

Universal Intelligence [Legg & Hutter 2007]

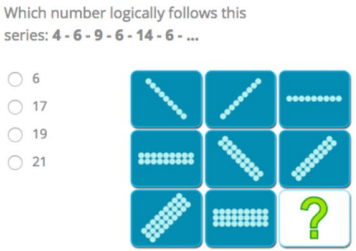
$$\sum_{\mu \in E} 2^{-K(\mu)} V_{\mu}^{\pi}$$

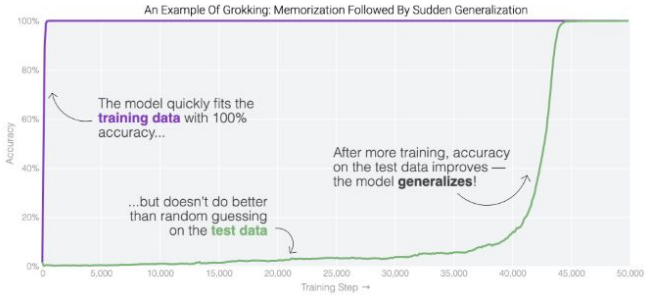
environment

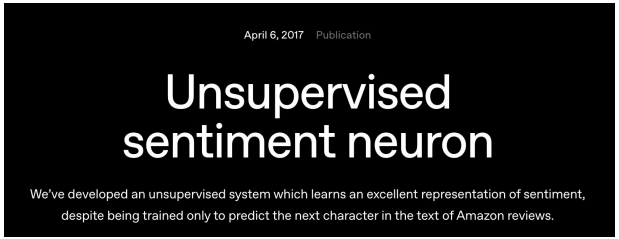
Google DeepMind

Solomonoff Induction: Prediction = Understanding

- "The best prediction is inferring the shortest program that reproduces the data." (Ilya Sutskever)
- Prediction = Understanding = Inferring the causal program of a phenomenon.
- Examples: IQ tests, "Aha!" moments, scientific discoveries, deep learning (Grokking), and physics (Phase Transitions).
    - Text prediction yields emotional understanding (OpenAI Sentiment Neuron), while video/audio prediction yields physical and emotional understanding (Google Veo3).
- Despite hardware differences, humans and AI models both build intelligence through prediction—mastering prediction means mastering understanding.





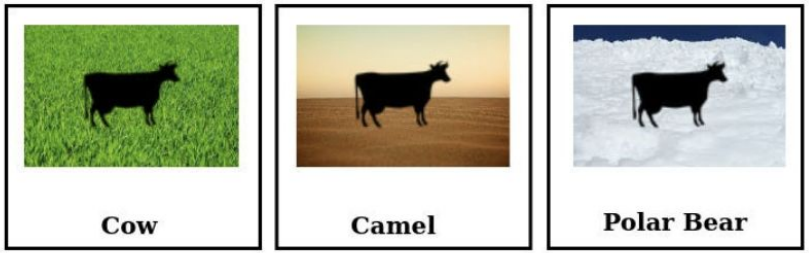


Causality: Understanding = OOD Generalization

- Invariant Risk Minimization (2019): non-causal spurious data biases prevent true OOD generalization in ML
- Generalization can be achieved by having data coming from different interventions of the true causal graph → “Diversity is all you need” → GPT-3, i.e. train to predict everything.

Invariant Risk Minimization

Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, David Lopez-Paz



Empowerment and 3 levels of prediction

- Level 1: Passively fit your world model on world
- Level 2: Actively fit your world model on world
- Level 3: Actively fit world to your world model
- Notations: s = future of the world, z = your actions
  - Empowerment: Daniel Polani 2005, Shakir Mohammed 2015

$$\mathcal{I}(s;z) \geq \mathbb{E}_{z\sim p(z),s\sim p(s|z)}\left[\log q_\phi(s|z) - \log p(s)\right]$$

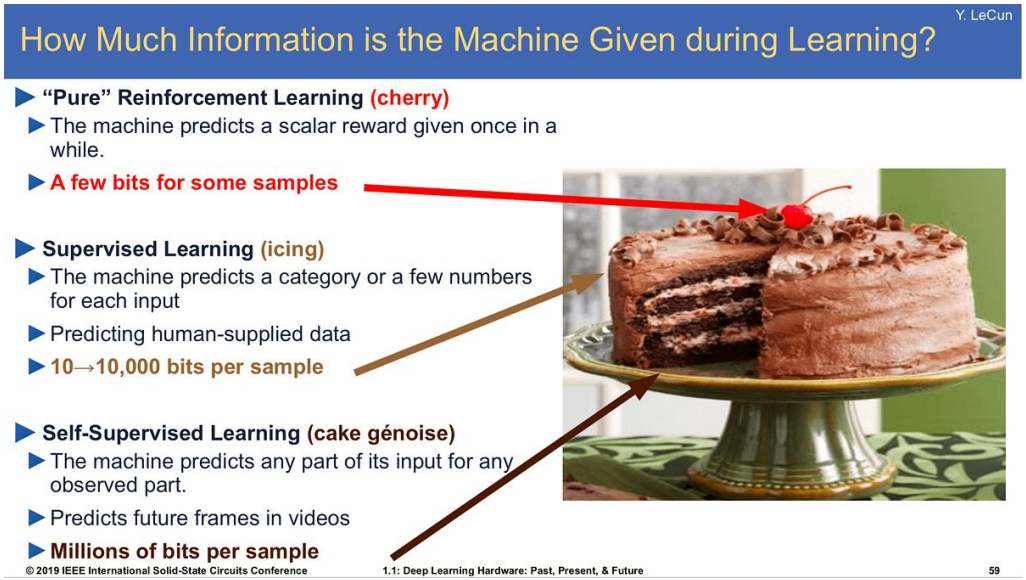
Variational Empowerment as Representation Learning for Goal-Based Reinforcement Learning

