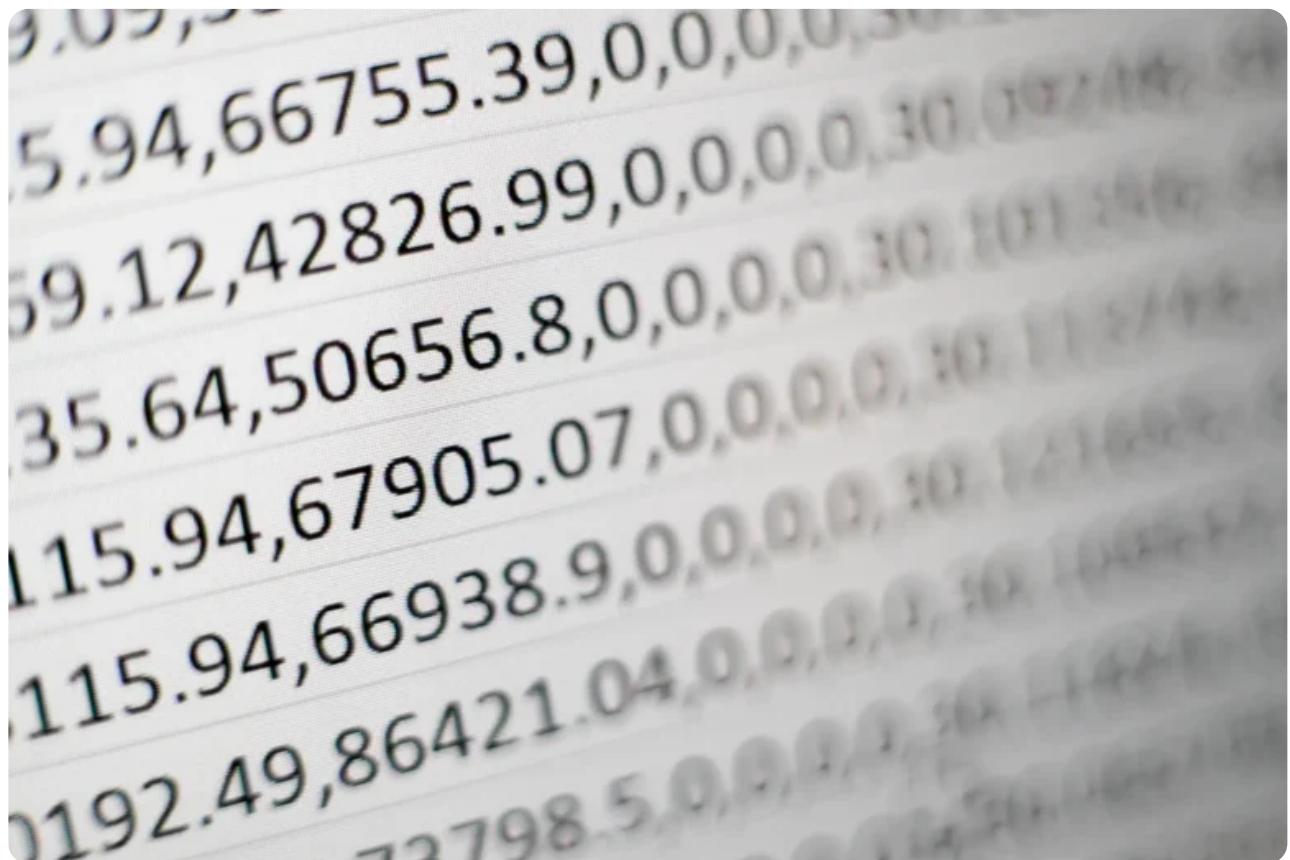


[Back to Tasq Blog](#)

