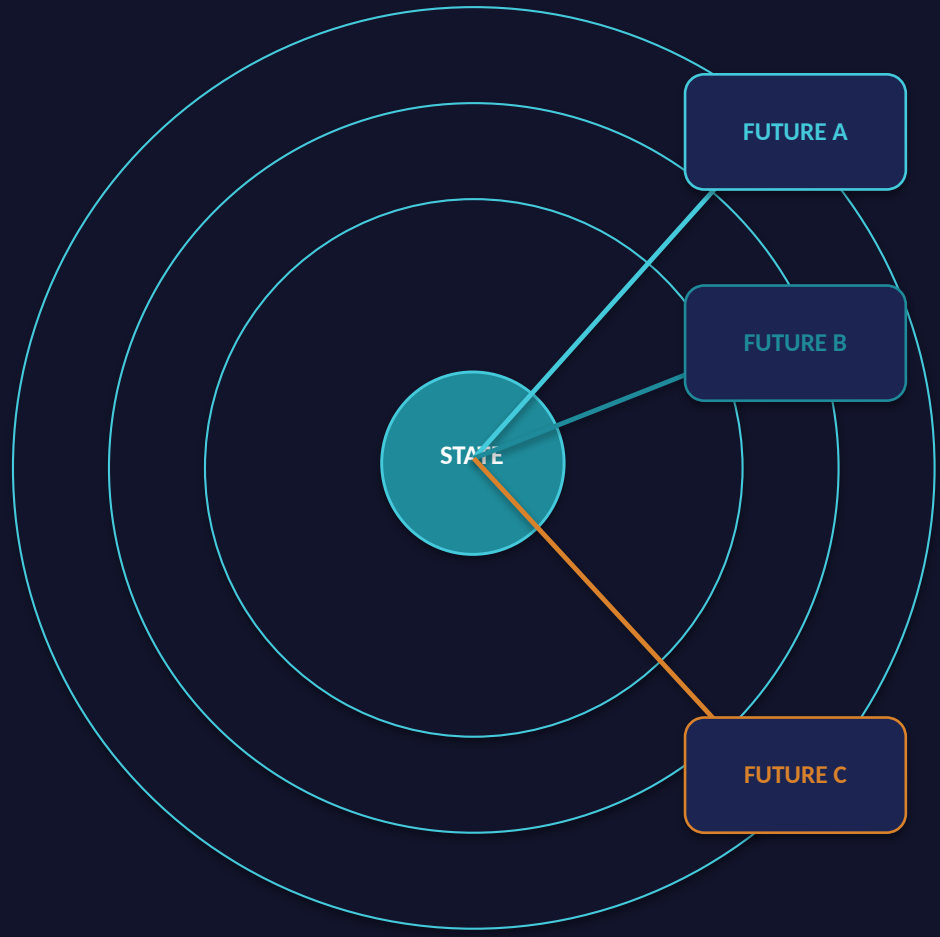
PUBLIC COMPANION · CURATED 10 SEPTEMBER 2026

World Models: The AI That Rehearses What Happens Next

A public primer on prediction, simulation, planning, and the systems required to serve them



STATE + ACTION → POSSIBLE FUTURES

Curated edition · 10 September 2026

Sources accessed 2026-09-10: Ha & Schmidhuber, World Models; World Labs, Atlas; Runway, GWM Worlds 2; NVIDIA, Cosmos 3

A world model predicts consequences, not just descriptions

A learned model of an environment's dynamics that predicts future states from current state, history, and optional actions.

1

PREDICT

What happens next?

2

SIMULATE

What could happen?

3

COUNTERFACT

What if we act differently?

4

PLAN

Which action reaches the goal?

Sources accessed 2026-09-10: Ha & Schmidhuber, World Models

Three curves converged: video, scalable architectures, and faster generation

1

VIDEO

Internet video, robot logs, driving data, game traces

2

MODELS

Transformers, diffusion, latent prediction, causal tokenizers

3

SERVING

Cached decoding, shortcut objectives, high-bandwidth accelerators

Result: learned environments moved from toy control tasks to interactive and persistent worlds.