---

Jongwook Choi [†1] **Archit Sharma** [†2] **Honglak Lee** [1 3] **Sergey Levine** [4 5] **Shixiang Shane Gu** [4]

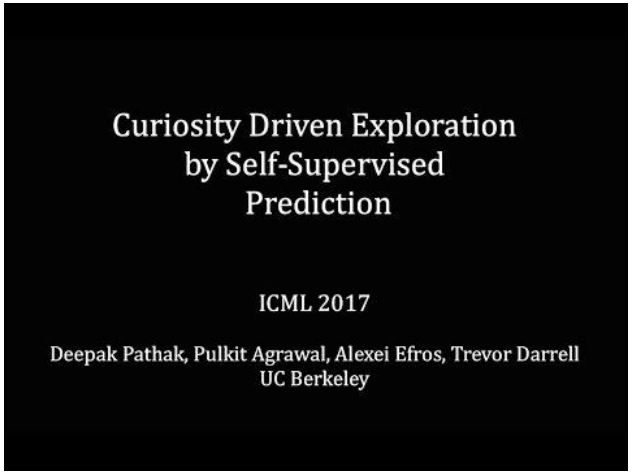
Level 1: Passively fit your world model on world

- Pre-training, Supervised learning, Self-supervised learning, Generative modeling
- Data distribution is **stationary** during predictive training

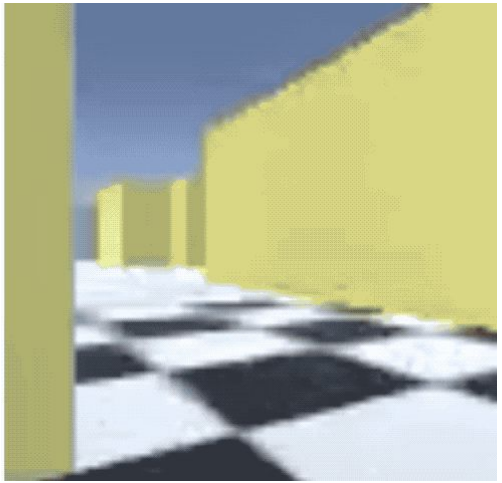


Level 2: Actively fit your world model on world

- Post-training, DAgger, GAIL, active learning
- Data distribution is **non-stationary** during predictive training



DPM (Stadie 2016) and ICM (Pathak 2017) incentivizes novelty



$$I\left(S_{t+1};\Theta|\xi_t,a_t\right)=\mathbb{E}_{s_{t+1}\sim\mathcal{P}(\cdot|\xi_t,a_t)}\left[D_{\mathrm{KL}}[p(\theta|\xi_t,a_t,s_{t+1})\,\|\,p(\theta|\xi_t)]\right]$$

$$r'(s_t,a_t,s_{t+1})=r(s_t,a_t)+\eta D_{\mathrm{KL}}[p(\theta|\xi_t,a_t,s_{t+1})\,\|\,p(\theta|\xi_t)]$$

VIME (Houthooft et al 2017) tackles noisy TV problem

Level 3: Actively fit world to your world model

- Perform actions in the world to make the world mode predictable wrt you / your own world model
  - Politicians
  - Financial firms
  - X Influencers
- **Difficult objective. Nobody has cracked this yet at scale.**

$$\mathcal{I}(s;z) \geq \mathbb{E}_{z\sim p(z),s\sim p(s|z)}\left[\log q_\phi(s|z) - \log p(s)\right]$$

Variational Empowerment as Representation Learning for Goal-Based Reinforcement Learning

Jongwook Choi [†1] **Archit Sharma** [†2] **Honglak Lee** [1 3] **Sergey Levine** [4 5] **Shixiang Shane Gu** [4]

Published as a conference paper at ICLR 2020

DYNAMICS-AWARE UNSUPERVISED DISCOVERY OF SKILLS

Archit Sharma[*], **Shixiang Gu, Sergey Levine, Vikash Kumar, Karol Hausman**
Google Brain
{architsh,shanegu,slevine,vikashplus,karolhausman}@google.com

Forward and inverse world models

What is a (World) Model?

Forward Models

$$s_{t+1} = F(s_t, a_t)$$

$$a = \Pi(s, F(s, a)) \quad \forall \quad s, a$$

$$a_t = \Pi(s_t, s_{t+1})$$

$$0 = Q(s_t, a_t, s_{t+1})$$

$$0 = Q(s, a, F(s, a)) \quad \forall \quad s, a$$

Inverse Models

* assume deterministic for simplicity

** will explain why I chose those symbols

Shooting vs Direct Collocation for Planning

Shooting:

$$\arg\max_{a_{t:t+T}} \sum_{i=t}^{t+T} r(s_i) \quad \text{where} \quad s_{i+1} = F(s_i, a_i)$$

