



PERSON WALKING

# Reliable Data Annotation for Autonomous Driving Systems Demands a Disciplined Methodology

**A**utonomous driving (AD) is a rapidly maturing market. While innovators in this space focus on varying levels of autonomy in driving, they share a common need for reliable data labeling to train their AI/ML systems. To ensure that data annotation achieves the desired standards, industry specialists need to develop methodologies that contribute to delivering consistent, high-quality results.

Demand for reliable data labeling will further accelerate in the future. A [recent report](#) says the autonomous driving (AD) market generated USD 21 billion in 2020. By 2026 it is projected to be worth almost USD 62 billion, a CAGR of 23%. According to [McKinsey](#), the AD market could be worth USD 300 billion to USD 400 billion by 2035.

The Society of Automotive Engineers (SAE) classifies AD systems using standard [SAE J3016](#). It defines six levels of driving, zero through five, ranging from no automation to full automation. For levels 0, 1, and 2, the motorist is in control. The motorist relinquishes control as level 3, 4, and 5 features are activated.

As McKinsey notes, “Growing demand for AD systems could create billions of dollars in revenue.”

Vehicles with lidar-based L2+ capabilities require about \$1,500 to \$2,000 in system components - and those with L3 and L4 capabilities require even more. Autonomous vehicles continuously deal with four basic questions:

- Where am I?
- What is around me?
- What will happen next?
- What should I do?

*An AD system must map a vehicle’s location and then scan its surroundings. It must anticipate what will happen next so that it responds successfully.*



*Accurate annotation of occluded objects is vital as it helps to determine the distance between them and the vehicle.*

## Basics of Data Labeling for Autonomous Driving

Data labeling or annotation manually classifies and segments objects in an image or video. Large, annotated datasets are essential to developing high-performing algorithms, which lead to autonomous driving success. Quality data sets train algorithms to infer absent data based on acquired experience.

To label a video or image, the annotator places anchor points at each of the object's edges. When part of the object is obscured or occluded, the annotator must estimate where the hidden edge may actually be. Accurate annotation of occluded objects is vital as it helps to determine the distance between them and the vehicle.

Another challenge is tracking objects that appear larger or smaller depending on their distance from the sensor. For a given application, should an algorithm be trained to track the change in size as distance changes, or should it estimate the actual size of the object?

Semantic segmentation is another kind of annotation that detects, classifies, and segments data. Some examples of classes include cars, bikes, pedestrians, traffic lights, and sidewalks. Semantic segmentation is more precise than other annotation techniques, as every pixel is associated with a class. The complicated relationship between pixels and successive frames makes semantic segmentation challenging.

## Importance of Expert Human Annotators

AD systems are trained to cope with the unexpected. The stakes are high, as a single egregious error among thousands of labels undermines the integrity of the entire process. In the AD industry, safety is always Job One. It relies on real-time communication between perception, decision-making, and other systems.

However, a lack of standardization complicates things. Manufacturers use varying sensors, data formats, and communication protocols, making it difficult for perception systems to communicate with one another.

Mimicking what humans deem appropriate is one of the greatest challenges in AD system development. When everything is in sync, an AD system makes correct decisions based on vehicle sensor data.

Data annotation is critical when dealing with edge cases and so-called “long-tail scenarios,” including those involving:

- Other vehicles
  - Vehicles running red lights
  - Oncoming vehicles approaching at high speeds
  - Vehicles parked illegally
  - Cross-traffic obscured by vegetation
- Pedestrians
  - Nonstandard pedestrians, like those carrying things
  - Pedestrians engaged in unusual behavior
- Signage and road markings
  - Invisible road signs due to adverse weather conditions
  - Poorly marked roads
- Atypical road events
  - Roadside emergencies
  - Construction zones

Also, both left and right turns present an array of issues, as do changing weather conditions. It is also essential for systems to appropriately respond to workers waving motorists through a construction zone, even when there’s a red traffic light. Or, consider the sudden appearance of a jogger as an autonomous vehicle begins a right turn.

Everything must be prioritized, so the AD system mimics the rational human response in every scenario imaginable.