Sources accessed 2026-09-10: Ha & Schmidhuber, World Models; Google DeepMind, Genie 3; Hafner, Yan & Lillicrap, Dreamer 4

“World model” now names four different products

LATENT PREDICTIVE — Future embeddings

Planning and control

V-JEPA 2 · Dreamer

GENERATIVE VIDEO — Future clips

Synthetic data and evaluation

Cosmos 3 generator

INTERACTIVE — Action-responsive frames

Playable training environments

Genie 3 · Worlds 2 · Oasis 3

PERSISTENT 3D — Meshes / splats / scenes

Digital twins and editable worlds

Atlas · HY-World 2.0

Sources accessed 2026-09-10: Assran et al., V-JEPA 2; Google DeepMind, Genie 3; World Labs, Atlas; Tencent Hunyuan, HY-World 2.0

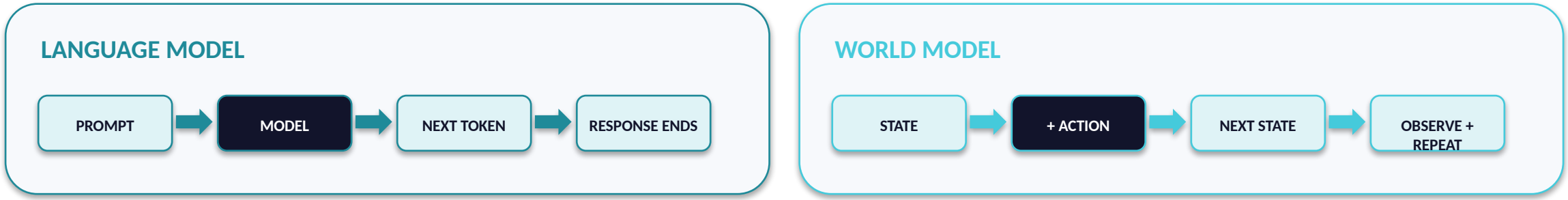
The September 2026 landscape spans previews, APIs, and open weights

MODEL	ORG	FAMILY	ACCESS	SIGNATURE
Atlas	World Labs	Omni spatial	Select-partner early access	Generate + reconstruct + simulate
GWM Worlds 2	Runway	Interactive A/V	Research preview	720p/24fps + 48kHz, company-reported
Cosmos 3	NVIDIA	Reason + world + action	Open models/tools	16B Nano / 64B Super
HY-World 2.0	Tencent	Persistent 3D	Open code + weights	Meshes / 3D Gaussian splats
Oasis 3	Decart	Physical-AI interactive	API preview	Action-conditioned multi-view worlds
Genie 3	Google DeepMind	Interactive	Experimental research prototype	720p, 20–24fps; minutes

Vendor/lab-reported unless independently reproduced · Lucy is not the Decart world-model entry.

Sources accessed 2026-09-10: World Labs, Atlas; Runway, GWM Worlds 2; NVIDIA, Cosmos 3; Tencent Hunyuan, HY-World 2.0; Decart, Oasis 3; Google DeepMind, Genie 3

Next-token prediction is open-loop; world modeling closes the loop



DIMENSION	LLM	WORLD MODEL
Target	symbol	environment state
Feedback	optional tool result	continuous consequence
Time	sequence order	metered dynamics
Unit	completion	session or rollout

Sources accessed 2026-09-10: Ha & Schmidhuber, World Models

Every implementation reduces to a state–action–future loop

ENCODE

Compress observations into state

PREDICT

Roll dynamics forward

SCORE

Estimate goal, value, or plausibility

ACT

Choose an intervention

OBSERVE

Correct the imagined trajectory

Feedback corrects the imagined trajectory after every real observation.

Sources accessed 2026-09-10: Ha & Schmidhuber, World Models; Assran et al., V-JEPA 2

The central design fork is whether anyone needs to see the future

GENERATIVE

frames → video tokens → dynamics → decoded frames

OPTIMIZES FOR

Watchable, playable, visually inspectable

TRADE-OFF

Rendering consumes capacity; drift is visible

LATENT

frames → encoder → dynamics → future embeddings

OPTIMIZES FOR

Planner-friendly, compact rollouts

TRADE-OFF

Harder to interpret; collapse must be prevented

Sources accessed 2026-09-10: Assran et al., V-JEPA 2; Google DeepMind, Genie 3

Reality needs a learned compressor before it can become context

TEXT

Discrete symbols already exist

Tokenizer is cheap and nearly lossless

VIDEO

Space × time × multiple sensors

Neural tokenizers decide what survives

WHAT THE COMPRESSOR CAN LOSE

- **fine motion lost**
- **geometry distorted**
- **rare objects erased**
- **history exceeds context**

Sources accessed 2026-09-10: Hafner, Yan & Lillicrap, Dreamer 4; Google DeepMind, Genie 3

Passive video teaches regularity; action data teaches causality

PASSIVE VIDEO	appearance, motion, recurring structure
ACTION-PAIRED DATA	control signals aligned to state changes
TASK SIGNAL	goals, rewards, success and failure

Instrument controls, interventions, and timestamps—not only outcomes.

