DIANE E. BAILEY AND INGRID ERICKSON

Selling AI: The Case of Fully Autonomous Vehicles

Advocates' claims about potential lives saved with self-driving vehicles are a misleading attempt to steer the discussion of traffic safety away from alternative approaches.

In the past several years, an array of technologists, economists, and technology pundits have predicted that advances in artificial intelligence (AI) are poised to revolutionize our lives, changing how we work, play, travel, shop, create, and more. The ensuing popular discourse often construes AI as the inevitable result of technological progress, against which we have no claim to stand. Promoters from multiple domains converge to inform us that AI is a socioeconomic boon, a superior alternative that can liberate human labor by replacing it with cheaper and more efficient computation. In other cases, promoters recast AI in more transformational terms as an innovative means to accomplish tasks beyond prior reach or as the only available or feasible solution for an intractable social problem. It is in this last instance that most promoters argue for the adoption of fully autonomous vehicles.

Specifically, the argument for the adoption of fully autonomous vehicles rests on a particular rhetorical strategy: that this AI-based technology constitutes the sole solution to the pressing social problem of motor vehicle deaths. AI, in this instance, not only provides the public with a ready solution to a recognized social problem, but comes to symbolize the recognized social good of safety. Unfortunately, promoters' persistent use of this rhetoric of safety has largely silenced the conception of, as well as the discussion of, alternative solutions to this problem. Moreover, it has focused our attention solely on motor vehicle deaths instead of encouraging broader conversations about the potential benefits of autonomous vehicles or considerations of helpful policies surrounding their deployment and use. Exploring this strategy raises fundamental questions about how society chooses to adopt and leverage AI technologies in the future.

The rhetoric of safety

In March 2017, a Tesla owner in California was killed in a crash while her car was in fully autonomous mode. In response, Tesla announced, "There are about 1.25 million automotive deaths worldwide. If the current safety level of a Tesla vehicle were to be applied, it would mean about 900,000 lives saved per year." In a similar vein, General Motors recently produced a 33-page report noting that 1.25 million people die globally each year in car crashes and asking us to imagine—as the company envisioned—a world with zero crashes. At the 2018 South by Southwest

Conference in Austin, Texas, John Krafcik, the chief executive at Waymo, Google's autonomous vehicle arm and the leader in this market, predicted that autonomous vehicles would reduce worldwide deaths by 1.25 million annually. In each instance, these automakers used the figure of 1.25 million to give weight to the claim that fully autonomous vehicles will have a huge, and therefore undeniably welcome, social impact. But this figure is an exaggeration.

From a technical standpoint, claiming that autonomous vehicles can save 1.25 million lives a year inflates the potential good of this technology because such a claim assumes that roads, driving conditions, and drivers in all countries match those in the key nations where the bulk of fully autonomous automotive technology is being developed (i.e., the United States, Germany, China, Sweden, the United Kingdom, South Korea, France, Italy, and Japan). Clearly, they do not. In the United States, fully autonomous vehicles are being designed to respond to visible road markers such as double yellow lines, curbs, and traffic lights, which means that they struggle in construction zones and other areas that lack the visible road markers that AI software applications such as lidar rely on for safe navigation. In India, where motor vehicle deaths rose between 1990 and 2015 to among the highest total in the world, many of the roads are in perpetual disrepair or under construction, few have road markers, and most are notoriously chaotic. Indian drivers are notably creative in navigating around these challenges, so much so that R. C. Bhargava, the chairman of Maruti Suzuki, the largest carmaker in India, was pessimistic about the potential of fully autonomous vehicles in his country. As he was quoted in *The Hindu*: "I think no technology will work here when nobody obeys any of the driving rules, nobody obeys any of the systems which are there." Similarly, past Uber CEO Travis Kalanick declared India as the last place one would want to develop fully autonomous vehicles.

Nor is India likely alone in having road conditions unfit for autonomous vehicles. In fact, India is not even among the top 25 countries in traffic deaths per capita. Using the figure of 1.25 million motor vehicle deaths per year to justify fully autonomous vehicles, therefore, is to argue that the technology can be applied in countries around the world without regard for local conditions. At best, this assumption appears unrealistic or naïve. It would not be a stretch to suggest that what works in the countries that are leading fully autonomous vehicle automotive technology are unlikely to work the world over.