*Mimicking what humans deem appropriate is one of the greatest challenges in AD system development.*

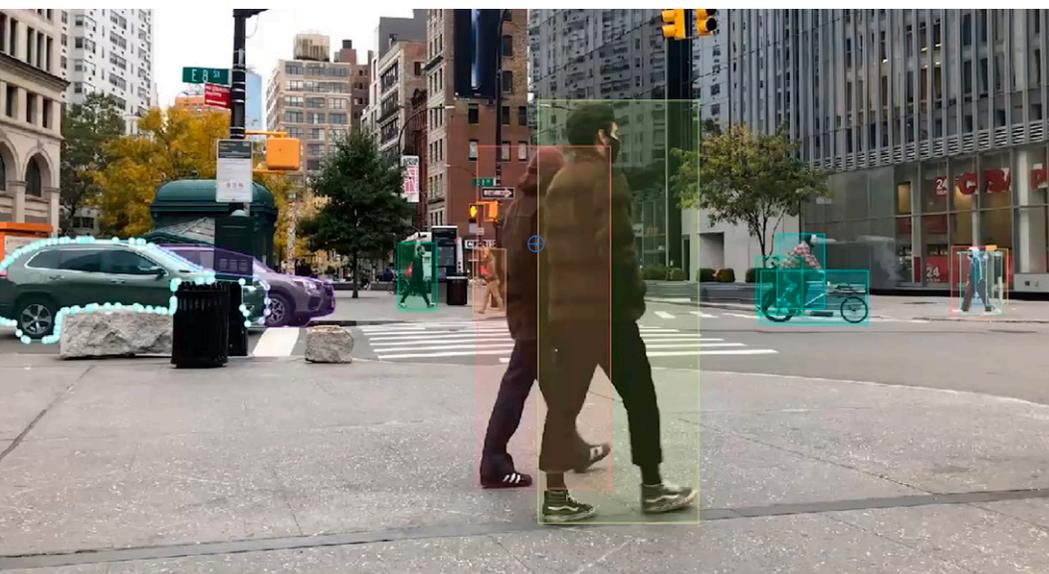
## Toward a Data Labeling Methodology

The quest for proficient data labeling requires systematically analyzing every part of the process. **Managing the data pipeline requires trained staff to identify and prioritize key performance indicators (KPIs).** The end goal is a data labeling process that optimizes the volume, quality, and timely completion.

Expert data labeling delivers better time-to-quality by deploying agile training data methods, advanced analytics, and, when possible, selective automation. Here are some key metrics in data labeling:

- **Standardized Format:** To do it right, annotations should be created using a standardized format. Overall, this makes them more accurate, complete, and consistent.
- **Category Accuracy:** Teams must track rates of false negatives and false positives and continually adjust in response. A false negative occurs when objects that should be annotated are not. Incorrectly grouped annotations also undermine quality. Keep in mind that a 96% accuracy rate means there's another 4% where it's essential to drill down and identify opportunities for further improvement.
- **Rate of Agreement:** Look at the work of multiple labelers, determine the degree of consensus, and score it. For example, divide the number of labels that agree by the total number of labels per asset.
- **Detecting Patterns:** For example, a review of the past month's performance reveals meaningful insights. The analysis may show a pattern or a clustering of issues to resolve.
- **Overall Team Performance:** This includes the performance of the labeler, the labeling team, and the project as a whole. Ensure the timely delivery of a corpus of labels that meets specifications.
- **Doing the Right Thing:** There is one final consideration. Ideally, a data labeling service engages with the ethical implications of doing annotations in a certain way. Sometimes, motorists must choose between less-than-ideal scenarios - and the same is true for AD systems. How can they be trained to identify the best choice under the most challenging circumstances?

To get the desired results, you must often customize metrics to better address each client's unique needs. In part, this is because priorities vary from one client to another. For example, there is a difference in emphasis between a client working on Level 3 versus Level 5 capabilities.



*Expert data labeling delivers better time-to-quality by deploying agile training data methods, advanced analytics, and, when possible, selective automation.*



*A labeler's deep knowledge of labeling specs is vital, as is the relevant and rigorous experience of project and operational managers.*

## Your Choice of Labeling Service Makes the Difference

AD developers need quality data labeling promptly delivered since delays may mean relinquishing a competitive edge. Here are some differentiators to use when assessing data labeling services.

### Selection of Annotation Tools

When it comes to which annotation tools to use, flexibility is critical. Ideally, the data labeling service is “tool agnostic.” That means it is open to any of three possibilities:

- Annotation tool recommended by the labeling firm
- Client's preferred data annotation tool
- Client's own custom annotation tool

Clients often want the experts to make an informed decision about the best tools for the task at hand.

