

World-models and the *twilight* of LLMs.

Gergely Várhelyi-Tóth
TÜV Rheinland - BUDAPEST · MAY 2026



A young child

910 days old - 2.5 half years

Can figure out letter A, E, R and
number 6

Consumed 850000 kCal ca. 1000kWh
so far



A freshly built AI datacenter

800 engineers put their average
20 years of experience into it

Consumed every media created by
humanity in the last 100.000
years. Multiple times.

Needed 300 GWh for training.
Consumes Costa Rica's yearly
energy in a week.

What should I do with a mug that has no bottom
and the top is sealed?



Flip it.



Be god.



Flip it.

What should I do with a mug that has no bottom and the top is sealed?

Fast answer

You've essentially described an object that can't do the one thing a mug is supposed to do—hold anything.

- **No bottom** → anything you put in falls out
- **Top sealed** → you can't even put anything in

So functionally, it's useless as a mug. At best, it's:

- A **paperweight** (if it has some heft)
- A **decorative object**
- A **conversation piece** ("look at this impossible mug")

Or, more bluntly: it's just a defective mug, so the most practical option is to recycle or toss it.

Unless this is a riddle—then the answer is simply: **you can't use it for anything as a mug.**

How can I turn into a mug?

Biological organisms cannot be transformed into ceramic or inanimate vessels. This is physically impossible under current biological and material science frameworks.

Human biology requires an open thermodynamic system for homeostasis, involving continuous matter and energy exchange. A mug is a static, closed-shape object that lacks the metabolic, neurological, or structural capacity to function as a living entity.

Be god.

Agenda

- I. The promise.
- II. The image of the promise.
- III. How to understand anything?
- IV. Coins and bets on the future.

Act I: The promise

A beautiful promise.

Transformers.

More data.

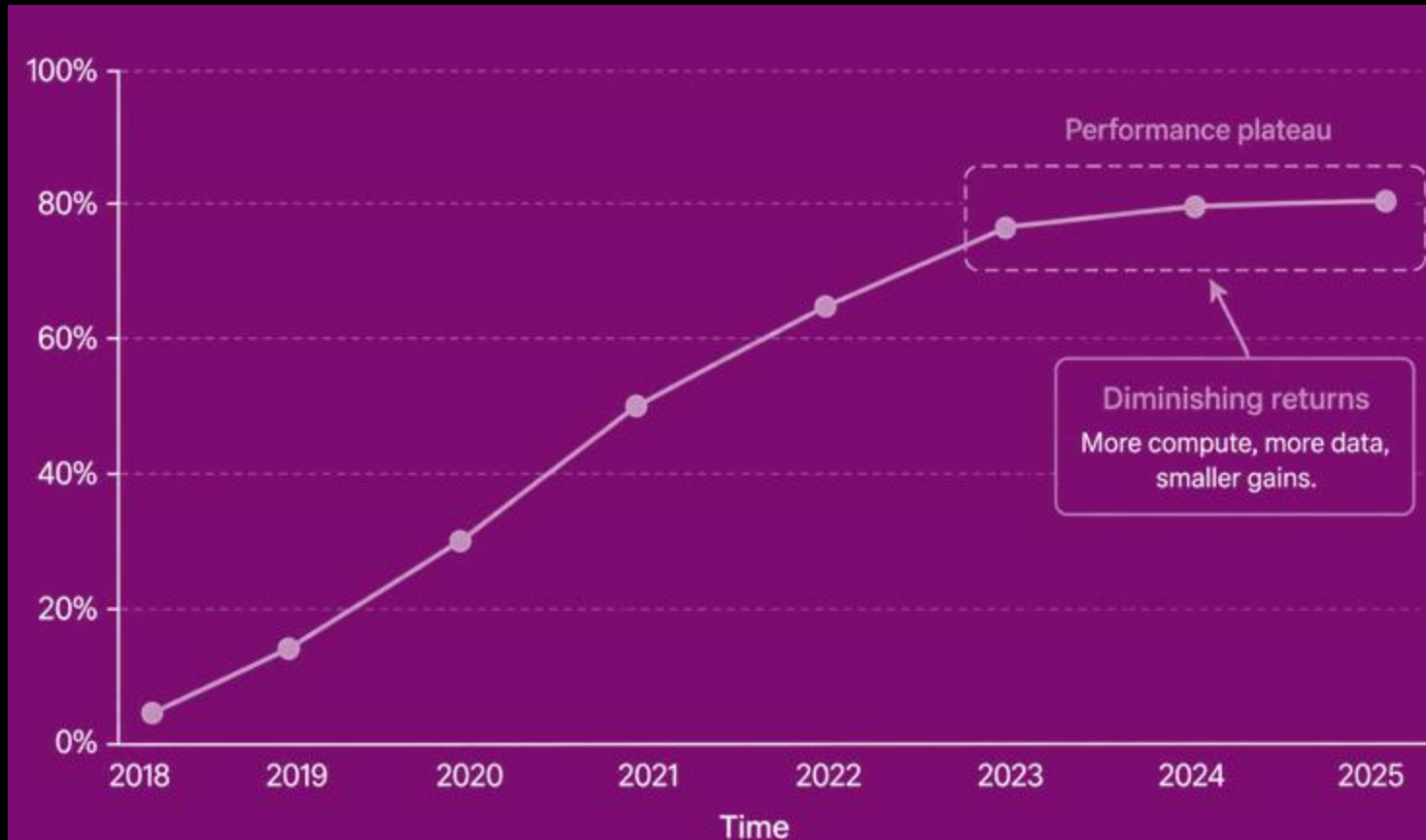
More parameters.