Sources accessed 2026-09-10: Assran et al., V-JEPA 2; Hafner, Yan & Lillicrap, Dreamer 4

Memory is not a longer prompt; it is a consistency system

FAILURE MODES

- geometry drifts
- objects respawn
- causes are forgotten
- camera returns to a different room

ARCHITECTURAL RESPONSES

COMPRESS

retain longer latent histories

RETRIEVE

bring back state tied to place or transition

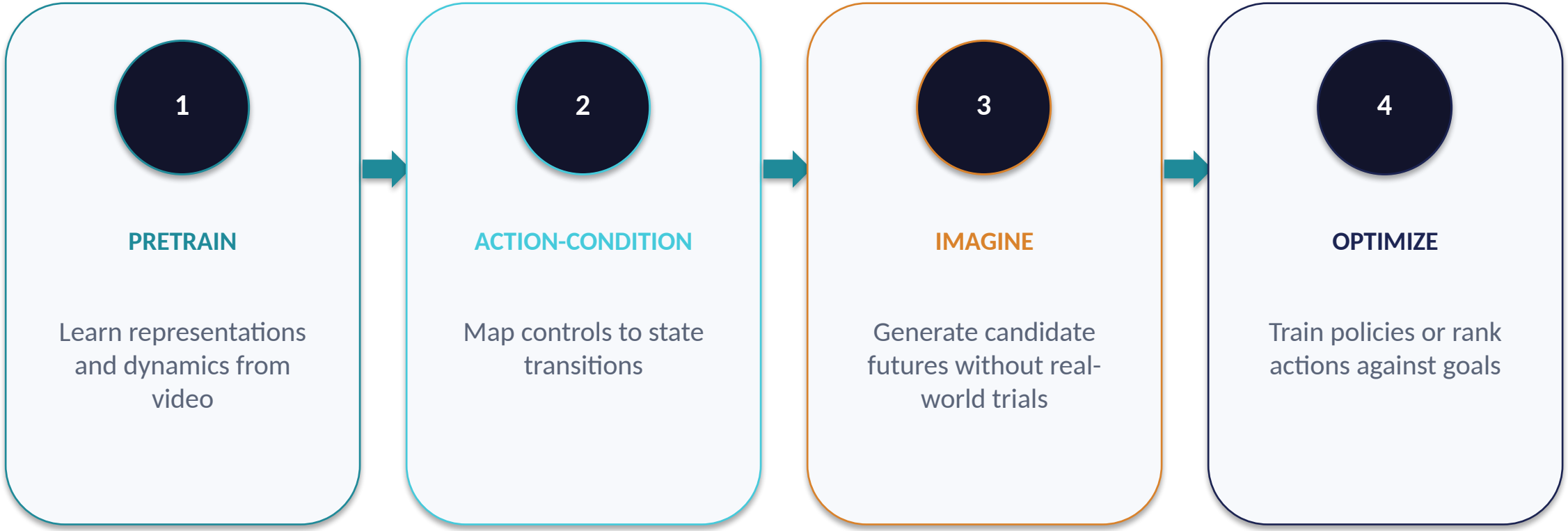
PERSIST

store the world as an editable 3D asset

Metric to track: Consistency horizon

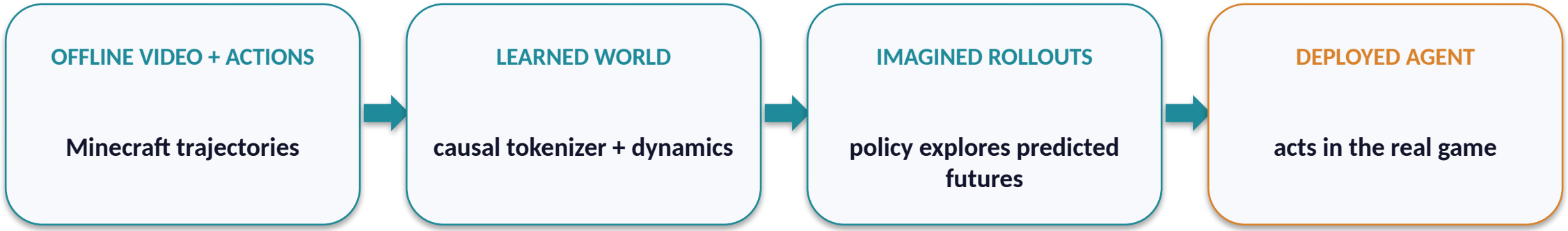
Sources accessed 2026-09-10: Google DeepMind, Genie 3; MemWM; Tencent Hunyuan, HY-World 2.0

Training adds a causal curriculum after representation learning



Sources accessed 2026-09-10: Assran et al., V-JEPA 2; Hafner, Yan & Lillicrap, Dreamer 4; NVIDIA, Cosmos 3

