The Trolley Problem and Ethical Dilemmas in AI.

Print

Title Annotation: *AI ethicist; artificial intelligence*
Author: *Kompella, Kashyap*
Article Type: *Column*
Geographic Code: *1USA*
Date: *Jul 1, 2020*
Words: *754*
Publication: *Information Today*
ISSN: *8755-6286*

Imagine an out-of-control trolley car hurtling down the track toward five people who are chained up and cannot get out of the way. You are out of harm's way and standing next to a lever that switches the trolley onto a different track. But on the side track, there is one person who is similarly chained. In this situation, your only options are a) do nothing, and let the trolley run over the five people, or b) pull the lever, and let the trolley run over one person. What's the right thing to do?

In ethics classes, the trolley problem (as this thought experiment is known)--or one of its many variants--is used to introduce ethical dilemmas and illustrate the differences between different schools of philosophy such as utilitarianism (which says that actions are right if they benefit the majority of people) and deontological ethics (in which the "rightness" of actions is not linked to such a cost-benefit analysis; actions are good according to a set of rules regardless of consequences). As you probably figured, what is good varies depending on moral philosophies.

REAL-WORLD EXAMPLE

Ethical dilemmas do not have perfect answers, and any choice is going to violate some ethical norms. In the world of AI, the trolley problem is often applied to self-driving cars. It is used as a starting point to illustrate decisions that a self-driving car may have to make in hypothetical accident scenarios. For example, if an accident is unavoidable, what should a self-driving car do? Swerve into a group with fewer people? Hit pedestrians or other vehicles? Try to save elderly pedestrians or children? People or animals? Should it prioritize the well-being of the car occupants or those on the road? Who decides? Who gets to weigh in? Should it be left to the car manufacturer? How will liability be assigned in the case of accidents?

Fully autonomous self-driving cars are, perhaps, at least several years away from hitting the roads in large numbers-so we still have some time to work through these issues. But there are other areas in which AI is already being used in the real world. For example, algorithms are often used to select who should receive an organ donation. AI systems can recommend to judges whether a person should get parole or how long a prison sentence should be. Algorithms can allocate scarce resources, such as public housing. The stakes can be quite high for those involved. In these (and more) scenarios, there are trade-offs and ethical questions.

ETHICAL BY DESIGN

If AI ethics are tricky to navigate, that's because there is no single objective function that can be optimized, and there aren't universal-decision rules that can be applied. You can't simply code your way out of this. What are some of the ways to ensure that ethics--and the trade-offs they imply--are explicitly considered in the creation and use of AI systems?

First, it should be sufficiently clear (or made clear) whose job it is to codify ethical choices into AI systems. As is happening now, it should not be left to the discretion of the system developers; we can't

expect them to figure it all out on their own. Input from key stakeholders such as ethicists, engineers, lawyers, and regulators has to be incorporated. We talk about human-in-the-loop AI systems, in which humans provide oversight for AI systems. We also need "humanities in the loop" to plug in a diversity of perspectives on ethical questions.

Second, for AI to be ethical by design, there are big gaps between theory and practice that need to be bridged. Some practical interventions that can help are training and/or workshops on ethics and ethical AI for engineers who are building AI systems. It is also helpful to offer training and/or workshops for ethicists, lawyers, regulators, and other nontechnical stakeholders on the technical aspects and architecture of AI systems.

Third, we have been building, using, and treating AI systems and applications as though they belong purely to the realm of technology. But AI systems contain multitudes-ethical priniciples, moral choices, human biases, and social mores--by default or by design. Combining AI technology advances with multidisciplinary expertise, deft regulatory approaches, and governance frameworks can help resolve thorny trolley problems.

KASHYAP KOMPELLA (linkedin.com/in/kashyapkompella) is the CEO of rpa2ai Research, a global AI industry analyst Arm, and a contributing analyst at Real Story Group. He is the co-author of Practical Artificial Intelligence. Send your comments about this column to ecletters@ infotoday.com or tweet us (@ITINewsBreaks).