More compute.

Capabilities would *emerge*.

AGI would *emerge*.

The plateau



For a while, every six months, the line moved exactly where the curve said it would.

IN ONE SENSE, THE BET PAID OFF.

A *trillion-dollar* industry got built on this line.

\$1000_{B+}

Annual AI capex, 2026 frontier
labs & hyperscalers

10_x

Parameter count growth, GPT-3 →
frontier 2026

800_M

Weekly active users of frontier
assistants

What is the *issue*?

Prediction happens in *data space*.

Imagine if from birth, all you've ever seen
were books, pictures and videos.
How much would you actually know about the
world?

Meaning cannot be learned from form alone.

Few other *caveats*.

ACTUALLY

An LLM is a *conditional distribution*.

THE WHOLE OBJECTIVE

$$p(x_t | x_{<t})$$

Given previous tokens, predict the next. Everything else – reasoning, planning, knowledge – is a downstream consequence of this single objective , applied at scale, on text humans wrote.

An *autoregressive* trap.

Each token is conditioned on tokens the model sampled itself.

Small errors do not cancel they *compound*.

There is no mechanism in the architecture to revise a commitment made earlier. The model cannot backtrack at the level of meaning only at the level of tokens.

$$\mathcal{L}_{\text{LM}}(\theta) = - \sum_{t=1}^T \log p_{\theta}(x_t \mid x_{<t})$$

Hallucination is not a bug.
It is the architecture
working as designed.

The model samples plausible continuations. Sometimes the most plausible continuation is true. Sometimes it isn't.

The model has no mechanism to tell the difference because the difference lives outside the data. It lives in the real world.

FOUR ABSENCES

What an LLM doesn't have.

01 · CAUSALITY

What causes
what.

Only co-occurrence in
text. Correlations
dressed as
explanations.

02 · PHYSICS

Objects,
forces,
time.

No intuitive simulator.
The room is not in
there.

03 · PERMANENCE

Entities
that
persist.

A character on page 3
is just tokens by page
30.

04 · PLANNING

Counterfactuals.

Can continue. Cannot
evaluate. Cannot search.

Text and media are a low-bandwidth, lossy projection of the world.

A system trained only on the projection cannot recover the source.

Act II: The image of the promise











Yann LeCun ✓

@ylecun



Let me clear a **huge** misunderstanding here.

The generation of mostly realistic-looking videos from prompts **does not** indicate that a system understands the physical world.

Generation is very different from causal prediction from a world model.