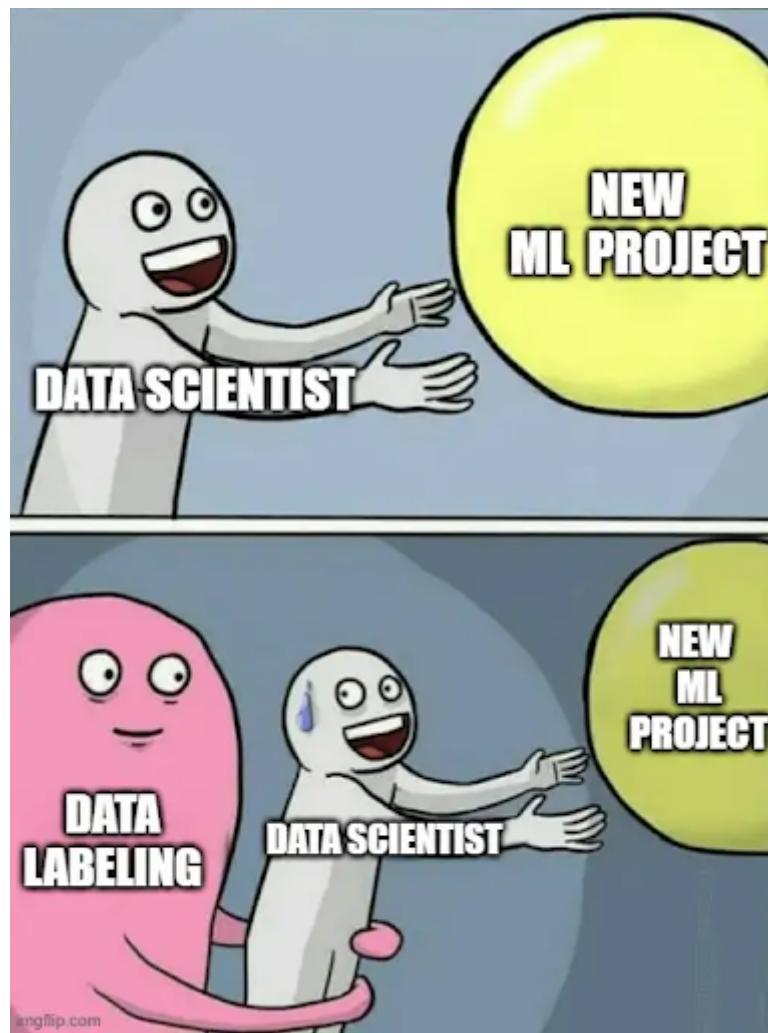
# Automated Data Labeling vs Manual Data Labeling: What's the Difference?

By: Nati Catalan

Dec 14 | 2022



*"The bedrock for supervised machine learning tasks is accurately labeled datasets".*



Machine learning systems heavily rely on data. These systems are capable of finding hidden patterns and providing necessary information pertaining to a certain task. But data must be well curated which is where tasks such as **structuring** and **labeling** the data come into the picture.

But how can we create the precise, reliable datasets that the tasks demand? Generally, there are two approaches by which we answer this question—**manual** and. **automated or AI-enabled data labeling**.

This article explores these two approaches i.e. manual vs auto labeling

and presents to you the pros and cons of each of them. Additionally, we will also briefly look at a case study.

## What is Labeled data?

The task of data labeling for machine learning is used in the task of supervised learning. In order for the model to learn effectively, the labels must be accurate and comprehensive. It must accurately reflect the relationships between the

features and the target variables. This allows the ML model to learn the underlying patterns and relationships in the data, and to generalize those patterns to new, unseen data.

In addition to providing the [ground truth](#) for machine learning algorithms, data labeling can also be used to improve the performance and reliability of the models. By carefully selecting and labeling the training data, it is possible to ensure that the models are trained on a representative and diverse sample of the data, which can help to improve their accuracy and generalizability.

Essentially, [data labeling](#) is an integral part of the machine-learning process and is crucial for ensuring the effectiveness and reliability of machine-learning models.

