The Ethical Dilemma of Self-driving Cars

(TED Ed, by Patrick)

Self-driving cars are already cruising the streets today. And while these cars will ultimately be safer and cleaner than their manual counterparts, they can't completely avoid accidents altogether. How should the car be programmed if it encounters an unavoidable accident? Patrick Lin navigates the murky ethics of self-driving cars.

This is a thought experiment. Let's say at some point in the not so distant future, you're barreling down the highway in your self-driving car, and you find yourself boxed in on all sides by other cars. Suddenly a large, heavy object falls off the truck in front of you. Your car can't stop in time to avoid the collision, so it needs to make a decision: go straight and hit the object, swerve left into the SUV, or swerve right into a motorcycle. Should it prioritize your safety by hitting a motorcycle, minimize danger to others by not swerving, even if it means hitting the large object and sacrificing your life, or take the middle ground by hitting the SUV, which has a high passenger safety rating?

So, what should the self-driving car do? If we were driving that boxed in car in manual mode, whichever way we'd react would be understood as just that, a reaction, not a deliberate decision. It would be an instinctual panicked move with no forethought or malice. But if the programmer were to instruct the car to make the same move, given conditions it may sense in the future, well, that looks more like premeditated homicide.

Now, to be fair, self-driving cars are predicted to dramatically reduce traffic accidents and fatalities by removing human error from the driving equation. Plus, there may be all sorts of other benefits: eased road congestion, decreased harmful emissions and minimized unproductive and stressful driving time.

But accidents can and will still happen, and when they do, their outcomes may be determined months or years in advance by programmers or policy makers. And they'll have some difficult decisions to make. It's tempting to offer up general decision-making principles, like minimize harm, but even that quickly leads to morally murky decisions. For example, let's say we have the same initial set up, but now there is a motorcyclist wearing a helmet to your left and another one without a helmet to your right. Which one should your robot car crash into? If you say the biker with the helmet because she is more likely to survive, then aren't you penalizing the responsible motorist? If, instead, you save the biker without the helmet because he's acting irresponsibly, then you've way beyond the initial design principle about minimizing harm, and the robot car is now meting out street justice.

The ethical considerations get more complicated here. In both in our scenarios, the underlying design is functioning as a targeting algorithm of sorts. In other words, it's systematically favoring or discriminating against a certain type of object to crash into. And the owners of the target vehicles will suffer the negative consequences of these algorithm through no fault of their own.

Our new technologies are opening up

Glossary

murky – dark, obscure, shady

to barrel - drive or move very fast, almost out of control

to swerve – to change direction suddenly

SUV – sport utility vehicle (внедорожник)

to prioritize – to treat as being most important

to sacrifice (life) – to give up (your life) for the benefit of others

deliberate – planned, well-considered

malice – rage, intention to harm

premeditated – planned

https://ed.ted.com/lessons/the-ethical-dilemma-of-self-driving-cars-patrick-lin#watch meting out street justice – верша уличное правосудие