Position: World Models of Physical Sites Must Discover Their Own State

MOVENDI | RESEARCH

REALITY MODELS

Luc H. Leroy Maximilien Leroy
Movendi Research · research@movendi.ai

POSITION

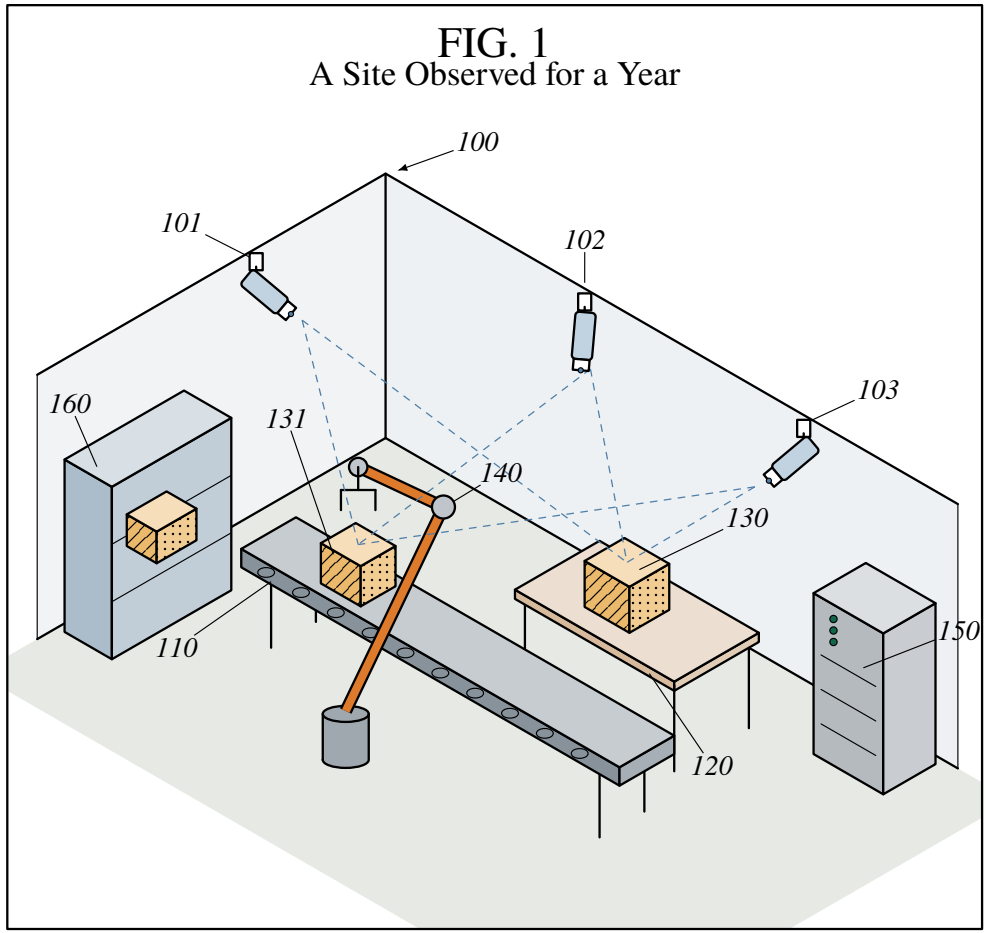
A world model of a specific physical site must discover its own state from evidence, instead of inheriting it from its sensors or from its designers.

We call such a model a **Reality Model**.

NeurIPS 2026 Workshop on World Models in Physical AI · Sydney, December 11, 2026

THE PROBLEM

A site observed for a year, not for an episode



- Cameras *101–103* get blocked and moved.
- Telemetry arrives late.
- Gripper *140* is replaced; the contact dynamics change.
- An operator's workaround enters the dynamics.

