

MR.ROBOT

EMBODIED AI • EGOCENTRIC DATA

DATA-AS-A-SERVICE

THE ONE-STOP SOURCE FOR ROBOT TRAINING DATA

Mr. Robot AI Labs is a managed data provider for robotics and embodied AI. You define the dataset you need — we collect it, annotate it, and deliver it.

OFFERING

Data-as-a-Service

TWO PRODUCTS

Embodied + Egocentric

MODELS ARE READY. THE DATA ISN'T.

Embodied AI is gated by training data. Building the collection pipeline in-house — robots, rigs, teleoperators, annotation, QA — is a capital project most teams shouldn't run. For labs outside the major robotics hubs, it's harder still.

01

Expensive

Robots, capture rigs, facilities and trained operators are a seven-figure standing cost before a single dataset ships.

02

Slow

Standing up a collection operation takes months — hiring, tooling, and process before the first usable demonstrations arrive.

03

Complex

Teleoperation, multi-modal capture, annotation and quality control are specialised disciplines, each with its own failure modes.

04

Off-core

Every hour spent running data ops is an hour not spent on the models and products that are your real advantage.

DATA-AS-A-SERVICE: YOU SCOPE IT, WE SHIP IT

No infrastructure to build, no operators to hire. Define the dataset — we handle collection, annotation, and delivery.

PRODUCT 01

Robot-generated

EMBODIED AI TRAINING DATA

Robot demonstration data, collected to your specification — teleoperated manipulation and task performance on humanoid hardware, annotated and structured for policy training.

AgiBot G2 · Robot Trainers · Genie Studio

PRODUCT 02

Human-generated

EGOCENTRIC DATA

First-person human-perspective data, collected to your specification — real tasks seen the way people see them, to train models on natural human behaviour.

Wearable & handheld capture · Flexible methods

START WITH A PILOT Tell us the task you need data for — we collect a sample to your specification so you can test it against your model before committing.

A DATA COLLECTION CENTER, BUILT TO SCALE

Mr. Robot AI Labs runs a purpose-built data collection center (DCC) — scaling robot-based collection, with egocentric data alongside it.

SCALING TO

20

AgiBot G2 units

Humanoid robots for embodied demonstration collection as we scale the floor.

DESIGNED FOR

24/7

Continuous by design

Built for around-the-clock collection as the floor scales up.

DATA STREAMS

2

Embodied + Egocentric

Robot-generated training data and first-person human data, under one roof.

MULTI-SCENARIO

The DCC can be configured to replicate the environment your robots work in — retail, pharmacy, warehouse and others. Whatever scenario you need, we can stage and collect for it.

EMBODIED AI TRAINING DATA

Robot demonstration data, collected to your specification on humanoid hardware and structured for policy training.

AGIBOT G2 · TELEOPERATED CAPTURE



WHAT IT IS

Teleoperated manipulation and full-task demonstrations — the ground-truth examples robots learn from.

HOW IT'S COLLECTED

AgiBot G2 humanoids, operated by trained **Robot Trainers** via teleoperation in our data collection center.

HOW IT'S ANNOTATED

Segmented, action-labelled and quality-graded in **Genie Studio**, our annotation pipeline.

WHAT IT TRAINS

Manipulation policies, vision-language-action (VLA) models, imitation & reinforcement learning for real-world tasks.

REQUEST TO DATASET, FULLY MANAGED

DEFINE — CONFIGURE — COLLECT — ANNOTATE — DELIVER ●

STEP 1

CLIENT REQUEST

Tasks, volume, format and acceptance criteria scoped together.

→ Dataset spec

STEP 2

DCC SETUP

Scene, hardware and task rig configured to the spec in our center.

→ Configured cell

STEP 3

TELEOP COLLECTION

Robot Trainers teleoperate AgiBot G2 to capture demonstrations at volume.

→ Raw episodes

STEP 4

GENIE STUDIO ANNOTATION

Segmentation, action labels and quality grading with human review.

→ Labelled data

STEP 5

DELIVERY

Structured dataset delivered in your format, against the agreed spec.

