# NVIDIA Human-1M-100K Opens a New Path to Robot Learning

**TechRounder PDF Edition**

Live article: https://www.techrounder.com/ai/nvidia-human-1m-100k-open-robotics-video-dataset/

By Vipin PG | Published November 10, 2026 | Updated November 10, 2026 | Format: Analysis | 7 min read

## In brief

NVIDIA’s Human-1M-100K offers 100,000 hours of annotated video for robot learning, but full-scale access and independent GR00T 2 results remain unverified.

NVIDIA Human-1M-100K is a release of 100,000 hours of human egocentric video annotated with language and motion data for robotics research. The dataset is available separately from GR00T 2, NVIDIA’s robot-learning model. Researchers can inspect human demonstrations, task descriptions and motion traces, while access, governance and independent reproduction of the reported results remain open questions.

NVIDIA’s GR00T 2 research page describes an open data subset within a larger training mix. It does not by itself settle whether the full collection is easy to retrieve or whether another team can reproduce the results.

For robotics researchers and embodied-AI teams, the immediate opportunity is to examine how rich human activity might complement robot demonstrations. The immediate work is to check access, schema, rights and reproducibility rather than infer those answers from the headline number.

## The dataset is a release, not a model download

NVIDIA introduced GR00T 2 at the Conference on Robot Learning (CoRL), alongside a description of the human data used in its training. The company says the model drew on human and robot data, including 100,000 hours of open-source egocentric video.

The Hugging Face card labels the broader Human-1M collection as one million hours and identifies the published Human-1M-100K as a 100,000-hour slice. Those figures refer to the larger collection and its open subset, not competing estimates of the same release. NVIDIA’s Human-1M-100K dataset card is the most useful starting point for understanding that distinction.

That separation matters to anyone evaluating the data on its own. A dataset may be available even when model weights, training code or a complete reproduction recipe are not.

## The open slice sits inside a much larger work corpus

NVIDIA characterizes Human-1M as egocentric video, footage recorded from a person’s point of view, with dense language labels and head and hand motion. The company says the collection focuses on real-world manipulation and includes both work in the field and controlled studio settings.

The dataset card associates the 100,000-hour slice with 3,405,573 episodes, 406,815 recorded task clusters, 41,288 collector IDs, 81,557 scene IDs and nine environment categories. These are figures published by NVIDIA, not independently audited counts, and they describe the slice rather than the company’s full Human-1M collection.

NVIDIA separately lists GR00T 2’s pretraining mixture as 402,730 hours of egocentric human video, 10,184 hours of handheld UMI-gripper demonstrations across 2,897 tasks, and 23,500 hours of robot teleoperation. The released 100,000-hour set is therefore one accessible portion of the human-video data, not the whole training mixture.

## The sample folder is not the full release

Researchers can approach the material through distinct surfaces: the Hugging Face dataset card, a preview configuration, sample episode archives, an Explorer Space and a linked storage bucket. These routes are not interchangeable, and the existence of a preview or a sample does not establish that the full collection is immediately downloadable.

The sample-download directory lists three episode archives: ‘stereo_desk_wrist.tar.gz‘ at 164 MB, ‘stereo_flower_wrist.tar.gz‘ at 43.7 MB and ‘stereo_repair.tar.gz‘ at 246 MB. Their combined displayed size is about 454 MB. NVIDIA’s sample episode archive listing offers a bounded way to inspect the format, but these examples are not evidence that the entire 100,000 hours sit in that folder.

The linked bucket shows an ‘oss_100k_v1‘ directory within ‘release-staging‘. Its displayed total size and file count change over time and apply to the bucket listing, not necessarily to the dataset subset itself. The visible page does not establish complete object coverage, transfer time or whether every file can be used without further steps.

For a first evaluation, researchers should record which route they used, what files were retrievable and whether sample episodes contain the expected metadata. That small audit is more informative than treating a storage listing as proof of a practical full-scale download.

## Each episode links a goal to timed actions and motion

The dataset card defines an episode as the atomic unit: one demonstrator completing a discrete, measurable physical-work goal, usually achievable within five to ten minutes. NVIDIA calls the overall goal T1 and the timed action descriptions T2.

T2 breaks an episode into contiguous segments of one to ten seconds, describing hand laterality, verbs and objects. In the card’s "Bundle flower stems" example, synchronized stereo and wrist-camera footage is paired with six timed actions, a stereo 3D reconstruction, head and hand tracks, and robot-motion retargeting.

The field reference describes LeRobot v3.0 shards arranged by vendor, batch and version. Frame-level Parquet data and MP4 video streams connect with episode metadata, task and subtask labels, environment categories and scene descriptions. Researchers can use the LeRobot v3.0 field reference to check how these components are represented.

The schema includes camera pose, hand keypoints, tracking masks and confidence values, while metadata describes camera intrinsics and extrinsics, the parameters that capture how a camera sees and is positioned relative to a scene. Views can include monocular or stereo head cameras and optional wrist cameras. Taken together, these fields could help teams study the relationship between what a person sees, the task they describe and how their hands move.