Beyond the issue of exaggerated claims lies the equally important issue of the neglect of alternative solutions to the social problem of motor vehicle deaths, which promoters of autonomous vehicles argue should be solved with the replacement of human drivers. In the United States, per capita motor vehicle fatalities peaked in 1937, and total fatalities peaked in 1972. In other words, proponents of fully autonomous vehicles may be selling a solution to a problem that was already in steady decline due to alternative solutions well before the introduction of this technology. To determine if that is true, we must consider how, if at all, AI accounts for the decline of deaths in recent decades.

As many news outlets reported in 2015, technological improvements, including seat belts, airbags, anti-lock brakes, rear-view cameras, electronic stability systems, and improved structural systems, account for most of the decline in motor vehicle deaths since 1972. Some of the items on this list reflect electronic technology and automation in vehicles, including sensors and controllers, but none of them reflect AI. Moreover, technological improvements in vehicles are only part of the story of reducing motor vehicle deaths. Changes in traffic infrastructure, such as replacing stop signs and traffic signals with roundabouts and using video cameras for red light violations, have helped. Laws, regulations, and related policies also have played a role. For example, mandatory seat belt laws and penalties for distracted driving have been important factors, as has a cultural campaign against drunk driving, stricter laws, and sobriety checkpoints. Regulatory solutions such as lowering speed limits, enacting helmet laws, and putting in place graduated licensing for young drivers have also contributed. Going forward, this combination of vehicle technology, traffic infrastructure, regulation, and policy could achieve near-zero vehicle mortality rates without the introduction of any AI—another fact that highlights the weakness in the argument that fully autonomous vehicles are the sole solution for the social problem of motor vehicle deaths.

Evidence to support the claim that motor vehicle deaths can be reduced significantly without turning solely to fully autonomous vehicles appears in the 2018 report *The Road to Zero*, published jointly by the Rand Corporation and the National Safety Council. The report outlines how the United States could reduce motor vehicle deaths to zero by the year 2050. Although the report includes the introduction of fully autonomous vehicles as one of the steps toward achieving this goal, it acknowledges that full fleet penetration will take decades and presumes that humans will still be driving in 2050. This recognition is important, particularly given that developers of fully autonomous vehicles such as Tesla, Uber, and Daimler have recently ended some of their autonomous vehicle programs or pushed back their timelines for having vehicles on the road. How, then, does the report project zero motor vehicle deaths by 2050?

It does so primarily by advocating the adoption of the "Safe System" approach. This approach originated under different names in Sweden, the Netherlands, and Australia in the 1990s, and was first implemented in the United States in the early 2000s in the states of Idaho, Minnesota, and Washington. Whereas promoters of fully autonomous vehicles note that human error is responsible for most motor vehicle deaths and conclude that the answer is to replace human drivers with AI, the Safe System approach, while acknowledging that human error is inevitable, shifts attention away from human drivers toward good road design. In other words, it holds system designers, not individual drivers, responsible for motor vehicle accidents.

According to a 2018 joint report entitled *Sustainable & Safe*, published by the World Resources Institute (WRI) and the World Bank's Global Road Safety Facility, the Safe System approach succeeds by addressing factors such as "land use and mobility planning—to reduce vehicle dependence and promote safe, healthy, and environment-friendly travel modes; comprehensive speed management to set safe speeds; intersection design to allow people to cross safely; road design that accounts for human error; improved public transport; safe vehicle design and technology; and better coordination and quality of post-crash emergency response and care." For example, roundabouts prove a better design than traditional right-angle intersections because they slow traffic, eliminate crossing conflicts, and facilitate crash angles that result in less severe injuries. According to a 2004 study by Maryland's State Highway Administration, fatal accident rates decreased by 100% at intersections where roundabouts were installed in that state. Similarly, barriers at the side of roads (that prevent vehicles from running off the road into fixed obstacles such as trees and poles) and in the center of roads (that prevent vehicles from running into oncoming traffic) reduce the number of deadly head-on crashes.

