

# CAN AUTOMATED VEHICLES PROVE THEMSELVES TO BE SAFE?

Fewer fatalities are central to the case for automated vehicles. Companies will need to cooperate more with each other and with government to fulfill that promise.

With more than 40 companies racing to put automated vehicles (AVs) on the road, why does commercial availability seem always just over the horizon? Safety is among the most compelling arguments for AVs, and yet safety remains a major hurdle for their development. In 2017 in the United States, more than 37,000 people died from car (and light truck) crashes and 2.7 million people were injured; another 5,500 pedestrians were killed by cars. Many collisions involving conventional cars involve some element of driver error. AVs, the argument goes, will improve safety because they replace flawed humans with something better: an intelligent machine that doesn't get tired, drunk, or distracted; can make superior split-second decisions and communicate with other vehicles; and can see and anticipate hazards that human drivers cannot.

The implication is that AVs will begin to deliver an aggregate public health benefit as soon as they can be shown to be safer than human-driven vehicles. But clearing that bar, and identifying those companies that clear the bar, is proving quite difficult. Different developers use different technologies or use similar technologies in different ways,

resulting in a diversity of approaches to measuring safety. Progress in measuring safety has been unclear, in part because there is no clear agreement on how to measure it meaningfully—tools and protocols providing evidence that can justify public confidence are still in development. We believe that a convergence to common approaches for measuring and communicating safety is needed, to both ensure AVs' safe operation and demonstrate it to policy-makers, as well as to the public who will use AVs, share roads with AVs, and whose well-being these vehicles will either safeguard or put at risk.

## Proving improved safety

AVs must prove themselves in the context of the public's expectations, which are grounded, of course, in people's experience with the safety of traditional automobiles. Automobile safety today reflects a history of steady improvement, with the occasional setback (for example, injuries related to airbag deployment) overcome by a safer, more trustworthy innovation. The maturation of automotive technology over decades and widespread compliance with national standards for crashworthiness and occupant

*Continued on page 74 →*



*Rathaus Steglitz, Berlin*



## Chris M. Forsyth: *The Metro Project*

Underground public transit systems play a pivotal role in the day-to-day functioning of cities around the world. They also contribute to a city's unique visual identity. Captivated by the overlooked beauty of his hometown transit system, Chris M. Forsyth began photographing the Montreal Metro in 2014. He went on to photograph the metro systems of Berlin, Munich, and Stockholm, creating *The Metro Project* series. He purposefully kept his photographs devoid of people to draw attention to the stations' art and architecture. In the midst of the COVID-19 pandemic, these images take on a new, unintended meaning. When will people in many parts of the world feel safe returning to mass transit?

From Paris to New York and from Moscow to London, every metro system has its own character and history. Montreal's system stands out because, since it opened in 1966, transit authorities have held competitions among Canadian architects for the design of each station, ensuring that every station is built by a different architect in a distinctive style. Stockholm's system, called "the longest art gallery in the world," features artwork in more than 90 of its 100 stations. Operating since 1902, Berlin's U-Bahn system is the oldest in Germany and is known for its vibrant yellow trains. Munich's system has a more subdued aesthetic, with art and bright color in some recently added stations. Instead of hurrying through transit stations, Forsyth hopes his photographs will inspire commuters to slow down and take time to appreciate them.

Chris M. Forsyth is a Montreal-based photographer who is currently working on a project titled *Keep Direction By Good Methods*, documenting the physical remains of the American transcontinental lighted airway system, which sought to visually guide pilots across the country at nighttime in the early days of aviation.





Stadion, Stockholm

protection provide a common, if imperfect, floor for conventional vehicle safety. Now, highly or fully automated vehicles change the picture. Achieving public trust will require equaling if not exceeding the expectation for vehicle safety implicit with conventional vehicles.

Traditional federal safety regulations, embodied in the Federal Motor Vehicle Safety Standards, do not address automated driving systems. As one of us (Fraade-Blanar) noted in a 2017 RAND Corporation report titled “Autonomous Vehicles and Federal Safety Standards: An Exemption to the Rule?,” an AV could be designed to fully conform to all federal safety regulations, yet it could “drive headfirst into the nearest wall, and such a vehicle would not require any exemptions to be sold in the United States. In other words, an AV could be completely compliant and still unsafe.” Safety researchers, with decades of historical data to draw from, can use crash rates per mile to assess vehicle safety or in-vehicle-technology safety for conventional automobiles. However, it would take hundreds of millions (if not billions) of AV vehicle miles traveled to create

statistically meaningful crash rates for AVs, because crashes, especially fatal crashes, occur relatively rarely for all types of vehicles. The safety argument for AVs cannot be made on the basis of compliance with current federal safety standards or through a simple comparison with the record of safety for conventional cars.