Direction Collocation (Constrained Optimization):

$$\arg\max_{s_{t:t+T}} \sum_{i=t}^{t+T} r(s_i) \quad \text{s.t.} \quad -|A| \leq \Pi(s_i, s_{i+1}) \leq |A|$$

$$\arg\max_{s_{t:t+T}, a_{t:t+T}} \sum_{i=t}^{t+T} r(s_i) \quad \text{s.t.} \quad Q(s_i, a_i, s_{i+1}) = 0$$

Direct Collocation: Contact Invariant Optimization (CIO)

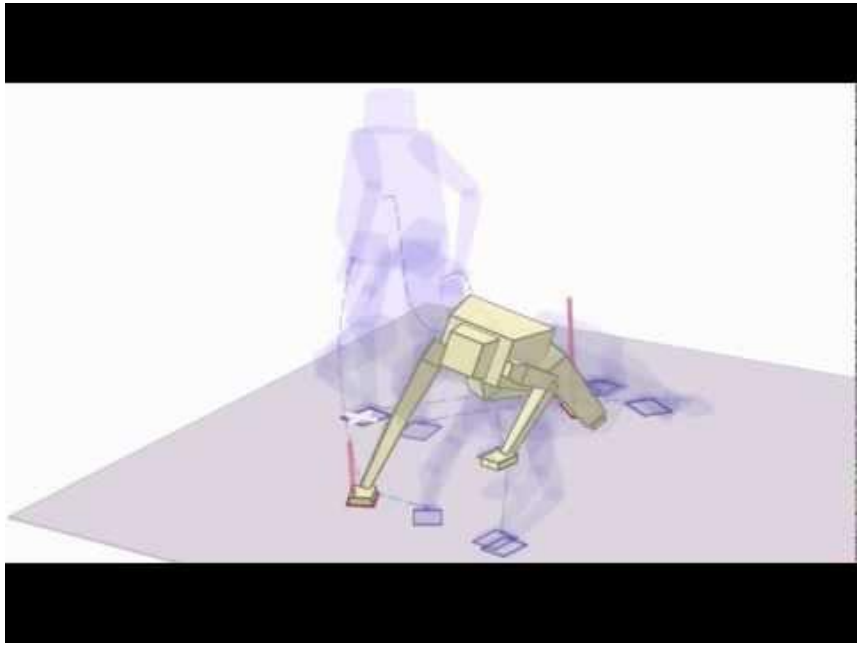
- Constrained optimization wrt dynamics constraints := relaxing (hacking) dynamics during planning
- “First solve task, then fix physics”

[Mordatch et al 2012 (SIGGRAPH)]

$$\tau\left(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}\right) = J\left(\mathbf{q}\right)^T \mathbf{f} + B\mathbf{u}$$

$$L_{\text{Physics}}\left(\mathbf{s}\right) = \sum_t \left\| J_t\left(\mathbf{s}\right)^T \mathbf{f}_t\left(\mathbf{s}\right) + B\mathbf{u}_t\left(\mathbf{s}\right) - \tau_t\left(\mathbf{s}\right) \right\|^2$$

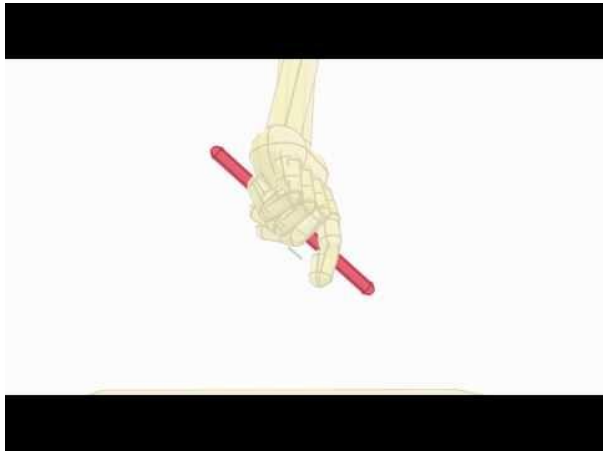
$$L_{\text{CI}}\left(\mathbf{s}\right) = \sum_t c_{i,\phi(t)}\left(\mathbf{s}\right)\left(\left\|\mathbf{e}_{i,t}\left(\mathbf{s}\right)\right\|^2 + \left\|\dot{\mathbf{e}}_{i,t}\left(\mathbf{s}\right)\right\|^2\right)$$



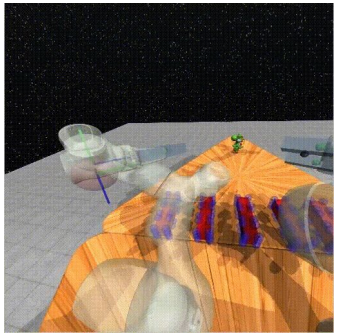
Direct Collocation: Contact Invariant Optimization (CIO)

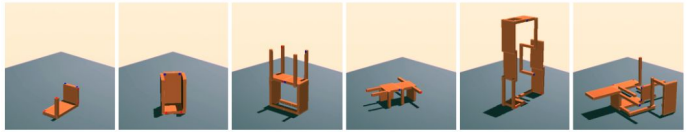
- Direct collocation solves contact-rich tasks with minimal reward shaping, i.e. dynamics relaxation provides appropriate “reward shaping”
- **Analogy:** autoregressive video diffusion vs bidirectional video diffusion