Dreamer 4 demonstrates the core promise: train behavior inside the model



First reported agent to obtain Minecraft diamonds from offline data alone

Paper-reported, domain-specific result; speedup multipliers omitted pending support.

Sources accessed 2026-09-10: Hafner, Yan & Lillicrap, Dreamer 4

Interactive worlds are sessions, not completions

LLM COMPLETION

- **request → response**
- elastic latency
- state can be recreated

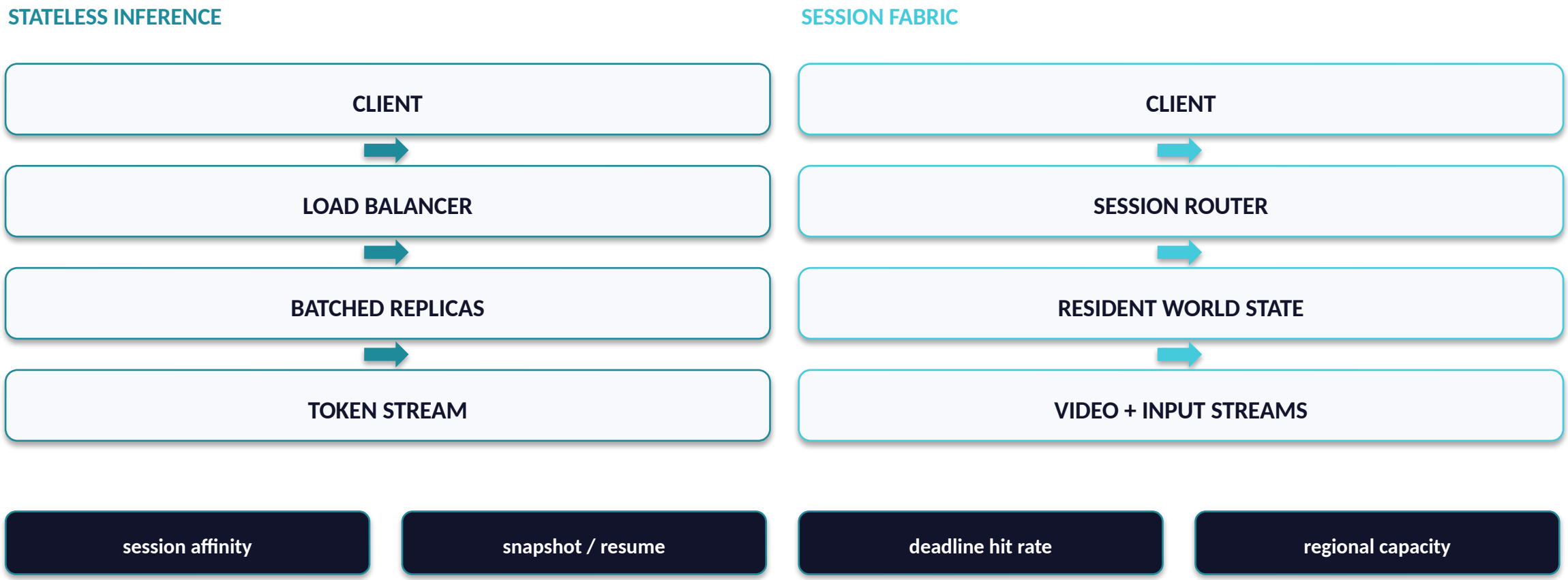
WORLD SESSION

- **input ↔ continuous frames**
- deadline-driven latency
- state must stay resident

At 24 frames per second, each frame has a 41.7 ms cadence—an output rate, not a guarantee that the entire motion-to-photon loop fits inside it.