## What are the different tasks in data labeling?

Data labeling consists of various tasks some common tasks include:

1. **Image classification:** Here, the goal is to assign a label or category to an image, such as "cat" or "dog".
2. **Object detection:** It usually pertains to identifying and locating objects within an image, such as cars or people.
3. **Semantic segmentation:** It deals with assigning a label to each pixel in an image, providing a detailed understanding of the image's contents. Some other segmentation task involves **panoptic** and **instance segmentation**.
4. **Named entity recognition:** In named entity recognition, the goal is to identify and classify named entities, such as people, organizations, and locations, within the text

These are just a few examples of the many tasks that can be performed in data labeling.

## What are the different types of data labeling?

There are several different types of data labeling, and the specific type used will depend on the data and the intended use of the labeled dataset. Some common types of data labeling include:

- **Binary data labeling:** It involves assigning a label of “true” or “false” or ‘1’ or ‘0’ to each piece of data. For example, a dataset containing images of animals might be labeled with “true” if the image contains a cat and “false” if it does not.
- **Multiclass data labeling:** It generally involves assigning labels to more than two classes in the given dataset. For example, a dataset containing images of animals might be labeled with the names of different animal species (e.g., “cat,” “dog,” “bird,” etc.).
- **Multi-label data labeling:** It deals with assigning multiple labels to each piece of data. For example, a dataset containing images of animals might be labeled with the names of the animals in each image, as well as the type of environment they are in (e.g., “cat in a grassy field”). This type of dataset is very popular in text-to-image generation.
- **Semantic data labeling:** This involves adding detailed annotations or descriptions to the data. For example, a dataset containing images of animals might be labeled with detailed descriptions of the animals’ appearance, behavior, and environment.
- **Structured data labeling:** It is labeling the data in a structured format, such as a table or database. For example, a dataset containing customer feedback might be labeled with the specific issues or topics that each piece of feedback addresses, as well as the customer’s name and contact information.
- **Unstructured data labeling:** This involves labeling data that is not organized in a structured format. For example, in NLP and speech recognition, a dataset containing audio recordings might be labeled with transcriptions of the words spoken in each recording.

## What are the different approaches to data labeling?

There are two approaches:

1. **Manual labeling** is the process of assigning labels to data by a human annotator. This can be time-consuming and may introduce errors if the person labeling the data is not careful or is not familiar with the task at hand.
2. **Automated labeling** is the process of using algorithms and software to automatically assign labels to data. This can be faster and more accurate than manual labeling, but it requires training the algorithms on a labeled dataset, which can be labor-intensive and time-consuming. Additionally, automatic

labeling may not always be as accurate as manual labeling, especially for complex tasks or data that is difficult to label.

It is important to note that the auto-labeling largely depends on the data the algorithm was trained for. If the consistency of the trained data does not match the consistency of the unlabeled data then the algorithm may not yield accurate results.

## Manual Data Labeling

As mentioned earlier, manual data labeling involves a human annotator. This process is typically done by a team of annotators who are trained to understand the task at hand and the specific labels that need to be assigned. The annotators generally review each data point and assign the appropriate label based on their understanding of the data, labeling guidelines, and the task in-hand.



Manual data labeling is often used when the data is complex or when there is a need for high-quality labels that require human interpretation and decision-making. For example, in tasks pertaining to medicine and healthcare. It can also be used when there are no existing algorithms or models that can accurately label the data.

## What are the advantages of manual data labeling?

One advantage of manual data labeling is that it allows for greater control and accuracy in the labeling process. Because the labels are assigned by humans who are trained to understand the data and the labeling guidelines, they are often more

accurate and consistent than those generated by automatic methods. This can be particularly important for tasks that require high-quality labels, such as image or audio classification tasks.

## What are the disadvantages of manual data labeling?

On the downside, manual data labeling can be time-consuming and costly, especially for large datasets. It also requires a team of trained annotators, which can be difficult to find and manage.