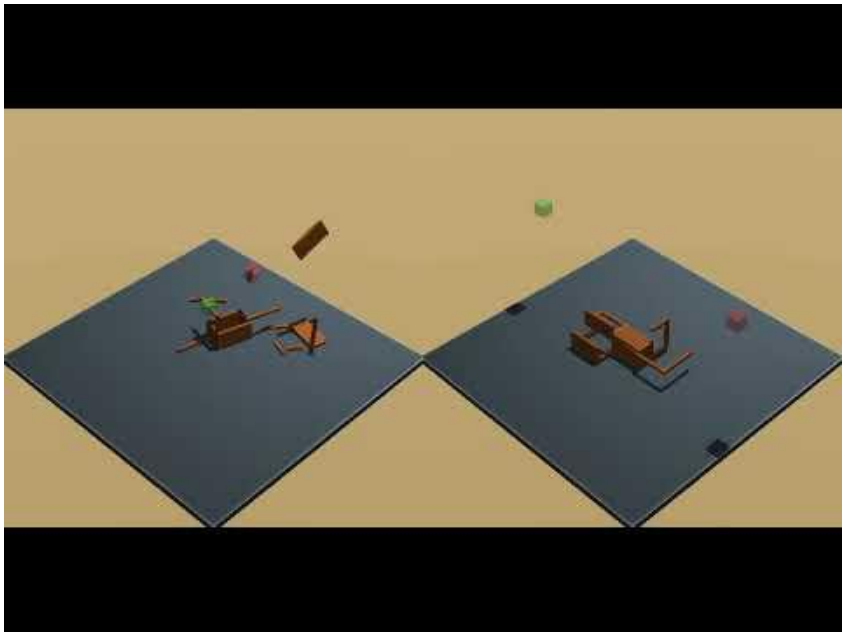
[Mordatch et al 2012+ (SIGGRAPH)]

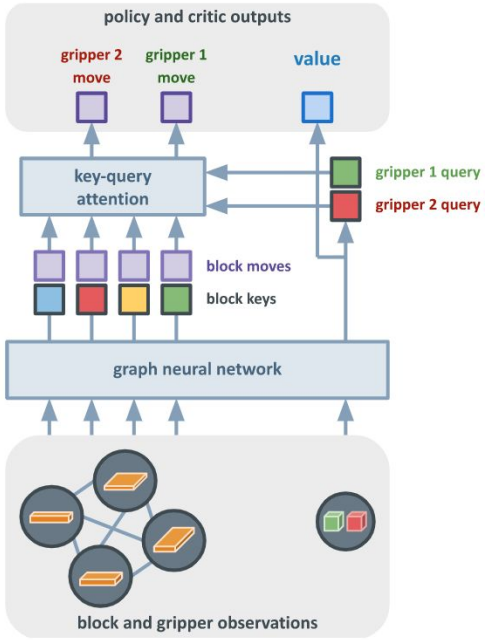


Dexterity with shooting method is hard









Kamyar Ghasemipour, Byron David, Daniel Freeman, Shixiang Shane Gu, Satoshi Kataoka, Igor Mordatch. "**Learning to Assemble with Large-Scale Structured Reinforcement Learning**." (2022)

Back to the notations…

$$s_{t+1} = F(s_t, a_t)$$

$$a_t = \Pi(s_t, s_{t+1}) \Rightarrow \text{a policy } |s_{t+1}$$

$$0 = Q(s_t, a_t, s_{t+1}) \Rightarrow \text{a Q-function } |s_{t+1}$$

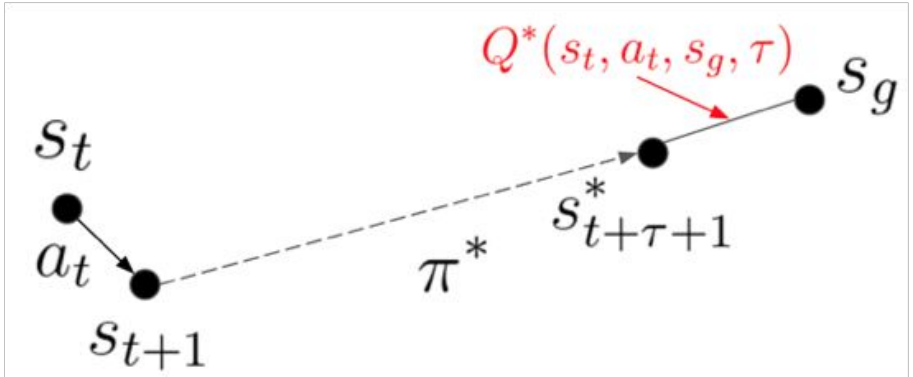
Temporal Difference Models (TDMs)

[Pong et al 2018 (ICLR)]

- Goal-conditioned optimal policy or Q-function as an implicit world model (temporally-extended)
- Use hindsight relabeling trick
- **Value functions are world models with different time scale and representation**

$$r_d(s_t, a_t, s_{t+1}, s_g, \tau) = -D(s_{t+1}, s_g)1[\tau = 0]$$

$$Q(s_t, a_t, s_g, \tau) = \mathbb{E}_{p(s_{t+1}|s_t,a_t)}[-D(s_{t+1}, s_g)\mathbb{1}[\tau=0] + \max_a Q(s_{t+1}, a, s_g, \tau-1)\mathbb{1}[\tau\neq 0]]$$



