Moral Machine Experiment

The infamous “trolley problem” was put to millions of people in a global study, revealing how much ethics diverge across cultures.

https://www.technologyreview.com/2018/10/24/139313/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem

MIT Technology Review

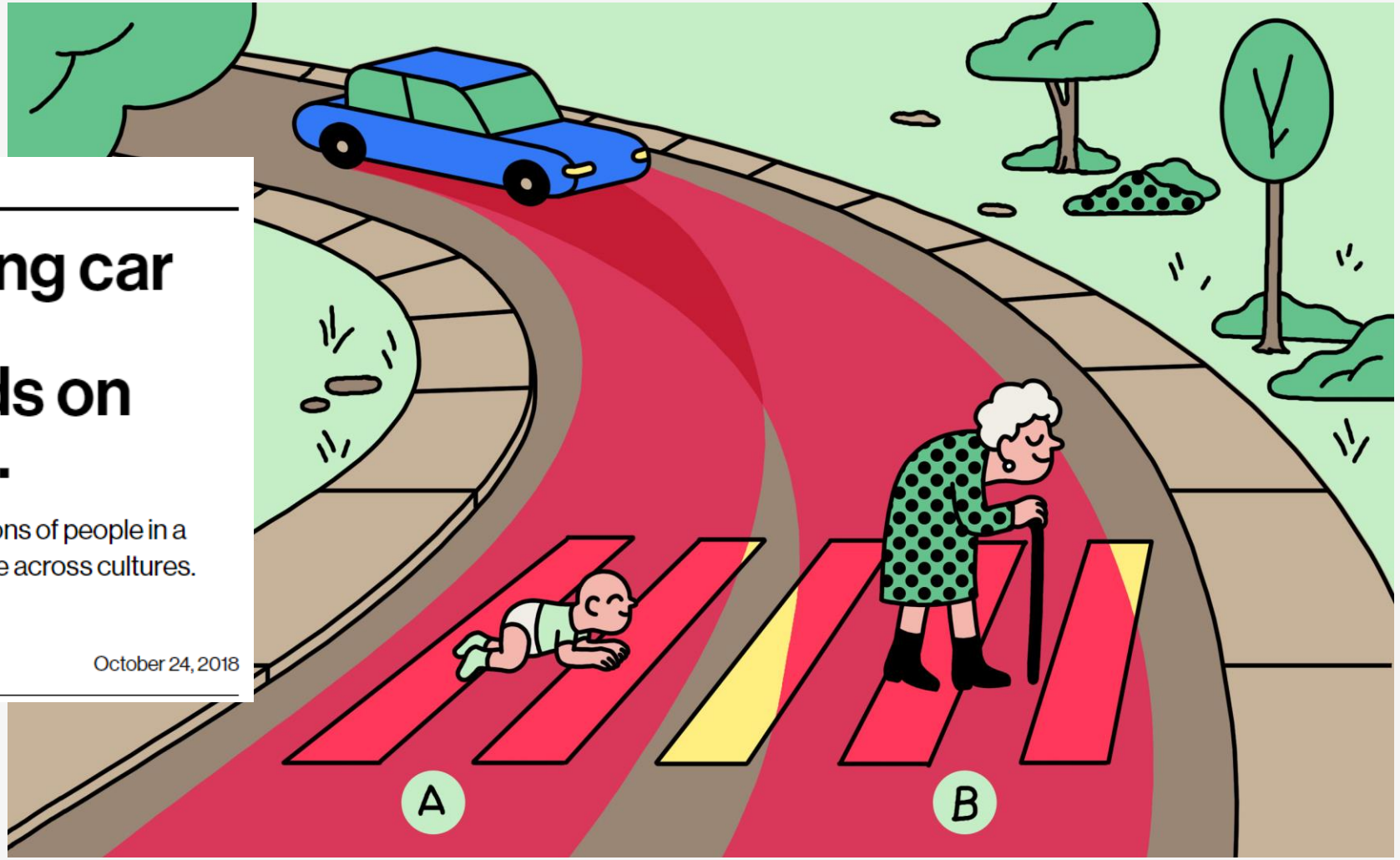
Moral Machine Experiment

Tech policy / AI Ethics

Should a self-driving car kill the baby or the grandma? Depends on where you're from.

