Every Tesla is providing
reams of sensitive data
about its driver's life

By MARK HARRIS

THE RADICAL SCOPE of TESLA'S DATA HOARD

YOU WOULDN'T HAVE SEEN A SINGLE TESLA in the Beidaihe District of Qinhuangdao, China, this summer. Officials banned Elon Musk's popular electric vehicles from this beachfront resort for two months while it hosted the Communist Party's annual retreat, presumably fearing what the cars' built-in cameras might capture and send back to the United States. • This might seem to be undue paranoia, but Tesla vehicles do collect an enormous amount of information. That's well illustrated by events in Florida, where Tesla faced a negligence lawsuit after two young men died in the fiery crash of a 2014 Model S. The father of the teen driver sued Tesla for producing a battery prone to ignite. As part of its defense, the company submitted a historical speed analysis showing that in the months before the crash the car had been routinely driven dangerously fast. This information was quietly captured by the car and uploaded to Tesla's servers.

Yaw Rate (degs)

Driver Front Ai
Driver Knee Ai
Driver Retracto
Driver Lap Pre
Driver Switcha
Driver Side Sea
Passenger Fron
Passenger Fron
Passenger Act
Passenger Kne
Passenger Ret
Passenger Lap
Passenger Swi
Passenger Side
Inflatable Curta
Inflatable Curta
Second Row R
Second Row L
Second Row R
Second Row R
Second Row S
HV Battery Dis

0x50a2
0x50a3
0x50a4
0x50a5
— ReR
— FrL
— FrR
— ReL

RAW

arbitration_id

2000
1500
1000
500
0

950ms   0ms   50ms   100ms   150ms
timestamp

g/ms

TESLA

CL versus HRL

ESP_wheelSpeed [km/h]
80
60
40
20
0

12s   13s   14s   15s   16s   17s   18s   19s
timestamp

0x50a2
0x50a3
0x50a4
0x50a5
— ReR
— FrL
— FrR
— ReL

FF

56

0000   FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
0028   FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
0056   FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
0084   FF FF FF FF FF FF FF FF 00 00 01 13 FF FF FF
0112   FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
0168   FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
0196   FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
0224   FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
0252   FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
0280   FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF

Driver Front Airbag Stage 1   Deployment Commanded
Driver Front Airbag Stage 2   Deployment Commanded
Driver Knee Airbag   Deployment Commanded
Driver Retractor   Commanded

LOG
140
120
100
80
60
40
20
0

:28:00   :28:30   :29:00   :29:30   :30:00   :30:30   :31:00
timestamp

S1

Acceleration (g)
40
20
0
-20
-40
-60
-80

-5   0   5   10   15   20   25   30   35   40   45   50
Time (ms)

Although there is no evidence that Tesla collects any data beyond what customers agree to in the company's terms of service, it's not been clear to outsiders exactly what data Tesla vehicles collect and how the company uses this information. So you really have to wonder whether owners or the company are in the driver's seat when it comes to accessing and exploiting those data.

These questions are relevant to you even if you don't own a Tesla. Every new production vehicle has a gaggle of sensors, often including cameras and radars that capture data about their drivers and their surroundings. There is now a worldwide connected car–data industry, trading in anonymized vehicle, driver, and location data aggregated from billions of journeys made in tens of millions of vehicles from all the major automotive equipment manufacturers. But none of those vehicles seem to store that information and send it back to the manufacturer as regularly, or in such volume, or have been doing so for as long, as those made by Tesla.

"As far as we know, Tesla vehicles collect the most amount of data," says Francis Hoogendijk, a researcher at the Netherlands Forensic Institute who began investigating Tesla's data systems after fatal crashes in the United States and the Netherlands in 2016.

While much still remains unclear, crash investigations by the U.S. National Transportation Safety Board (NTSB), reports from the U.S. National Highway Traffic Safety Administration (NHTSA), and Tesla's own documents help to reveal the data these vehicles collect and what the company does with them. Here's a short accounting of what's known.

T O   S T A R T , T E S L A S H A V E event data recorders (EDRs), as is true for most new vehicles. To assist in crash investigations, these "black box" recorders save a scant 5 seconds of information collected before an accident, including speed, acceleration, brake use, steering input, and the operation of automatic-brake and stability controls.