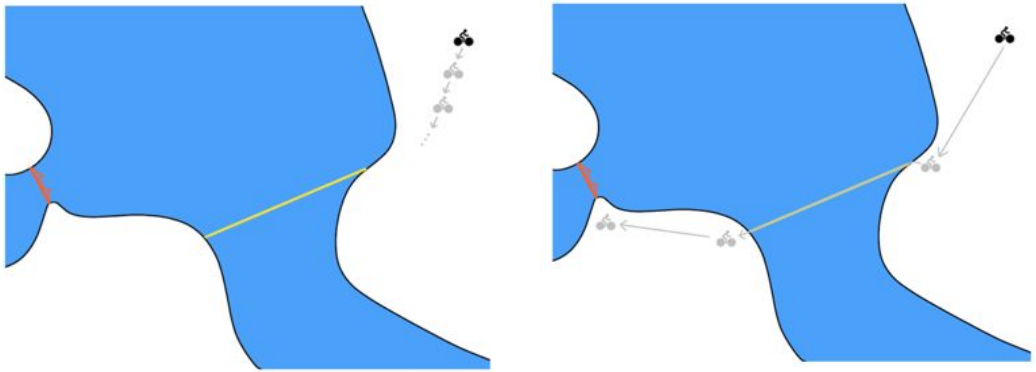
Direct Collocation with TDMs

[Pong et al 2018 (ICLR)]

- Hierarchical RL with goal-conditioned Q-function + direct collocation

$$a_t = \underset{a_{t:K:t+T}, s_{t+K:K:t+T}}{\operatorname{argmax}} \sum_{i=t,t+K,\ldots,t+T} r_c(s_i, a_i)$$

$$\text{such that } Q(s_i, a_i, s_{i+K}, K-1) = 0 \;\forall\; i \in \{t, t+K, \ldots, t+T-K\}$$



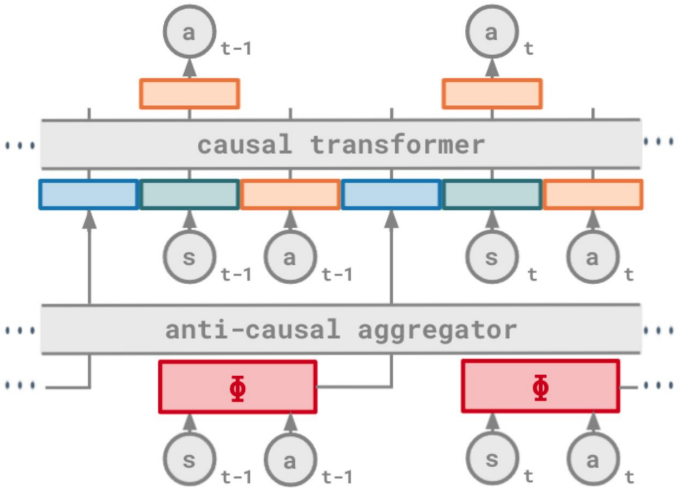
Direction Collocation (Constrained Optimization):

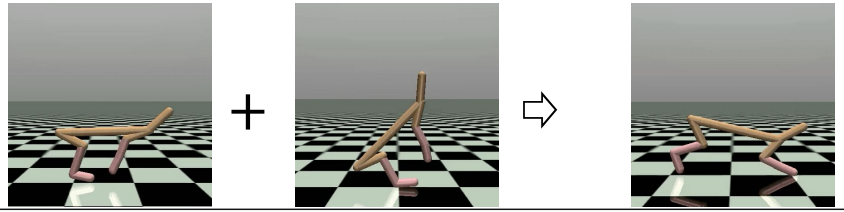
$$\arg\max_{s_{t:t+T}} \sum_{i=t}^{t+T} r(s_i) \quad \text{s.t.} \quad -|A| \le \Pi(s_i, s_{i+1}) \le |A|$$

$$\arg\max_{s_{t:t+T}, a_{t:t+T}} \sum_{i=t}^{t+T} r(s_i) \quad \text{s.t.} \quad Q(s_i, a_i, s_{i+1}) = 0$$

Generalized Decision Transformers

- Training: hindsight BC (behavioral cloning) wrt “any future statistics”
- Test-time: provide unseen “future” and ask it for “generalization”



