Self-driving cars are being put on a data diet

A Google self-driving car.

reader comments

19 with

For self-driving-car developers, like many iPhone and Google Photos users, the growing cost of storing files on the cloud has become a nagging headache.

Early on, robocar companies pursued a brute-force approach to maximize miles and data. “We could take all the data the cars have seen over time, the hundreds of thousands of pedestrians, cyclists, and vehicles, [and] take from that a model of how we expect them to move,” said Chris Urmson, an early leader of Google’s self-driving project, in a 2015 TED Talk.

Urmson spoke at a time when autonomous vehicle prototypes were relatively few and the handful of companies testing them could afford to keep almost every data point they scooped up from the road. But nearly a decade later, Google’s project and many others have fallen far behind their own predictions of the timeline for success. Growing fleets, fancier sensors, and tighter budgets are forcing companies working on robotaxi and robofreight services to get pickier about what stays on their servers.

The newfound restraint is a sign of maturity for an industry that has begun moving people and goods without drivers in a

few cities when the weather’s good and streets are relatively clear, but is yet to generate profits. Figuring out which data to keep and which to discard could be key to expanding service to more locations as companies train their technology on the nuances of new areas.

Advertisement

“Having tons and tons more data is valuable to some extent,” says Andrew Chatham, who oversees the computing infrastructure at the Google driverless tech spinout Waymo. “But at some point, having more interesting data is important.” Rivals including Aurora, Cruise, Motional, and TuSimple are also keeping closer watch on their data stores.

The trend could spread at a time that driverless projects are facing pressure to control spending after years of losses. Companies ranging from General Motors, which owns robotaxi service Cruise, to Waymo-owner Alphabet are in the midst of wide-ranging cost-cutting this year—including mass layoffs—as sales in core businesses slow due to a shaky economy. Meanwhile, cheap and easy funding is drying up for autonomous vehicle startups.

Naturally, all spending is under scrutiny. Amazon Web Services charges about 2 cents per gigabyte monthly for its popular S3 cloud storage service, a price that adds up quickly on data-intensive projects, and doubles in some cases when factoring in bandwidth costs to transfer data. Intel estimated in 2016 that each autonomous vehicle would generate 4,000 gigabytes of data per day, a volume that would cost about $350,000 to store for a year at Amazon’s current prices.

Chucking data might sound perverse for the tech industry. Companies like Google and Meta have long been ridiculed and even penalized for collecting everything they can—including users’ locations, clicks, and searches—with the idea that greater understanding of behavior leads to better-designed services. The mantra created a culture of collecting data

despite any clear application. For instance, Google CEO Sundar Pichai acknowledged in 2019 that only “a small subset of data helps serve ads.”

Self-driving-car developers initially held a similar philosophy of data maximization. They generate video from arrays of cameras inside and outside the vehicles, audio recordings from microphones, point clouds mapping objects in space from lidar and radar, diagnostic readings from vehicle parts, GPS readings, and much more.

Some assumed that the more data collected, the smarter the self-driving system could get, says Brady Wang, who studies automotive technologies at market researcher Counterpoint. But the approach didn’t always work because the volume and complexity of the data made them difficult to organize and understand, Wang says.

https://arstechnica.com/?p=1938865