ETHICS UNWRAPPED™

TEXAS McCombs | CENTER FOR LEADERSHIP AND ETHICS

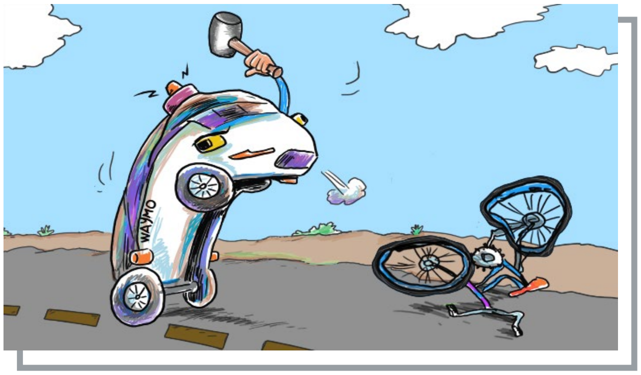
CASE STUDY

"We have built computers that makes errors, but we have not built machines that understand or suffer the consequences of those errors." — Neil Lawrence

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 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?" (p.137)

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 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

Discussion Questions

1. For both ethical and policy reasons, the law of products liability has long imposed accountability in the form of civil liability upon the sellers (e.g., manufacturers, wholesalers, retailers) of products that caused injury to consumers and others. The key drivers of this approach are a desire to encourage these sellers to design, manufacture, and sell products that are safe for use and to compensate parties injured when they are not. Some applicable legal theories apply when sellers are careless. Others impose liability without regard to fault on the seller's behalf. Do you think that the standards for accountability should be different for sellers of an autonomous vehicle than for sellers of a traditional vehicle? Why or why not?

2. Criminal liability is rarely imposed upon sellers of products. These sellers are typically companies that have "no soul to damn, no body to kick," and therefore can be punished only via monetary fines, which often doesn't seem worth the trouble. Nonetheless, occasionally such cases are brought and won. Again, do you think that the standards for criminal accountability should be different for sellers of an autonomous vehicle than for sellers of a traditional vehicle. Why or why not?

3. Anthropologist Webb Keane suggests that to be accountable, an AI "machine must, literally, be answerable, that is able to give a response if we were to ask 'why?' [it made a given decision]." (p. 139) Keane also quotes computer scientist Stuart Russell who "says the way to make AI safe is to have machines 'check in with humans—rather like a butler—on any decision.'" (p. 109) Do either of these suggestions sound like viable approaches to maintaining AI accountability? Why or why not?

4. Shadbolt and Hampson write:

> *"There simply is a fundamental accountability difference between a human and a machine, arising from all the other differences. Alexa or Siri can have a face painted on it, be put in smart clothes, and be set up to recognize you as you walk up to the bar, buy you a drink and ask about your day at the office,*

> *yet this will not affect you in the same way as a human performing exactly the same set of acts. A full, thick description of the difference would include dozens of dimensions. Prime among them is that even where we care a lot about a decision an AI makes about us, we don't care at all what private opinion the machine may have about us, not in the way we are affected by human opinions. Nor do we feel, reciprocally, that we should be diplomatic in how we treat the machine." (p. 114)*

Does the fact that humans react differently to AI tools that make certain decisions than to humans who make similar decisions (about how to maneuver a car or which prisoners to parole) justify differential accountability/liability judgments when injuries occur?

5. Looking at the issue of accountability from a different angle, Vallor suggests that "opaque AI decision systems are highly attractive tools for those in power; they offer a virtually bulletproof accountability shield." (p. 119) In other words, like "the dog ate my homework excuse," a political actor can say: "I didn't make that unpopular or disastrous decision, the algorithm did." This is especially true if the "model was trained using deep learning and other opaque techniques; even the software engineers and data scientists who created it will not know exactly how or why it works in a given case." (p. 127) Do you agree with Vallor's point as a factual matter? As a policy concern? Explain.