The space of plausible videos is very large, and a video generation system merely needs to produce **one** sample to succeed.

NEWS FROM THE LAST SEVEN WEEKS

The Sora app shut down. The team dispersed.

CNN · 17 MARCH 2026

OpenAI is shutting down its Sora video app.

TECHCRUNCH · 8 APRIL 2026

Bill Peebles and Kevin Weil exit OpenAI.

Act III:

How to understand anything?

The true test of intelligence is not
how much we know how to do, but how
we behave *when we don't know what to
do.*

*John Holt – based on Jean Piaget
The pioneers of developmental
psychology*

A world-model *mechanically*

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

Given the state of an environment and an action , predict the next state. In practice: an encoder compresses observation into latent state, a predictor rolls it forward.

A world model is a learned function that takes the current state of some environment, and an action, and predicts the next state.

Predict. Plan. Imagine.

01 · PREDICT

What happens next?

One step forward. The minimum useful behaviour.

Trained by self-supervision.

02 · PLAN

Which action is best?

Search over candidate actions. Score each rollout. Reasoning, but as *search*.

03 · IMAGINE

Train on a future that never happened.

Generate counterfactual rollouts. Use them instead of real experience.

How to do this?

World Models: What they actually mean

	Yann LeCun (AMI Labs, ex-Meta)	Jim Fan (NVIDIA)	Fei-Fei Li (World Labs)
The bet	Work in latent space Plan over abstractions; rendering detail wastes capacity.	Generative video as the simulator Scale it into a simulator; predict next-frame physics, skip the VLA.	Simulation is the bridge Unify the planner, renderer and simulator
Predicting pixels is	A dead end	The roadmap	Fine, if geometry is explicit & checkable
Flagship model	V-JEPA 2 (Meta lineage) AMI Labs scaling it, pre- product	Cosmos 3 (omnimodel, mixture-of- transformers)	Marble (splats + collision meshes)

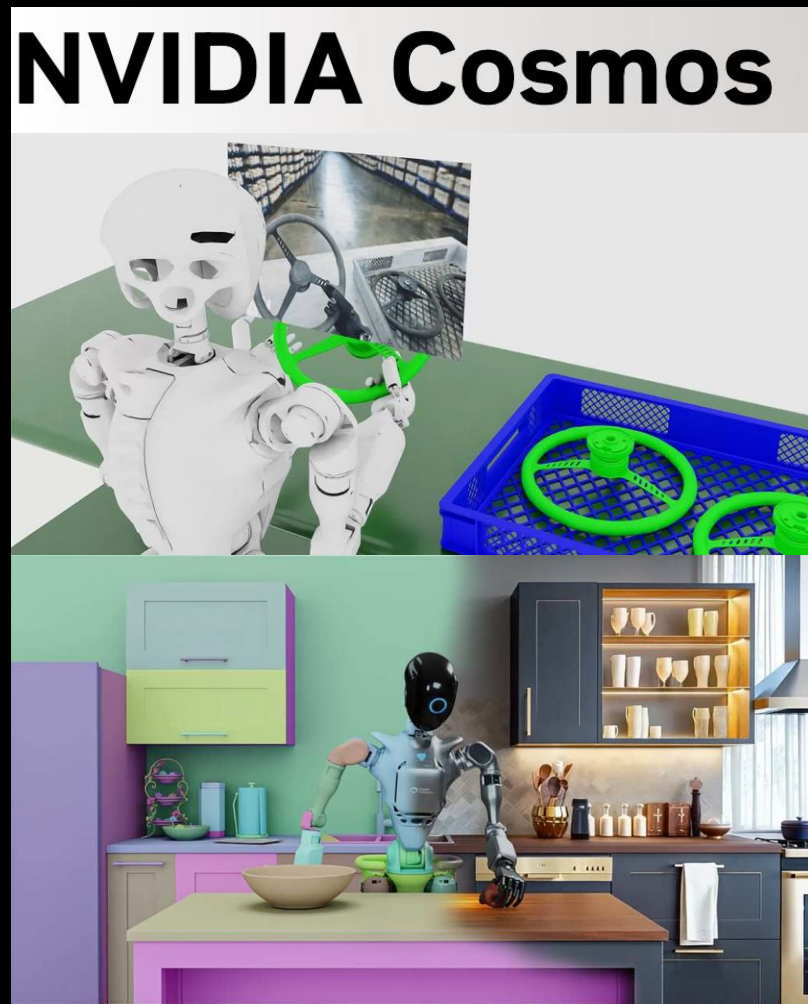
Source: *A Functional Taxonomy of World Models* (Fei Fei Li), NVIDIA, JEPA/V-JEPA 2 papers