Such strong results for Safe System implementations appear to be the norm, not the exception. Presenting the results of WRI's analysis of data from 1994 to 2015 across 53 countries, the *Sustainable & Safe* report revealed that those countries that had adopted a Safe System approach had both the lowest number of motor vehicle deaths per 100,000 inhabitants and the fastest rate of decline in those rates. In the United States, the three states that adopted a Safe System approach fared much better than those that did not. For example, Minnesota saw its motor vehicle deaths drop by 40% over a 10-year period. Similar improvements have been noted in New York City, which adopted the Safe System approach in 2013. Perhaps the most striking feature of the Safe System approach is its applicability in the low- and middle-income countries with high motor vehicle death rates and poor road conditions. The amenability of the Safe System approach stands in stark contrast to the problems that fully autonomous vehicles currently face in these countries.

The results of the Safe System approach as documented in the *Sustainable & Safe* report, in addition to the technical improvements in vehicle design and regulatory changes, call into question the rhetoric of safety proffered by the designers and producers of fully autonomous vehicles, such as the leaders of Tesla, GM, and Waymo who position fully autonomous vehicles as the main, if the not the sole, pathway to zero motor vehicle deaths. Automakers, however, are not alone in utilizing this rhetorical strategy. Investors, consultants, university professors, think-tank researchers, and even government officials have become perhaps the strongest voices advocating fully autonomous vehicles as a solution to motor vehicle deaths.

Not surprisingly, these additional promoters use many of the same tactics as the automakers in their rhetorical salvos. Adrienne LaFrance, writing for *The Atlantic* in 2015, noted, "Globally, there are about 1.2 million traffic fatalities annually, according to the World Health Organization. Which means driverless cars are poised to save 10 million lives per decade—and 50 million lives around the world in half a century." That quick calculation rests, again, on the faulty assumption that fully autonomous vehicles will work in all countries with equal efficacy. Investor Shahin Farshchi invoked a similar safety argument in his May 2018 *Forbes* article: "Studies show that many billions of miles need to be driven by autonomous vehicles until we can statistically prove that they are safer than humans. Unfortunately, thousands of lives will be lost at the hands of human drivers while we wait for those billions of miles to be driven autonomously."

The studies to which Farshchi referred were actually singular, not plural: a study conducted by two Rand Corporation researchers who developed quantitative models to calculate motor vehicle deaths for fully autonomous vehicles over many potential future conditions and policies. Rand's blog discussion of this study was picked up by news outlets across the country, many of which included a quote in the blog from Mark Rosekind, at the time head of the National Highway Traffic Safety Administration (NHTSA). Speaking at the Automated Vehicle Symposium in San Francisco in 2016, Rosekind said, "We can't stand idly by while we wait for the perfect. We lost 35,200 lives on our roads last year.… If we wait for perfect, we'll be waiting for a very, very long time. How many lives might we be losing if we wait?" The adoption of the safety argument was not a random speaking point by the head of this government agency at this symposium; on its website, NHTSA lists safety as the first benefit of fully autonomous vehicles. Of course,

contrary to what Rosekind implied, no one is standing idly by, most especially not the engineers and policymakers working on new non-AI solutions, including those in the Safe System approach.

Consulting firms such as McKinsey & Company, KPMG, J. D. Power and Associates, Boston Consulting Group, and Deloitte have also been strong advocates of the safety argument. In the course of helping their clients predict the market for fully autonomous vehicles, these firms have generated a slew of white papers, many of which prominently frame the safety benefits of fully autonomous vehicles. These papers are cited not only in countless media articles but also in the publications of university transportation research centers. For example, drawing on statistics from KPMG and McKinsey & Company, a 2017 publication from the University of Michigan's Center for Sustainable Systems stated that fully autonomous vehicles have the potential to reduce motor vehicle crashes by up to 90%, thereby saving lives.

