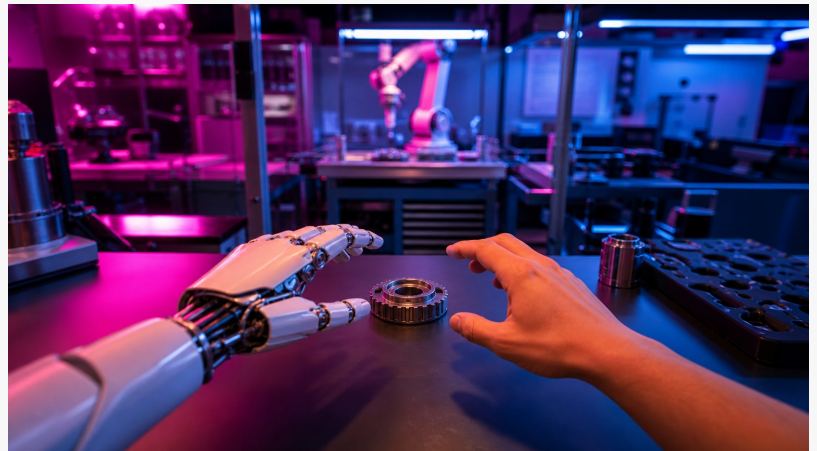


Flikforge Launches EmbodiedAI Family, a Two-Tier Training Data Family for Physical AI

Moving beyond raw video volume, Flikforge delivers rights-cleared, sensor-synchronized training data for frontier world models and advanced robotics.

SAN FRANCISCO, CA, UNITED STATES, July 9, 2026 /EINPresswire.com/ -- Flikforge, the Data Foundry for AI, today announced the Flikforge EmbodiedAI family, a two-tier curated training data family for physical AI, built to help frontier labs and robotics teams move beyond raw video volume toward rights-cleared, model-ready training signal.



Conceptual visualization of human and robot interaction data used to train physical AI systems.

The launch introduces two highly optimized data tiers:

“

Physical AI is a coverage problem, not a volume problem. The market keeps trying to sell both customers the same pile of raw footage. Flikforge EmbodiedAI gives each exactly what their models need.”

Jeff Allen, Co-Founder and CEO, Flikforge

Flikforge EmbodiedAI-500k: A broad, curated corpus of more than 500,000 hours of rights-cleared human activity video, built for frontier labs training robot foundation models and world models.

Flikforge EmbodiedAI-75k: A 75,000-hour engineered depth tier drawn from that corpus and paired with purpose-built robot-native capture, delivering more than 3 million training-ready segments engineered around 1,000+ specific robotic primitive skills for focused robotics development.

The launch comes as robotics investment accelerates and physical AI teams confront a glut of egocentric video. But

raw first-person footage alone does not solve the training data problem. Public datasets are often too small, narrow, or not commercially usable, while bulk footage can be redundant, misaligned, or missing the labels, provenance, sensor context, rights, and delivery formats

needed for production workflows. For many teams, the hidden cost is not capture; it is filtering, validating, and converting unstructured video into training data that can actually improve model behavior.



Flikforge

The Flikforge EmbodiedAI family is built on a different premise: every hour must be curated, scored, labeled, rights-cleared, and aligned to what a model needs to learn. Every segment is consented, commercially licensed, and delivered with Flikforge's patent-pending AI Rights Certificates, carrying provenance, licensing terms, and chain-of-title verification through the data layer itself.

Research supports this shift from volume to coverage. The ICLR 2025 study "Data Scaling Laws in Imitation Learning for Robotic Manipulation" found that policy generalization tracks diversity of environments and objects, not raw demonstration count alone. Follow-on research points in the same direction: task diversity and paired human-video plus robot-native data matter more than repeating the same action at scale. The emerging physical AI training pattern is layered: broad human activity video for pretraining, paired with targeted robot-native and teleoperated data for embodiment-specific fine-tuning.

"Physical AI is a coverage problem, not a volume problem," said Jeff Allen, Founder and CEO of Flikforge. "Frontier labs need breadth: the widest possible window into how humans actually work across environments, geographies, and the long tail of everyday tasks. Robotics teams need depth: every variation of the specific skills their robots must perform. Those are different products, and the market keeps trying to sell both customers the same pile of raw footage. Flikforge EmbodiedAI gives each exactly what their models need."

EmbodiedAI-500k: Breadth for Frontier Models provides the broad human-video foundation layer increasingly used in robot foundation model and world-model pipelines. Sourced through Flikforge's content partners and directed capture network, the corpus is deduplicated, quality-filtered, coverage-scored, densely labeled with customizable metadata, and delivered with full provenance and chain-of-title verification. Each hour is selected to contribute new environments, geographies, task families, demonstrators, and real-world variation.

EmbodiedAI-75k: Depth for Robotics Teams provides the engineered skill layer. Its more than 3 million segments are organized around 1,000+ robotic primitive skills, such as opening a screw-top jar, opening an appliance door, or picking a part from a bin, across Household, Commercial, Public Service, Industrial, and Entertainment & Companionship domains. For each skill, Flikforge specifies variation axes, including objects, surfaces, environments, lighting, handedness, skin tone, clutter, failure modes, recovery behavior, and task transitions. The same primitive catalog

is designed to support composed long-horizon task sequences as robots move from single actions to full jobs.

The EmbodiedAI-75k tier mirrors how leading labs train physical AI systems: a pretraining layer of coverage-scored human egocentric segments with synchronized sensor data and wrist-camera coverage for fine manipulation, and a teleoperation layer of roughly 450,000 segments with joint data, LiDAR, haptic force, and tactile feedback for embodiment-specific fine-tuning. That structure supports both broad pretraining and skill-specific policy development.

Curated Signal, Not Raw Hours

Every segment in the Flikforge EmbodiedAI family passes through Flikforge's data engineering pipeline: coverage and diversity scoring, sensor synchronization, normalization, dense sub-action indexing and labeling, and delivery in MCAP, LeRobot, or RLDS format. The result is training-ready data, not generically tagged raw video.

"The market has proven it can generate millions of hours of footage in months, and it has also proven that undesigned volume does not equal usable training data," Allen said. "Correctly selected, enriched, and packaged against what a model actually needs to learn, the same hour of video becomes something entirely different. The value was never in the pixels. It was in the fit."

Availability

EmbodiedAI-500k and EmbodiedAI-75k are available to qualified frontier labs, robotics, humanoid, autonomous systems, industrial automation, healthcare AI, defense, and world-model developers. Buyers can license the full corpus, a single tier, a domain, or a specific skill list built to their specifications.

To explore the Flikforge EmbodiedAI family and request a custom data schema, visit the [Flikforge EmbodiedAI product page](#). For additional context on the egocentric video market shift, read Flikforge's companion Field Notes analysis, "[The Great Robot Data Glut](#)."

[About Flikforge](#)

Flikforge is the Data Foundry for AI. It sources, rights-clears, segments, densely labels, and delivers video training data for robotics, autonomous systems, healthcare AI, industrial automation, defense, world models, and other video-native AI.

By combining large-scale video acquisition, dense labeling, provenance, chain-of-title verification, rights management, licensing, and delivery in a single platform, Flikforge enables AI developers to acquire commercially deployable training data with confidence, speed, and verifiable rights.

Media Contact
Susanne Estrada

info@flikforge.com
☎(415) 578-0905☎
flikforge.com

###

Susanne Estrada
Flikforge
+1 415-578-0905
[email us here](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/925394547>

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.