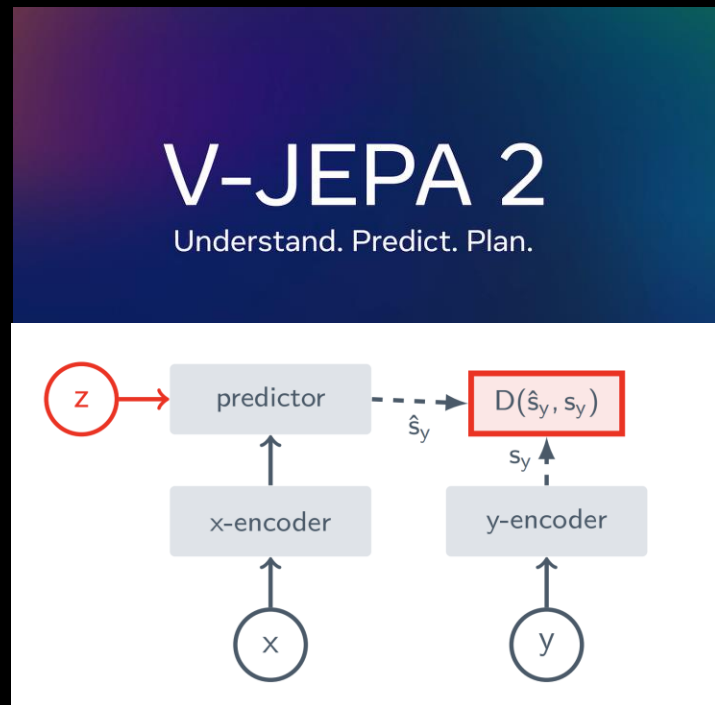
Interactive generative
world models



Diffusion Transformer video
world models



Latent predictive
architectures



Genie shipped to Google AI Ultra. Waymo built its own.

DEEPMIND · 29 JANUARY 2026

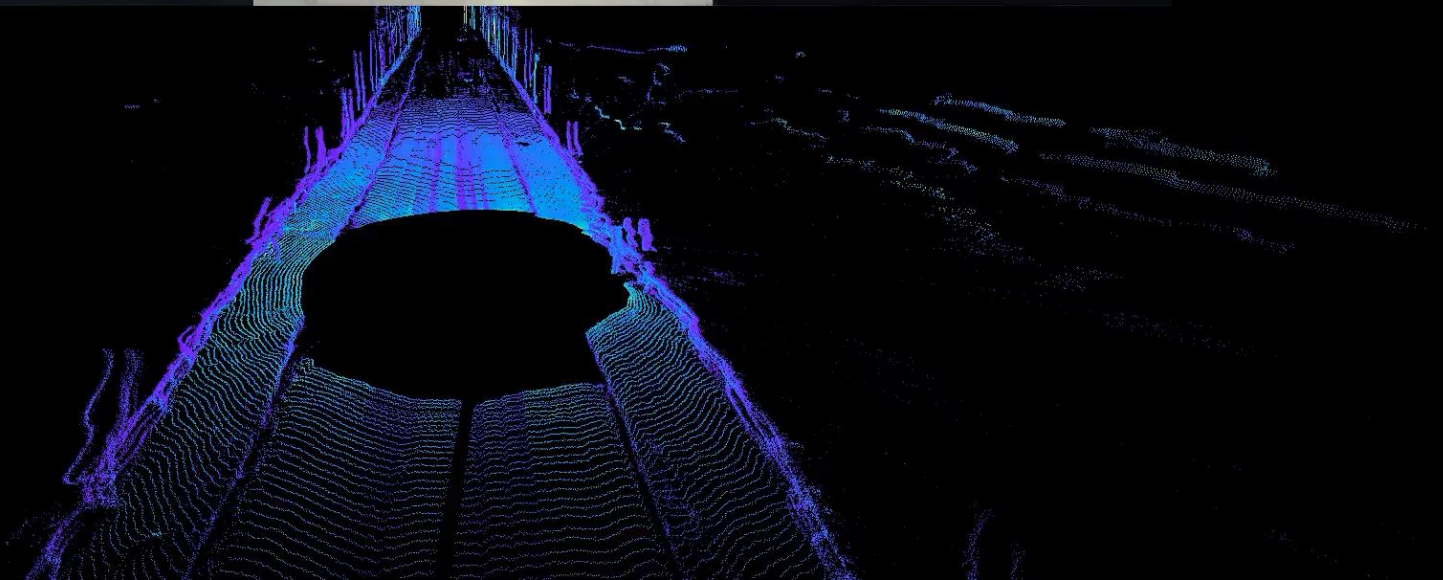
Project Genie launches to Google AI Ultra subscribers.

