How Tesla and Waymo are tackling a major problem for self-driving cars: data

Sean O'Kane   April 19, 2018

Transportation

Autonomous cars won't happen without tons of data, but Tesla and Waymo have a big head start

By Sean O'Kane@sokane1 Apr 19, 2018, 8:00am EDT
Illustration by William Joel

There's a race happening right now that stretches from Silicon Valley to Detroit and back: who can make a self-driving car that behaves better than a human driver? It's a far harder task than it sounded even a few years ago because human drivers know a lot — not just about their cars but about *how people behave* on the road when they're behind the wheel. To reach that same kind of understanding, computerized cars need lots of data. And the two companies with the most data right now are Tesla and Waymo.

Both Tesla and Waymo are attempting to collect and process enough data to create a car that can drive itself. And they're approaching those problems in very different ways. Tesla is taking advantage of the hundreds of thousands of cars it has on the road by collecting real-world data about how those vehicles perform (and how they *might* perform) with Autopilot, its current semi-autonomous system. Waymo, which started as Google's self-driving car project, uses powerful computer simulations and feeds what it learns from those into a smaller real-world fleet.

It's possible — and proponents certainly claim — that self-driving technology would lower the number of yearly deaths in the US that result from car crashes, a staggering 40,000 people. But there's also a huge financial incentive to apply all this data-driven tech to the road as quickly as possible. Intel believes autonomous vehicles could generate $800 billion per year in revenue in 2030 and $7 trillion per year by 2050. Last summer, Morgan Stanley analyst Adam Jonas said in a note that data might be more valuable to Tesla than something like the Model 3. "There's only one market big enough to propel the stock's value to the levels of Elon Musk's aspirations: that of miles, data and content," he wrote in June.

Still image from a Tesla video demonstrating Autopilot in action.
*Image: Tesla*

Tesla is developing towards autonomy by using customer-owned cars to gather that all-important data. The company has hundreds of thousands of customers, many of whom use Autopilot on streets around the world every day, and Tesla — according to its privacy policy — collects information about how well the feature performs. It's a familiar strategy for anyone who's followed another of Elon Musk's companies: SpaceX. Musk has quietly tested equipment on real rocket launches and even sold some of the company's test launches.

It's hard to pin down exactly how many miles of data Tesla's gotten from Autopilot because the company doesn't make many public statements about it. In 2016, the then-head of Autopilot told a conference crowd at MIT that Tesla had logged 780 million miles of data, with 100 million of those miles coming while Autopilot was "in at least partial control" according to *IEEE Spectrum*. Later that summer, Musk said that Tesla was collecting "just over 3 million miles [of data] per day." As of last July, though, the total number of fleet miles driven had jumped to 5 billion. As Tesla sells more cars, the amount of data that can be collected increases exponentially.

"Tesla's customers have driven billions of real-world miles"
Not all of those miles are from Autopilot, and Autopilot is still only a semi-autonomous feature. But Tesla also collects data about how Autopilot would handle different driving scenarios even when the feature isn't used. Tesla cars can log instances where the Autopilot software would have taken an action, and that data eventually gets uploaded back to Tesla. This so-called "shadow mode" of collection means that Tesla could be simulating full Autopilot data across many of those billions of miles that are driven.

The only other company working with similar amounts of data is Waymo, which announced earlier this year that it has simulated 5 billion miles of autonomous driving. The company also said it has notched 5 million self-driven miles on public roads. That's more than basically every other company testing self-driving vehicles *combined*, if the recent reporting figures in the state of California — the biggest hotbed for autonomous testing so far — are any indication.

