

AUTONOMOUS VEHICLES: THE RACE IS ON



Self-driving cars are capturing news headlines and people's imaginations. Is it really possible to read a book or watch TV while the self-driving car commutes to work? Is there a need to be in the driver's seat at all? Confusion reigns for good reason. There are significant capability differences between the assists built into today's cars and a true self-driving car of the future. But, one thing is certain:

THE RACE IS ON.

Standards created by SAE International measure the self-driving capability of a car on a scale of zero to five, where zero represents no automation and five is defined as "full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver."¹ Audi's 2019 A8, which features level three autonomous technology, will allow the car to start, accelerate, steer and brake on any road where there is a central barrier between traffic directions.² Tesla anticipates its vehicles will have level five autonomy in about two years³ while, in November 2017, Waymo achieved a major milestone when it became the first company to have autonomous vehicles on U.S. public roads with no human in the driver's seat.⁵ Independent of the timing question, dozens of companies, from Volvo, BMW, Mercedes Benz, Chrysler and Ford to Bosch, Uber, Apple, and Intel are working alone or in collaboration in the highly competitive, and rapidly advancing, autonomous vehicle race.

This AV race has several mountainous challenges: engineering, regulatory, lack of industry-standardized technology and tools, consumer trust and acceptance, to name a few. At each progressive level of autonomy the challenges become more difficult. But, among the most mountainous of challenges is data. Underlying an automobile's autonomous capabilities is volumes and volumes of data, required for both training its AI systems and also for real-time decision making once those same systems are deployed.

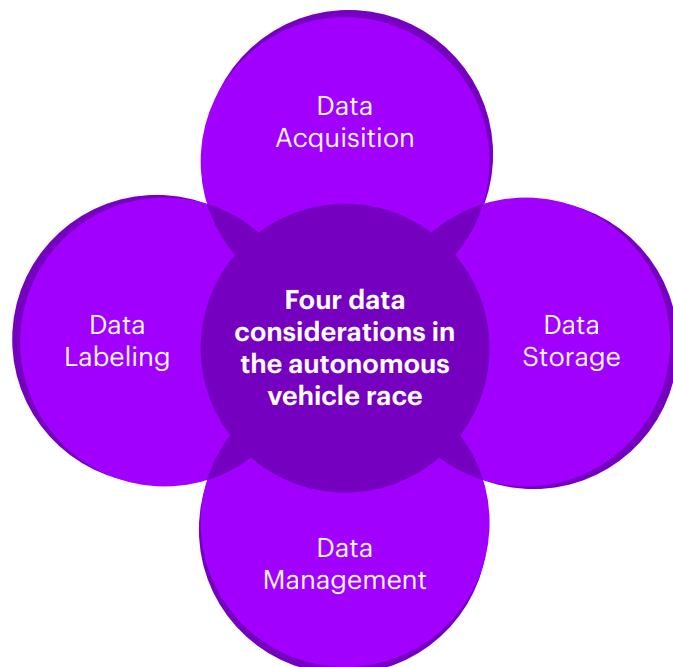
FOUR DATA CONSIDERATIONS IN THE AUTONOMOUS VEHICLE RACE

While there are different approaches to training autonomous vehicle computer vision models, many companies choose deep learning. For vehicles to advance to higher levels of autonomy through deep learning, their models need volumes of data produced from sensors, such as camera, radar, LiDAR, and ultrasonic data. This creates an acute challenge.

In one day, just one test autonomous vehicle produces as much data as the Hubble Space Telescope produces in a full year.⁶ Acquiring the data is difficult, storing it takes massive space and labeling and annotating it accurately takes tremendous resources. Given that a failure of an autonomous vehicle could result in serious injury or loss of life, high-quality training data is vital to the AV's mission-critical computer vision systems' ability to learn patterns and operate safely.

It behooves companies to consider data-related processes and infrastructure needs early in research and development to pre-empt the complex issues that arise as operations scale. Without efficient data management, the sheer resources the process will consume can dramatically slow innovation.

Here are four areas to consider when developing methods to manage extraordinary amounts of data for use by AVs.