Advanced by this diverse group of promoters, the safety argument undergirds appeals for the immediate development and testing of fully autonomous vehicles. The recent slight increase in US motor vehicle deaths between 2011 and 2016 has become a call to arms to address what promoters paint as an escalating problem. This rallying cry is perhaps most stridently, and somewhat ironically, made in the *Road to Zero* report, which argues that "The more than 37,000 people killed in crashes in 2016 represent a troubling reversal in previous progress. For the past several decades, all the important measures of roadway deaths—the total number, the number per population, the number per miles driven—were going down as a result of several factors, including changes in driving patterns, increased seat belt use, improvements in vehicle design, more-forgiving roadway designs, and stronger graduated driver's licensing programs for teen drivers. After reaching an all-time low in 2011, these trends began reversing in 2015, and got even worse in 2016."

The advocates make their case by a cleverly misleading presentation of the data. They choose the period 1985-2011 and characterize it as a time of steadily declining fatalities, and then describe 2011-2016 as a time of steadily increasing fatalities. The reality was not so neat. There were years of increase and decrease in both periods, and the increase during 2011-2016 is not proof of an emerging crisis. The recent rise in motor vehicle deaths may simply reflect the normal variation in these numbers as they have declined, steadily but unevenly, over time. In fact, the US Department of Transportation reports that motor vehicle deaths in 2017 declined 1.8% from their 2016 value, suggesting just the kind of variation that is obscured by the simplified data used by the advocates.

The trolley problem

One final aspect of the safety argument bears discussion. As early as 2015, writers of various hues, including technology pundits, ethicists, philosophers, scholars, and journalists, began to debate the "trolley problem" confronting fully autonomous vehicles. Originated in the 1970s as a moral dilemma, the trolley problem asks you to imagine an approaching trolley that is speeding toward a group of people (often 50, but sometimes as few as 5) who are tied to the rails and who will die upon impact. As an onlooker, you are offered the hypothetical chance to pull a lever that will divert the trolley to an alternative track to which (typically) a single person is tied. Will you pull the lever? Recast for the current situation, the trolley problem sparks discussions about how designers should design fully autonomous vehicles to handle the moral dilemmas the vehicles could face in the course of everyday operation. Literally hundreds if not thousands of articles online and in print have taken up the ethics of fully autonomous vehicles in the context of the trolley problem.

From the perspective of selling AI using the rhetoric of safety, what merits consideration about this interest in the trolley problem is that it presumes that deciding who an autonomous vehicle should strike and kill in a crisis driving situation is a central design question demanding immediate attention, and it takes for granted the very idea that AI should and will replace human drivers. Although many writers tackle the trolley problem with gravity, far fewer question why the promoters of fully autonomous vehicles are using a rhetorical strategy based on safety in the first place or, more broadly, how autonomous vehicles figure in the larger conversation surrounding the relationship between AI and humans.

Granted, the rhetoric of safety that promoters of fully autonomous vehicles use to sell AI to the public may be merely the latest attempt to link the promises of technology to pressing social problems, little different from such attempts for a long line of technological advances put forward in the past. But to dismiss this current strategy as little more than run-of-the-mill techno-optimist hype would be to ignore the problematic fact that this narrow focus on motor vehicle deaths limits and obscures a broader discussion, both nationally and internationally, that ought to be had about the wide array of possible individual and social effects that futurists predict will be ushered in with the adoption of fully autonomous vehicles.

For example, at the individual level fully autonomous vehicles may ease the transportation burdens that individuals with physical or cognitive disabilities face, helping them not to miss medical appointments (a change that could save the health care industry billions of dollars) and making it possible for them to work outside the home. Similarly, autonomous vehicles may prove advantageous

for the elderly, whose own mobility may have suffered when they gave up or curtailed their own driving. Working mothers stretched for time might find compelling the idea of autonomous vehicles that chauffeur their children to after-school activities. Although these and many other imagined individual uses portend positive outcomes, a recent review of academic studies of autonomous vehicles published by London's Department of Transport warns that limited research has been undertaken to study the willingness, desire, and ability of the disabled, the elderly, or parents to employ autonomous vehicles. Moreover, not all individual use cases may be positive. For example, scholars contemplating the impact of autonomous vehicles on urban tourism have entertained the possibility that prostitution and drug use might shift from "hotels-by-the-hour" and street corners to roaming autonomous vehicles.