Watch Video At: https://youtu.be/B8R148hFxPw
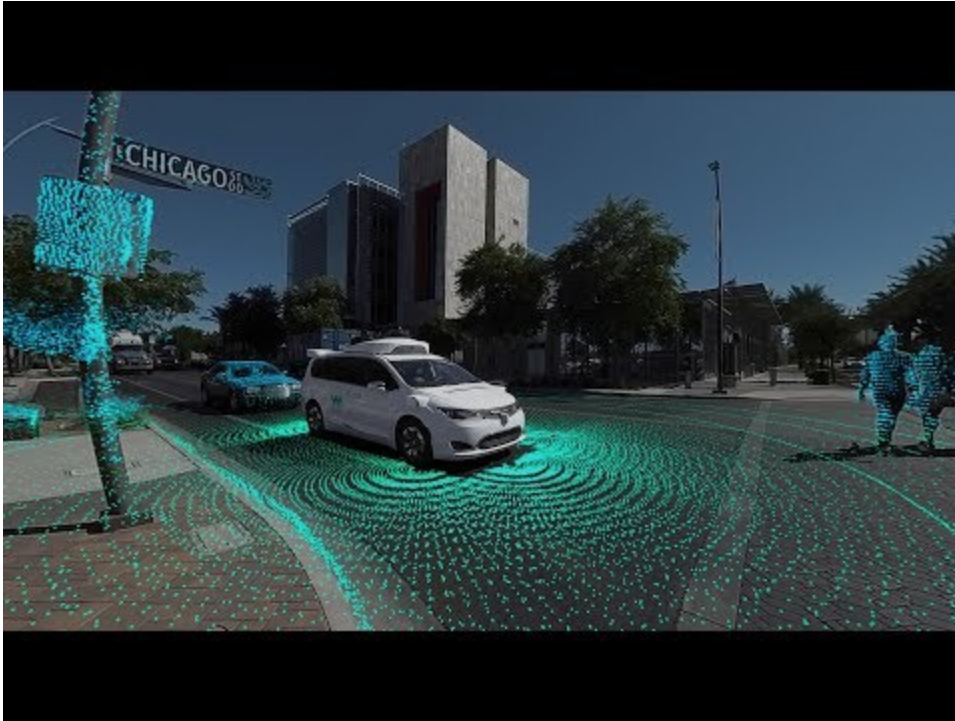
Waymo is constrained by the fact that it is only gathering real-world data via a fleet of about 500 to 600 self-driving Pacifica minivans. Tesla has over 300,000 vehicles on the road around the world, and those cars are navigating far more diverse settings than Waymo — which is currently only in Texas, California, Michigan, Arizona, and Georgia. But Tesla is only learning from those real-world miles because even when Autopilot *is* engaged, the current version is only semi-autonomous.

This balance will also change. Waymo plans to add "thousands" more Chrysler minivans are to its fleet starting at the end of this year. And it recently announced a partnership with Jaguar Land Rover to develop a fully self-driving version of the all-electric I-Pace SUV from the ground up. Waymo says it will add up to 20,000 of these to its fleet in the coming years, and it will be able to handle a volume of 1 million trips per day once all those cars are on the road.

Until then, Waymo relies heavily on its simulations, and computers can't always come up with every strange real-world scenario. That's why it matters that Tesla is leading in real-world miles now, argues analyst Tasha Keeney, who covers the company for Ark Invest. "I

feel like everyone agrees Waymo's technology is the best right now, but I think a lot of people are underestimating the power of the dataset that Tesla has," she says.

Photo by Amelia Holowaty Krales / The Verge

TYPES OF DATA

Not only are these two companies collecting data at different scales, they're also collecting different data. Waymo's self-driving minivans use three different types of LIDAR sensors, five radar sensors, and eight cameras. Tesla's cars are also heavily kitted out: eight cameras, 12 ultrasonic sensors, and one forward-facing radar.

But Tesla doesn't use LIDAR. LIDAR is a lot like radar, but instead of radio waves, it sends out millions of laser light signals per second and measures how long it takes for them to bounce back. This makes it possible to create a very high-resolution picture of a car's surroundings, and in all directions, if it's placed in the right spot (like the top of a car). It maintains this precision even in the dark since the sensors are their own light source. That's important because cameras are worse in the dark, and radar and ultrasound aren't as precise.

LIDAR can be expensive and bulky, and it also involves moving mechanical parts (for now, at least). Musk recently called the technology a "crutch," and argued that while it makes things easier in the short term, companies will have to master camera-based systems to keep costs down.

"A huge chunk of the industry agrees that LIDAR is necessary, but Musk disagrees"

If Tesla can develop autonomous cars without that tech, Keeney says that would be a huge advantage. "It's a riskier strategy but it could pay off for them in the end," she explains. "If Tesla solves [self-driving cars without LIDAR], everyone else is going to be kicking themselves."

That's a huge "if." Without LIDAR data, Tesla may find itself at a disadvantage, according to Raj Rajkumar, the co-director of General Motors-sponsored connected and autonomous driving research lab at Carnegie Mellon University. (CMU is a school so famous for its robotics chops that Uber poached dozens of staffers in 2015.)