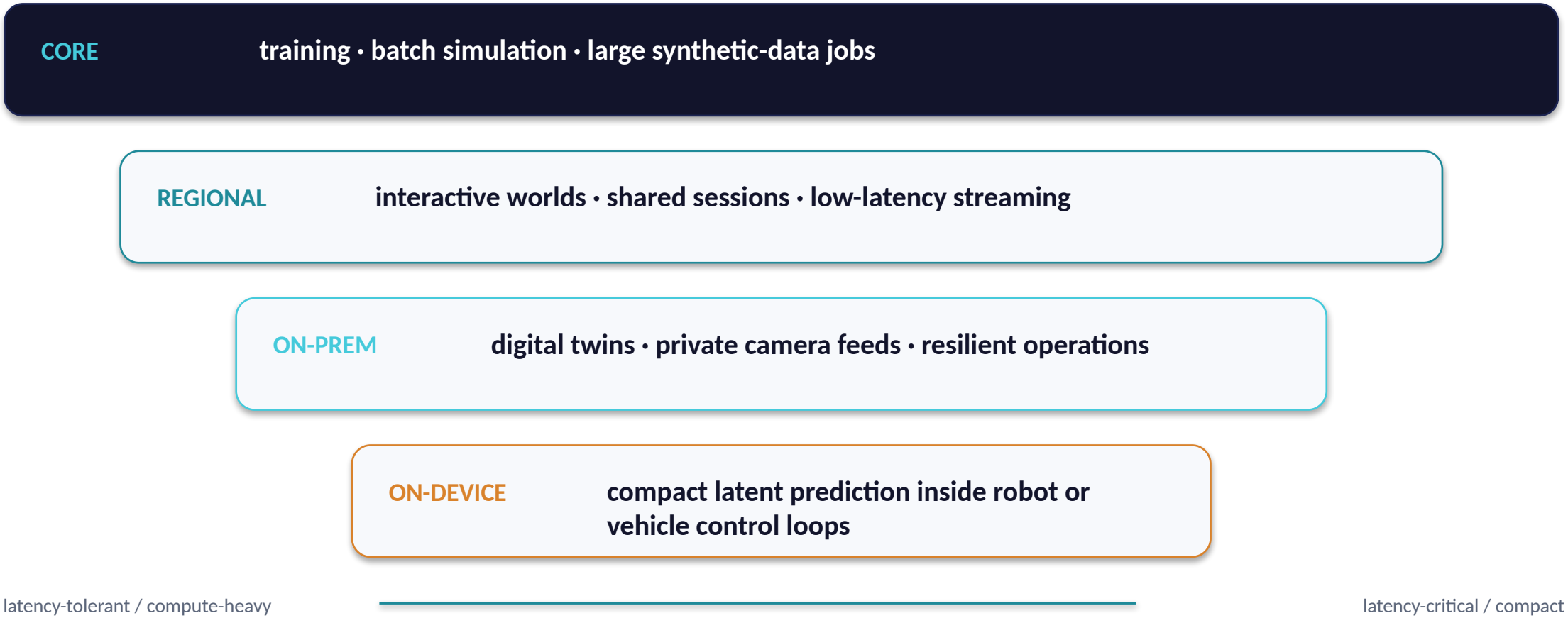
Sources accessed 2026-09-10: Google DeepMind, Genie 3; Runway, GWM Worlds 2

Serving becomes a stateful session fabric



Sources accessed 2026-09-10: Runway, GWM Worlds 2; Google DeepMind, Genie 3

Cloud imagines; edge acts



Sources accessed 2026-09-10: NVIDIA, Cosmos 3; Assran et al., V-JEPA 2; Tencent Hunyuan, HY-World 2.0

