

NVIDIA releases Cosmos3-H-Dreams surgical simulator trained on clinical procedures

Source copy kept by wheresthe.ai for the story "NVIDIA releases Cosmos3-H-Dreams surgical simulator trained on clinical procedures" (<https://wheresthe.ai/d/intl-nvidia-releases-cosmos3-h-dreams-surgical-simulator-tra-2026-10/>).

Source: <https://huggingface.co/blog/nvidia/cosmos3-h-dreams>

Text saved 20 Oct 2026 (the page as published).

SHA-256 of the saved text: d3217c9fb8eaa6df18ff23d577d20d0747075e2c2a8786eb76c352d30b5f362b

[Back to Articles](#) a]:hidden">

NVIDIA Cosmos3-H-Dreams: Bringing Clinical Surgery to Life Through Generative Simulation

Enterprise + Article Published October 20, 2026

Upvote -

Lukas Zbinden lzbinden Follow

nvidia

Javier Gamazo javirk1 Follow

nvidia

Nigel Nelson nigelN Follow

nvidia

Mostafa Toloui mtoloui Follow

nvidia

Sean Huver shuver Follow

nvidia

:last-child]:mb-0">

Picture moving the Cadiere grasper, grasping the gallbladder, and lifting it, all within a generated surgical scene that responds in real time. Today, we take the next step and introduce Cosmos3-H-Dreams, bringing interactive generative simulation to the visual and procedural complexity of clinical surgery with our most convincing in-body imagery to date.

Our first Cosmos-H-Dreams release, built on Cosmos-Predict2.5, brought interactive generative simulation to dVRK tabletop suturing, a foundational proving ground for surgical robotics. As world foundation model (WFM), Cosmos-H-Dreams demonstrated an alternative to numerical simulation: rather than explicitly modeling scene geometry, tissue deformation, and physical interactions, it learned visual dynamics directly from synchronized video and robot kinematics, yielding interactive tabletop task simulation with high realism.

Today, Cosmos3-H-Dreams advances this capability by learning from video and robot kinematics recording during clinical procedures on the CMR Surgical Versius platform. To our knowledge, this is the first real-time generative surgical simulator trained on clinical cholecystectomy recordings to support live kinematic control. Powered by Cosmos 3 and accelerated with FlashDreams, NVIDIA's accelerated streaming-inference library, Cosmos3-H-Dreams runs interactively on a single NVIDIA RTX PRO 6000 GPU.

1. From Tabletop Tasks to Clinical Scenes

In a controlled tabletop environment like that of Cosmos-H-Dreams, instruments operate against a relatively simple background. Inside the human body, however, small instrument movements can affect an entire scene: tissue stretches under traction, reflections shift, structures become occluded, and smoke changes visibility. Given an initial image and supplied tool actions, a generative simulator must capture these responses while preserving the anatomy and instruments as it generates each new block of video. Small errors can accumulate over time,

gradually changing instrument geometry or anatomical structure, making temporal consistency essential.

Visual realism is therefore a starting point for validation, not proof of biomechanical accuracy. The generated scene must also follow the controls: translation, rotation, and gripper commands should produce coherent anatomical consequences. A plausible-looking rollout that ignores those commands cannot serve as an interactive environment.

Maintaining this fidelity and responsiveness in a live control loop requires low latency generation. Cosmos3-H-Dreams addresses this setting through a model specialized for clinical cholecystectomy and distilled for real-time generation, bringing action-conditioned interaction into scenes with complex anatomy, tissue motion, and occlusion.

2. What is New in Cosmos3-H-Dreams?

2.1 A Cosmos 3 Foundation and Richer Surgical Controls

Cosmos3-H-Dreams builds on Cosmos3-H-Embodied-Nano , a 16-billion-parameter model pretrained on Open-H-Embodiment , and specializes a bidirectional teacher on curated cholecystectomy video and kinematics. Sample selection targets two visible, engaged instruments, a left Cadierre grasper and a right monopolar hook, and limits camera commands to emphasize instrument-driven interaction.

The model uses 28 action dimensions: pose and gripper information for both arms, plus energy controls, thumbstick axes and thumbstick buttons. Training uses 832 x 480 video with actions aligned at 10 Hz; that sampling rate is distinct from runtime generation throughput.

2.2 From Teacher to Four-Step and Two-Step Students

Cosmos3-H-Dreams combines a teacher-to-student training pipeline designed for long autoregressive rollouts with accelerated streaming inference through FlashDreams.

We first adapt the bidirectional teacher on 24 future frames, then extend the horizon to 72 frames, 2.4 and 7.2 seconds at 10 Hz. This gives the teacher longer sequences over which to learn surgical dynamics.

