

The Training Data Landscape 2026

99

COMPANIES MAPPED

6

SUB-SECTORS

\$3bn

TOTAL CAPITAL RAISED*

~\$3bn

OF IT, VENTURE FUNDING

76%

US-HEADQUARTERED

Funding is modest and almost entirely venture (**\$3.1bn of \$3.2bn**), striking for a layer every model depends on. The movement is upmarket: commodity labelling still leads (40 companies), but expert data and RL environments for frontier labs (25) are the fastest-emerging cluster, where value shifts from volume to expertise.

Data Labelling & Annotation Services40



AI Data Innovations



AlxBlock



Alegion



Anote

ALSO LLM Evaluation & Testing

ACQUIRED



Appen

ALSO Expert Data & RL Environments for Frontier Labs

ACQUIRED



Argilla



Aris Data



Basic AI



Centaur



Centific

ALSO Expert Data & RL Environments for Frontier Labs



Cloudfactory

ALSO AI Consulting & Implementation



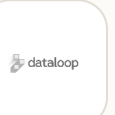
Co-One

ALSO LLM Evaluation & Testing



Dataforce

ACQUIRED



Dataloop

ALSO RAG & Retrieval Infrastructure



Dataocean



Datasaur AI

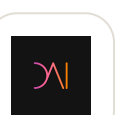
ALSO Enterprise AI Platforms & Operating Systems



Datum



Datum



Defined.ai



Digital Divide Data

ACQUIRED



FigureEight



HumanSignal

ALSO LLM Evaluation & Testing



Humyn Labs

ACQUIRED

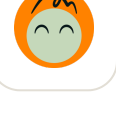


iMerit



Innodata

ALSO Expert Data & RL Environments for Frontier Labs



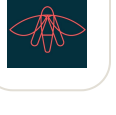
Kili



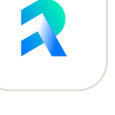
Labellerr



LXT

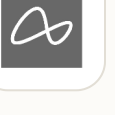


Peroptyx



Rapidata

ACQUIRED

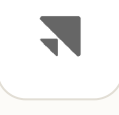


refuel



Sama

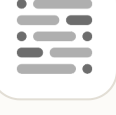
ACQUIRED



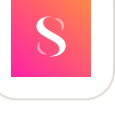
Scale AI

ALSO Expert Data & RL Environments for Frontier Labs

ACQUIRED



Shaip

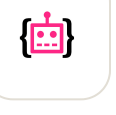


SuperAnnotate



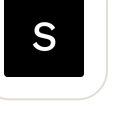
Superb.ai

ALSO MLOps & ML Lifecycle Platforms

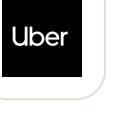


Supervisely

ALSO Computer Vision Platforms



Surge



Uber AI Solutions



Welo Data

ALSO Expert Data & RL Environments for Frontier Labs

Expert Data & RL Environments for Frontier Labs25



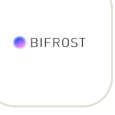
Abundant



AfterQuery



BeatpulseLabs



Bifrost



Datacurve



David AI

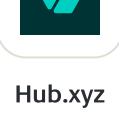


Deccan AI

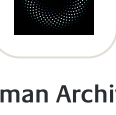
ALSO LLM Evaluation & Testing



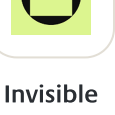
Deeptune



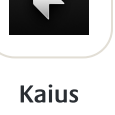
Hub.xyz



Human Archive



Invisible

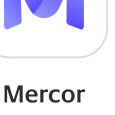


Kaius



Labelbox

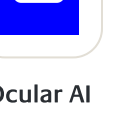
ALSO Data Labelling & Annotation Services



Mercor



Micro1

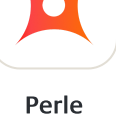


Ocular AI



Patronus AI

ALSO LLM Evaluation & Testing



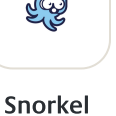
Perle



Poindexter Labs

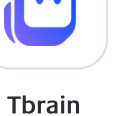


Protege



Snorkel

ALSO Training Data Curation & Management



Tbrain



Theta Software



Toloka

ALSO Data Labelling & Annotation Services

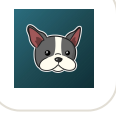


Turing

Training Data Curation & Management16



CUBIG



DagsHub

ALSO MLOps & ML Lifecycle Platforms

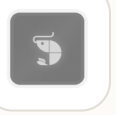


Datachain



Datology AI

ACQUIRED



Deasy Labs



DVC



Encord

ALSO Data Labelling & Annotation Services



Graviti



K2View

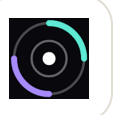
ALSO Synthetic Data Generation



Lightly



Luel



Origin Lab



Sahara Labs

ALSO Agent & Workflow Orchestration Frameworks



Structured Labs



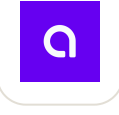
Voxel51

ACQUIRED

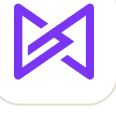


Xethub

Synthetic Data Generation11



Aindo



DataMasque



DiffuseDrive

ACQUIRED



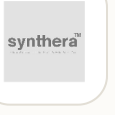
Gretel

ACQUIRED



Hazy

ACQUIRED



Mindtech

ACQUIRED



Mostly



Secludy

ALSO AI Data Privacy & Confidential Computing



Simmetry.ai



Tonic



Ydata

ALSO Data Quality & ML Monitoring

Document Processing & Extraction (IDP)5



Hercules



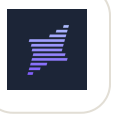
Knowledgegator



Landing AI



Super.ai



Upstage

ALSO Foundation Model Labs

Hyperscaler / SaaS Giant2



Amazon Web Services



Google

*Funding from Crunchbase, 80 of 99 companies matched. "Total capital raised" includes all funding types (venture, debt, post-IPO, grants); "venture funding" excludes debt, private equity, post-IPO and grant rounds. Companies may appear in more than one sub-sector. Logos via logo.dev.

/aipeople

© AI People 2026 • AI People proprietary research • Training Data