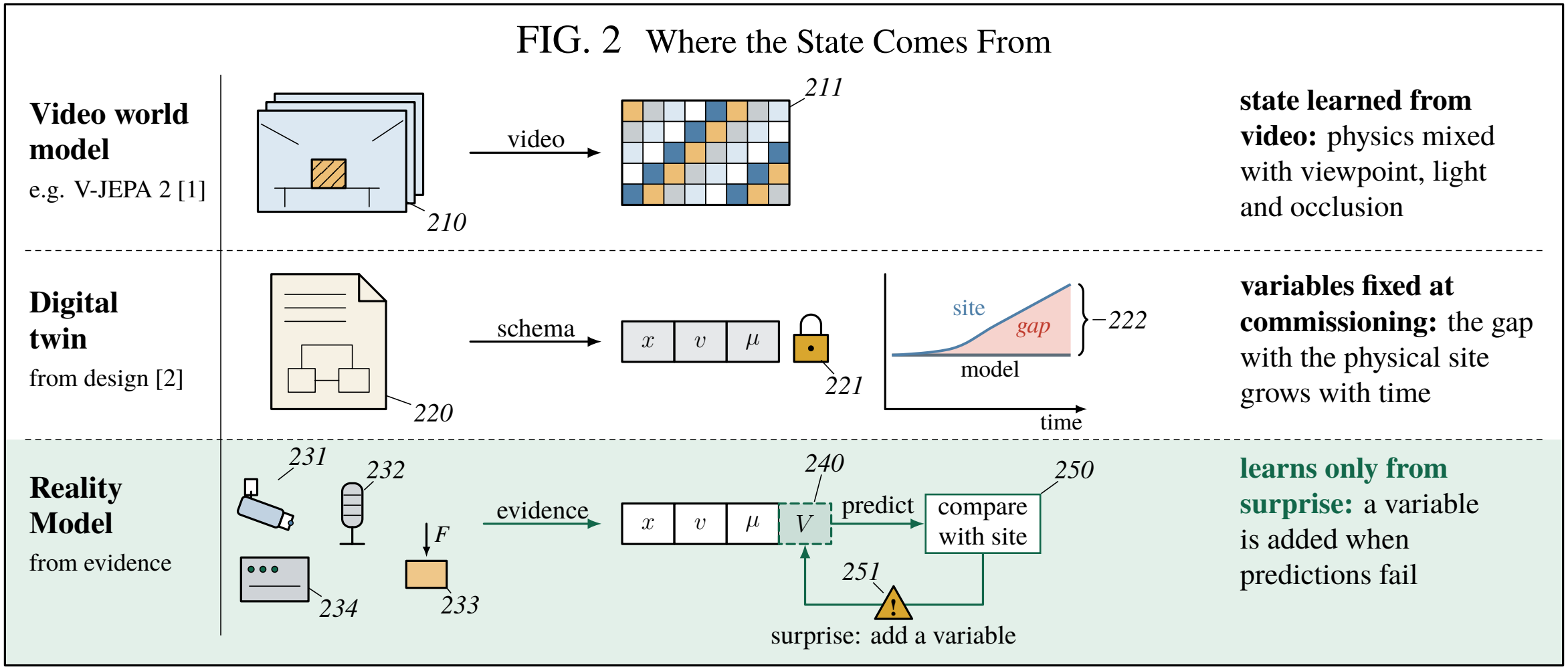
The model must tell a late measurement from a physical change, and its own state error from a bad dynamics model.

A plausible future video is not enough.

MAIN IDEA

Where does the state come from?

FIG. 2 Where the State Comes From



DEFINITION

State revision, defined by behavior

1 Detect

A residual persists that no variable explains.

2 Change

By any means: larger latent, new slot *240*, new mode, or reused capacity.

3 Pass H0

The new variable passes the H0 tests.

A model that only predicts better has adapted, not revised.

A Reality Model is defined by what it passes, not by how it is built.

H0 — did the model find the cause? A revised variable is decodable by a small probe, keeps prediction accurate under a held-out intervention, and acts causally when set in the state [3, 4]. Control: same model and budget, fixed state structure.

REQUIREMENTS

What a Reality Model requires

Sensor-open

Any time-stamped sensor. No single sensor family defines the model.

Prediction-compressed

Keep a variable because it improves prediction, not because it is easy to see.

Queryable geometry

Distance, contact normal, free space. Not appearance.

Persistent and attached

Identity survives occlusion, delay, and sensor swaps. Every prediction is checked against the site.

PROGRAM

Hypotheses

H0 Did the model find the cause? Decodable, transfers under intervention, causal in the state.

H1 Sensors exposing hidden variables improve observability, if calibration is estimated as state.

H2 They also shorten the belief: lower entropy over probe-decoded physical variables.

H3 Separate state, memory, and shared parameters beat context-only adaptation at matched context.

H4 Attributed error beats one undifferentiated predictive distribution.

H5 Multi-timescale updates recover faster from drift and maintenance, with less forgetting.

FALSIFIABILITY

What would prove us wrong

A general pre-trained model matches grounded models on persistence, calibration, and recovery.
Or models that predict better routinely fail H0's tests.
Then the term should go.

WHAT WE ASK

- Report physical-state metrics, not perceptual similarity.
- Build long-run benchmarks from instrumented sites.
- Treat calibration and timing as estimated.
- Report pretraining and site-adaptation cost apart.

THANK YOU

Questions and discussion

Paper: openreview.net/forum?id=DKNjWlwroi
Contact: research@movendi.ai

This work was funded by Movendi.

[1] Assran et al., V-JEPA 2, arXiv 2025. [2] Kapteyn et al., Nature Comput. Sci. 2021. [3] Geiger et al., NeurIPS 2021. [4] Elazar et al., TACL 2021.