In 2015, an estimated 38,300 people were killed and 4.4 million injured in motor vehicle crashes in the US alone, with damage costs exceeding $400 billion. In contrast, there were no commercial aviation fatalities and 136 civil aviation deaths in the US in 2015. More than 90 percent of auto accidents result from human impairment such as drunk driving or road rage, errant pedestrians, or just plain bad driving. This has motivated the development of automotive safety systems that use sensors, cameras, lasers, and radar to monitor the vehicle and its surroundings and to prompt either the vehicle or driver to take corrective action to avoid an accident. These systems include forward-collision warning and automated braking, lane-departure warning, rear cross-traffic alerts, drowsiness detection, adaptive headlights, pedestrian detection, and automated parking assistance.

Autonomous vehicles (AV) promise to further reduce collisions dramatically by removing the human element altogether. More than 30 tech companies and auto manufacturers are working on AVs and expect to start deploying them within the next three to five years. Some safety experts have cautioned against allowing AVs on the road until collision-avoidance systems achieve perfection, though this wouldn't necessarily eliminate all accidents. However, even less-than-perfect systems could be compelling enough if they substantially reduce accidents, considering that every percentage point improvement equates to 400 lives saved and 40,000 injuries avoided annually.

It could be some time before AVs are commonplace. Although the physics associated with driving is well understood and sensor technology continues to improve, machines must learn human behavior and norms, which can be complex. In addition, the human interface in AVs is still evolving as manufacturers work to enhance the software and its security. Nevertheless, federal transportation officials appear receptive to self-driving

Equity, Safety, and Privacy in the Autonomous Vehicle Era

Vasant Dhar, New York University

Big data from onboard vehicular systems can be used to help determine liability in accidents, streamline insurance pricing, motivate better driving practices, and improve safety—all with potentially minimal impact on privacy.
technology and recently published first-ever guidelines on safety best practices, existing and potential new regulatory tools, and the adoption of uniform practices across states (www.transportation.gov/AV).

Regardless of whether and when vehicles become fully autonomous, the data and analytics provided by current-day onboard systems are sufficient to determine fault in many accidents, and will improve with time. This capability, which already exists in AVs but is also installable on human-operated vehicles, demands that we revisit existing insurance laws and practices regarding liability and compensation, which have been in place for decades and are based on outdated assumptions about available evidence. Equally important, the data obtained from this technology can yield novel insights into how to prevent accidents and save lives, especially as human-operated and driverless vehicles begin to share the road. At the same time, we must ensure that vehicles’ new data-gathering abilities don’t lead to the violation of individual privacy, specifically the fundamental human right “to be left alone” as articulated by Samuel Warren and Louis Brandeis more than a century ago.²

**AV DATA AS THE BASIS FOR INSURANCE EQUITY AND PRICING**

Auto insurers rely on various forms of individual and demographic data—for example, your age, sex, and marital status; what type of vehicle you drive; where you live; how often and far you drive; your driving record; and your credit history—to assess customer risk factors and price insurance efficiently. However, when it comes to determining who’s at fault in an accident, which insurer should pay for it, and what the amount should be, often there are no easy answers based on the available evidence. Data traditionally used to establish fault such as physical damage, police reports, and testimonies from victims and witnesses are often insufficient to achieve incontestable results.

Consequently, a complex patchwork of insurance laws has emerged across the US in which the determining factors for financial liability can vary considerably. Most states have “at-fault” or tort laws that let accident victims sue drivers at fault for medical expenses, material damages, lost wages, and sometimes pain and suffering, with different monetary limitations across states. Given that the data used to assign fault is often incomplete, resolving disputes about the degree of liability and the amount of compensation can be lengthy, costly, and frustrating for all parties. This problem led a dozen states to adopt “no-fault” laws that allow cases to bypass the litigation process by requiring the parties’ insurers to cover their own policyholders’ expenses regardless of who’s ultimately held responsible. Such laws ensure quick payments (subject to monetary thresholds), but they don’t address fairness, nor do they discourage bad driving behavior. They also lead to fraudulent claims by drivers in collusion with other unscrupulous agents, which can be difficult to disprove and raise premiums for law-abiding drivers.

Because fairness in insurance law has been elusive, economists and policymakers have focused largely on pragmatic steps to reduce the number and severity of accidents and ways to spread their cost equitably through society. However, quantifying the total costs of accidents, including preventive measures,³ is difficult with limited data, and the larger issues of fairness and liability remain unaddressed.