Tesla vehicles keep a permanent record of these data—and many more—on a 4-gigabyte SD or 8-GB microSD card located in the "infotainment" computer, part of the car's media control unit (MCU). These time-stamped "gateway log" files also include information about seatbelt use and various parameters of the car's self-driving system, including whether drivers had their hands on the steering wheel.

Because the gateway logs use data from cars' standard Controller Area Network (CAN) buses, they can include the unique vehicle identification number. But there is no evidence to suggest that these logs include information from the car's GPS module or cameras.

When an owner connects a Tesla to a Wi-Fi network—for instance, to download an over-the-air update that adds new features or fixes bugs—the gateway logs are periodically uploaded to Tesla. Judging from that Florida lawsuit, the company must have ways to link that data to the originating vehicle. (Tesla did not respond to requests for clarification on this or other issues.)

These gateway log files are just the tip of the data iceberg, though. Tesla's Autopilot computer takes inputs from vehicle cameras to handle driver-assistance functions, including cruise control, lane-keeping, and collision warnings. If owners plug their own USB thumb 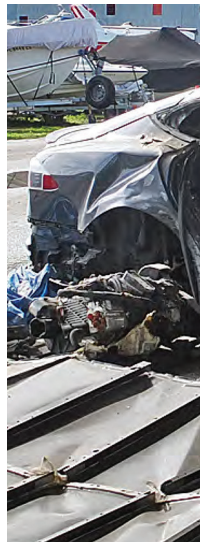drives into the car, they can make live dashcam recordings and set up Sentry Mode to record the vehicle's surroundings when parked. These recordings do not appear to be uploaded to Tesla.

But there are occasions in which Tesla vehicles do store images and (in 2016 and later models) videos that are then shared with the company. These Autopilot "snapshots" can span several minutes and consist of up to several hundred megabytes of data, according to one engineer and Tesla owner who studied Tesla's data-collection process using salvaged vehicles and components. This engineer tweets about his findings using the pseudonym Green (@greentheonly). These snapshots also include high-resolution log data, similar to that captured in the gateway logs but at a much higher sampling rate—up to 50 times per second for wheel-speed information, notes Hoogendijk.

Such data snapshots are triggered when the vehicle crashes or when certain other conditions are met. These can include anything that Tesla engineers want to learn about, such as particular driving behaviors, or specific objects or situations being detected by the Autopilot system.

According to Green, GPS location data are always captured for crash events and sometimes included for other snapshots. Like gateway log data, snapshots are uploaded to Tesla when the car connects to Wi-Fi, although those triggered by crashes will also attempt to upload over the car's 4G cellular connection. Green indicates that once a snapshot has been successfully uploaded, it is deleted from the Autopilot computer's onboard 32-GB storage.

In addition to the snapshots, the Autopilot computer also records a complete trip log every

leading EV automaker doing with all that information? Tesla doesn't say exactly, but it's not hard to surmise.

I N 2019, ELON MUSK stood up at a Tesla Day event that was devoted to automated driving and said, "Essentially everyone's training the network all the time."

He was referring to Tesla's suite of assistive and semi-autonomous technologies, collectively known as Autopilot, which is the most widely deployed—and most controversial—driver-assistance system on the road today. While many drivers love it, the technology has been involved in hundreds of crashes, some of them fatal, and is currently the subject of a comprehensive investigation by the NHTSA.

