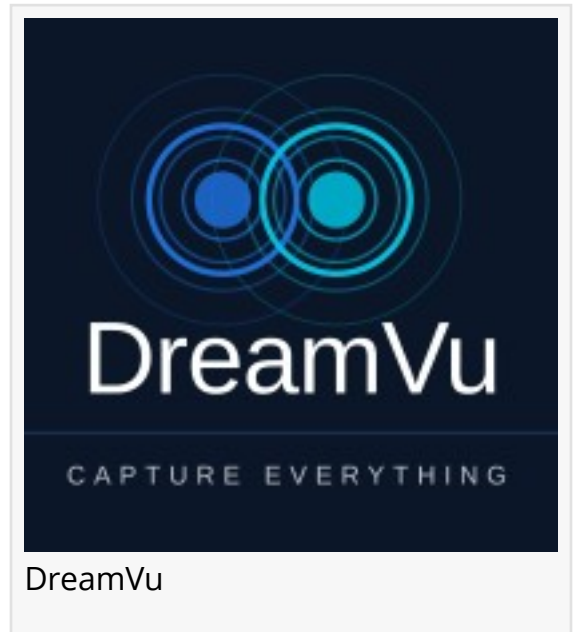


DreamVu Research on NVIDIA Cosmos3 Overturns Core Assumption in Robot Training — Overlooked Camera Angle Beats More Data

DreamVu's Cosmos3-Nano study: a single wide-view overhead camera matched or beat training on all first-person footage combined — and it used just half the data.

PHILADELPHIA, PA, UNITED STATES, July 2, 2026 /EINPresswire.com/ -- DreamVu today released [RetailSMV](#) (Retail Synchronized Multi-View), the first retail video dataset to capture real store staff at work from two synchronized perspectives at once: a head-mounted camera showing what the worker sees, and DreamVu's Alia 360° camera observing the entire scene. Alongside the dataset, the company published research — demonstrated on NVIDIA's Cosmos3 world model — with a finding that cuts against years of conventional wisdom in robotics: when adapting a world model to a real retail environment, the wide scene view — not the first-person view the field has long favored — did nearly all the work.



For years, embodied AI research has centered on egocentric data: first-person footage from cameras mounted on a person's head or a robot's body. The reason is historical — the field's benchmark tasks are manipulation tasks, grasping and handling objects up close, where the first-person view is the natural fit. Exocentric data — the third-person view of the whole scene — has been treated as a supporting signal at best.

“

First-person suited manipulation. Store robots need a model of the whole environment — that signal lives in the wide view. Our results show it's been dramatically undervalued. It's what Alia captures.”

Rajat Aggarwal, Co-Founder and CTO of DreamVu

DreamVu's results invert that hierarchy. Fine-tuning on exocentric footage alone matched or beat training on the full combined dataset on six of seven metrics — while using half the clips. For world models, which must learn how an entire environment behaves — how carts move through aisles, how fridge doors swing, how people and

objects share space — the right camera viewpoint turned out to matter more than raw data volume.

“The field standardized on first-person data because it standardized on manipulation tasks — and for a hand grasping an object, that’s the right view. But a robot working a store doesn’t just need to know what its hands are doing. It needs a model of how the whole environment behaves. That signal lives in the wide view, and our results show it’s been dramatically undervalued. It’s also exactly the view Alia was engineered to capture: the entire scene, in 3D, from a single camera.”

— Rajat Aggarwal, Co-Founder and CTO of DreamVu

A dataset built around the people robots will replace-in-task, not the customers they’ll serve. Unlike prior retail datasets, which are built around the shopper’s experience, RetailSMV captures the staff side of retail: stocking shelves, weighing produce, carrying crates, pushing supply carts, and scanning at checkout — the tasks embodied robots will actually be asked to perform.

Measured on a leading world model. Using LoRA — a lightweight fine-tuning technique that adapts a large model without retraining it from scratch — the team adapted NVIDIA’s Cosmos3-Nano world model on RetailSMV. Validation loss, the paper’s core training measure, fell 2.8X. Generated video improved on every one of 200 held-out test clips, with the statistical gap between generated and real footage shrinking by up to 33.5%.

From plausible video to deployable video. Before adaptation, the base model produced the kind of failures that make generated video unusable for robotics: hands passing through crates, fridge doors swinging through people’s bodies, aisle walls opening onto sections of the store that don’t exist. Adaptation on RetailSMV eliminated these deployment-blocking errors. The gains were largest in the 0.5-to-2-second prediction window — precisely the horizon embodied agents rely on to plan their next action and check that a policy is safe before executing it.

The paper also sets a new statistical bar for video-generation research: every result is reported with paired tests, win-rates, and p-values — a level of rigor largely absent from current literature, where results are typically reported as raw averages.

Captured on Alia. The exocentric stream at the heart of these results was recorded on Alia, DreamVu’s proprietary omnidirectional 3D camera. A single fixed unit captures the full 360° scene in stereo, in one shot, with no stitching — an optical design protected by 32+ patents and hardened over eight years of production deployment. Synchronized with head-mounted cameras on store staff, Alia produces the dual-stream, scene-plus-hands footage that RetailSMV is built from — and that, per this research, world models most need.

RetailSMV extends DreamVu’s Physical AI data platform from perception to world-model simulation: PRISM (March 2026) taught vision-language models to reason about physical scenes; SABER (May 2026) trained robot action models on real human demonstrations; RetailSMV now

grounds the world models that let robots imagine and verify what happens next. The full paper is available on [arXiv: https://arxiv.org/abs/2607.00310](https://arxiv.org/abs/2607.00310)

ABOUT DREAMVU

DreamVu builds the data infrastructure for Physical AI and humanoids. Founded from breakthrough computational imaging research at IIIT Hyderabad, DreamVu pairs its proprietary capture platform with an end-to-end annotation pipeline and simulation-ready data outputs to power the training of next-generation humanoid robots and embodied AI systems. RetailSMV follows DreamVu's SABER dataset (May 2026), which delivered a 2.19× improvement on NVIDIA GR00T N1.6, and PRISM (March 2026), which achieved a 66.6% error reduction on NVIDIA Cosmos-Reason2. The company is headquartered in Philadelphia, PA, with offices in Palo Alto, CA and R&D in Hyderabad, India.

Sanju Pillai

DreamVu

+1 267-914-5213

sanju@dreamvu.ai

Visit us on social media:

[LinkedIn](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/923877629>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2026 Newsmatics Inc. All Right Reserved.