6. In writing about autonomous weapons systems (AWS), guns and the like that can make their own decisions about when and upon whom to fire, Eggert writes: "Free from human limitations [AWS] promise the prospect of a world without abuses like [human soldiers have often committed]. They do not succumb to anger or fear or vengefulness. And they can process vast amounts of information at superhuman speed. But, also unlike humans, they have no conscience to wrestle with." (p. 7) Egger then asks: "How should we weigh the promise of AWS to reduce harm to innocent people against the value of accountability?" (p. 13) In terms of accountability, how do you feel about the use of AWS without humans "in the loop"? Explain.

7. Several observers (including Harari, Wynn-Williams, and Lawrence) have written extensively about the damage that Facebook's engagement-maximizing algorithms inflicted in Myanmar in 2016-2017 by inciting anti-Rohingya violence leading to genocide. As Harari noted:

   > *"In 2016-2017, Facebook's algorithms were making active and fateful decisions by themselves….The algorithms could have chosen to recommend sermons on compassion or cooking classes, but they decided to spread hate-filled rumors." (p. 197-198)*

   Who is accountable for the genocide? Facebook's algorithm for spreading inciting information? The company? The engineers who developed the algorithm to maximize engagement without regard to potential dangers or costs? Those who turned the inciting information into violence? Is this type of situation an argument for always having humans in the loop? Is that even feasible? Explain.

8. As AI agents become more active, they are likely to not only produce many more good results, but also to cause more damage. A general requirement for criminal liability is criminal intent (mens rea). Floridi and Sanders suggest that AI agents "may be causally accountable for a criminal act), but only a human agent can be morally responsible for it." Do you agree with their distinction on this key issue of accountability? Why or why not?

9. When we think about AI accountability, do we need to be keeping users in mind as well? In what ways do we need to be considering their responsibility? What considerations do you personally try to have top of mind when you interact with AI?

Sources

Jana Schaich Borg et al., Moral AI and How We Get There (2024).

Ben Chester Cheong, "Transparency and Accountability in AI Systems: Safeguarding Well-being in the Age of Algorithmic Decision-Making," in Frontiers in Human Dynamics, Vol. 6 (2024).

Luca Collina et al., "Critical Issues About A.I. Accountability Answered," California Management Review Insights, Nov. 6, 2023, at https://cmr.berkeley.edu/2023/11/critical-issues-about-a-i-accountability-answered/.

Jovana Davidovic, "On the Purpose of Meaningful Human Control of AI," Big Data, Vol. 5, Jan. 2023, at https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2022.1017677/full.

Virginia Dignum, "Responsibility and Artificial Intelligence," in The Oxford Handbook of Ethics of AI (Markus Dubber et al., eds. 2021).

Linda Eggert, "Autonomous Weapons Systems and Human Rights," in AI Morality (David Edmonds, ed, 2024).

Luciano Floridi & J.W. Sanders, "On the Morality of Artificial Agents," Minds and Machines Vol 14 (2004), pp. 349-379.

Yuval Noah Harari, Nexus: A Brief History of Information Networks from the Stone Age to AI (2024)

Webb Keane, Animals, Robots, Gods: Adventures in Moral Imagination (2025)

Joshua Kroll, "Accountability in Computer Systems," in The Oxford Handbook of Ethics of AI (Markus Dubber et al., eds. 2021).

Neil Lawrence, The Atomic Human: Understanding Ourselves in the Age of AI (2025).

Theodore Lechterman, "The Concept of Accountability in AI Ethics and Governance," in The Oxford Handbook of AI Governance (Justin Bullock et al., ed. 2024).

Claudio Novelli et al., "Accountability in Artificial Intelligence: What It Is and How It Works," AI & Society, Vol. 39 (2024), pp. 1871-1882.

Jason Scholz & Jai Galliott, "The Case for Ethical AI in the Military," in The Oxford Handbook of Ethics of AI (Markus Dubber et al., eds. 2021).

Nigel Shadbolt & Roger Hampson, As If Human: Ethics and Artificial Intelligence (2024).

Shannon Valor, The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking (2024).

Sarah Wynn-Williams, Careless People: A Cautionary Tale of Power, Greed, and Lost Idealism (2025).