

<https://www.wsj.com/tech/ai/self-driving-cars-safety-ai-a43d8fd0>

How Will Self-Driving Cars Learn to Make Life-and-Death Choices?

Carmakers are grappling with the dilemmas AI will face on the road

By Brett Berk

Aug. 28, 2024 9:00 am ET

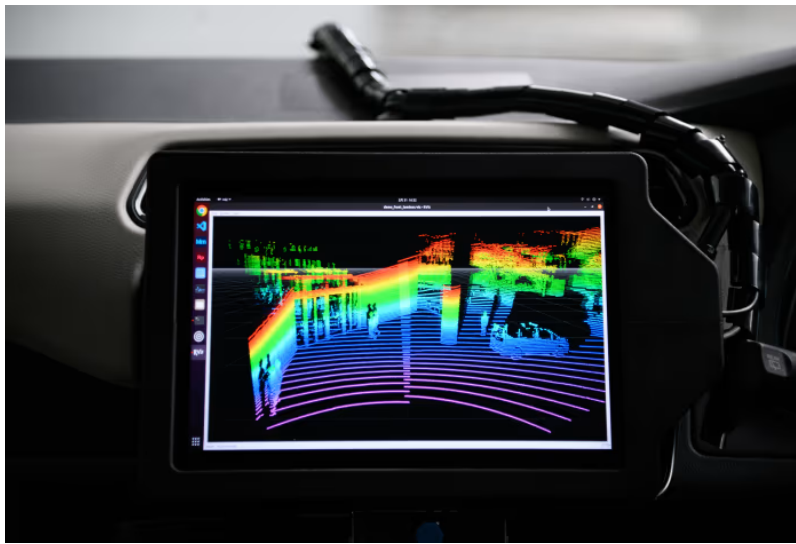
What will the roadway scruples of AI look like?

Artificial intelligence will hold sway on our roadways as we edge closer to fully self-driving cars—a shift that many carmakers say will make driving safer.

David Margines, director of product management at Waymo, Alphabet's self-driving subsidiary, says putting AI in charge precludes many of the factors that cause accidents, such as speeding, exhaustion, distraction and intoxication—preventing problems before they occur.

Others see different considerations. “The expectations for autonomous vehicles are much higher than for individuals,” says Melissa Cefkin, an anthropologist and lecturer in the department of general engineering at Santa Clara University who has worked with Nissan and Waymo on the interaction between humans and autonomous vehicles. “When an individual has an incident, we can say, ‘They’re only human. They tried their best. It was an accident,’ ” she says. “We’re not going to be as forgiving with a company, and we probably shouldn’t be.”

Here are some of the top questions engineers, programmers and bioethicists are grappling with as the world of AI-driven vehicles approaches.



A monitor in a Nissan Leaf electric vehicle that is equipped with autonomous-driving technology.

PHOTO: BLOOMBERG

Seeing humans

To avoid hitting pedestrians, autonomous vehicles use an array of AI-enabled features: perception, classification, prediction, path planning. But it can still be a challenge to distinguish between humans and humanoid delivery robots, retail mannequins or even street-level ads featuring images of people.

Biased training data and biases in the way electronic cameras “see” also make some AI systems less reliably detect dark-skinned pedestrians, says Peter Singer, emeritus professor of bioethics at Princeton University’s University Center for Human Values.

Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) systems will allow autonomous vehicles to communicate with each other, and with traffic lights and cameras.

Some envision the development of vehicle-to-pedestrian (V2P) communications as well. For instance, a cellphone’s broadcasting ping could signal a human presence to an autonomous vehicle’s AI, says Sam Abuelsamid, principal mobility analyst at transportation research firm Guidehouse Insights. But this brings potential concerns about privacy and personal liberty—and a disadvantage for people without phones. Is it conscionable to expect someone to have a cellphone in order to protect their own safety?

Autonomous vehicles themselves could be more expressive—using arrows, lights or text to communicate what they are seeing and what decisions they are making, suggests Cefkin, the anthropologist. “People come to realize there’s more of a two-way street here, because a vehicle will now tell them something more.”

Animal values

In a collision, moose and deer pose an existential risk to vehicles and their occupants. Smaller animals such as hedgehogs, or cats and dogs, present less of a risk. Is it morally acceptable for AI to weigh the lives of these animals differently?

For large animals, Waymo gives priority to “reducing injury-causing potential” for humans, through avoidance maneuvers, Margines says. When it comes to small animals, such as chipmunks and birds, Waymo’s AI “recognizes that braking or taking evasive action for some classes of foreign objects can be dangerous in and of itself,” he says.

How might the equation differ for midsize animals such as porcupines or foxes? Or for animals that might be pets?