Discussions about AV safety are confounded by a surprising problem: there is no standard way to define safety, in general or for transportation. The industry, as well as government officials and members of the public, need sensible measures and a common language. Marketing language—such as calling a feature an “autopilot” or “chauffeur”—often fails to explain what the system can actually do or not do. Any given AV can be operated safely under certain conditions and certain expectations of human involvement (e.g., being alert and ready to take control at any time). Since these are model-specific, what does autopilot really mean? Explaining the safety qualities of a vehicle to the public and policy-makers will require much more careful and consistent characterization.



Sherbrooke, Montreal

To begin, the safety of an AV cannot be determined by design specifications alone. Safety involves the AV's behavior, and thus extends into a realm formerly pertaining not to the car but to the ability of human drivers to behave safely on the road. AVs will be safe when they can travel without mishap in mixed traffic on public roads, an ecosystem that includes other vehicles—which for the foreseeable future will primarily be driven by people—as well as pedestrians, bicyclists, and other road users. Many factors combine to make AV safety in mixed traffic challenging, including the extent of communication between vehicles and between vehicles and objects around them (such as traffic signals), and AVs' ability to both avoid crashes and avoid causing a crash. One response to these challenges has been to build conservatism into automated driving systems. The most common AV response to confusion or surprise is to stop. However, simply stopping can have unintended consequences, as more than half of AV crashes reported in California in 2018 involved a rear-end collision. Thus, ensuring AV safety is up against

the twin challenges of complex automation design and the absence of a commonly agreed-upon definition of safety to use as a target.

### Accentuating the positive

Given that AVs offer potentially huge economic rewards, the open frontier for their development and testing has attracted lots of companies. As the National League of Cities put it, the federal government's vision for regulating AVs, which took shape between 2013 and 2017, "embraces a permissive environment marked by regulatory restraint and heavy trust in AV developers.... There is ... an expressed desire to include as many actors in this process as possible, meaning that AV testing privileges ... are open to a wide group of public and private organizations." However, safety advocates are troubled by lagging policy. The situation is complicated by the multiple levels of government (and policy-making) associated with AVs as well as the lack of information sharing between an innovating industry and potential regulators.



The characteristics of conventional cars are controlled by the manufacturer and regulated by the federal government. This includes crashworthiness and occupant protection that are covered by federal safety standards. Although many of those qualities and associated standards are not sensitive to a shift from a human driver to an automated driving system, some assume a forward-facing, actively driving human in a particular location in the vehicle. Those assumptions have been both challenged by AV developers and acknowledged, at least in principle, by the federal government. But in their current form, the federal standards do not speak to automation of the driving function, limiting the current standards' ability to govern AV safety.

The US Department of Transportation (DOT) has developed and disseminated a series of "vision documents" that have incrementally encouraged the development of AVs, promoted a voluntary and self-regulatory approach for industry, and fostered an integrated approach to automation across different modes of transportation. Specifically, DOT is promoting voluntary safety self-assessments, which companies have used to document their various approaches and philosophies around safety. For example, Nvidia's self-assessment stresses holistic command of hardware and software. Waymo emphasizes deep experience with and linkages between simulation, range, and real-world driving. Uber focuses on how it integrates its responses to the safety elements flagged by DOT. These safety self-assessments accentuate the positive, as one would expect from a voluntary, nontechnical, public document. The National Transportation Safety Board, as part of its follow-up to recent accident investigations, has called for safety assessments to be required of all developers and for more DOT oversight of the deployment of automated driving systems.

Congress, which establishes the authorities under which DOT acts, has struggled to legislate over the last couple of years. In the Senate, the ambitiously named American Vision for Safer Transportation Through Advancement of Revolutionary Technologies (AV START) Act attracted considerable attention and vigorous debate. (As did a House counterpart, the Safely Ensuring Lives Future Deployment and Research Vehicle Evolution or SELF DRIVE Act, in 2017.) The debate demonstrated how hard it is for lawmakers to understand the issues and establish mandates, and to balance encouraging an emerging technology with protecting constituents, many of whom are lukewarm on AVs.



Solna Centrum, Stockholm







## Public interests, private motives

Systematic safety testing and verification, a public need, can be at odds with the competition between AV developers, a private race. But competing on safety would incentivize companies to put private gain ahead of public good, as safety measures and safety testing become trade secrets.