## Automatic data labeling

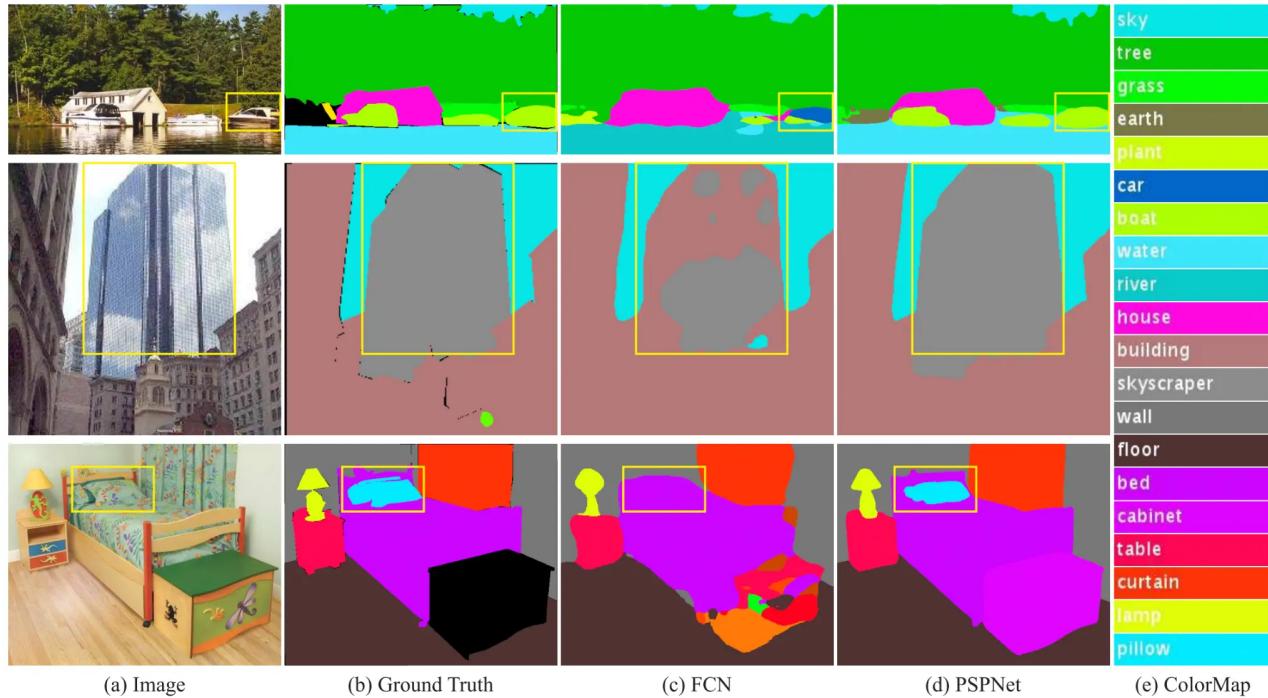
Automatic data labeling processes or AI-assisted data annotation is a process where artificial intelligence (AI) algorithms are *used to assist human annotators* in labeling data. In some sense, it is hybrid or semi-automated data labeling. This can be done in a few different ways, such as by providing suggestions for labels based on the data (something like a recommendation system), or by automatically generating labels that can be reviewed and corrected by human annotators.

AI-assisted data annotation is often used to improve the **efficiency** and **accuracy** of the labeling process. By using AI algorithms to assist in the labeling process, human annotators can focus on more complex or nuanced tasks, while AI algorithms handle the more routine or repetitive aspects of the labeling process. This can help to reduce the amount of time and resources required for labeling, while also improving the quality of the labels.

## What are the advantages of AI-assisted data annotation?

One advantage of AI-assisted data annotation is that it can provide a balanced approach that combines the strengths of both human and machine learning. Human

annotators can provide the expertise and interpretation skills needed for complex tasks, while AI algorithms can handle the more routine aspects of labeling, such as image recognition or text classification. This can help to improve the accuracy and consistency of the labels, while also reducing the time and resources required for labeling.



### [Source](#)

*The image above compares the two deep neural network's performance*

## What are the disadvantages of AI-assisted data annotation?

However, AI-assisted data annotation also has its limitations. It relies on the availability of labeled training data to train the AI algorithms, which may not always be available or may be difficult to obtain. It also requires the development and implementation of AI algorithms that can assist in the labeling process, which can be a time-consuming and complex process.