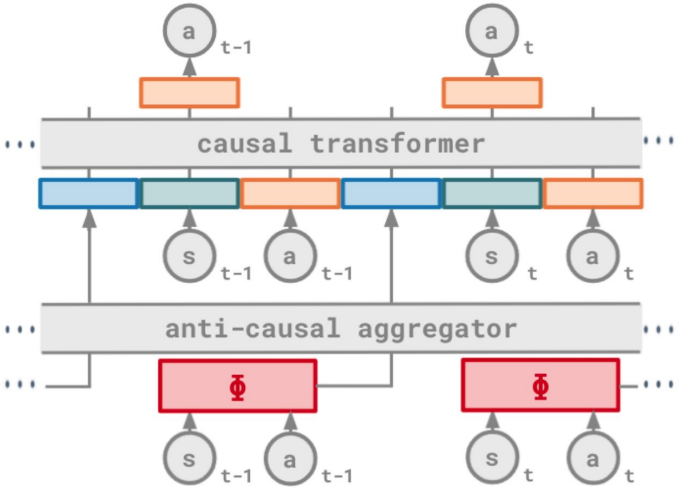


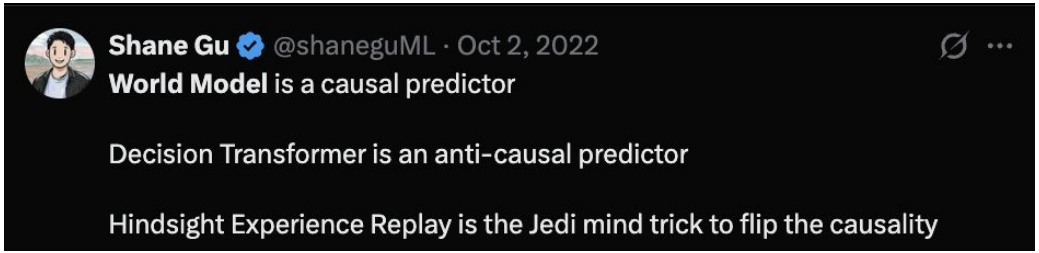
Method	Algo. Type	Training	$\mathbf{I}^{\Phi}(\tau)$	Architectures
Andrychowicz et al. (2017)	RL	Online	ϕ_T	MLP
Pong et al. (2018)	RL	Online	ϕ_T	MLP
Chebotar et al. (2021)	RL	Offline	ϕ_T	CNN
Li et al. (2020)	RL	Online	$\arg\max \sum_t \gamma^t r(s_t, a_t, \cdot)$	MLP
Eysenbach et al. (2020)	BC/RL	On/Offline	$\arg\max \sum_t \gamma^t r(s_t, a_t, \cdot)$	MLP
Lynch et al. (2019)	BC	Offline	ϕ_T	Stochastic RNN
Ghosh et al. (2021)	BC	Online	ϕ_T	MLP
Srivastava et al. (2019)	BC	Online	$\sum_t \gamma^t r_t$	Fast Weights
Kumar et al. (2019)	BC	Online	$\sum_t \gamma^t r_t$	MLP
Janner et al. (2021)	BC	Offline	$\sum_t \gamma^t r_t$ or ϕ_T	Transformer
Duan et al. (2017)[3]	BC	Offline	τ	MLP + LSTM
Generalized DT (ours)	BC	Offline	Any	Transformer
DT (Chen et al., 2021a)	BC	Offline	$\sum_t \gamma^t r_t$	Transformer
Categorical DT (ours)[4]	BC	Offline	histogram(r_t, γ)	Transformer
Bi-Directional DT (ours)	BC	Offline	τ	Transformer

Hiroki Furuta, Yutaka Matsuo, Shixiang Shane Gu. "**Generalized Decision Transformer for Offline Hindsight Information Matching**". ICLR 2022 **[Spotlight, 5.2% of 3391 submissions].** [Arxiv] [Project Page] [Github]

Direct Collocation with GDTs (untested)

- Search for a “reachable future” using the policy function





Direction Collocation (Constrained Optimization):

$$\arg\max_{s_{t:t+T}} \sum_{i=t}^{t+T} r(s_i) \quad \text{s.t.} \quad -|A| \le \Pi(s_i, s_{i+1}) \le |A|$$

$$\arg\max_{s_{t:t+T}, a_{t:t+T}} \sum_{i=t}^{t+T} r(s_i) \quad \text{s.t.} \quad Q(s_i, a_i, s_{i+1}) = 0$$

Physical and symbolic world models

2022: AGI Year 0

- “LLMs can reason” → Symbolic AGI is achievable through LLMs
- “ImagenVideo / DreamFusion” → Physical AGI is achievable through video models

Large Language Models are Zero-Shot Reasoners

Takeshi Kojima
The University of Tokyo
t.kojima@weblab.t.u-tokyo.ac.jp

Shixiang Shane Gu
Google Research, Brain Team

Machel Reid
Google Research*

Yutaka Matsuo
The University of Tokyo

Yusuke Iwasawa
The University of Tokyo

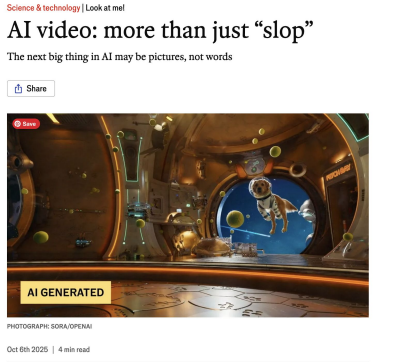
(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: ***Let’s think step by step.***

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓

2025: Video model as the missing foundation model

- World consists of symbols, space, and time
- LLMs: reasoning with symbols
- Video models: reasoning in space and time
- Chain-of-Frames (CoF) vs Chain-of-Thoughts (CoT)