The infamous "trolley problem" was put to millions of people in a global study, revealing how much ethics diverge across cultures.

by **Karen Hao** October 24, 2018



Source: https://www.technologyreview.com/2018/10/24/139313/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem

Moral Machine Experiment | Intro

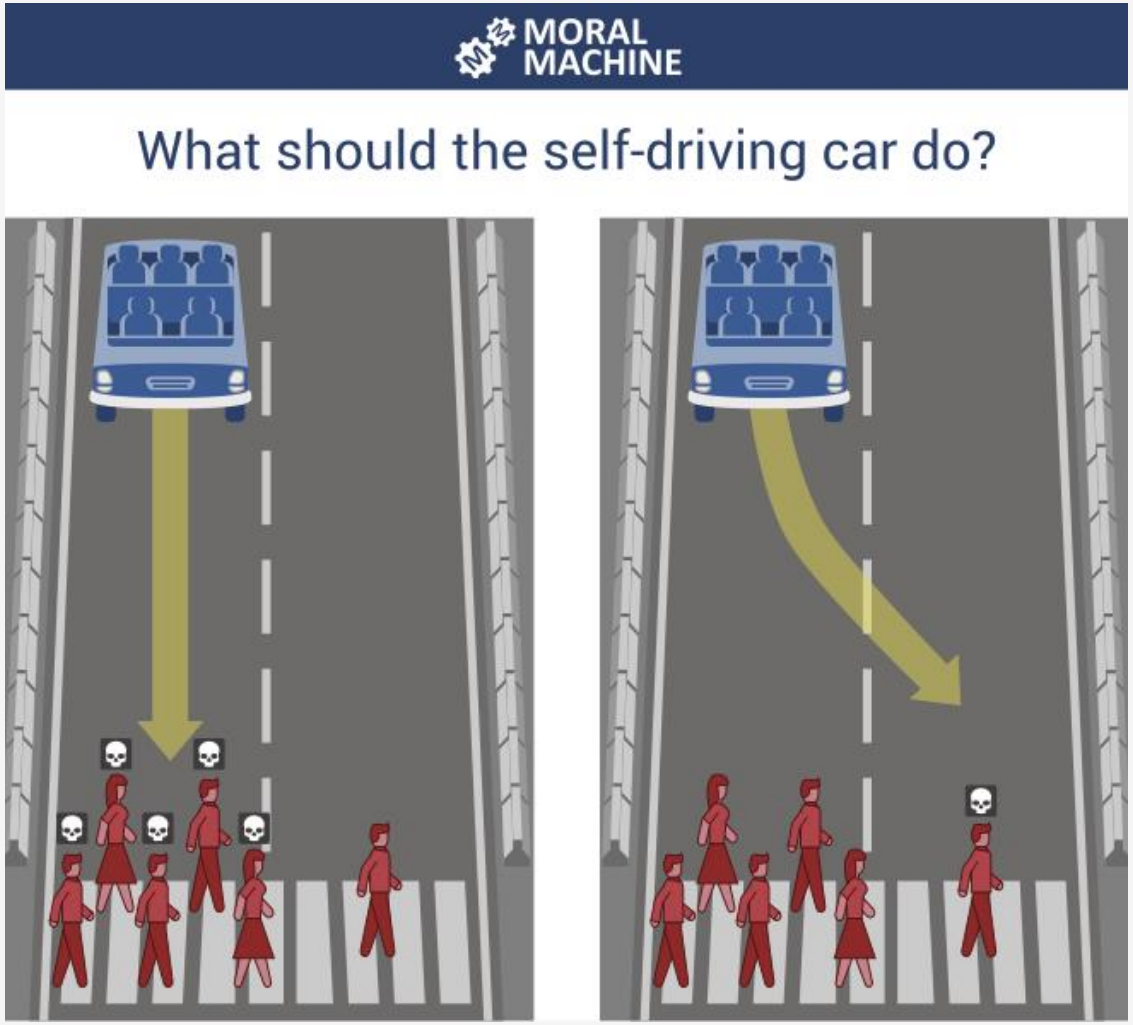
Researchers at the MIT Media Lab designed an experiment called Moral Machine. The idea was to create a game-like platform that would crowdsource people's decisions on how self-driving cars should prioritize lives in different variations of the "trolley problem." In the process, the data generated would provide insight into the collective ethical priorities of different cultures.

The researchers never predicted the experiment's viral reception. Four years after the platform went live, millions of people in 233 countries and territories have logged 40 million decisions, making it one of the largest studies ever done on global moral preferences.