Most companies working on automated driving rely on a small fleet of highly instrumented test vehicles that are outfitted with high-resolution cameras, radars, and laser-ranging devices. Some of these vehicles have been estimated to generate 750 megabytes of sensor data every second, providing a rich
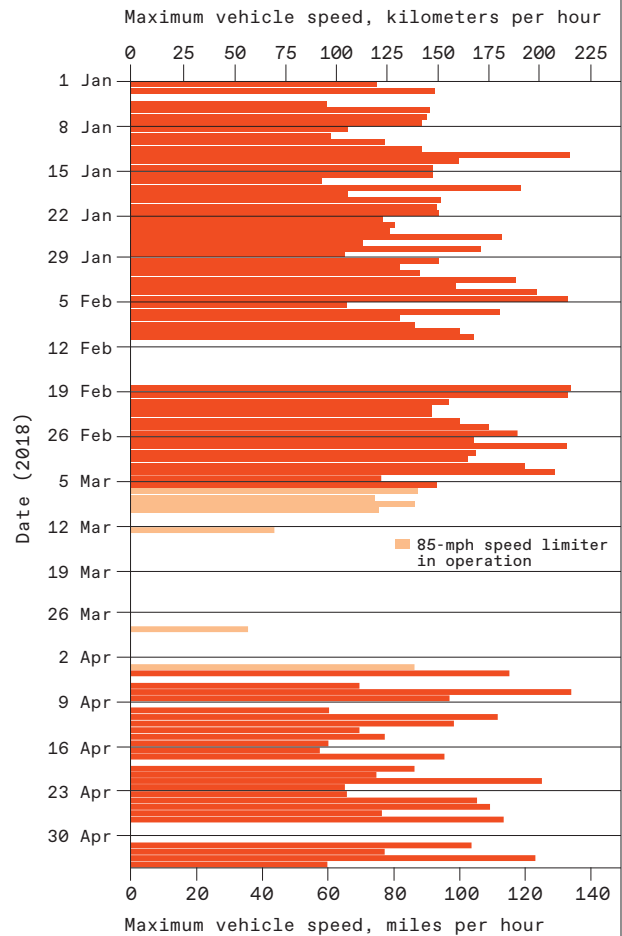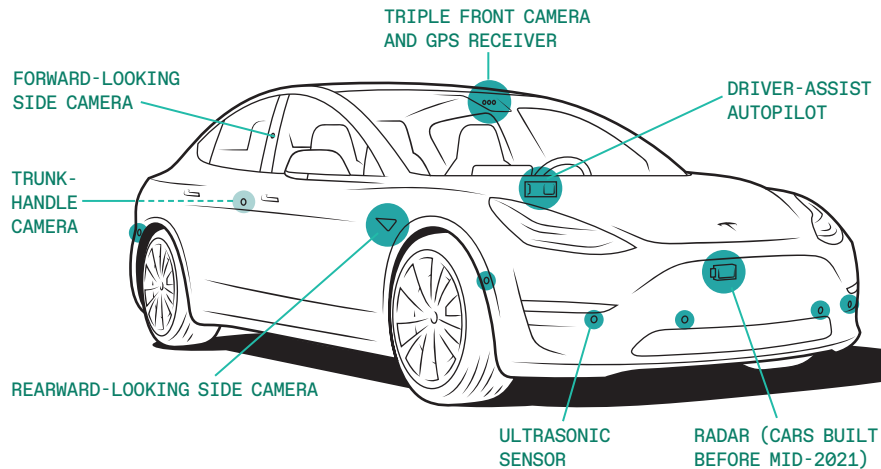
time a mid-2017 or later Tesla is shifted from Park to Drive, says Green. Trip logs include a series of GPS coordinates of the path followed, speeds, road types, and when or whether Autopilot was activated. Green says that trip logs are recorded whether or not Autopilot (or the car's full self-driving mode) is used. Like the snapshots, trip logs are deleted from the vehicle's computer after being uploaded to Tesla.

What happens to this treasure trove of data? Tesla has sold about 3 million vehicles worldwide, the majority of which are phoning home daily. They have provided the company with billions of kilometers of real-world driving data and GPS tracks, and many millions of photos and videos. What is the world's

Maximum vehicle speed, kilometers per hour

85-mph speed limiter in operation

Maximum vehicle speed, miles per hour

TESLA'S MOUNTAIN OF DATA

The company's vehicles carry eight cameras: three front-facing, two on each side, and one on the rear-trunk handle. A GPS module and various sensors add to the data being generated and recorded.

TRIPLE FRONT CAMERA AND GPS RECEIVER

FORWARD-LOOKING SIDE CAMERA

DRIVER-ASSIST AUTOPILOT

TRUNK-HANDLE CAMERA

REARWARD-LOOKING SIDE CAMERA

ULTRASONIC SENSOR

RADAR (CARS BUILT BEFORE MID-2021)