### High-Performing Labeling Team

How responsive and expert is the team assigned to a given client? A labeler's deep knowledge of labeling specs is vital, as is the relevant and rigorous experience of project and operational managers. Brainstorming and swapping information are essential to success, and checking in daily maintains quality across a remote workforce.

The quest for quality requires supervisory talent with the requisite expertise. It also demands discipline at every level - the labeler, the team, and the process. Meeting established benchmarks requires consistent, hands-on supervision - and it includes diligent review followed by comprehensive training and re-training.

Annotation teams are particularly important in the quest for higher levels of AD performance as defined by [SAE J3016](#). Ideally, they employ the best practices noted below to achieve two vital goals: 1) enhance self-driving performance and 2) ensure compliance with AD industry standards.

## Better Dashboards

Teams optimize results by embracing data labeling best practices. For example, streamlined, more intuitive dashboards minimize cognitive load and reduce context switching. Make it easy for labelers to access concise examples of correct and incorrect labeling, particularly edge labeling.

## Consistent Label Auditing

Another best practice is randomly sampling labels to ensure accuracy. In AD data labeling, sample percentages in the 10-20% range are often ideal. Use the results to update procedures as needed.

## Consultative Engagement

Consultative engagement is another vital value proposition. To what degree does a labeling service collaborate and consult with the client's team? At the same time, the client remains free to devote collaborative resources only as preferred. Not every client has the resources to devote to ongoing collaboration. The degree to which there's an attitude that "we're all in this together" may be an intangible metric, but it's essential nonetheless. A truly symbiotic relationship leads to success for all the parties involved.

Those involved in the autonomous driving industry often approach consultative engagement differently. Newer EV startups often source their self-driving software stack primarily from outside vendors, although they might still embrace team collaboration when expanding specific AD/ADAS features.

By contrast, long-established automakers might turbocharge their AD research via acquisitions. They'll buy up relevant vendors and integrate the arriving teams with their own internal teams. Consider General Motors' billion-dollar [purchase](#) of San Francisco-based Cruise Automation, for example.