## Current trends of data labeling

There are several current trends in data labeling, including:

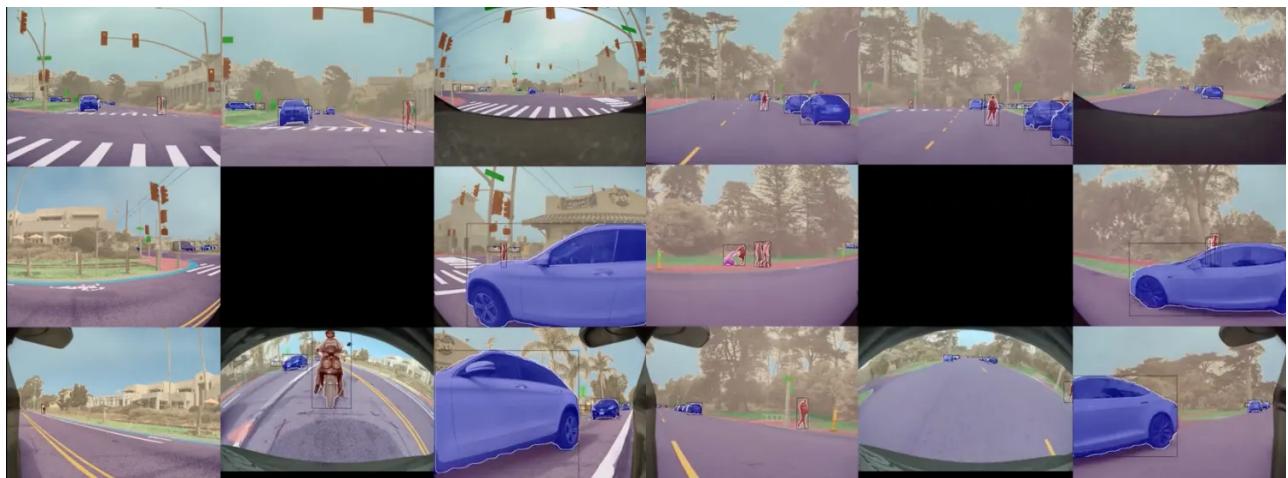
- **The use of AI and machine learning algorithms to automate the labeling process:** Many organizations are using AI and machine learning algorithms to

automatically label their data, reducing the need for manual labor and speeding up the labeling process.

- **The development of new tools and technologies for data labeling:** There are many new tools and technologies being developed for data labeling, including user-friendly interfaces, visual aids, and quality control mechanisms. These tools are making it easier to label data accurately and efficiently.
- **The increased emphasis on the quality and accuracy of data labels:** With the growing importance of machine learning and AI, there is an increased emphasis on ensuring that data labels are accurate and reflect the true meaning of the data. This is important for training high-quality machine learning models and achieving good performance.
- **The use of crowdsourcing and other collaborative methods for data labeling:** Many organizations are using crowdsourcing and other collaborative methods to label their data, allowing them to tap into the knowledge and expertise of a large community of annotators.
- **The growing importance of data labeling in a wide range of industries:** Data labeling is being used in an increasingly diverse range of industries, from healthcare and finance to retail and transportation. As more organizations recognize the value of labeled data for machine learning and AI, the demand for data labeling services is growing.

## Case study: How Tesla labels their data?

Tesla uses a variety of [methods to label the data](#) for their self-driving cars. They employ a team of human annotators who manually label the data, and they also use AI and machine learning algorithms to automatically label the data.



[Source](#)

To label the data, Tesla's annotators use a specialized tool that allows them to view the data (such as images or videos captured by the car's cameras) and add labels or annotations. The annotators may label objects in the data, such as other vehicles, pedestrians, or traffic signs, or they may label events or actions, such as lane changes or turns.

In addition to manual labeling, Tesla also uses AI and [machine learning algorithms](#) to automatically label the data. These algorithms are trained on a large dataset of labeled data, and then they are used to label new data. The algorithms make the final decision on the labels and annotations, without any human input.