Source: https://www.technologyreview.com/2018/10/24/139313/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem

Moral Machine Experiment

The Moral Machine took that idea to test nine different comparisons shown to polarize people: should a self-driving car prioritize humans over pets, passengers over pedestrians, more lives over fewer, women over men, young over old, fit over sickly, higher social status over lower, law-abiders over law-benders? And finally, should the car swerve (take action) or stay on course (inaction)?

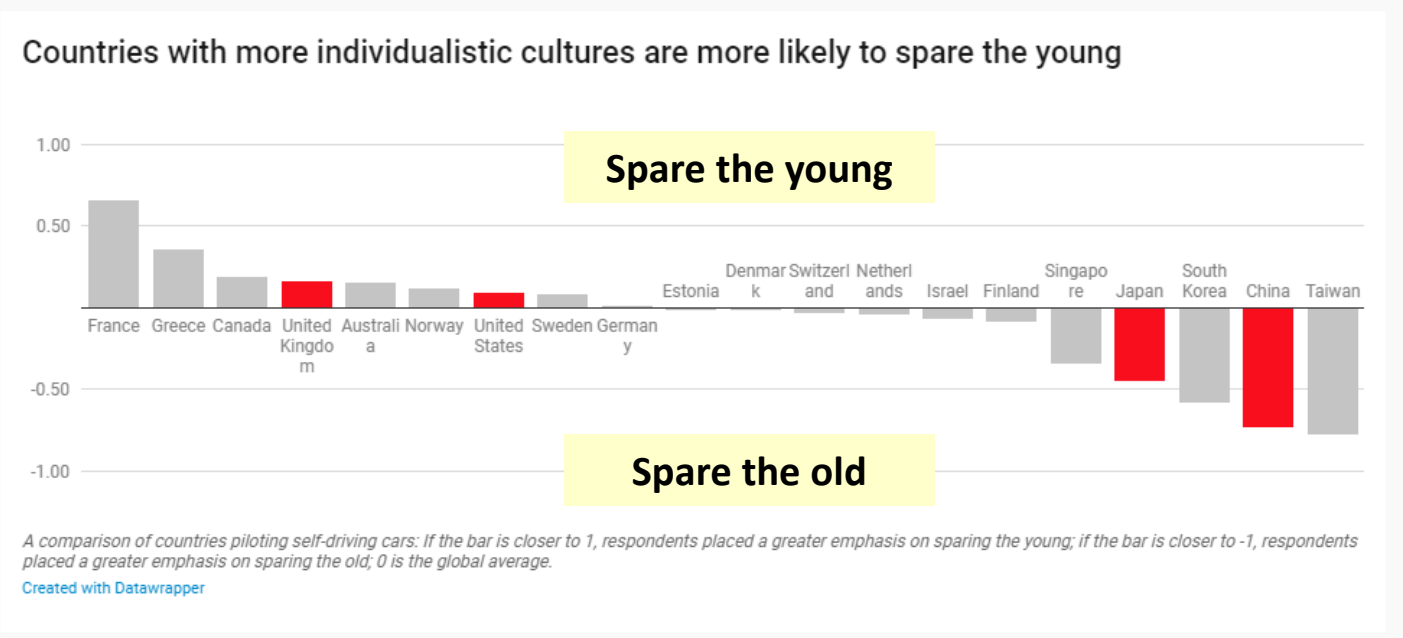


Source: https://www.technologyreview.com/2018/10/24/139313/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem

Countries with more individualistic cultures are more likely to spare the young

Rather than pose one-to-one comparisons, however, the experiment presented participants with various combinations, such as whether a self-driving car should continue straight ahead to kill three elderly pedestrians or swerve into a barricade to kill three youthful passengers.