Developers adopt different approaches to measuring safety, and make different selections of component technologies or, when using similar technologies, implement them differently. These developer-created measures of safety, especially the ones used in simulations and on test-tracks, are not disclosed publicly. Absent either a regulatory requirement or the emergence of new standard practices, AV companies seeking a competitive edge will likely exercise discretion with the information they develop about improving and measuring safety. Such information is valuable intellectual property, created at considerable expense to the company. Conversely, releasing information that shows problems or failure may help create an

## **Discussions about AV safety are confounded by a surprising problem: there is no standard way to define safety, in general or for transportation.**

existential threat—a product line or even a company could cease to exist. The public's safety, as well as the success of AV developers, requires established measures and formal guidelines for safety performance and testing.

Historically, government officials and researchers have used tools including crash rates per mile to assess vehicle or in-vehicle-technology safety, although these approaches have limited usefulness for AVs since the technology is so new. Once vehicles begin to be tested on public roads, standard reporting practices for passenger vehicle collisions and associated investigations can shed some light on the behavior of various AVs on the road and in a collision. What is measured and known publicly now includes cumulative mileage and collisions (at least whether they occurred and typically their circumstances). California goes further, requiring reporting of events called “disengagements”—instances when a human safety driver monitoring AV operations and the roadway takes control from the automated driving system. All disengagements occurring on public roads in California must be reported to the state Department of Motor Vehicles, but standardization of this information is lacking

because each AV developer defines, implements, and reports disengagements slightly differently. In addition, disengagement rates do not take into account the driving environment. The number of disengagements on one mile of a rural interstate highway is treated the same as for one mile of city driving. Consequently, disengagement rates cannot be used to assess performance within a company over time or between companies.

Going forward, key private-sector companies, for-profit and not-for-profit groups, and government organizations concerned about AV development need to agree on a shared set of practices based on a framework for measuring AV safety that is neutral regarding specific technologies and companies. With a well-defined, broadly applied set of safety measures, companies would be able to effectively assess safety of AVs under development, whether the testing takes place during simulation, on a test track, or on public roads, without needing to drive tens or hundreds of millions of miles. Coalitions are beginning to emerge, such as the Automated Vehicle Safety Consortium and the Partnership for Automated Vehicle Education. The groups provide a forum for diverse AV developers to talk with one another, as well as with government and the public, to coordinate and collaborate on measurement issues.

## Roadmanship

In the absence of adequate assessment methods, the strong need for safety testing is giving rise to new measures and approaches, including one that our research group at RAND has recommended: an integrated measure that we call roadmanship. Is the AV a good citizen of the roadway? Roadmanship is the ability to drive on the road safely without creating hazards and to respond well to the hazards created by others. Roadmanship should be based as much as possible on quantitatively measurable physical traits or behaviors, and should be directly interpretable in terms of official and unofficial rules of the road. It should also make a number of distinctions that have not thus far been relevant for car safety. Such a measure should distinguish the initiator from the responder; no one should get in trouble for simply trying to get out of the way (an automobile safety measure that does not currently exist). Time to collision or rapid acceleration or deceleration may provide one potential component of a broader roadmanship criterion.

Another evolving approach is the concept of a safety envelope—an imaginary buffer zone around the front, back, and sides of a vehicle. This concept builds on experience from aviation, which takes extreme care in developing and testing safety-critical software using practices that historically have set aviation apart from other transportation industries. Preliminary work on adapting safety envelopes to AVs counts how often the





*De La Savane, Montreal*

safety buffer is breached, who is at fault, the time to collision, and how quickly the buffer is restored. Intel's Responsibility-Sensitive Safety model and Nvidia's Safety Force Field are two promising, preliminary implementations of a safety-envelope approach. This approach is promising enough to be the basis of a new standards-setting effort through the Institute of Electrical and Electronics Engineers.

A complementary approach adopts a process similar to driver's license testing. Behavioral testing would assess the ability of the AV to operate in the road environment for which it was designed. This testing would be more complex for an AV than for a human driver because most current AVs are designed to operate only within narrowly defined environments, known as operational design domains (ODDs), that driver's licensing tests have not had to consider. The California PATH research organization, at the request of the California Department of Motor Vehicles, proposed in 2016 a three-part process for doing just that. A manufacturer would submit a safety plan covering the circumstances in

which the AV would be operated and associated behavioral competencies. A third-party tester would select driving scenarios for the AV to perform on a test track and then, assuming satisfactory performance, on public roads.

One challenge to assessing safety stems from the fact that some changes in driving performance will likely be due to machine learning. This sort of "black-box" improvement will be harder to demonstrate or explain to the public and governments than improvements in design or engineering. Accordingly, we have recommended formal demonstrations at regular intervals during the AV development process, when developers would show regulators, safety advocates, and potential consumers how the AV performs. Such demonstrations could allow for marking improvement over time and, if the same testing is done across companies, comparisons that are meaningful and understandable.