(1) The vehicle's event data recorder will save measurements taken in the seconds before an accident, creating gateway logs. The car will attempt to upload this information to the company using the cellular network.
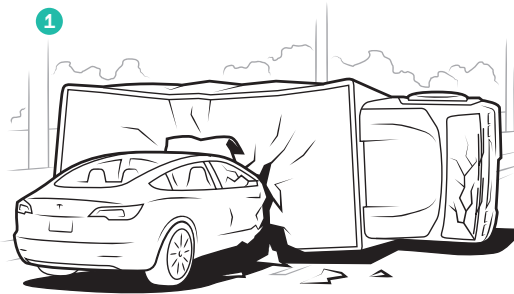
(2) When connected to Wi-Fi, say, while charging at home, the vehicle may periodically upload gateway logs and Autopilot snapshots.
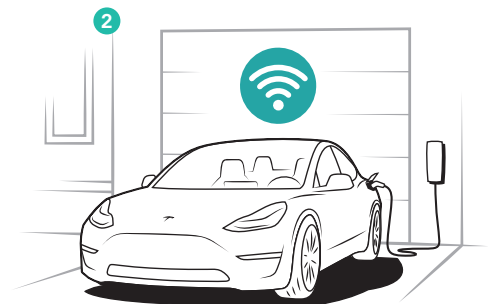
(3) Using their own USB storage devices, owners can capture dashcam video and images recorded in Sentry Mode while the car is parked.
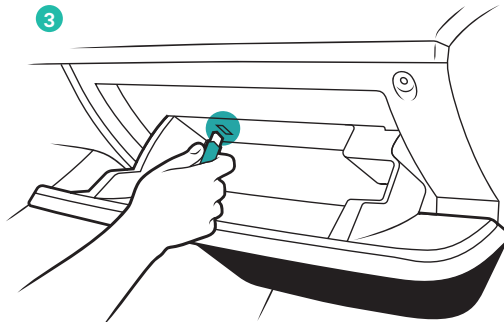
(4) Tesla may program the car to save video and other sensor data when the car encounters rare events of interest to the company's engineers.
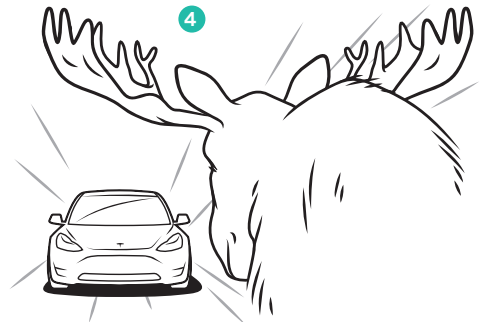
seam of training data for neural networks and other machine-learning systems to improve their driving skills.

Such systems have now effectively solved the task of everyday driving in a variety of weather conditions and road types, says Henry Liu, director of Mcity, a public-private mobility-research partnership at the University of Michigan.

"But right now, automated vehicles are one to two magnitudes below human drivers in terms of safety performance," says Liu. "And that's because current automated vehicles can't handle the curse of rarity: low-frequency, long-tail, safety-critical events that they just don't see enough to know how to handle."

Tesla's bold bet is that its own customers can provide the data needed to boost self-driving cars to superhuman levels of safety. Many are happy to do so—willing participants in the development of technology that they have been told will one day soon

allow them to sit back and enjoy being driven by the car itself.

In presentations over the past few years, Musk and Tesla's former head of AI, Andrej Karpathy, detailed the company's approach, including its Shadow Mode. Here the car's Autopilot computer is not controlling the car, but it is simulating the driving process in parallel with the human driver. When its own predictions do not match the driver's behavior, the Autopilot computer might trigger the recording of a data snapshot for later uploading to Tesla.

Tesla engineers can then use these results to better train its neural networks for autonomous driving. Or they may notice through these snapshots that their system is failing, for instance, to properly identify road signs obscured by trees.

In that case, engineers can create special software known as a detector and download it to some Tesla vehicles. If that

Illustration by James Provost

detector thinks it spots such a road sign, it will capture images from the car's cameras for later uploading. Tesla engineers would then soon receive thousands of images, which they would use to improve the detector and eventually roll it out to all production vehicles. "I'm not exactly sure how you build out a data set like this without the fleet," said Karpathy.

LIU IS BULLISH on Tesla's approach to leveraging its ever-growing consumer base. "I don't think a small...fleet will ever be able to handle these [rare] situations," he says. "But even with these shadow drivers—and if you deploy millions of these fleet vehicles, that's a very, very large data collection—I don't know whether Tesla is fully utilizing them because there's no public information really available."