Overall, Tesla's data labeling process is designed to produce a labeled dataset that is accurate, consistent, and comprehensive, and that can be used to train their self-driving car models.

## Conclusion

Data labeling is an important task as it enables machine learning algorithms to yield accurate results. The approaches to data labeling can vary depending on the task at hand. For instance, manual data labeling is an important approach when it comes to areas such as medical and healthcare. For tasks pertaining to genome sequencing or radiology where the data is extremely complex and fragile manual annotation is a must or at least to a point where algorithms can surpass human annotators. But certainly, for tasks involving autonomous vehicles where the area of interest is quite large automated labeling will suffice provided human intervention.

Data labeling is certainly seeing a shift from a manual approach to automatic approach. Maybe in the coming 5 years, most industries will be using automated labeling processes as the pace at which AI is evolving is quite certain.

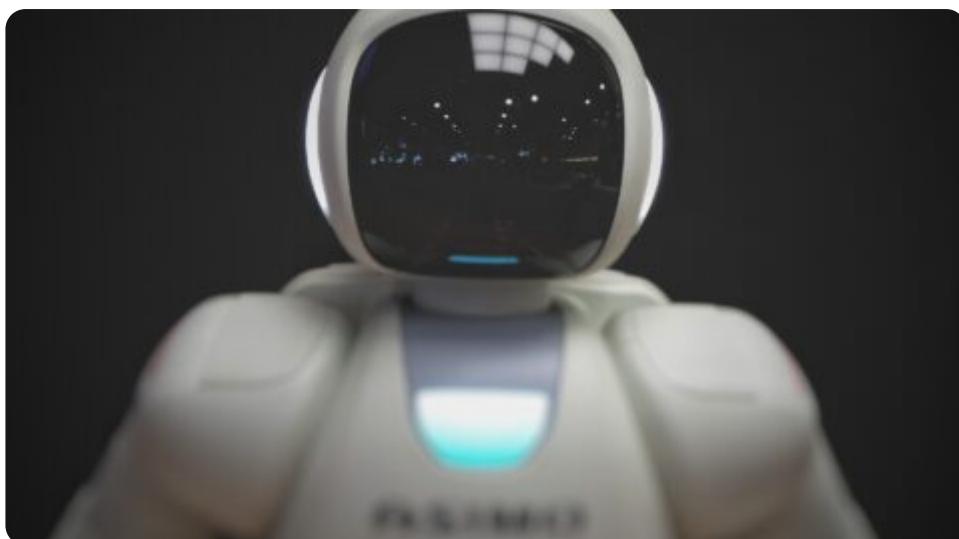


Nati Catalan

## Related Articles



The Benefits of Auto Labeling For ML



Unlocking the Power of Machine Learning Data Catalogs



# How to Choose Your Best Data Labeling Outsourcing Vendor

Sign up to our newsletter

## Solutions

- Computer vision labeling
- Image Annotations
- Model Validation & Tuning
- Ranking & Surveys
- Similarity & Comparison
- Synthetic Data Validation
- Transcription
- Video Annotations
- Content Moderation

## Industries

- Drones & Robotics
- Autonomous Vehicles
- Agriculture
- Sports, Media & Entertainment
- E-commerce & Retail

## Company

- About Us
- Why Tasq
- Blog
- Webinars & Videos
- Glossary
- Frequently Asked Questions
- Careers
- FAQ

## Case Studies

- "Finding a Tiny Pest in a Big Field"
- Frictionless Checkout for International Retail Chains
- How fast does it take to label an image with 3,000 objects?

**Compare**

- 
- [Tasq vs Scale](#)
  - [Tasq vs Appen](#)
  - [Tasq vs Labelbox](#)
  - [Tasq vs Lionbridge](#)
  - [Tasq vs Competitors](#)

**Contact**

- 
- [Book a Demo](#)
  - [Contact Us](#)

[Terms of use](#)      [Privacy Policy](#)