Ethical algorithms

October 14 2019

Duke Computer Science and ECE professor Cynthia Rudin. Credit: Duke University

Nearly forty thousand people lost their lives in car crashes last year in the U.S. alone. We can only presume that many of those fatalities were owed to our uniquely human frailties: distracted driving, driving under the influence, or plain inexperience. It makes sense to get human drivers off the roads as quickly as possible, and let machines do the driving.

That's one argument, anyway. There's also a compelling argument to stop and consider the ethical issues that this new technology surfaces. Whether it's self-driving cars or a selfie-sharing app with questionable privacy protections, the rush to deliver innovations to market often brushes ethical considerations aside—but several Duke ECE professors are pushing back.

Duke ECE professor Missy Cummings is a former Navy fighter pilot who now directs the Humans and Autonomy Lab, or HAL. Her research focuses on maximizing human and machine cooperation for optimal performance and outcomes, and she is an outspoken "techno-realist" when it comes to the idea that we're nearly ready for highly autonomous cars to hit the roads.

"Driverless cars could dramatically reduce roadway deaths, but currently, computer vision systems on these cars are extremely brittle, and not ready for unsupervised driving," said Cummings. "We know that long shadows, low sun angles, and even a quarter-inch of snow can cause these systems to fail, sometimes in catastrophic ways—so we are still 10 or more years away from achieving full driverless capabilities."

Manufacturers have spent around $80 billion on autonomous vehicle research and development to date. The size of that investment comes with pressure of equal magnitude; the investments need to pay off, and there is a clear interest in hustling the technology to an eager market. Yet, the shortcomings of current AV systems are well documented. They are vulnerable to hackers. They are not good at tasks of inference—for example, knowing that a ball that rolls into the road will probably be followed by a child chasing it. These types of knowledge and skill errors, noted Cummings, would cause a human driver to fail a driver's license test—but no equivalent "computer vision" test currently exists that examines the reasoning abilities of driverless cars.

Despite the dubious capabilities of autonomous systems and the lack of processes for testing and certifying highly autonomous vehicles, however, they have already taken to our roads—in what are essentially large-scale experiments involving the public without its explicit consent.

Cummings said that wanting to achieve fully autonomous vehicles is necessary to making the incremental improvements that will get us there, eventually. But pushing the technology out into the world before it can be effectively regulated, she warned, is both irresponsible and dangerous.

It's a problem that extends far beyond the AV sector.

Professor Cynthia Rudin runs Duke's Prediction and Analysis Lab, and she is a machine learning expert—specifically, she is an expert at building interpretable machine learning algorithms, in a world increasingly obsessed with black box models.

"A black box predictive model is a model that's too complicated for a human to understand, or a formula that's proprietary, meaning it's hidden by a company," said Rudin. Black box algorithms are commonly used in low-stakes applications like retail, where your age, income, occupation, purchase history, and a hundred other bits of data inform the decision of whether to show you an advertisement for airline tickets or vitamins.

More problematic are black box models used in high-stakes decisions, like evaluating credit risk and setting parole. Those decisions can profoundly affect people's lives, stressed Rudin, and it's difficult for someone who has been denied parole to challenge the decision if it's impossible to see how the decision was reached.

Rudin's lab specializes in developing simple, interpretable models that are more accurate than the black box models currently used by the

justice system. According to Rudin, you don't even need a calculator to compute them.

"There's sort of a widespread belief that because a model is a black box, it's more accurate," said Rudin. "And that, as far as I can tell, is wrong. I've worked on many different applications—in medical care, in energy, in credit risk, in criminal recidivism—and we've never found an application where we really need a black box. We can always use an interpretable model for a high-stakes decision problem."

The enthusiasm for black box models, said Rudin, should be tempered by careful consideration of the possible ramifications.

"Often the academic community doesn't train computer scientists in the right topics," said Rudin. "We don't train them in basic statistics, for instance. We don't train them in ethics. So they develop this technology without worrying about what it's used for. And that's a problem."

This year, Duke Engineering established the Lane Family Ethics in Technology Program, which will embed ethics education across the engineering and computer science curricula. The program supports faculty-led course content, extracurricular activities and an annual symposium focused on ethics in technology.

Stacy Tantum, the Bell-Rhodes Associate Professor of the Practice of Electrical and Computer Engineering, will lead one of the program's first courses this fall. Tantum will work with Amber Díaz Pearson, a research scholar at Duke's Kenan Institute for Ethics, to integrate ethics-focused modules into ECE 580, Introduction to Machine Learning.

Three elements of ethical algorithm development will be emphasized in the course, said Tantum. First is transparency, or why others should be able to easily evaluate all aspects of algorithm design, from the input

training data and algorithmic assumptions, to the selection of algorithmic parameters, to the process by which predicted performance is evaluated. Second is algorithmic bias—the conditions that are likely to result in bias, but which are often overlooked, or deemed unimportant. And third is unintended use-cases of algorithms—the potential pitfalls of re-purposing algorithms for use-cases other than those for which they were designed.

"Our goal is to lead students to incorporate ethical considerations as a natural part of algorithm development, not an afterthought to be considered only after an unintended or unanticipated consequence arises," said Tantum.

Provided by Duke University