

Is this robot data trainable?

Two public robot-learning datasets, measured before anyone trains on them. Both ship with defects their published documentation does not mention. In one, **8.3% of demonstrations never finish the task they claim to teach**. In the other, **90.8% of episodes are physically out of reach** of the arm they would be retargeted to. Every figure below is measured, and every method is stated.

Subjects `LightwheelAI/leisaac-pick-orange`, `fpvlabs/stera-10m` Measured 2026-09-09
Target embodiment `SO-ARM101`

8.3% of leisaac episodes never complete the task 5 OF 60	23.7% of leisaac frames have the arm effectively still 8,613 OF 36,293
90.8% of Stera episodes exceed the arm's workspace 184 EPISODES SAMPLED	73.5% of Stera motion recoverable by segmentation 52,978 OF 72,113 STEPS

SUBJECT 1

leisaac-pick-orange

Sixty teleoperated demonstrations of an SO-101 moving three oranges into a plate. Published by LightwheelAI under Apache 2.0. The audit reads the joint channel only, no video and no model, and runs in seconds.

EPISODES	FRAMES	DURATION	RATE	ACTION	CAMERAS
60	36,293	20.2 min	30 fps	6 joints	2 × 480p

Five demonstrations never complete the task

The environment scores success as all three oranges in the plate. Five episodes transport fewer than three. Episode 28 transports **one**, and wanders while doing it, at a path efficiency of 0.056 against a corpus median of 0.739. These episodes cannot teach the scored behaviour, and a policy trained on them learns to stop early.

Method. Gripper closures detected by Schmitt trigger over the active span, then counted as a carry only when the tool translates at least 5 cm while closed. A closure that moves nothing carried nothing.

EPISODE	ORANGES CARRIED	REGrips	LENGTH	PATH EFFICIENCY
28	1	1	17.6 s	0.056
31	2	0	15.2 s	0.199
4	2	3	20.4 s	0.505
5	2	2	18.4 s	0.708
38	2	3	17.3 s	0.772

1 orange		1 ep
2 oranges		4 eps
3 oranges		26 eps
4 carries		19 eps
5 carries		8 eps
6 carries		2 eps

The 29 episodes above three carries are not bonus work. Three oranges is the whole task, so a fourth carry means one was dropped and re-picked. Just under half the corpus contains a recovery from a failure the demonstration does not label.

FINDING 02 MATERIAL

Nearly a quarter of all frames have the arm effectively still

Across the corpus, **8,613 of 36,293 frames** move the tool slower than 2 cm per second. Per episode the median is 22.9% and the worst is 35.1%. These are operator pauses, and they are sampled at training time like any other frame, teaching the policy that holding position is a valid response to the scene.

Method. Tool speed from forward kinematics on the published URDF, differenced frame to frame at 30 fps.

ep 26		35.1% idle
ep 16		32.4% idle
ep 49		30.1% idle
median		22.9% idle
best		15.3% idle

FINDING 03 MATERIAL

Most episodes contain fumbles at the jaws

A regrip is a gripper closure that carries nothing. Only **7 of 60 episodes** are free of them. The median episode contains two, and eight contain three. Each one teaches the policy to close on empty space.

0 regrips		7 eps
1 regrip		16 eps
2 regrips		29 eps
3 regrips		8 eps

FINDING 04 ADVISORY

The published metadata describes the wrong robot and the wrong units

The dataset card presents this as real-world deployment, because `robot_type` reads `so101_follower`. It is simulation, recorded in IsaacLab with a real leader arm driving a simulated follower. Separately, the joint columns are named `*.pos` with no unit given. They are LeRobot normalised percent, not degrees, and no calibration file ships with the dataset. A team assuming degrees gets joint angles wrong by both a scale factor and an offset, with no error raised.

Method. Unit inferred from saturation. `shoulder_lift` bottoms at exactly `-100.00` and `wrist_flex` tops at exactly `100.00`, which is the signature of a normalised range rather than a physical one.

Workspace and grasp geometry are clean

Every frame sits inside the arm's 0.5297 m reach, with a maximum radius of 0.410 m, so nothing needs discarding on kinematic grounds. Grasp quality is also directly readable. The jaws close to a median of 25.2 when an orange is between them and to about 2 when empty, because the fruit blocks them, which is what makes grasp success measurable from the joint channel with no video at all.

MEASURE	MIN	MEDIAN	MAX	VERDICT
Tool radius from base	0.304 m	0.357 m	0.410 m	within 0.530 m
Transport per carry	0.082 m	0.187 m	0.251 m	consistent
Gripper value while holding	19.3	25.2	35.5	separable
Episode length	11.6 s	18.4 s	35.1 s	3.0× spread

SUBJECT 2

Stera-10M

Human egocentric video with hand reconstruction, proposed as pre-training material for robot policies. The audited slice is 143 sessions selected for manipulation content. The question asked here is whether a fixed tabletop arm can reproduce what the camera recorded.