LIDAR is seen by many in the industry as an essential tool for creating cars that can drive themselves, and Rajkumar says there is heavy skepticism about Tesla's approach. "We don't think the hardware will be sufficient to do that, and I don't think Tesla is particularly anywhere close to getting to [fully] driverless operation," he says.

It's also not clear what data Tesla is collecting to begin with. Tesla has access to data about the car's speed, acceleration, braking, battery use, and can save "short video clips" during accidents, according to the company's privacy policy. This data can be collected remotely or during service appointments. But with specific regard to Autopilot, the privacy policy only states that Tesla can access "information regarding the use and operation" of the feature.

Tesla declined to comment on what data is being collected from which sensors, or the quality of that data. It could be all the video from the car, from just some of the cameras at certain moments (like crashes), or data from the ultrasonic sensors without video. And, Rajkumar says, and it's also unclear whether it's the full frame-rate video or something with less fidelity.

Keeney agrees. "The Waymo data set is way more detailed just by the fact that they're using LIDAR, which pulls in so much more information than you'd get off of cameras alone," she says.

PROCESSING CHALLENGES

Collecting data is one thing. But even Musk has noted that processing the data is also a difficult task. "It's actually quite a challenge to process that data, and then train against that data, and have the vehicle learn effectively from the data, because it's just a vast quantity," Musk said on an earnings call last summer.

Waymo, comparatively, sounds more confident about its simulations. The company re-creates full computer models of the cities it's testing in, and sends 25,000 "virtual self-driving cars" through them each day, according to a report in *The Atlantic* from last summer.

This helps Waymo create a tight feedback loop by recreating real-world driving data on the computer, where "thousands of variations" of a scenario can be run. The data is then downloaded back into Waymo's test cars. Waymo has also built a dedicated test facility in

California, where it can build out particular street features or stage scenarios that seem to give its vehicles the most trouble.

"Waymo has a more obvious loop between its simulations and its real-world test fleets" This closed loop, Rajkumar says, "has come at the expense of incredible investments, resources, time, and effort — which Waymo of course clearly has plenty of because of its parent company." He says it would be hard for Tesla to match this. "Tesla would have to spend a lot more on it, and go through a highly labor-intensive process."

In his second "master plan" for Tesla, published two years ago, Musk said he believed it would take about 6 billion miles to gain "worldwide regulatory approval" of true self-driving technology. Tesla has likely passed that mark by now in real-world miles, and yet its cars still aren't able to fully drive themselves. A demonstration run of a Tesla driving itself from LA to New York that was supposed to take place in 2017 has been delayed, and the target for a rollout of the ultimate version of Autopilot keeps moving.

Meanwhile, Waymo is near that 6 billion-mile figure on the simulation side, and the company is racking up virtual miles faster than ever, with thousands more test cars waiting in the wings. It plans to launch a commercial ride-hailing program with its self-driving minivans later this year, something it is already trialing in Arizona, which could further bolster that data feedback loop.
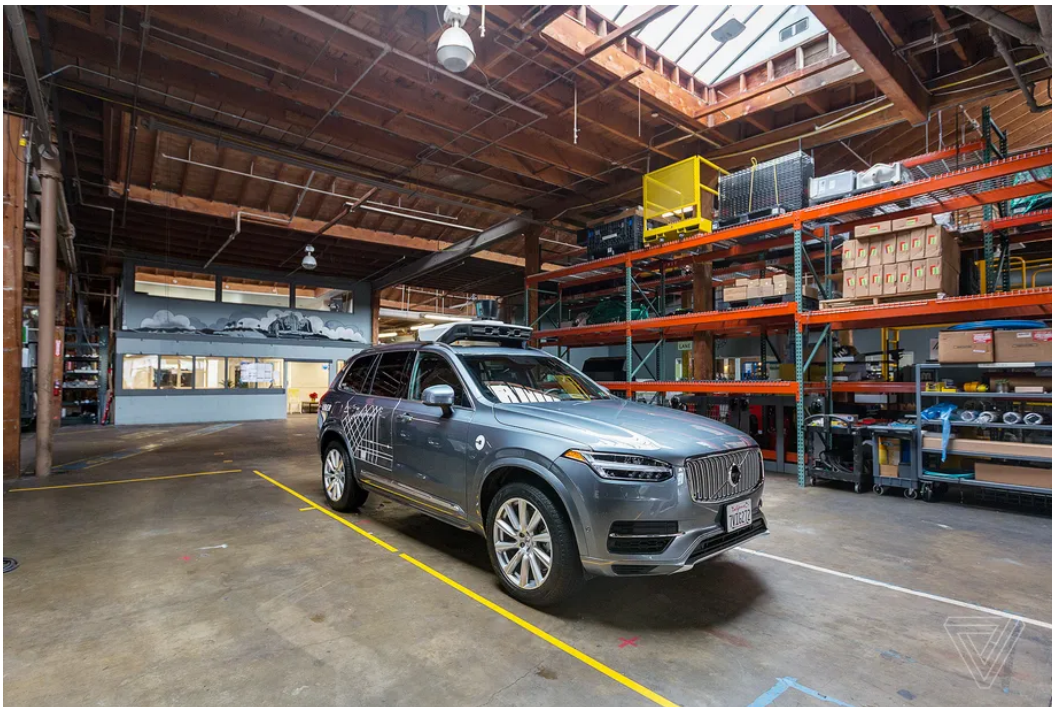