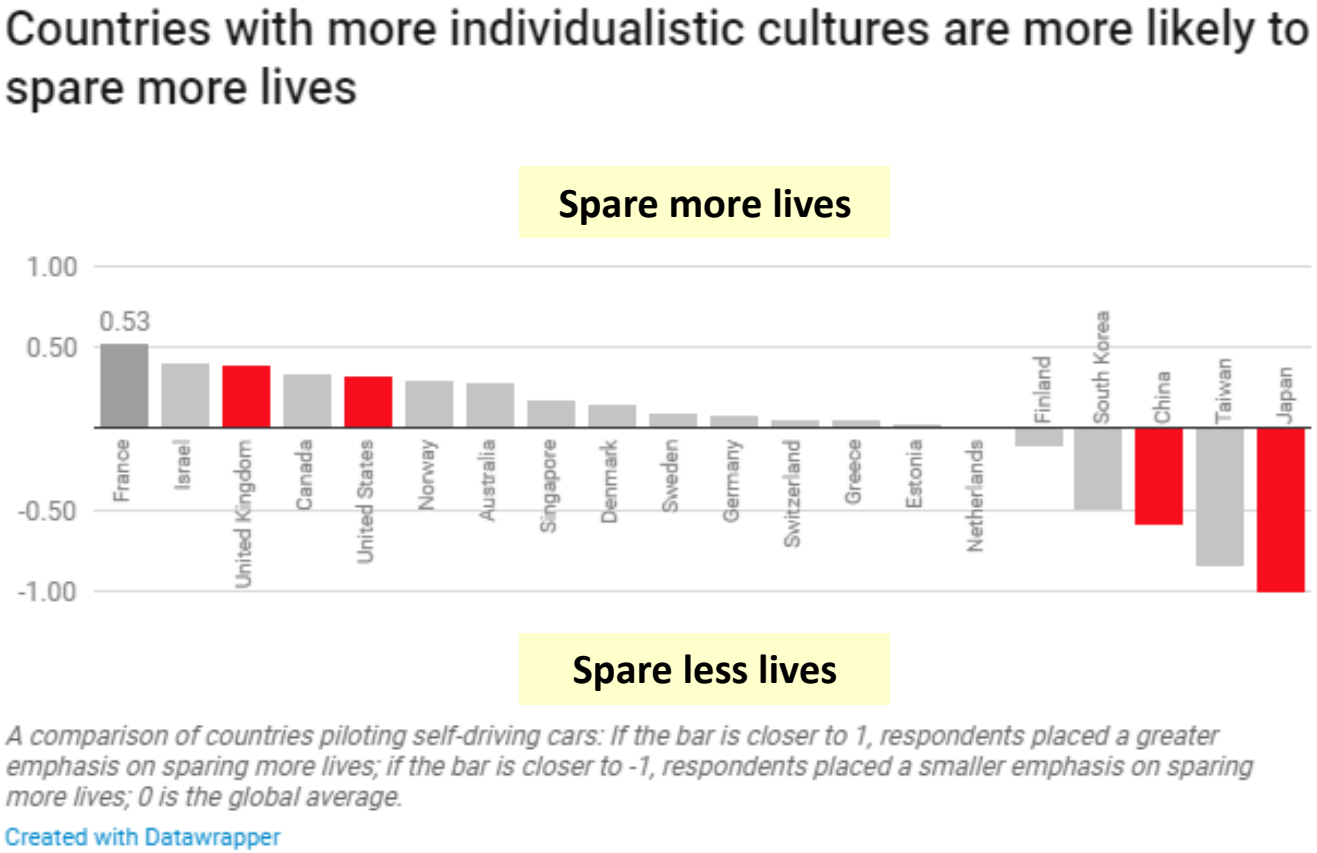
The researchers found that countries' preferences differ widely, but they also correlate highly with culture and economics. For example, participants from collectivist cultures like China and Japan are less likely to spare the young over the old—perhaps, the researchers hypothesized, because of a greater emphasis on respecting the elderly.