The availability of big data obtained from onboard vehicular systems and roadway sensors changes the calculus of liability. Such data provides sufficient details about the circumstances of an accident—ranging from physical

---

If AV data can be used to establish fault in accidents with high precision, it isn’t too much of a stretch to envision many liability decisions based on this data becoming largely automated.
to avoid practices such as selling personal data or arbitrarily assigning a higher level of risk to policyholders to charge a higher premium, there would have to be explicit restrictions on the gathering and appropriate use of such data, including the granting of policyholder consent.

How can we enable the desirable uses of AV technology without intrusions on our privacy?

SAFETY AND PRIVACY TRADEOFFS

To understand the privacy risks associated with AV data, it’s important to understand the nature of the data and what its analysis might yield. Onboard vehicular systems continuously record data, which can be translated into “instances” describing the vehicle’s movement such as turning, accelerating, braking, changing lanes, and so on. Combining such instances can reveal unsafe behaviors by the vehicle operator—for example, weaving, following too closely, or “rolling through” stops. Combining instances across many drivers can reveal general patterns such as “tired drivers have more accidents,” in which the degree of tiredness is estimated from the low-level sensor data.

Driving patterns can serve as useful coaching aids—for example, to inform a driver about the risks of driving late at night or changing lanes frequently, and ways to ameliorate those risks. However, if the instance-level data is in the hands of an insurer, auto manufacturer, or other entity, the driver might have little or no control over its “secondary use” or sale to third parties, which could have undesirable consequences. For example, if you collide with another driver, could that driver subpoena data from your insurer to support his claim that you appeared “tired” prior to the accident? What if your insurer sells data to a disreputable marketer or its servers are hacked?

Fundamental to addressing these concerns is where the data is stored and the analytics are conducted. If privacy is a top priority, one solution is to retain data on the onboard “black box” device and perform the analytics locally as well. In this case, to carry out risk assessment and premium pricing, insurers could transmit aggregate-level driving statistics to the device, which could then extract individual driver data for comparative analysis. This would ensure privacy and prevent unauthorized secondary use or sale of the data. In expectation of a lower premium, drivers could, however, elect to share onboard vehicular data with their insurer—preferably stripped of sensitive components such as location. The precedent for this type of privacy-preserving cryptographic protocol exists in several European countries, where drivers pay mileage-based road taxes without revealing personal data. Of course, sharing instance-level data exposes the driver to potential data misuse.

Some scholars argue that a decision to keep this and other types of data private could be negatively viewed by stakeholders as a signal of high risk. For example, law professor Scott Pettet suggests that so-called voluntary data disclosure implicitly erodes privacy.

In an economy with robust signaling, those with valuable credentials, clean medical records, and impressive credit scores will want to disclose those traits to receive preferential economic treatment. Others may then find that they must also disclose private information to avoid the negative inferences attached to staying silent. This unraveling effect creates new types of privacy harms, converting disclosure from a consensual to a more coerced decision.

Although some policyholders such as new drivers might feel pressure to share their detailed driving data with insurers as proof of safe driving practices, aggregate statistics that are compared with a broader population along a few key dimensions should be sufficient for pricing the large majority of drivers fairly. For those who elect to share details, driving data is arguably less sensitive than, say, medical or financial data, which can be indicative of a chronic health condition or high credit risk, respectively. In contrast, drivers can readily alter their driving practices, and those who demonstrate safer driving behavior could probably limit data sharing and its associated risks with little economic penalty.

AFTERSHOCK

There will undoubtedly be significant trial and error before AVs are a common feature on the road. But regardless of the degree of autonomy vehicles are ultimately entrusted with, onboard systems will have increasingly powerful data-gathering capabilities that could be used to significantly improve the transportation ecosystem. Indeed, the analytics from today’s fairly primitive systems already have the capacity to help determine liability in accidents, streamline insurance pricing, motivate better driving practices, and improve vehicle safety.

Despite the fine-grained nature of the data they collect, onboard vehicular systems need not expose drivers to serious privacy risks as long as we regulate the use of such data carefully. Technical and legal solutions are available to anonymize the data and prevent its unauthorized use. Just as wearable devices are enabling individuals to improve their health and fitness, vehicles could become yet another rich source of information for personal knowledge discovery by the “quantified self,” while also making the world a safer place for everyone to navigate.

REFERENCES


VASANT DHAR is a professor of information systems in the Leonard N. Stern School of Business and director of graduate studies at the Center for Data Science at New York University, as well as editor in chief of Big Data. Contact him at vdhар@stern.nyu.edu.