Individual use, illicit and otherwise, of autonomous vehicles may increase motor vehicle trips as people consider taking trips that they otherwise would have avoided (perhaps because they no longer lose time spent driving), thus causing concerns for rising greenhouse gas emissions that would constitute a negative societal outcome. Many other societal outcomes surface when futurists consider a switch from the model of individual vehicle ownership that exists today to a fleet-based scenario in which autonomous vehicles operate as a shared service. These outcomes include fewer vehicles in service, decreased traffic congestion, reversed urban sprawl, less urban space devoted to parking, lower insurance costs, fewer and less expensive vehicle repairs, reduced oil dependency, increased economic development, improved access to retail and jobs, and higher worker productivity. As at the individual level, however, not all envisioned societal outcomes are positive. Among potential negative outcomes are downward pressure on the earnings of commercial truck drivers, reduced revenue for small cities that rely on payments from traffic violations, fewer symbols for human navigation (i.e., road signs), and concentrated power in the hands of autonomous vehicle fleet owners.

In short, fully autonomous vehicles have a range of favorable and unfavorable potential outcomes for individuals and society, some of which futurists have begun to envision and others of which they have not. Moreover, these outcomes will affect a wide variety of constituent actors (e.g., riders, drivers, people in a range of occupations and professions, cities, and infrastructures) in direct as well as indirect ways. As a public, we should be keen to understand the complexities of autonomous vehicle adoption and use, and be eager for the opportunity to discuss these complexities imaginatively and responsibly. Further, we need to investigate, on the basis of such a discussion, the types of public policy that would best support the adoption and use of autonomous vehicles.

Wittingly or unwittingly, however, promoters of fully autonomous vehicles stymy this important discussion and investigation by inveigling us rhetorically into thinking that the realities we will soon encounter can best be understood and contained within the narrow focus of an AI-enabled solution to motor vehicle deaths.

We do not contest the inherent or potential safety of fully autonomous vehicles, despite recent accounts of deaths during testing or use of them on public roads. Rather, we challenge the rhetorical strategy of "selling" this particular AI-based technology using a safety argument. This argument is riddled with problems of logic that become more than evident when weighed against the relevant facts. As citizens, policy-makers, and lawmakers we should demand better before it is too late. Specifically, we should demand a broader discussion about autonomous vehicles that includes consideration of the full array of possible use scenarios, outcomes, and relevant and necessary policies.

Diane E. Bailey *is an associate professor in the School of Information at the University of Texas, Austin.*
Ingrid Erickson *is an assistant professor in the School of Information Studies at Syracuse University.*

Recommended reading

Clemence Cavoli, Brian Phillips, Tom Cohen, and Peter Jones, *Social and Behavioural Questions Associated with Automated Vehicles: A Literature Review* (London, UK: UCL Department for Transport, 2017).

Liisa Ecola, Steven W. Popper, Richard Silberglitt, and Laura Fraade-Blanar, *The Road to Zero: Executive Summary* (Santa Monica, CA: RAND Corporation, 2018).

Insurance Institute for Highway Safety Highway Loss Data Institute, "Low-Hanging Fruit: Existing Countermeasures Merit Renewed Attention," *Status Report* 46, no. 7 (2011).

International Transport Forum, "Urban Mobility System Upgrade: How Shared Self-Driving Cars Could Change City Traffic," OECD/ITF (2015).

Niddhi Kalra and David G. Groves, *The Enemy of the Good: Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles* (Santa Monica, CA: RAND Corporation, 2017).

Michael Sivak and Brandon Schoettle, "Mortality from Road Crashes in 193 Countries: A Comparison with Other Leading Causes of Death," University of Michigan Transportation Research Institute, UMTRI-2014-6 (Sept. 2017).

Ben Welle, Anna Bray Sharpin, Claudia Adriazola-Steil, Amit Bhatt, Saúl Alveano, Marta Obelheiro, Celal Tolga Imamoglu, Soames Job, Marc Shotten, and Dipan Bose, *Sustainable and Safe: A Vision and Guidance for Zero Road Deaths* (Washington, DC: World Resources Institute, 2018).