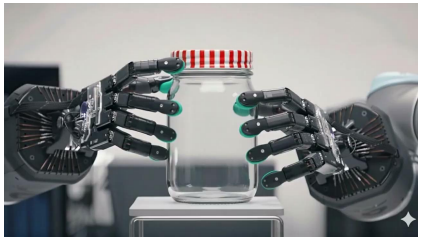


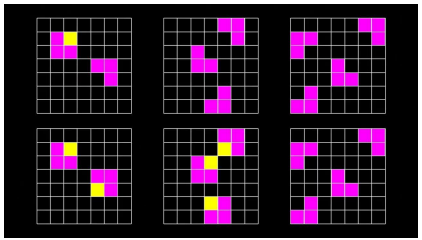
Google DeepMind 2025-9-25

Video models are zero-shot learners and reasoners

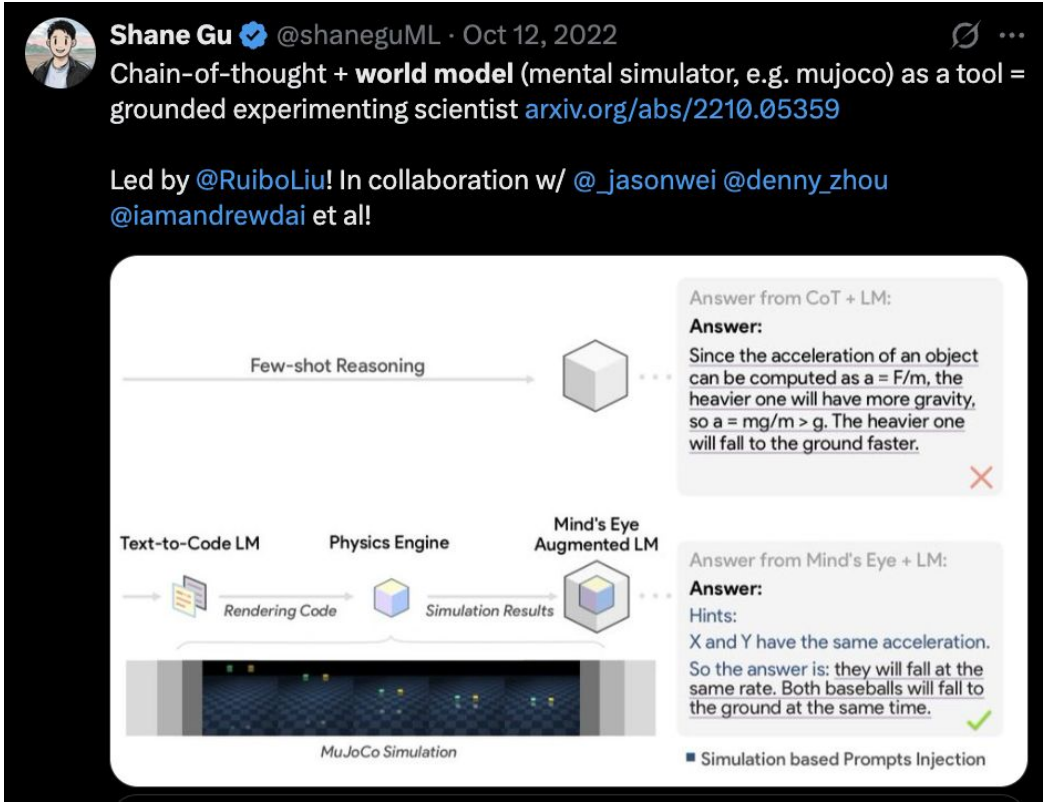
Thaddäus Wiedemer*[1], Yuxuan Li[1], Paul Vicol[1], Shixiang Shane Gu[1], Nick Matarese[1], Kevin Swersky[1], Been Kim[1], Priyank Jaini*[1] and Robert Geirhos*[1]
[1]Google DeepMind

The remarkable zero-shot capabilities of Large Language Models (LLMs) have propelled natural language processing from task-specific models to unified, generalist foundation models. This transformation emerged from simple primitives: large, generative models trained on web-scale data. Curiously, the same primitives apply to today’s generative video models. Could video models be on a trajectory towards general-purpose *vision* understanding, much like LLMs developed general-purpose *language*





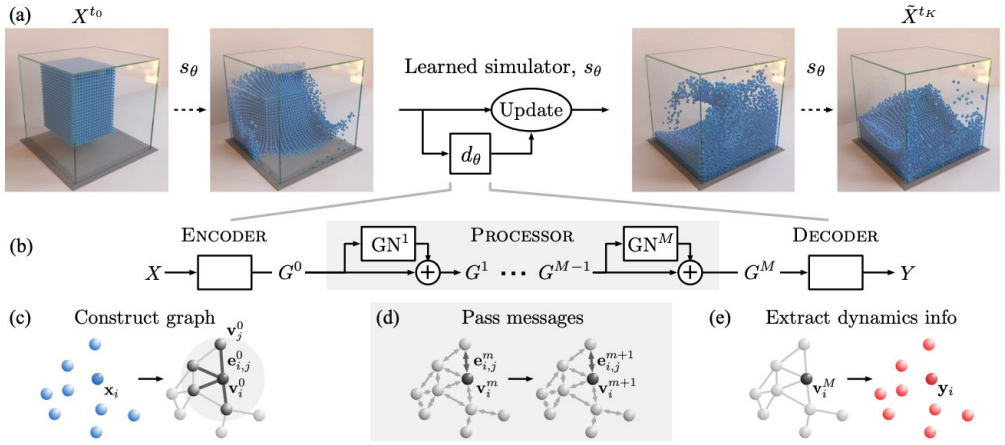
2025: Video model as the missing foundation model