Data Acquisition

Companies are investing significant time, effort and money in both deploying car fleets that gather real world data via various sensors and also in developing simulated environments to complement the real world. Rightly so. AI must be exposed to a huge diversity of scenarios to identify patterns and learn what the AV could encounter on the road. Data from various topographies, urban and rural areas, weather conditions (rain, fog, snow, cloudy, sunny), road types, and country variances such as left- and right-side driving are all needed to train AVs. In aggregate this is a lot of data. For instance, as of December 2016, Tesla had collected more than 1.3 billion miles of data from Autopilot-equipped vehicles operating under diverse road and weather conditions around the world.⁷

Companies can gain speed and efficiency in data acquisition by optimizing the data requirements and collection approach.

**A company's data acquisition should consider
a balance of three factors:**



A plan that balances these three elements will eliminate data redundancy and help ensure data acquisition meets comprehensive needs while running as fast and efficiently as possible given available resources.

Data Storage

Test AVs today generate between 4 and 6 TBs of data per day, with some producing as much as 8-10 TBs depending on the number of mounted devices and their resolution.⁸ By comparison, the typical person's video, chat and other internet use averages about 650 MB per day. That means, on the low end, the data generated from one test car in one day is roughly the equivalent to that of nearly 6,200 internet users.⁹ If not sufficiently considered, technical decisions associated with storing such high volumes of sensor data of differing formats, sizes and characteristics can halt a project.

Researchers commonly embark on data storage by purchasing hardware or cloud storage within the department. While having disparate teams create point solutions is easy and expedient in the short term, this approach can't scale successfully or economically as the storage challenge mounts. Further, having mission-critical data stored in a distributed and unsecured manner introduces risk. Therefore, it's smart to engage a company's centralized IT team early before distributed teams get too far down a difficult path.

Considerations when developing a robust and scalable data storage strategy include:

- 1** Will you use on-premise or cloud infrastructure? If you leverage a hybrid infrastructure how will you connect on-premise and cloud?
- 2** How will you offload data from the data collection vehicles? Such high volumes and varied terrains mean on-vehicle storage is needed. How will you move data from the vehicles to the storage infrastructure?
- 3** How will you secure the data at each stage of the collection, annotation and usage process?
- 4** How will you understand what data you have that is usable and not usable? For example, if a camera lens cracks on a data collection vehicle half way through the day, or the lens fogs due to rain, some amount of a 10-to-12 hour video stream may not be usable, but how will you know that and accommodate it?

Data Management

Companies developing AV functions such as lane departure warning, auto emergency braking or parking assist have unique data annotation and labeling requirements based on each of their specific models. Different teams may pull from the same data lake to create datasets to train models that support various functions. This is where tracking the data's origin, what happens to it and where it moves over time becomes an important issue. A single data set could be broken into smaller ones based on various criteria. One image in a data lake could require different annotation types (bounding boxes, segmentation masks, polylines, etc.) to support a different function that will be saved as multiple independent files.

To understand and track what the data contains – and therefore gain full leverage from it – companies need a strategy, approach, policy and data platform for longitudinal data management. Information on the source data such as the locations where it was collected, what streets were covered, what intersections were recorded, whether the data is from day or night, or sun or rain all needs to be recorded and associated with the data to aid in scene selection and to ensure the full portfolio of data requirements are being met. Scene selection is particularly important for supporting sensor fusion, where researchers combine data from different sensors and sensor types to use the combined information to perceive the environment more accurately. Information on the data's journey over time through various annotations, labeling needs, and training uses also must be tracked to maintain data integrity and usability.

Finally, consider: how will you educate and communicate to all research teams where the source data resides, what it contains and how it can be accessed?

Data Labeling

As car fleets traverse the roads, they collect many different types of data. Many vehicles have multiple sensors (radar, ultrasound, LiDAR, cameras), each gathering different, complementary data. In just one frame from one camera there can be hundreds of objects to label accurately. By some estimates each hour of data collected takes almost 800 human hours to annotate.¹⁰ The massive scale of this challenge is impeding many companies from moving as quickly as they would like.

There are a few important considerations when annotating and labeling AV data.

PROVIDE CLARITY ON WHAT TO CAPTURE

In a simple traffic intersection there may be hundreds of different possible objects to identify, so creating guidelines on what and how to annotate and label them is critical for efficiency and consistency. Guidelines should define what objects are considered qualified (e.g., passenger vehicles that are >50% visible), and the capture criteria for them (e.g., does the annotation cover “enough” of the object to be acceptable).

