

Data Annotation and Labelling Services for Training AI Models

Image, video, text and audio labelled to clear guidelines and checked in review, so your models learn from consistent, accurate training data.

OVERVIEW

Data Annotation & Labelling

We label images, video, text and audio for teams training AI models: boxes and masks, entities and intents, transcripts and ratings of LLM answers.

A model is only as good as its labels, so quality control is built into every batch. Each project starts with a pilot set you review, then moves to full volume with second-pass checks, agreement scoring between annotators and a gold set of known answers that catches drift early.

We work in tools such as CVAT and Label Studio or inside your own labelling platform, and deliver in the format your training pipeline expects. Annotators sign confidentiality agreements, access is limited to your project, and sensitive data can be redacted before labelling starts.

BUILT FOR

- AI and ML teams training custom models
- Computer vision product companies
- Teams building chatbots and LLM applications
- Speech and voice product teams
- Researchers with large unlabelled data sets

FEATURES

What you get with Data Annotation & Labelling

01

Image classification

Images sorted into your categories or tagged with attributes such as colour, condition or product type, for recognition, search and recommendation models.

02

Bounding boxes

Tight boxes around every object of interest, with class and attribute labels, for object detection models such as YOLO and Faster R-CNN.

03

Segmentation and keypoints

Pixel-level semantic and instance segmentation, polygons, and landmark or keypoint labels for road scenes, faces, poses and medical images.

04

Video annotation

Objects tracked across frames with consistent IDs, events marked with timestamps, and clips categorised by scene or activity.

05

Text classification and NER

Reviews, posts and documents classified by topic, and people, organisations, places, dates and amounts tagged as named entities.

06

Intent and sentiment

Messages labelled by what the user wants to do and by positive, negative or neutral sentiment and its strength, for chatbots and opinion mining.

07

Transcription and speaker labels

Audio and speech transcribed to text, speakers identified, recordings classified by language, and subtitles and captions timed to video.

08 LLM response rating

Model answers rated, ranked and corrected against your rubric, producing preference and instruction data for fine-tuning and evaluation.

09 Quality control

Multi-step review, inter-annotator agreement checks and gold-standard samples on every batch, with a quality report delivered alongside the labels.

WHY IT MATTERS

The difference it makes to your business

01 Labels a model can learn from

Consistent labels across annotators and batches mean the model learns the pattern, not the noise between people.

02 Edge cases decided up front

Written guidelines with examples settle the awkward cases before labelling, and are updated as new ones appear.

03 Volume that matches your schedule

A dedicated team scales up for a large training set and down for ongoing top-ups, without you hiring or training annotators.

04 Your data stays confidential

Access controls, encryption, background-checked staff and confidentiality agreements protect the data you share.

HOW WE DELIVER

How we deliver Data Annotation & Labelling

01 Scope the task

We review sample data, agree the label set, the volume, the output format and how quality will be measured.

02 Write the guidelines

Clear instructions with examples and edge cases are drafted with you, so every annotator labels the same way.

03 Run a pilot batch

A small batch is labelled and reviewed by your team. We adjust the guidelines until the labels match what you expect.

04 Annotate at volume

Trained annotators label the full data set, with second-pass review, agreement checks and gold samples on every batch.

05 Deliver and iterate

Labels are delivered in your format with a quality report, and we re-label or extend the set as your model reveals gaps.

Questions about Data Annotation & Labelling

What is data annotation?

Data annotation is adding labels to raw data so a machine learning model can learn from it: boxes around objects in images, tags on words in text, transcripts for audio. The labels show the model the right answer for each example.

Which types of annotation do you offer?

Image classification, bounding boxes, semantic and instance segmentation, keypoints, video object tracking, text classification, named entity recognition, intent and sentiment labelling, audio transcription, speaker identification, captioning and rating LLM responses.

How do you make sure the labels are accurate?

Through written guidelines with examples, a pilot batch you approve, second-pass review, agreement checks where two annotators label the same item, and gold-standard samples with known answers mixed into each batch. Each delivery comes with a quality report.

Do you use AI to pre-label data?

Where it saves time, yes. A model proposes labels and a person checks and corrects every one, so the speed comes from the model and the accuracy from the review. We measure pre-labelled batches against the gold set like any other, and switch it off for tasks where correcting takes longer than labelling.

How long does an annotation project take?

Scoping and guidelines take about two weeks, and the pilot one to two more. After that, throughput depends on the data type, label complexity and quality level, and the quote sets a delivery schedule by batch.

Which tools and output formats do you use?

We commonly use CVAT and Label Studio, and can work in your own platform. Labels are delivered as COCO, YOLO, Pascal VOC, JSON, CSV or a custom schema that loads straight into your training pipeline.

Can you label data for large language models?

Yes. We rate and rank model responses, write and correct reference answers, and label intent and entities, producing data for fine-tuning, preference training and evaluation sets.

How is sensitive or personal data handled?

Annotators work under confidentiality agreements with individual accounts, data is encrypted, and personal details can be redacted before labelling. We can also work inside your own environment so the data never leaves it.

How is annotation priced?

It depends on the data type, label complexity, volume and quality level, so prices are set per item or per hour after we see sample data. The pilot batch shows the real effort, and we quote a fixed price for the full set from it.

NEXT STEP

Start with a pilot batch

Send us a sample of your data and the labels you need. We will draft the guidelines, label a pilot set for your review and quote the full volume.

Book a free consultation: <https://appsin.ai/book-a-consultation/>

WhatsApp or call: [+91-9025555584](tel:+91-9025555584)

About AppsinAi

AppsinAi Private Limited is an AI development company in Coimbatore, India. We build generative AI assistants, AI agents, chatbots, machine learning models, computer vision and data platforms on our clients' own data, from a first proof of concept to production, with monitoring and support after launch.

Contact

sales@appsin.ai

+91-9025555584

Housing Unit, Sf No. 393, CMS School Backside, 3rd Street, Nehru Nagar, Ganapathy, Coimbatore 641006, India

<https://appsin.ai>