Any effort to create a common safety-measurement framework must be enhanced by establishing a common approach to describing the operating conditions under

which an AV is specifically designed to function. Today each AV model under development is not expected to have the capability of driving on all types of roadways under all types of conditions. From the blindingly sunny byways of the Sonoran Desert to the dense bustle of New York City, the roadway environment is simply too diverse for current technology. Instead, as noted above, each AV model has a unique operational design domain, the environment and conditions in which the AV can operate autonomously. An ODD evolves in idiosyncratic ways as the technology and automated driving system mature and gain capability for driving in different kinds of conditions. ODDs can pertain to distance from a key location, weather (the AV may be unable to drive at night or in heavy rain or fog), roadway type (such as highways only, or no one-way roads), maneuvers (for example, may need to avoid left turns), infrastructure (for example, may need to avoid roundabouts), and so on.

Devising a common way to describe ODDs would help establish a basis for comparing AVs and tracking their development. AV developers have begun to collaborate with one another on this challenge, and both the Automated Vehicle Safety Consortium and the federal government through the National Institute of Standards and Technology have begun to examine how ODDs are characterized. State and local governments might also encourage a common approach to defining ODDs as a condition for a developer to test in a given area. Providing clarity around ODDs could also help to manage consumers' expectations so that they would know exactly in what conditions and to what destinations their particular AV may drive.

## Safety in numbers

Testing on public roads, along with the accumulating safety incidents and the public scrutiny and local regulation that grow in response, has fed an emerging sense of shared fate among AV developers. They have increasing interest in finding common ground—albeit through different approaches. For example, Uber has published a flow chart of qualitatively described processes that show evidence of safety, to support what it terms a safety case framework. Waymo has released a detailed discussion of how first responders and law enforcement can handle an AV post-crash. Both frameworks could be adopted by others. A coalition including Intel, Volkswagen, BMW, Audi, Baidu, and Aptiv has articulated qualitatively what an AV's safety and performance capabilities should be and what can be demonstrated in different settings. These emerging clusters of developers suggest the possibility of convergent approaches, but whether they translate into measurable progress toward an agreed-upon concept of safety remains to be seen.

Capitalism thrives on competition, and competition has advanced the state of AVs. Competition can continue to

thrive even if safety-relevant information is shared. Some companies such as Waymo and its Open Dataset initiative are beginning to experiment with such transparency. Because accidents and other safety incidents are rare—limiting the usefulness of data on these incidents—we have recommended that AV companies report the circumstantial details of each accident involving one of their vehicles to other companies and regulators as case studies from which everyone can learn. The aviation industry's practice of voluntary, sanction-free reporting of safety-relevant information could also be worth adopting. A common framework for measuring and communicating about safety is essential for comparing across developers and communicating with the public.

The commercial success of AVs will require public trust, which in turn will require the kinds of safety assurances that a good system of measurement and analysis can provide. Industry can voluntarily move in that direction through greater collaboration and transparency, and government (at multiple levels) can do much to encourage industry to do so. The world of AV safety is likely to get only more complex and challenging over time as the overall vehicle fleet and roadway system coevolve. Although future crashes are inevitable, having a common framework for assessing safety beforehand can help minimize them, in turn improving prospects for public acceptance—and public benefit.

*Marjory S. Blumenthal is a senior policy researcher at the RAND Corporation with a focus on emerging technologies and related policy. Laura Fraade-Blanan is an associate policy researcher at RAND with a focus on injury epidemiology and transportation technology.*

## Recommended reading

- Automated Vehicle Safety Consortium, "AVSC Best Practice for Describing an Operational Design Domain: Conceptual Framework and Lexicon," SAE Mobilus (April 2020).
- Laura Fraade-Blanan, Marjory S. Blumenthal, James M. Anderson, and Nidhi Kalra, "Measuring Automated Vehicle Safety: Forging a Framework," RAND Corporation (2018).
- Laura Fraade-Blanan and Nidhi Kalra, "Autonomous Vehicles and Federal Safety Standards: An Exemption to the Rule?" RAND Corporation (2017).
- Nidhi Kalra and Susan M. Paddock, "Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?" RAND Corporation (2016).
- National League of Cities, "Autonomous Vehicle Pilots Across America: Municipal Action Guide," NCL Center for City Solutions (2018).
- University of California PATH Program, "Peer Review of Behavioral Competencies for AVs" (Feb. 2016).