## The dataset license does not answer model access

The Hugging Face card identifies the dataset license as Creative Commons Attribution 4.0 (CC BY 4.0) and says the data is intended for commercial and non-commercial use. That is a statement about the dataset release; researchers should still read the linked license and assess their own use case.

It does not establish the license or availability of GR00T 2's weights or source code. NVIDIA describes the model as a diffusion transformer that predicts future video and robot actions together, with smaller per-robot heads translating its plan into motor commands. The research page and linked whitepaper material reviewed for this article did not verify the brief's claimed December code-release timing, so that date should not be treated as confirmed.

This distinction prevents a common leap: an open dataset can enable investigation without making the model that trained on it reproducible. Until a current primary release notice identifies code, weights, terms and instructions, the data and model should be treated as separate offerings.

## NVIDIA reports transfer results under specific conditions

NVIDIA says GR00T 2 succeeded on 40% of 50 tasks that were not seen on the target robot, using knowledge from human video or other robots. The company also reports that five demonstrations per task raised success to 95% after a 30-minute fine-tune on one RTX 5090.

The condition matters: the reported 95% result is not zero-shot performance, and it includes both demonstrations and fine-tuning. NVIDIA's CoRL blog reports related results, 40 of 50 tasks completed at least once without task-specific demonstrations, 71% average success with one demonstration and 95% with five, but the blog and research page come from the same company and should not be counted as independent confirmation. See NVIDIA's CoRL results and robotics blog.

The research page also describes experiments transferring human demonstrations to robot tasks. These figures are company-reported evaluations. They are worth examining, but the pages reviewed here do not constitute an independent reproduction of GR00T 2's training or results.

## Independent studies support the question, not this model's results

Other research helps explain why teams are testing human video as a source of robot-learning data. It does not verify NVIDIA's model-specific claims, because the studies use different datasets, methods and evaluation setups.

The EgoVLA preprint describes a pipeline that predicts human wrist and hand actions from egocentric video, maps those actions into robot control through inverse kinematics and retargeting, then fine-tunes with a small set of robot demonstrations. Its reported experiments use a simulation benchmark. The paper itself emphasizes the gap between human and robot camera positions, hand shapes and visual appearance, making alignment part of the method rather than a solved problem. The EgoVLA preprint details that approach.

A 2026 preprint, HumanScale, reports a controlled comparison between egocentric video and real-robot trajectories for embodied pretraining. Its authors report better validation loss and task success for egocentric pretraining under their protocol. That result supports studying human video as a training source, but it does not test Human-1M-100K or GR00T 2. The HumanScale preprint reports the comparison.

## Scale leaves quality and governance questions open

NVIDIA describes curation criteria around language and timing, task visibility, diversity and 3D tracking. Such criteria are useful, but a large episode count does not show that every annotation is complete or equally reliable. A researcher should examine the field definitions, sample notes and failure cases before assuming uniform quality.

The schema includes anonymous vendor, collector and scene identifiers, as well as a ‘capture_privacy‘ field. The public pages reviewed here do not explain contributor consent, collection provenance or privacy review in enough detail to resolve those governance questions. They also do not show whether the samples are representative of all released hours across occupations, places and recording conditions.

Access is another open issue. A bucket route and sample archives are visible, but the sources do not establish full-scale download performance, complete repository coverage or the practical requirements for retrieving the full slice. Nor does the independent research reviewed here reproduce GR00T 2’s specific transfer results.

## Treat NVIDIA Human-1M-100K as a research resource to audit

Human-1M-100K gives robotics teams a concrete dataset to inspect: human video tied to task language, camera geometry and hand motion. Its potential value lies in broadening the evidence available for learning manipulation, while the schema may help researchers connect visual context to actions.

The next useful checkpoint is not another headline metric. It is a documented trial that inventories a representative part of the release, checks annotation completeness and records the actual retrieval workflow. Teams considering commercial use should also review the license and seek answers about provenance, consent and privacy before building downstream products.

For claims about robot transfer, the bar is higher: a reproducible method, accessible model artifacts and independent testing across embodiments. Until those pieces are available, NVIDIA’s results are promising company reports, not a general guarantee that human video alone will teach a robot a new skill.

### References


1. research.nvidia.com - labs / gear - https://research.nvidia.com/labs/gear/gr00t-2/
2. huggingface.co - datasets / nvidia - https://huggingface.co/datasets/nvidia/PhysicalAI-Human-1M-100K/blob/main/README.md
3. huggingface.co - datasets / nvidia - https://huggingface.co/datasets/nvidia/PhysicalAI-Human-1M-100K/tree/main/samples/downloads
4. huggingface.co - datasets / nvidia - https://huggingface.co/datasets/nvidia/PhysicalAI-Human-1M-100K/blob/main/SCHEMA.md
5. blogs.nvidia.com - blog / robotics-simulation-advancements-corl - https://blogs.nvidia.com/blog/robotics-simulation-advancements-corl/
6. arxiv.org - html / 2507.12440v3 - https://arxiv.org/html/2507.12440v3
7. arxiv.org - html / 2606.20521v1 - https://arxiv.org/html/2606.20521v1