One obstacle is the sheer cost. Karpathy admitted that having a large team assess and label images and video was expensive. He said that Tesla was working on detectors that can train themselves on video clips captured in Autopilot snapshots. It seems the company must now have this capability, because in June Tesla laid off 195 people working on data annotation.

For all the promise of Tesla's fleet learning and the enthusiastic support of many of its customers, Autopilot has yet to prove that it can drive as safely as a person can, let alone be trusted to operate a vehicle without supervision. And there are other difficulties looming. Karpathy left Tesla in mid-July, and the company continues to face the possibility of NHTSA issuing a recall for Autopilot. This would be a terrible blow for the company but would likely not halt its harvesting of customer data nor prevent the continued deployment of Autopilot outside the United States.

Tesla's use of fleet-vehicle data to develop Autopilot echoes the user-fueled rise of Internet giants like Facebook, Google, and Twitter. The more its customers drive, so Musk's story goes, the better the system performs.

But just as other tech companies have faced scrutiny for mining customer data, Tesla, too, is beginning to see a backlash. People are beginning to wonder about all the information these cars collect and what happens when other entities, including the government, seek access to it.

FOR TESLAS BUILT since mid-2017, "every time you drive, it records the whole track of where you drive, the GPS coordinates and certain other metrics for every mile driven," says Green. These trip logs and the data snapshots captured by the Autopilot system are stripped of vehicle-identification numbers and given a temporary, random ID number when uploaded to Tesla, says Green. But he notes that temporary IDs can persist for days or weeks, connecting all the uploads made during that time.

Given that some trip logs will also likely record journeys between a driver's home, school, or place of work, expecting anonymity is unrealistic, says John Verdi, senior vice president of policy at the Future of Privacy Forum: "If an entity is collecting, retaining, [and] sharing historical location data on an individualized level, it's extraordinarily difficult to de-identify that, verging on impossible."

"Individuals ought to think about their cars more like they think about their cellphones."

—JOHN VERDI, senior vice president of policy at the Future of Privacy Forum

Tesla, like all other automakers, has a policy that spells out what it can and cannot do with the data it gets from customers' vehicles, including location information. This policy states that while the company does not sell customer and vehicle data, it can share that data with service providers, business partners, affiliates, some authorized third parties, and government entities according to the law.

Owners can buy a special kit for US $1,400 that allows them to access data on their own car's event data recorder, but this crash-related data represents just a tiny subset of the information the company collects. Owners living in California and Europe benefit from legislation that requires Tesla to provide access to more data generated by their vehicles, although not the Autopilot snapshots and trip logs, which are nominally anonymized.

Once governments realize that a company possesses such a trove of information, it may be only a matter of time before they seek access to it. "If the data exists...and in particular exists in the domain of somebody who's not the subject of those data, it's much more likely that a government will eventually get access to them in some way," says Bryant Walker Smith, an associate professor in the schools of law and engineering at the University of South Carolina.

This is not necessarily a terrible thing, says Smith, who suggests that such rich data could unlock valuable insights into which roads or intersections are dangerous. The wealth of data could also surface subtle problems in the vehicles themselves.

In many ways, the data genie is already out of the bottle, according to Verdi. "Individuals ought to think about their cars more like they think about their cellphones," he says. "The auto industry has a lot to learn from the ways that mobile-phone operating systems handle data permissions.... Both iOS and Android have made great strides in recent years in empowering consumers when it comes to data collection, data disclosure, and data use."

Tesla permits owners to control some data sharing, including Autopilot and road-segment analytics. And if they want to opt out of data collection completely, they can ask Tesla to disable the vehicle's connectivity altogether. But this would mean losing features such as remote services, Internet radio, voice commands, and Web-browser functionality, even safety-related over-the-air updates.

Green says he is not aware of anyone who has successfully exercised this nuclear option. The only real way to know you've prevented data sharing, he says, is to "go to a repair place and ask them to remove the modem out of the car."

Tesla almost certainly has the biggest empire of customer and vehicle data among automakers. But even though Tesla dominates the discussion around connected cars, others are not far behind. Elon Musk's insight—to embrace the data-driven world that our other digital devices already inhabit—is rapidly becoming the industry standard. And when our cars become as smart as our phones, it is hardly surprising that they suffer the same challenges around surveillance, privacy, and accountability. ■