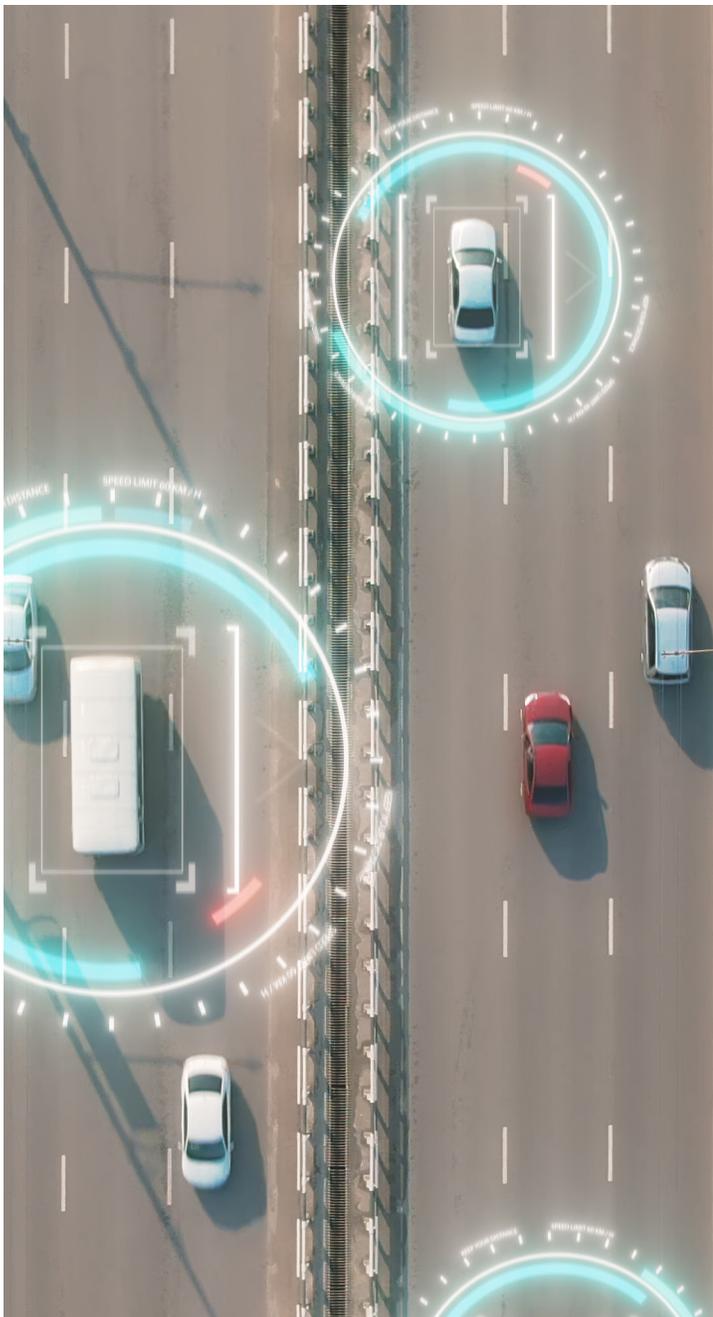
## Capacity to Customize

Quality metrics often require customization, as accuracy requirements vary by use case. For example, consider the tracking of a moving object. Capturing the edge with five-inch accuracy might be appropriate in one application, while another might require two-inch accuracy. Or, the accuracy of location data might be relatively more important than cuboid size.

Increased error lists and more detailed attributes increase accuracy but also cost time and money. For optimum efficiency, the level of annotation should align with a client's needs.

It is always important to contemplate whether the given labeling service has the expertise to address a client's unique needs.

*For optimum efficiency, the level of annotation should align with a client's needs.*



*Ideal outcomes require a steadfast commitment to a continuous cycle of annotation, training, more annotation, and re-training.*

## Performance Management

Success in data labeling requires ongoing performance management. Ideal outcomes require a steadfast commitment to a continuous cycle of annotation, training, more annotation, and re-training. This requires the right company culture, one that cultivates a relentless quest for operational excellence. Management must constantly search for ways to do more in less time without compromising quality.

Advances in AD are a product of [state-of-the-art ML](#) made possible by ongoing data collection and performance management applied in the field. For example, performance management facilitates Waymo's robotaxi fleet expansion from the broad, wide, gridded streets of Phoenix to more complex environments in San Francisco and Los Angeles. In the latter, Waymo will begin by [mapping multiple neighborhoods](#): Santa Monica, Miracle Mile, Koreatown, Westwood, and West Hollywood. Waymo led the way in California with [AD-permitted miles](#) with a human safety operator. The 628,838 miles Waymo recorded in 2020 soared to 2.3 million miles in 2021. Performance management helps deliver the scalability needed in an era of rapid growth in the autonomous driving industry.

## Beyond the Analytic to the Strategic

Again, safety is paramount in the verification and validation of AD systems. Scenario-based and simulation-based validations are integral to the development of AD systems.

Methods vary across data labeling firms. Some understand their task more narrowly, others more broadly. The narrower mindset simply focuses on meeting client specifications. By comparison, the broader perspective adds value whenever possible. It is about doing the right thing and helping the client to prosper.

With the narrower approach, a data labeling enterprise gets a client’s specifications and analyzes how to fulfill them. With a more comprehensive approach, the company’s effort extends beyond technical fulfillment. It employs thinking that is at once creative, continuous, and collaborative.

The high-level, comprehensive approach is ideal in the highly dynamic AD marketplace.

## Continuous Improvement

In a very real sense, the data labeling process is analogous to the act of autonomous driving itself. There is a continuous cycle of review and re-training. And this often requires a remote workforce.

Data labeling best practices vary according to the needs of each AD project. Customization borne of collaboration yields optimum results.

Going forward, it is incumbent upon data labeling enterprises to combine quality metrics and customization to be responsive to each client’s needs. The ultimate goal is continuously improving the AI data lifecycle, from data collection and preparation to tuning and maintaining the models.

*The ultimate goal is continuously improving the AI data lifecycle, from data collection and preparation to tuning and maintaining the models.*

### About Digital Divide Data (DDD)

Digital Divide Data (DDD) specializes in data labeling for autonomous driving. Clients value DDD’s data labeling teams because they:

- Embrace technological innovation at every opportunity
- Are nimble and responsive as project requirements evolve
- Remain with your project, improving quality and efficiency over time
- Essentially become an extension of your in-house workforce

DDD’s years of experience and collaboration capacity, combined with our tool-agnostic approach, give us a competitive edge. We meet benchmarks with a potent combination of expertise, discipline, and strategic thinking. And we welcome the opportunity to learn the details of your next project. Request a [consultation](#) today!

Learn more at [digitaldividedata.com](https://digitaldividedata.com).



Changing How  
the World Works.

### Contact US

Our team of experts welcomes the opportunity to discuss your project requirements. Please [contact us](#) today!