DETERMINE THE TOOLSETS NEEDED TO BEST LABEL AND ANNOTATE OBJECTS ACROSS DATA FORMATS

The value of using the right tools for each annotation task can't be overstated. For instance, you might need to draw bounding boxes for object localization and detection. Or you may require the ability to apply text labels and draw cuboids for metadata attribution. Or to create polylines to outline road and lane markings. The same tools you use for these annotation types may not work for segmentation masks, which require outlining overlapping objects and objects that share boundaries with 100% precision.

CONSIDER ECONOMIES OF SCALE

As companies move from research to prototype to production, the scale of data annotation needs increases exponentially, and the risk associated with bad data increases in parallel. Training data needs are outstripping any single company's ability to hire enough internal resources to address them at scale, even with expected improvements in algorithmic labeling of data sets. For many companies the unit economics associated with building the expertise and capacity just don't make sense. In these cases, the most viable option may be turning to third parties that able to build and manage a workforce of annotators at scale to serve multiple entities' needs.

CHARTING A PATH TO WIN THE RACE

Whether just embarking on aggregating data for AVs or well into the race, companies that are aware and ahead of these data challenges will circumvent issues that could potentially halt their progress.

For those early in the data collection process, consideration of one's data approach and thoughtful decision-making regarding relevant tradeoffs will help ensure an action plan that is both executable and expeditious. For those where data collection is becoming increasingly precarious, a careful retrofit that leverages what is already in place can take the organization to a more secure, accessible and sustainable data approach.



Together Accenture and Mighty AI are deeply experienced in understanding and addressing AV data management challenges. Mighty AI works with perception and research teams to get them the high-quality training data they need to ensure their vehicles are safe. Accenture brings its broad set of skills and capabilities to bear as companies think through data strategy, collection, storage and use.

FOR A MORE IN-DEPTH DISCUSSION CONTACT

Michael Stephan

Mighty AI
michael@mty.ai

Paul Lalancette

Accenture | Industry X.0
paul.lalancette@accenture.com

Andrea Regalia

Accenture | Products OEM
andrea.regalia@accenture.com

Martin Stoddart

Accenture | Webscale
martin.stoddart@accenture.com

Matthew Quinlan

Accenture | Webscale
matthew.h.quinlan@accenture.com

NOTES

- 1 For more background on the SAE definition of each level of autonomy in advanced driver-assistance systems, see https://www.sae.org/misc/pdfs/automated_driving.pdf.
- 2 Vijayenthiran, Viknesh, "Audi A8 to be first with 'Level 3' self-driving capability, but regulations holding back tech," Motor Authority, April 25, 2017
- 3 Lambert, Fred, "Elon Musk clarifies Tesla's plan for level 5 fully autonomous driving: 2 years away from sleeping in the car", electrek, April 29, 2017.
- 4 Salesky, Bryan, "A Decade after DARPA: Our View of the State of the Art in Self-Driving Cars," Medium, October 16, 2017.
- 5 Hawkins, Andrew, "Waymo is first to put fully self driving cars on US roads without a safety driver", The Verge, November 7, 2017.
- 6 "A Look at the Numbers as NASA's Hubble Space Telescope Enters its 25th Year," NASA, May 12, 2014.
- 7 Hull, Dana, "The Tesla Advantage: 1.3 Billion Miles of Data," Bloomberg, December 20, 2016.
- 8 Maloney, Kieran, "The data management implications of the Federal Automated Vehicles Policy," IoT News, January 3, 2017.
- 9 Beres, Damon, "Each autonomous car will one day generate more data than thousands of people", Mashable, August 17, 2016.
- 10 Burke, Katie, "Humans help train their robot replacements," Automotive News," August, 27, 2017, August 27.

ABOUT ACCENTURE

Accenture is a leading global professional services company, providing a broad range of services and solutions in strategy, consulting, digital, technology and operations. Combining unmatched experience and specialized skills across more than 40 industries and all business functions—underpinned by the world's largest delivery network—Accenture works at the intersection of business and technology to help clients improve their performance and create sustainable value for their stakeholders. With more than 435,000 people serving clients in more than 120 countries, Accenture drives innovation to improve the way the world works and lives. Visit us at www.accenture.com.

ABOUT MIGHTY AI

Mighty AI's platform enables autonomous vehicle technology-makers to continuously tune and modify their datasets based on the needs of their AI models. We pair machine learning with human intelligence to revolutionize how companies manage sensor data, generate datasets, and validate their computer vision models. Visit www.mty.ai to learn more, and follow us at @mighty_ai.

Copyright © 2018 Accenture All rights reserved.

Accenture, its logo, and High Performance Delivered are trademarks of Accenture.