Photo by Vjeran Pavic / The Verge

OTHER COMPETITORS

Tesla and Waymo are two of the most advanced companies testing this tech, but they're not alone. One of the most visible competitors in this space has been Uber. Compared to Tesla and Waymo, Uber took a more haphazard approach with its self-driving testing, which is typical for the company that has epitomized the "move fast and break things" motto of Silicon Valley.

After starting testing in Pittsburgh in 2016, Uber put early versions of its modified semi-autonomous Volvos on the streets of San Francisco without obtaining the necessary state permits. When the company got busted, they moved testing to Arizona. Uber eventually acquiesced to California's basic requirements, but its scraps with lawmakers there put the company behind competitors like Waymo in real-world miles driven.

Once it was set up with test fleets in three states, Uber quickly clicked off miles. It reached 2 million miles driven nationwide by November 2017, according to *The New York Times*. It's unclear how many miles Uber has simulated, though, and the quality of its technology has come into question after one of its test cars killed a pedestrian in Arizona in March. Uber CEO Dara Khosrowshahi has said the company remains "absolutely committed" to the program, but its testing efforts remain suspended for now.

The only other company doing similar-quality work to Waymo or Tesla when it comes to self-driving cars, Keeney says, is a more old-fashioned one: General Motors. GM has been developing self-driving Bolt EVs with the help of a company it acquired called Cruise Automation, and it recently announced plans to trial its own limited commercial self-driving service in 2019.

GM is designing all-electric Chevy Bolts with no steering wheel or pedals, and will launch a commercial trial with the Bolts it has retrofitted with Cruise Automation's technology in 2019. Image: GM

GM is following in Waymo's footsteps by generating and processing the data required to teach cars how to drive themselves with small test fleets. But Keeney believes GM strength is its production scale. "Waymo has this deal with Jaguar, and that can turn into something in the future, but they're not actually producing the cars in-house. I think that there's an advantage to having a vertical strategy," she says. "With an autonomous sensor set, when you build it from the ground up, you have a better handle on what the production should look like and how you can optimize everything."

GM, like Tesla, also has a semi-autonomous product in customer cars that are on the road right now. But that product — Super Cruise — is limited to one Cadillac model, and there are no signs that it will spread to other models anytime soon.

In Keeney's eyes, that's another missed opportunity. "That's what they're missing, and that's what every other automaker is missing," she says. "Why has no one put sensors on their customer cars that collect data like Tesla has?"

WHAT GOOD IS SIMULATION, ANYWAY?

There's a dark horse in the race: Nvidia. It may not be racking up the billions of miles that Tesla and Waymo boast, but Nvidia's technology is being used by hundreds of companies — Tesla included — in the self-driving space. Last month, Nvidia began selling what it calls

"Drive Constellation," which is essentially a ready-made simulator for other company's self-driving projects. In other words, it's a commercial version of the simulations Nvidia was already using to test and also validate its own self-driving software and hardware.

""There's no way we can possibly drive around and capture all the crazy stuff that happens on the roads.""
Access to good simulation is crucial to developing autonomous vehicles, says Danny Shapiro, the senior director for automotive at Nvidia, in an interview with *The Verge*. "There's no way we can possibly drive around and capture all the crazy stuff that happens on the roads. There are trillions of miles that are driven, [but] a lot of those, the majority of those are very boring miles," he says. "After a certain point, you've mastered that."