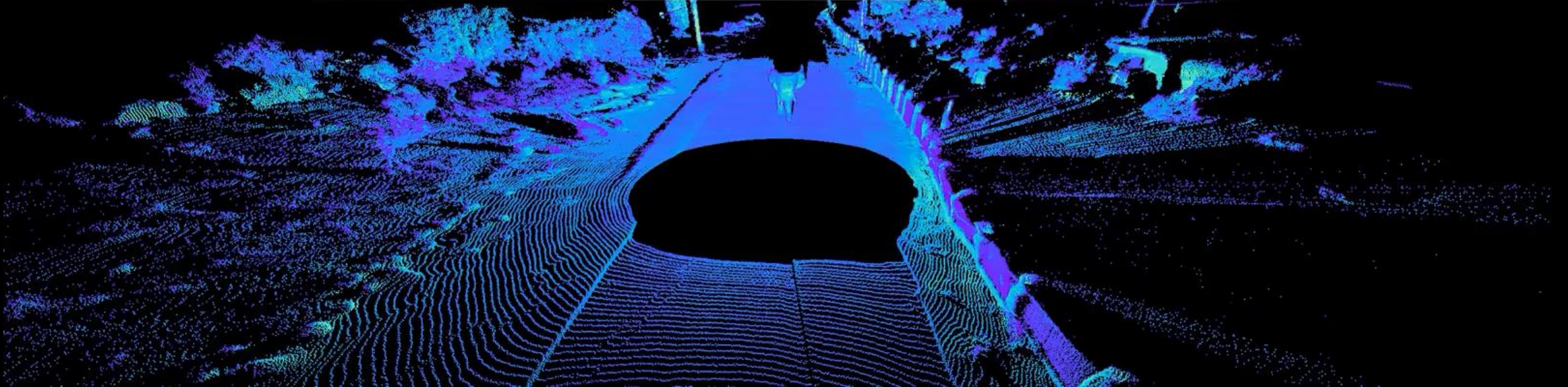
WAYMO · FEBRUARY 2026

Waymo World Model: a Genie-3-derived simulator for autonomous-driving edge cases.

WAYVE · ARXIV 2503.20523 · MARCH 2025

GAIA-2 — large-scale driving simulation goes open.





COSMOS.