Countries with more individualistic cultures are more likely to spare more lives

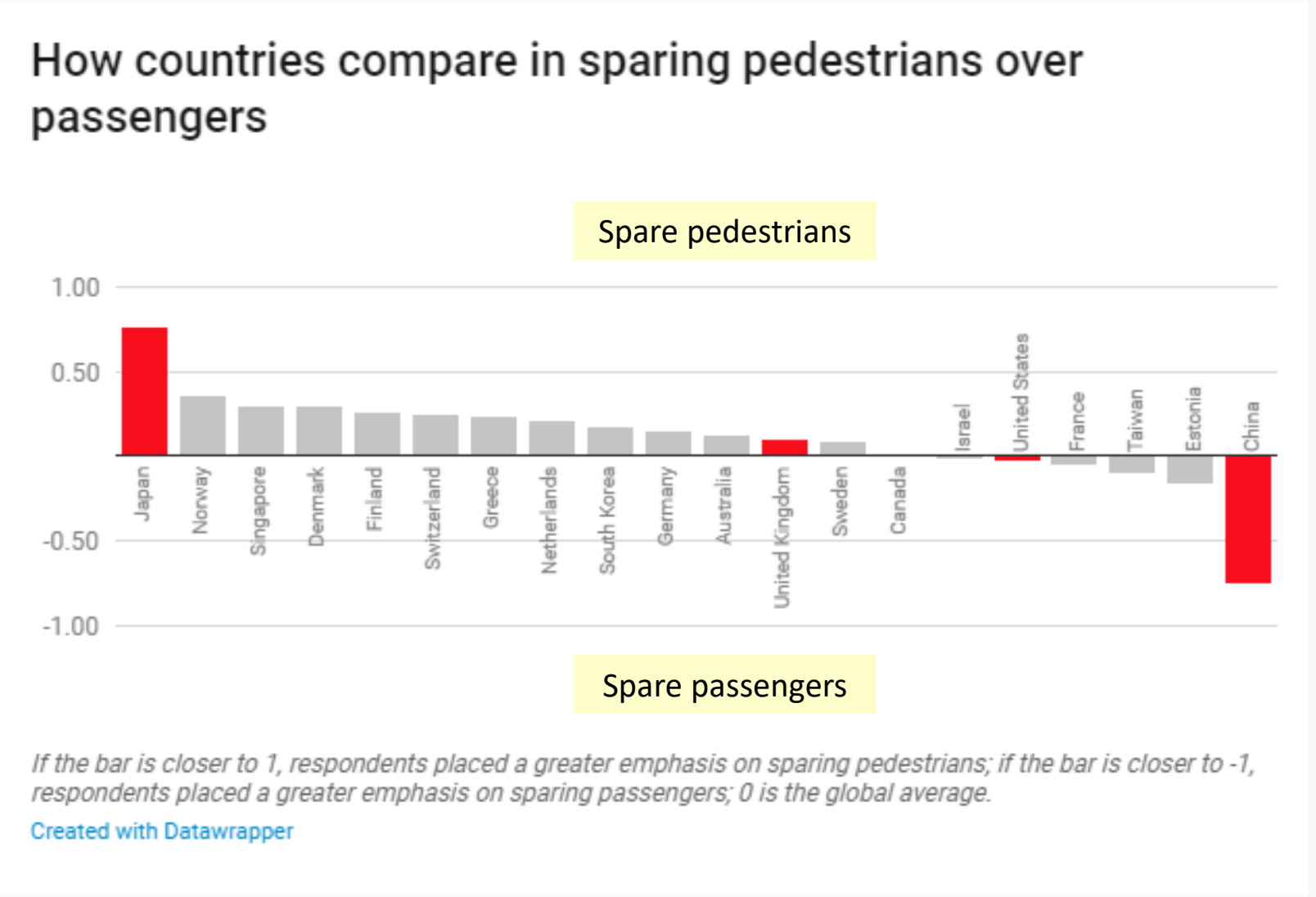
Similarly, participants from poorer countries with weaker institutions are more tolerant of jaywalkers versus pedestrians who cross legally. And participants from countries with a high level of economic inequality show greater gaps between the treatment of individuals with high and low social status.

And, in what boils down to the essential question of the trolley problem, the researchers found that the sheer number of people in harm's way wasn't always the dominant factor in choosing which group should be spared. The results showed that participants from individualistic cultures, like the UK and US, placed a stronger emphasis on sparing more lives given all the other choices—perhaps, in the authors' views, because of the greater emphasis on the value of each individual.



Source: https://www.technologyreview.com/2018/10/24/139313/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem

How countries compare in sparing pedestrians over passengers



Some Conclusions

The researchers acknowledged that the results could be skewed, given that participants in the study were self-selected and therefore more likely to be internet-connected, of high social standing, and tech savvy. But those interested in riding self-driving cars would be likely to have those characteristics also.

The study has interesting implications for countries currently testing self-driving cars, since these preferences could play a role in shaping the design and regulation of such vehicles. Carmakers may find, for example, that Chinese consumers would more readily enter a car that protected themselves over pedestrians.

Source: https://www.technologyreview.com/2018/10/24/139313/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem

And you?

Think more deeply about the ethics of AI beyond self-driving cars. How these results could translate into the more ethical design and regulation of AI is something he hopes to study more in the future.

You are the AI Engineer responsible to design the solution. How would you deal with the different situations and different scenarios?

What is your opinion? What is your decision?

(1) Spar the young or the old?

(2) Spar more lives or fewer lives?

(3) Spare pedestrians or passengers?

How to make AI fair, considering those situations and scenarios?