Singer, the bioethicist, says the goal should be “to minimize the harm that is caused” for all animals—human and nonhuman. He worries that AI programmers will be “speciesist”—biased toward humans and human-companion animals. As a pragmatist, he favors such a compromise over “having no concern for the interests of nonhuman animals.” Still, moral decisions like these aren’t something you should just leave to the marketplace, he cautions. Instead, expert input and regulation should provide guidance.



A Waymo autonomous vehicle in San Francisco. PHOTO: GETTY IMAGES

The good trip

Is taking the fastest route the core metric that should guide autonomous vehicles, or are other factors just as relevant? Focusing on getting to the destination quickly would allow self-driving ride-share vehicles to make more trips and more profit, but might result in more danger. Giving priority to safety alone could slow and snarl traffic. And what about choosing routes that let passengers enjoy the journey?

Such decisions “become ethical when you think about whose interests are you serving,” Cefkin says. “Passengers? Robotaxi services? Other roadway users?”

The trade-off between safety and speed is “the one thing that really affects 99% of the moral questions around autonomous vehicles,” says Shai Shalev-Shwartz, chief technology officer at Mobileye, a unit of Intel that develops autonomous-vehicle software and hardware and supplies some of the world’s largest auto companies. Shifting parameters between these two poles, he says, can result in a range of AI driving, from too reckless to too cautious, to something that seems “natural” and humanlike.

Mobileye has designed a logic system with five tenets for AI decision-making: Don’t hit someone from behind. Don’t cut in recklessly. Right of way is given, not aggressively taken. Be cautious in areas with limited visibility. And, if you can avoid a crash without causing another one, you must. Shalev-Shwartz says company simulations found that if all drivers—human or autonomous—abided by these rules, no accidents would happen.

The software can also be calibrated to allow different driving styles, he says. So, for example, an autonomous sports car might drive more aggressively to enhance a sense of performance, an autonomous minivan might put the biggest emphasis on safety and an off-road vehicle might default to taking a scenic route.

The Swedish automaker Volvo wants to “build trust” between the public and self-driving cars. “Our autonomous vehicles will, by design, drive calmly and comfortably, within set speed limits,” says Martin Kristensson, Volvo’s head of product strategy. “This has the potential to influence traffic patterns, encourage collaboration, reduce accidents and, in time, positively change behaviors on the road.”

Such behaviors are contagious and can tamp down the anxiety of drivers, passengers and pedestrians, says Cefkin, whose research points to similar findings. “There’s evidence to suggest that there’s some magnanimous effects in traffic...when people see other calmer reactions,” she says.



An EV equipped with Mobileye technology for customer testing, on a drive in Zagreb, Croatia. PHOTO: MOBILEYE

Building to bend rules?

Tesla recalled vehicles equipped with its so-called Full Self-Driving system after the National Highway Traffic Safety Administration said the technology allowed some cars to violate local traffic laws and make moves such as rolling through stop signs.

Similar questions of strict legality can be applied to the ways many human drivers merge, change lanes, complete left turns or navigate around standing delivery vehicles. Should autonomous vehicles be held to different standards?

In Mobileye's AI technology, data derived from suites of sensors positioned around the vehicle allows it to note and distinguish local driving styles, using metrics such as speed, gaps between vehicles and turn-taking. "We can tune our system to how human drivers drive in every place," Shalev-Shwartz says. "But just because we can do it, doesn't mean that we should." He places an onus on regulators, suggesting that they should "figure out if they like how people are driving in their city today, or if they would like if traffic behaved a bit differently."

A broad rethinking of our roadways and rules is needed for the AI age, to promote safety and ease and expand usability, Cefkin says. Replicating the status quo might deny us opportunities to make deeper shifts. "Roads long predated cars, and they were used for many other things," she says. "It may be time for us to take that back."

Fleet mentality

Imagine if tech companies deploy fleets of self-driving vehicles to deliver humans, cargo, food and more. Would they gang up on us?

"Autonomous vehicles are designed to learn from driving experiences, together, as a community," says Volvo's Kristensson. "This drastically increases their rate of learning in comparison to an individual driver."

V2V and V2I communications will allow these vehicles to speak to each other, and this AI teamwork might help self-driving cars avoid incidents. But when members of competing or collaborating fleets encounter one another, how should they behave? Should these car "communities" be allowed to favor their compatriots over other vehicles in crashes, maximize efficiencies in chain-like convoys or dominate routes?

Cefkin wonders if self-interest would modulate the decisions the system makes. "Whoever is providing the product or the service, or both, they might want to weight certain things more or differently," she says.

“There should be no difference between who owns a car, who is not in the car, who is driving the other car,” Shalev-Shwartz argues, suggesting that all cars on the road should be held to the same standards, and treat each other equally.

However, the bottom line is that so far no regulation exists on this. Companies could decide that cars should favor others in the same fleet when a conflict arises, says Abuelsamid. “We need to have a discussion as a society about what are we going to preference?” he says. “Are we going to preference the needs of companies, or what’s best for the people that live in a society?”

Write to future@wsj.com

Videos

