



CHOOSING THE RIGHT DATA ANNOTATION TOOL

Since the performance of your ML model is as good as the data that trains it, understanding the tools used for annotating this data becomes especially important. These tools determine the quality of the data and can have implications on the success or failure of your model. Plus, with the advent of commercial tools in the market lately, there is always a build vs buy dilemma for tools. If you don't have the time, surety about model iterations, expense capacity or workforce to build tools from the scratch, or inclination to develop and maintain tools built on open source and freeware, then going in for commercially available tools could be the immediate choice.

While exploring the tool market, you might want to look at dataset management, which includes support for importing large volumes of data and exporting to file storage targets (local/network/cloud) that suit best with your operations workflow. Also, depending on your model's anticipated future needs, you may want to go in for tools that specialize in a particular labeling use case or are more of an all-rounded platform.

With the ongoing AI use and research, annotation platforms have come out with automatic pre-labeling features, which can be further fine-tuned with data labelers taking corrective actions to annotated data.

Although automated labeling does the job quite well, the need for human-in-the-loop services is imperative to its success when it comes to edge cases and building high confidence models. APIs and SDKs may be available in a few tools for developers to tweak how data interactions materialize with tools doing the pre-labeling work.

Best in class tools should provide workforce management dashboards with job time analysis, and QC control. Additionally, depending on the deployment model (on-premise/cloud), you can look out for modus operandi vis-a-vis secure file access; record of annotation details (timestamps, author, etc), and viewing rights of the annotator.

Choosing the right data annotation tool is not as easy. When evaluating the tools, you must determine the following:

What all formats the tool supports viz. images, videos, text, sensory data, or audio.

If the tools available are configurable as per your model iterations.

If the tool provider is open to an option of you partnering with a workforce provider of your choice. Inherently, tool providers have partnerships with third-party labelers, whose knowledge base and working style may not be the best fit for your organization.

Whether the QC capabilities are as per the use case.

And, if the tool service provider invests in R&D to continuously develop their tooling capabilities, precision, and throughput.

At Alsmartz, we work with teams who are vetted, trained, and managed by project managers who will be constantly in touch with you to match your working style and processes. Although our teams can work on virtually any tool in the market, we would like to share our recommended annotation tools that we generally use:



LabelBox is an end-to-end platform to create, manage, and process high-quality training data. It is a seamless, collaborative tool for machine learning teams. It is a complete solution for data marking, data management, and data science tasks. Some of its key features include bounding box image annotation and text classification

Amazon SageMaker Ground Truth helps you create highly accurate training datasets for machine learning. It provides built-in labeling workflows for text and image datasets and allows labelers to create custom workflows as well. These workflows support a raft of use cases including 3D point clouds, video, images, and text.

SageMaker Ground Truth enhances data label accuracy via annotation, consolidation, and audit workflows. It is easy to use and comes with automated labeling features such as 'auto-segment', automatic 3D cuboid snapping, and 'sensor fusion with 2D video frames'.



SuperAnnotate is a complete data annotation tool for computer vision projects. Right from data engineering to model creation, the platform supports a variety of formats such as image, video, LiDar, text, audio, etc.

SuperAnnotate claims that their platform can speed up annotation tasks by at least three times by leveraging their more advanced features like automated predictions, transfer learning, and data and quality management.

Dataloop offers human-operated tools and plug-ins for images, videos, and more right from object classification to pixel-level segmentation. The tool allows users to build an infinite number of datasets with guidelines and ontology. Algorithms, instruction instruments, and templates ultimately minimize labeling time, while quality assurance components render visibility to the annotation process.