SESSIONS	EPISODES	DURATION	SIZE	RATE	SOURCE
143	1,708	29.5 h	286.4 GB	15 fps	human hands

Nine per cent of episodes fit inside the arm's workspace

A fixed-base arm has to reach every point in a trajectory, so the binding constraint is the smallest sphere containing a whole episode. The median episode needs **1.371 m** against an SO-101 reach of 0.530 m, which is about **2.6 times** the arm. Only 9.2% of episodes fit. This is not a scaling error. A Stera episode is a person working across a kitchen for tens of seconds, and the target is a 0.53 m tabletop arm. They are different sizes of activity.

Method. Per-episode enclosing radius of the reconstructed wrist path, 184 episodes across 12 sessions. An earlier session-level statistic suggested the motion fit with 2 mm to spare. It averaged over whole sessions and was wrong, which is exactly the error a per-episode measure catches.

FRAME	MEDIAN ENCLOSING RADIUS	P95	EPISODES INSIDE 0.530 M
Camera	1.371 m	2.077 m	9.2%
World	1.792 m	5.689 m	2.2%

Most of the motion is recoverable, but only after segmentation

Cutting each episode into the longest stretches that stay inside the arm's reach keeps **73.5% of steps**. Segments run to a median of 3.7 s and a maximum of 74.7 s, and every one of the 52,978 retained steps was accepted by the retargeter. Across the full 143-session slice that is roughly **21.7 executable hours out of 29.5**.

The practical reading is that the dataset is usable and the naive read of it is not. A team that skips this step trains on trajectories the arm can never follow, and the failure appears later as a policy that does not work, with nothing pointing back at the data.

Method. Greedy maximal segmentation. Extend a window while the enclosing radius stays under the reach limit, then cut. Fixed two-second windows retain a similar 77.7% but cut through the middle of reaching motions.

STAGE	STEPS	HOURS	RETAINED
Manipulation footage in	72,113	1.34	—
Inside reach after segmentation	52,978	0.98	73.5%
Accepted by retargeter	52,978	0.98	100%

The retargeting step itself

Both datasets have to be converted into joint commands before an arm can execute them. That conversion is where a silent and expensive failure lives, so it was measured directly rather than assumed.

FINDING 08

BLOCKING

Position-only inverse kinematics points the tool the wrong way

An SO-101 has five joints against a pose's six degrees of freedom, so it cannot hit an arbitrary orientation at an arbitrary point. A solver that optimises position alone still returns a confident answer. It places the tool to within seven microns and leaves it pointing a median of **137.8 degrees away** from the commanded orientation, with a 95th percentile of 175.8, which is very nearly backwards. Nothing errors. The result is a grasp that approaches correctly and never closes on anything.

Seeding matters as much as the solver. From a cold start the same solver fails **18.7%** of reachable targets by landing in a local minimum. Seeded from the previous frame, which is what a controller running at 15 or 30 fps actually does, it is exact on every target tested.

Method. 300 randomised targets, damped least squares, fixed seed. Orientation error is the angle of the residual rotation. Tests A and D use targets generated by the arm's own forward kinematics, so every one is reachable by construction and any failure belongs to the solver.

TEST	TARGET	SEED	POSITION ERROR	ORIENTATION ERROR	WITHIN TOLERANCE
D	Reachable	Previous frame	0.005 mm	0.004°	100.0%
A	Reachable	Mid-travel	0.006 mm	0.005°	81.3%
B	Free orientation	Mid-travel	61.8 mm	59.8°	1.3%
C	Free orientation	Position only	0.007 mm	137.8°	0.0%

Row C is the dangerous one. By the position number it is the best result in the table. It is also the only one that is completely unusable, and a pipeline reporting position error alone would rank it first.

What this audit covers, and what it does not

Every number above is measured from the published data. Nothing is estimated, extrapolated from a sample, or carried over from another dataset, except where a figure is explicitly labelled as extrapolated across the full session list.

- **Covered.** Demonstration completeness, grasp and transport geometry, operator idle time, workspace feasibility against a named target arm, unit and metadata correctness, and retargeting fidelity.
- **Not covered.** Any claim about trained policy performance. No success rates appear in this document because no policy has been evaluated on these datasets yet. Training on curated subsets is in progress and will be reported separately.
- **Not covered.** Image quality, camera calibration, and language annotation accuracy, none of which were part of this pass.

The audit reads the joint channel only. It decodes no video and runs no model, which is why it completes in seconds on 2.6 MB of data and why every finding can be checked against a named episode by anyone holding the dataset.