Next, distillation turns the bidirectional teacher into a causal student that can generate video in short blocks as new actions arrive. Causal teacher forcing prepares the student for sequential generation using recorded context. Self-Forcing then trains it on its own generated context, guided by the teacher and a learned critic. This helps address the accumulation of errors during interactive rollouts.

The first released student uses four denoising steps per generated block. We then train that student further with a two-step schedule to produce the second release checkpoint. The two-step model is separately trained, not simply the four-step model run with fewer sampling steps.

Fewer denoising steps results in faster inference time at the expense of lower visual quality. Both checkpoints let developers explore the balance between visual quality and inference cost.

2.3 Streaming through FlashDreams

FlashDreams supplies the accelerated streaming runtime. Long rollouts also require managing the model's attention history. The Cosmos3-H-Dreams integration uses an adapted FadeMem module that consolidates the growing key/value cache during long rollouts. This complements Self-Forcing: distillation addresses how the model learns from its generated history, while cache consolidation addresses how that history is stored during streaming generation.

3. Exploring Unseen Trajectories

During a clinical procedure there may be complications or some room for the surgeon to perform “novel” movements. Therefore, a simulator must be able to generalize beyond the training data, and produce coherent frames with unseen movements. The demonstrations below show Cosmos3-H-Dreams generating different cholecystectomy scenes conditioned on instrument actions recorded with a VR headset.

The two videos begin with the same initial frame but follow different action sequences illustrating how instrument control leads to distinct continuations of the same scene with physically accurate tissue deformations.

Furthermore, these rollouts allow us to inspect whether instruments follow the controls, whether tissue responds

coherently to interaction, and whether anatomical and instrument appearance remain consistent over time.

4. Adapting to Your Own Data and Controller

Cosmos3-H-Dreams includes four-step and two-step checkpoints for cholecystectomy. The system can be adapted to other datasets and embodiments with synchronized video and compatible action representations. To train a real-time student model on your own dataset, we provide a complete recipe for teacher fine-tuning and Self-Forcing distillation in our step-by-step guide . It describes data preparation, public teacher-training tools, and the distillation stages. Controller integrations must use the same action conventions and normalization as training.

5. What is Next

The advance in visual realism opens several research directions:

Model the full procedure. Expand coverage and temporal consistency so all phases of cholecystectomy can be explored systematically, with clear task boundaries and success criteria.

Connect generative and numerical simulation. Combine learned visual realism with explicit geometry, contact and tissue mechanics. Numerical models and measured data could help constrain and validate deformation, making simulation behavior inspectable as well as convincing.

Evaluate with surgeons. Assess anatomical accuracy, instrument behavior and tissue response, turning expert feedback into reproducible benchmarks.

Support more instruments. Move beyond two controlled tools to richer coordination, occlusion and interaction.

Develop autonomous surgical policies. Couple world action models to Cosmos3-H-Dreams in a closed loop, so policies can learn defined cholecystectomy subtasks.

These are research directions, not established clinical capabilities. Cosmos3-H-Dreams is a research simulator, not a validated system for patient care. It offers a foundation for studying surgical interaction in an increasingly realistic environment.

6. Get Started Today

Explore the models, data, and runtime behind Cosmos3-H-Dreams:

Cosmos3-H-Dreams code and examples: [GitHub repository](#)

Cosmos3-H-Dreams model: [Hugging Face checkpoint](#)

Cosmos3-H-Embodied-Nano: [Hugging Face model](#) and [GitHub repository](#)

Adapt to your own data: [Step-by-step recipe for teacher fine-tuning and self-forcing distillation](#)

Open-H-Embodiment: [Hugging Face dataset](#)

FlashDreams: [GitHub repository](#)

NVIDIA Cosmos 3: [GitHub repository](#)

Bring your controllers, your trajectories, and your research questions. We look forward to seeing what the community builds inside these new surgical worlds.

Models mentioned in this article [2](#)

Datasets mentioned in this article [1](#)

More from this author

[Scaling Indic Pretraining: 2.8 Trillion Tokens and an Open Data Generation Pipeline](#)

47 October 16, 2026

[Fine-Tune ASR for Your Domain with NVIDIA Speech Agent Skills](#)

13 October 13, 2026

[Community](#)

Edit Preview

Upload images, audio, and videos by dragging in the text input, pasting, or clicking here . Tap or paste here to upload images

Comment · Sign up or log in to comment

Upvote -

Models mentioned in this article 2

Datasets mentioned in this article 1