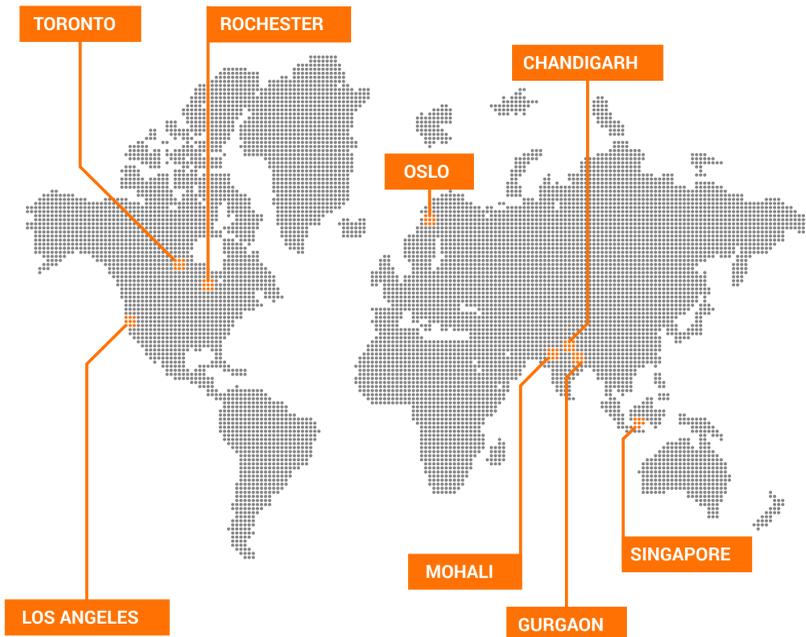
Datasaur is a great text labeling tool for natural language processing. Using smart models and improved workforce management, Datasaur helps annotators better manage their team and enhance the quality of their training data.

At Alsmartz, we understand that better data is the secret to better AI products. We leverage the best-in-class tools and technologies and a highly-skilled workforce to provide the quality annotation of text, images, audio, and video data sets for complex models. We have worked with global clientele on use cases including Medical AI, Autonomous Vehicles, Geospatial Imagery, Customer Service, Retail, and E-commerce. Drop us a line to know how we can materialize your AI vision.

ABOUT NETSMARTZ



Founded in 1999, Netsmartz is a privately held company, with a dedicated AI division Alsmartz (incepted in 2015). It operates across the globe with multiple offices, and is a CMMi3 & ISO 9001:2008 certified company. Formed over 20 years ago, we currently have more than 1000 employees, a rich experience of successfully executing 2000+ projects and working with small companies to Fortune 1000 clients.



Rochester, NY (Headquarters) Chandigarh (Global Delivery Center)

332 Jefferson Road
Rochester, NY 14623
Phone: 1-585-340-1166

Plot no. 10, Rajiv Gandhi Chandigarh
Technology Park, Chandigarh -
160101

Irvine, CA (Sales Office)

200 Spectrum Center Dr
Irvine,
CA 92618

Mohali (Global Delivery Center)

IT-C-6, Sector 67,
Mohali – 160062,
Punjab, India

Toronto (Sales Office)

545 King St W, Toronto
ON M5V 1M1, Canada
Phone: 1-647-542-0114

Gurgaon (Global Delivery Center)

9th Floor, JMD IT Megapolis
Sector 48, Sohna Road
Gurgaon - 122001

Oslo (Sales Office)

Gjekstadveien 36
3218 Sandefjord, Norway
Phone: +47 97653546

Singapore (Sales Office)

16 Raffles Quay, #33-03 Hong leong
Building, Singapore 048531
Phone: +65-9450-4800



+1 585 633 7555



kanav.jain@netsmartz.com



<https://www.aismartz.com/label-services.php>

**Like to find out more about how you can
Build Your Team@Netsmartz?**

If you need to know more about how we can help you get your dedicated team, do get in touch. We can set up a meeting at a time suitable for you.



ROCHESTER | LOS ANGELES | TORONTO | OSLO | SINGAPORE | CHANDIGARH | MOHALI | GURGAON

HQ: 1250 Pittsford-Victor Road, Ste 190, Pittsford, NY 14534, Phone: 1-585-340-1166