The interface changes from prompting a model to inhabiting a state

SEED

text · image · video · scan

ENTER

camera and agent take a viewpoint

ACT

keys · controls · robot commands

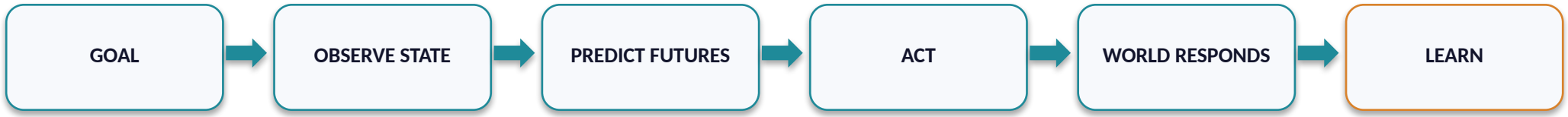
BRANCH

rewind and compare alternatives

The prompt creates the initial conditions. Interaction creates the product.

Sources accessed 2026-09-10: Runway, GWM Worlds 2; Google DeepMind, Genie 3; World Labs, Atlas

Agents can rehearse in generated worlds—but transfer remains unproven



EVIDENCE

DeepMind demonstrated SIMA 2 orienting and acting toward goals in unseen Genie 3 environments.

BOUNDARY

This is a research demonstration in virtual worlds, not evidence of broad physical-world transfer or safety.

Sources accessed 2026-09-10: Google DeepMind, SIMA 2; Google DeepMind, Genie 3

Planning spends inference compute comparing futures

1. Encode current state
2. Sample candidate action sequences
3. Roll each sequence through the world model
4. Score predicted outcomes against the goal
5. Execute only the first action; observe and repeat

replan after every executed action

V-JEPA 2-AC

V-JEPA 2-AC shows this pattern on Franka-arm tasks, but uses image goals, is sensitive to camera placement, and has no pixel decoder.

Useful evidence ≠ universal control

Sources accessed 2026-09-10: Assran et al., V-JEPA 2

The first value pools replace expensive trials and handcrafted simulators

ROBOTICS

policy rehearsal and synthetic edge cases

AUTONOMOUS SYSTEMS

rare traffic and sensor scenarios

GAMES + MEDIA

playable generated environments

DIGITAL TWINS

persistent, branchable operational scenes

DESIGN + VFX

editable 3D assets and camera-controlled worlds

TRAINING

hazardous or costly procedural practice

Sources accessed 2026-09-10: NVIDIA, Cosmos 3; Decart, Oasis 3; World Labs, Atlas; Tencent Hunyuan, HY-World 2.0

Cosmos 3 collapses reasoning, generation, and action into one open stack

REASONER

interpret text, image, and video; analyze physical context

GENERATOR

produce video, ambient sound, and action sequences

Nano · 16B

Super · 64B

Edge · announced / coming

Open models, code, tools, Hugging Face artifacts, and NIM deployment paths

Sources accessed 2026-09-10: NVIDIA, Cosmos 3

Document RAG can inform a world model; it cannot supply the physics

DOCUMENT RAG

- **retrieve text**
- add instructions or facts
- useful for digital procedures

ENVIRONMENT GROUNDING

- **capture state + actions**
- learn or retrieve transitions
- anchor geometry and live telemetry

CONDITION	RETRIEVE STATE / TRANSITIONS	POST-TRAIN	GROUND IN A PERSISTENT TWIN

RAG is a component, not the data strategy.

Sources accessed 2026-09-10: Yang et al., Retrieval-Augmented World Models; MemWM; Tencent Hunyuan, HY-World 2.0

A credible evaluation asks whether the rehearsal changes real outcomes

PLAUSIBILITY

Do objects and motion obey constraints?

CONTROLLABILITY

Do different actions produce appropriate divergence?

CONSISTENCY

How long does state remain coherent?

UTILITY

Does an agent trained or tested inside perform outside?

RISKS TO HOLD BESIDE EVERY SCORE

simulation bias · coverage gaps · unsafe transfer · privacy / likeness · vendor-reported benchmarks

Sources accessed 2026-09-10: Assran et al., V-JEPA 2; Hafner, Yan & Lillicrap, Dreamer 4; Google DeepMind, Genie 3

The emerging stack gives language a physics engine

LANGUAGE goals · instructions · explanations

WORLD state · dynamics · counterfactual futures

ACTION policy · control · feedback

The important shift is not from text to video. It is from answering about the world to rehearsing what happens next.

Sources accessed 2026-09-10: World Labs, Atlas; Runway, GWM Worlds 2; NVIDIA, Cosmos 3; Assran et al., V-JEPA 2