Pour the cup into the box



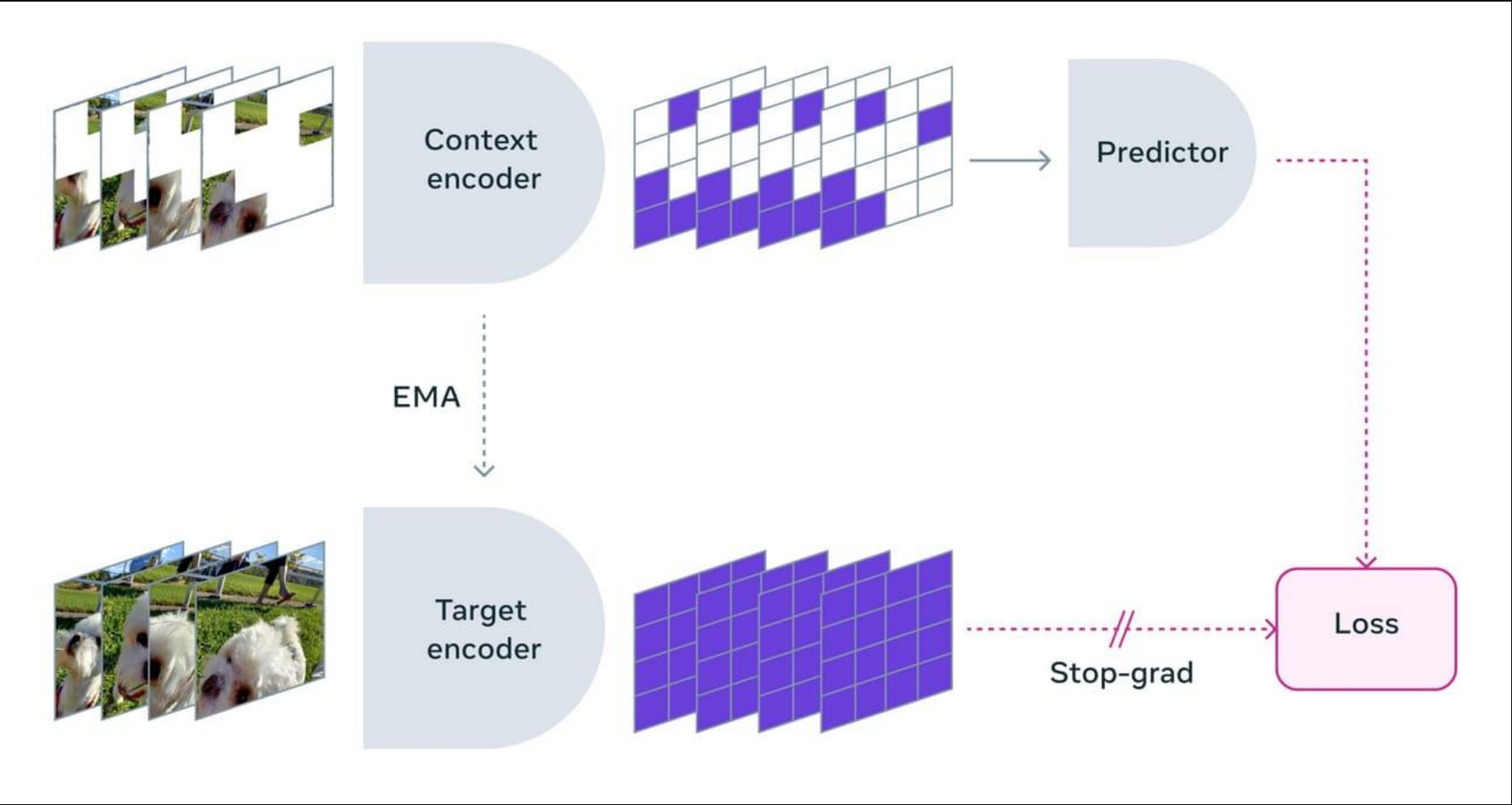
Dream



Real

JEPA.