→ Training-ready

EGOCENTRIC DATA

WHAT IT IS

The world as a person sees and acts in it — first-person video of real tasks and everyday activity.

HOW IT'S SOURCED

Wearable & handheld rigs, plus **potential partnerships** with gig-worker networks and retail stores to reach high-volume, real-world human data.

WHY IT'S VALUABLE

Natural human behaviour at scale — far cheaper to capture than robot demonstrations, and rich in intent.

WHAT IT TRAINS

Human-action understanding, intent & affordance models, and large-scale pretraining for embodied and assistive AI.

First-person human-perspective data — natural human behaviour, collected to your specification.



SAME MANAGED MODEL, METHOD-FLEXIBLE

Collection methods are still being defined —
so this flow stays deliberately general and
adapts to each dataset.

01

DEFINE

Scope the egocentric dataset together —
activities, environments, volume and intended
use.

→ DATASET SPEC

02

COLLECT

Capture first-person data using the method
best fit to the use case — wearable,
handheld or hybrid rigs.

→ RAW CAPTURE

03

DELIVER

Curate, annotate and quality-check, then deliver
the dataset to spec in your chosen format.

→ TRAINING-READY

HOW WE SOURCE EGOCENTRIC DATA AT SCALE

Our own collection network — first-person data captured in real commercial workplaces, not lab or simulated settings.

01

MANAGED WORKFORCE, REAL WORKPLACES

We can activate a large, managed workforce to record first-person-view data inside live commercial environments — warehouses, retail floors, F&B and similar.

02

SOUTHEAST ASIA COVERAGE

Reach across multiple Southeast Asian markets brings variety of environment, task and cultural context that existing robotics datasets underrepresent.

03

FLEXIBLE CAPTURE METHODS

Client-provided hardware, mobile-device capture, or fully managed turnkey collection — we adapt to your technical spec rather than forcing one method.

CONSENT & GOVERNANCE

Per-session worker consent, PDPA-aligned data handling and a full audit trail — egocentric data that is legally clean by design, activated at volume across real environments.

THE SCALE BEHIND THE DATA

Through our sourcing network we can activate a large, verified workforce across Southeast Asia — already recruited, trained and managed on-site. That reach is what lets us collect egocentric data at real volume.

110,000+

Verified workers in the network we can activate across Southeast Asia

440,000+

Managed field hours the network coordinates every month

30+

Real-world task and environment types the network covers

500+

Enterprise programmes the network has run at scale

Figures reflect the managed sourcing network we can activate – the collection capacity we can put behind your egocentric dataset.

FROM FIELD RECORDING TO ROBOT TRAINING

We own and govern the raw data. You receive a clean, structured dataset, ready to plug into your training pipeline.

WE OWN	WE OWN	OPTIONAL	HANDOFF	YOU
01 GOVERN & CAPTURE Workplace permissions and per-session worker consent. FPV footage recorded in real commercial settings, handed over securely. → Worker consent → Raw video	02 STORE & QUALITY CHECK Secure storage with GPS and timestamps, and a full PDPA-aligned audit trail. Batches QC'd against quality metrics. → PDPA-aligned → Audit trail	03 PRE-PROCESS & ANNOTATE QC filtering, frame extraction and annotation. Atomic task segmentation, object pose and success/fail labels to agreed spec. → QC filtering → Task metadata	04 HANDOFF TO YOU Packaged in your preferred format — MCAP, HDF5, MP4 + JSON. Licence agreement executed; you receive a clean dataset. → Licence signed → Reproducible	05 ROBOT TRAINING Dataset feeds your imitation-learning or RL pipeline. Feedback loops back to us to refine recording requirements. → Imitation learning → Feedback loop
WE OWN THE RAW DATA & GOVERN ALL CONSENT		OPTIONAL — AGREED PER PROJECT		YOU OWN WHAT IS DELIVERED

WHAT GOOD DATA LOOKS LIKE

The quality-control guidelines behind every hour of egocentric footage we deliver — quality and scale benchmarks.

DATA QUALITY

What we measure to ensure every hour of footage is usable

>85%

Usable Data Rate

Per delivery batch, after QC review and re-recording of rejected clips.

