

A.I. & Accountability: Who's at the Wheel?

Self-driving cars have been in development since the 1920s and tested on the road with varying degrees of success. In 1995, Carnegie Mellon University's self-driving car crossed the United States, but only the steering was automated (the gas and brakes were controlled by a person). The first license for a fully autonomous vehicle was issued in Nevada in 2012. Since then, autonomous vehicles of all types and sizes in countries from China to Germany to Japan have become more and more common.

In 2018, the first recorded case of a pedestrian being killed by a self-driving car occurred in Tempe, Arizona. The pedestrian, 49-year-old Elaine Herzberg, was walking her bike across a road when she was hit by a self-driving Volvo sport utility vehicle traveling at 39 miles an hour in its autonomous mode. Uber was testing the car when, unfortunately, its sensors and artificial intelligence (AI) detection system failed to detect Herzberg. The car's "safety driver," Rafaela Vasquez, was streaming an episode of The Voice on her phone when the accident happened.

AI expert Neil Lawrence has explained how the AI failed Elaine Herzberg: "Modern autonomous vehicles classify objects in the roadway using neural networks. The systems need to know whether objects are people, vehicles or bicycles because each of these will behave in a different way. Vehicles and bicycles travel with the traffic, pedestrians walk across the road. If the system can't decide what the object is, it says 'other'. Unfortunately, in this system, Elaine fell into a [gap... something the system had not seen before]: she was a pedestrian pushing a bicycle. As the Uber vehicle approached Elaine, it decided she was a motor vehicle, then it decided she was an 'other'. The neural networks switched between vehicle and 'other' until two and a half seconds before impact. At that point, it finally decided Elaine was a bicycle. The car, knowing that bicycles travel along the road, pulled to the right to go

Case study written by: Robert Prentice, J.D. Department of Business,
Government and Society McCombs School of Business The University of
Texas at Austin

around Elaine, but unfortunately Elaine was walking across the road, not along it, and the car drove straight into her. Only one second before impact the system decided a collision would occur and began to brake, but it was too late for Elaine. The system made a mistake, but the accident occurred not only because of the mistake but because of how the car responded to the gremlin of uncertainty. Humans make errors, but when we're confused by what we see we tend to pause. If a human had been confused by an object crossing the road, they wouldn't have ploughed on regardless, they would have slowed down. Delaying action is one of the ways we respond to the gremlin of uncertainty. The computer did not pause, it ploughed on. We have built computers that makes errors, but we have not built machines that understand or suffer the consequences of those errors."

This tragic incident put a human face to a controversy that had already been discussed in hypothetical terms in ethics circles – especially AI ethics circles – for several years. Who is accountable in an accident involving a self-driving vehicle? Who decides how the vehicle will respond if an accident is unavoidable? Is it ethical to program the vehicle to maneuver to spare the passenger (and perhaps kill others)? The list of important questions is long, but the answers are much harder to parse. So how important is accountability and how should it be conceived of and applied in situations involving AI? In the case with Elaine Herzberg, Borg and her co authors ask (in *Moral AI and How We Get There*): "Who was responsible for Herzberg's death? Herzberg? Vasquez? The car? Uber? Safety managers at Uber? Engineers at Uber? Arizona government officials who allowed Uber to test their cars in Tempe? All of them? None of them?"

Accountability in the world of AI is extremely important. It is also extraordinarily challenging. But because autonomous AI-guided cars can cause wrecks and kill people, and autonomous weapons systems can kill

Case study written by: Robert Prentice, J.D. Department of Business,
Government and Society McCombs School of Business The University of
Texas at Austin

innocents, and AI-derived guidelines can inject bias into sentencing decisions for prisoners, and so on and so forth, it is essential that everyone carefully consider the matter of accountability. Whether we're creating or using AI products or services, we need to spell out who is accountable and for what before an unforeseen (or predictable) tragedy occurs.

Case study written by: Robert Prentice, J.D. Department of Business,
Government and Society McCombs School of Business The University of
Texas at Austin