That's when engineers have to study so-called corner cases, or scenarios that don't happen that often. There are tons of these when it comes to driving, Shapiro says: cars running red lights, road rage, hazardous weather, harsh sunlight at sunrise or sunset. Do enough real-world driving with test cars, and you'll certainly come across these events and scenarios, but not frequently enough to learn how to handle them. For example, in the real world, you only have a few minutes every day to drive a particular road as the sun goes down. In simulation? "We can drive every road 24 hours a day at sunset, and stage all kinds of [other] potential hazards," he says.

This is why any company simulates autonomous miles in the first place. By lowering the barrier to entry, though, Nvidia has made it easier for companies without the kind of fleet size or financial backing that Tesla and Waymo boast to enter this space. What's more, Nvidia's business model as a supplier of autonomous technology could help create a de facto industry standard for self-driving simulation — if it's widely adopted.

Creating standards for self-driving simulation could be a major step for the technology, because right now it's difficult to evaluate the quality of simulations being done by private companies, according to Nidhi Kalra, senior information scientist for the nonprofit research organization RAND Corporation.

"The problem with any simulator is that it's a simplification of the real world," Kalra says. "Even if it stimulates the world accurately, if all you're simulating is a sunny day in Mountain View with no traffic, then what is the value of doing a billion miles on the same cul-de-sac in Mountain View? I'm not saying that's what anyone's doing but without that information we can't know what a billion miles really means."

Kalra has co-authored a number of studies for RAND about self-driving technology, including one in 2016 that tried to determine how many real-world miles would need to be driven to prove that autonomous cars are safer than humans. Kalra and co-author Susan M. Paddock came to the conclusion that self-driving cars will need to be driven "hundreds of millions of

miles and sometimes hundreds of billions of miles" to make any statistically reliable claims about safety. Because of this, they wrote, companies need to find other ways to demonstrate safety and reliability.

""When a company says 'we've driven this many miles in simulation,' I think, 'Well, I'm glad you've got a simulator.'""
Simulations could serve that purpose, Kalra says, but there needs to be more context surrounding those mileage claims. "If I tell you I've played a billion miles of *Grand Theft Auto*, it doesn't make me a good driver," she says. "When a company says 'we've driven this many miles in simulation,' I think, 'Well, I'm glad you've got a simulator.'"

Kalra says it's important to be skeptical of any "simulated miles driven" milestones that companies share unless they offer more detail about what's being simulated. "Real-world miles still really, really matter. That's where, literally, the rubber meets the road and there's no substitute for it," she says.

Photo: Sean O'Kane / The Verge

Knowing that Tesla and Waymo have racked up the most miles in both simulation and in the real world helps set the table for the discussion about who has the "most" data. But that knowledge isn't enough on its own to really determine who has the ultimate advantage. If Tesla does crack full self-driving without LIDAR, it could theoretically push a software update to its customers that flips the switch.

But how will the company prove that it's safe? Tesla does have its own small fleet of test cars registered with the California DMV, but they drove zero miles in 2017. And for all the miles the company has racked up with the current version of Autopilot on the road via its customer

fleet, most of those have been spent gathering data about real-world application of semi-autonomous tech — tech that is once again under investigation by the National Transportation Safety Board after another driver died using the feature.

Waymo might be in a better position to prove safety via real-world miles once it has a fleet of cars in the thousands, but that could be difficult since it's still limited to a handful of locations. Even in the current lax regulatory environment for self-driving testing, progress in expanding those efforts will take time.

"There's no perfect metric or definition for how "safe" these cars are"
Another problem is how to define "safety" to begin with. the only common metric applied to all these companies equally is something called "disengagements," which tracks how many times safety drivers have to regain control over a car's autonomous systems. It's an imperfect metric, too: it's only consistently cataloged by the California DMV, and it's been proven easy to fudge because it has such a loose definition.

When it comes time for these companies to prove to regulators or customers that they've developed fully self-driving tech, the most likely metric that will be used to judge whether a company has developed a full-stop fully self-driving car is whether or not they're as safe or safer than human driving. How to define that — the rate of crashes per X miles, injuries per X miles, or even deaths per X miles — is another question.

As Kalra and Paddock point out in their study, this will be hard to prove in real-world terms. But Kalra thinks it can't be proven by simulation alone — at least not without a more thorough and open understanding of the quality and rate of data being collected. "We're probably going to see this technology deployed before we have conclusive evidence about how safe it is," she says. "This is the rub. We can't prove how safe self-driving cars are until we all decide to use them."