10 min

Idle Time Threshold

Maximum allowable pause within a session. Exceeded, the clip is flagged and rejected.

100%

Privacy Pass Rate

Zero PII tolerance at submission. Any face or document capture auto-rejects.

<15%

Re-record Rate

Proportion of clips flagged for reshoot before a batch is cleared for delivery.

1 week

Submission-to-Delivery

Time from worker upload to client-ready batch after QC sign-off.

HARDWARE OPTIONS FOR FPV RECORDING

Depending on the dataset and the site, we recommend or procure the right setup — assessed on data suitability for robotics and how disruptive it looks on-site.

DATA SUITABILITY

ON-SITE INTRUSIVENESS

BEST FOR



MeGo Gripper

Handheld multi-camera rig with live preview

High

Multi-camera capture with depth cues and on-device preview — purpose-built for manipulation and grasp data.

Medium

Operator-carried handheld unit. Reads as professional equipment; no headgear required.

High-fidelity manipulation and grasp datasets where framing and camera placement matter.



MeGo View

Head-mounted headset with hand-tracking cameras

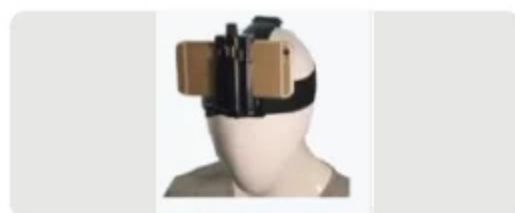
High

True first-person view plus dedicated hand cameras — natural egocentric perspective with fine hand-object detail.

Medium

Lightweight headset, worn hands-free. Visible but unobtrusive across a full shift.

Hands-free, full-task egocentric capture where the worker needs both hands free.



Mobile Phone Mount

Chest or forehead phone holder

Medium

Video quality varies by device; 0.6× ultra-wide required. Battery heats and depletes fast on long shifts.

High

Looks bulky and unusual. May make clients or customers uncomfortable.

Quick-start pilots where workers use their own devices. Not recommended for client-facing environments.

THE ENGINE BEHIND THE DATASET

A purpose-built collection facility, humanoid hardware, a trained operator team and an in-house annotation pipeline.

01

Data Collection Center

Purpose-built DCC facility

02

AgiBot G2

Humanoid collection hardware

03

Robot Trainers

Skilled teleoperation team

04

Genie Studio

Annotation & QA pipeline

Our purpose-built facility, humanoid hardware and in-house annotation pipeline are the backbone of our embodied-data collection.

BUILT FOR DATA, BY DESIGN

01



PURPOSE-BUILT FACILITY

A dedicated data collection center — not a lab retrofit. Throughput is the design goal.

02



PREFERRED AGIBOT ACCESS

A JV with AgiBot means priority humanoid hardware and AI — hard for others to match.

03



FULL DATA OWNERSHIP

You own the dataset outright — full rights, no lock-in, yours to train on however you choose.

04



HUMAN-IN-THE-LOOP QA

Every dataset is reviewed by people — quality graded, not just auto-labelled.

05



FLEXIBLE SCOPING

From a pilot batch to a standing program — scope, format and volume shaped to your need.

WHO BUYS THE DATA

WHAT IT TRAINS

01 Manipulation policies

02 VLA & robot foundation models

03 Imitation & reinforcement learning

04 Human-action & intent models

05 Benchmarking & evaluation

WHO IT'S FOR

ROBOTICS COMPANIES

Building and shipping real-world robots that need task data fast.

AI & FOUNDATION-MODEL LABS

Training embodied and multimodal models at scale.

RESEARCH INSTITUTIONS

Academic and applied robotics groups — across Europe and beyond.

MR.ROBOT

/ START WITH A PILOT

LET'S BUILD YOUR DATASET.

Start with a scoped pilot — a small dataset to prove fit and quality.
Tell us the need; we'll deliver the data.

EMAIL

eason@mr-robot.tech

PHONE

+60 12-246 4384

WEB

www.mr-robot.tech