Joint Embedding Predictive Architecture

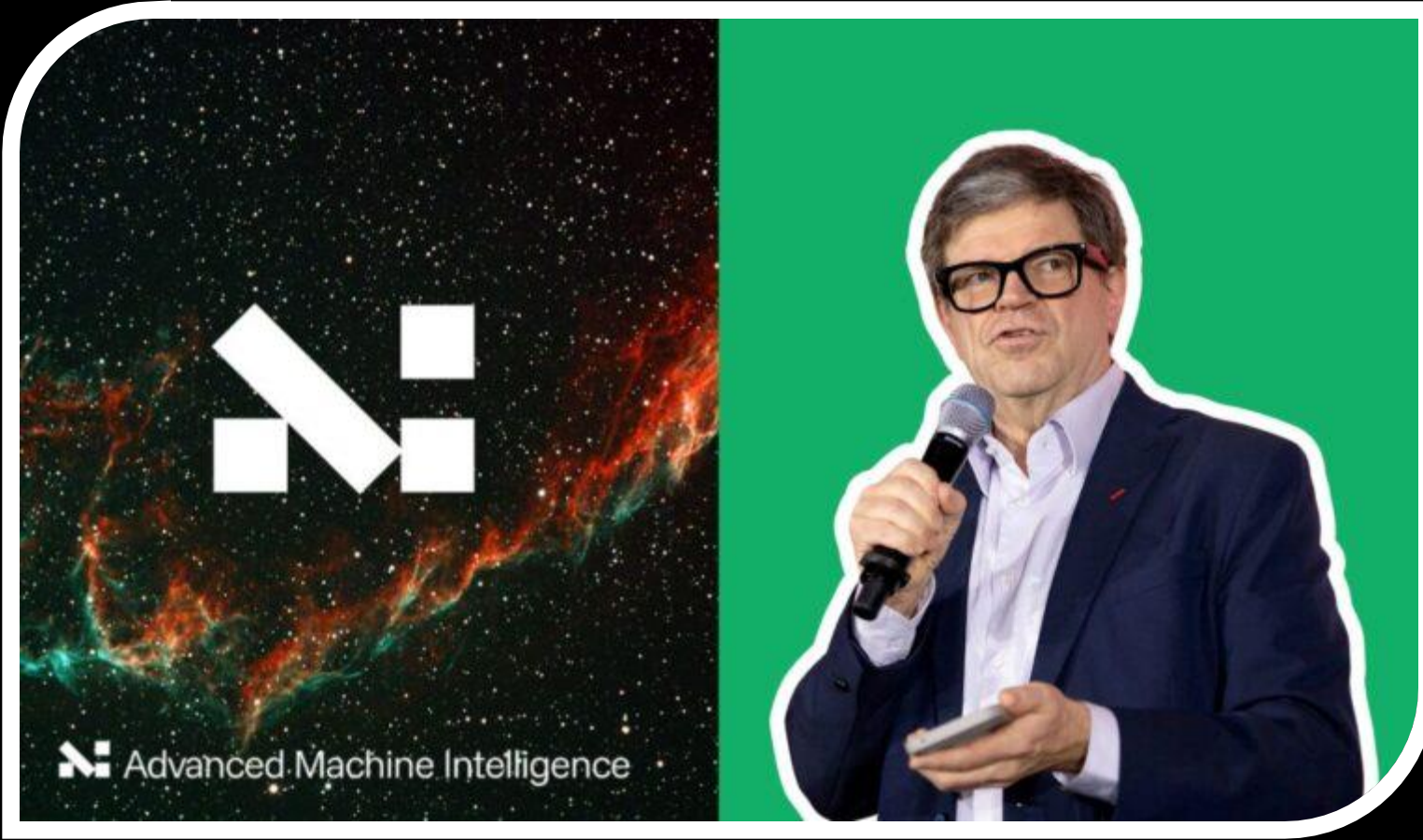


I-JEPA

V-JEPA

V-JEPA2

Le-JEPA



Yann LeCun left Meta FAIR. Founded **AMI Labs** .
Raised **\$1.03 B.**

Act IV:

Coins and bets on the future

Robotics.

Robotics is where world models cannot just look plausible.

They must work.

They function in the open-world.

They will face new challenges that were not in the training data.





Conclusions

Will LLMs vanish soon? **No**. They are extremely useful in many scenarios.

Will AGI/ASI/StrongAI emerge from LLMs? **No**. All empirical data points to this.

Will the need for safety and security and the fear of hallucination force world-models? **Yes**. The economic drive for robotics will skyrocket and they need to be secure.

Dawn always follows
the twilight.

Thank you.

Gergely Várhelyi-Tóth
Advanced Technology Scientist
@