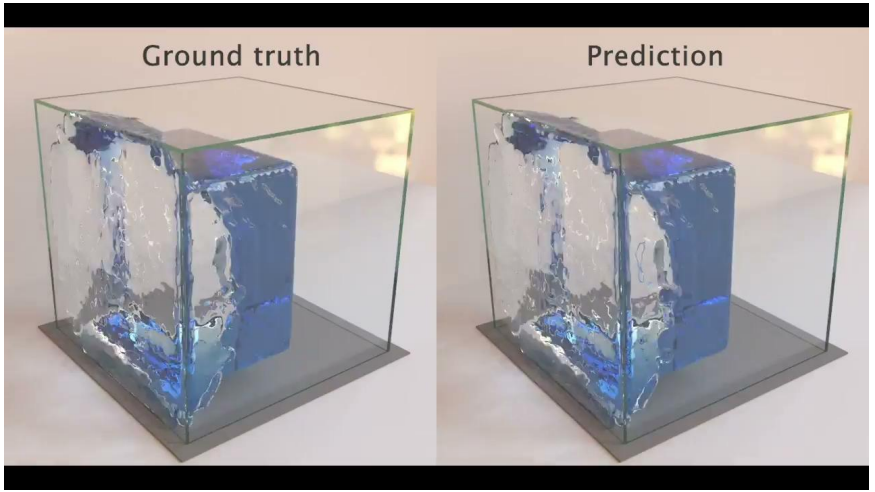


Older work:

[Sanchez-Gonzalez et al 2020]

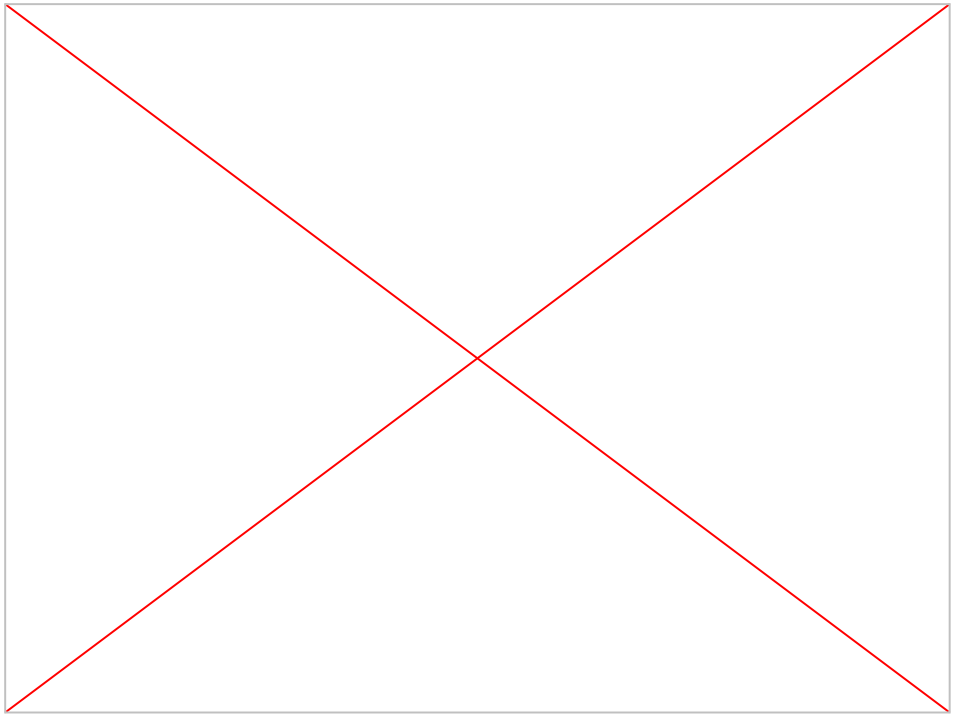
- Particle-based simulation = message passing on graphs
- **We unlikely won't need graph nets / NeRF etc. Just video models.**
  - **AlphaFold?**





2026+: Modeling humanity

Simile: model 8B people



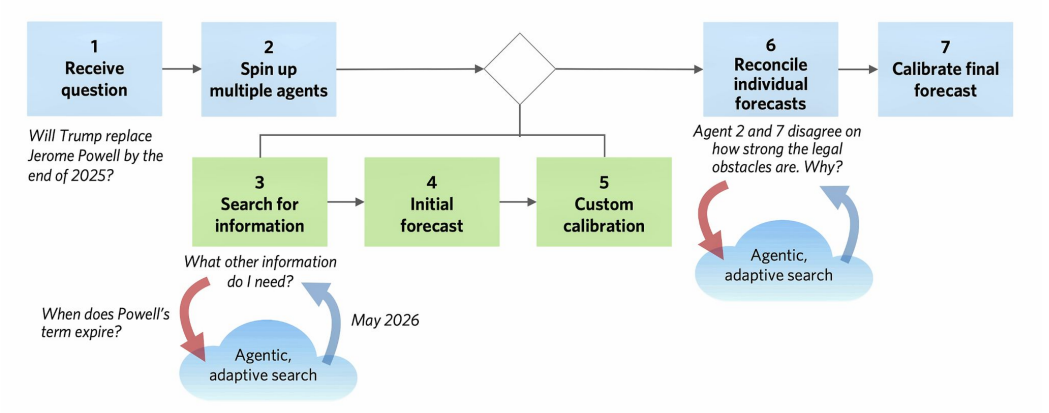
2026+: Modeling financial market

arXiv:2511.07678v1 [cs.AI] 10 Nov 2025

AIA Forecaster: Technical Report

Rohan Alur*, Bradly C. Stadie*, Daniel Kang, Ryan Chen, Matt McManus,
Michael Rickert, Tyler Lee, Michael Federici, Richard Zhu, Dennis Fogerty,
Hayley Williamson, Nina Lozinski, Aaron Linsky, Jasjeet S. Sekhon

\addr Bridgewater AIA Labs
New York, NY



Thank you! Remember the Level 3.

X (English): shanegumlㅤ@

X (Japanese): shanegjpㅤ@

Email: shanegumlㅤ@gmail.com