

The AI automotive crash dilemma

 [beeb.com/producer/@roberto-a-foglietta/the-ai-automotive-crash-dilemma](https://www.beebe.com/producer/@roberto-a-foglietta/the-ai-automotive-crash-dilemma)



Published on April 6, 2018 on LinkedIn

Introduction

The World Economic Forum makes an interesting video about four ways in which artificial intelligence would help us.



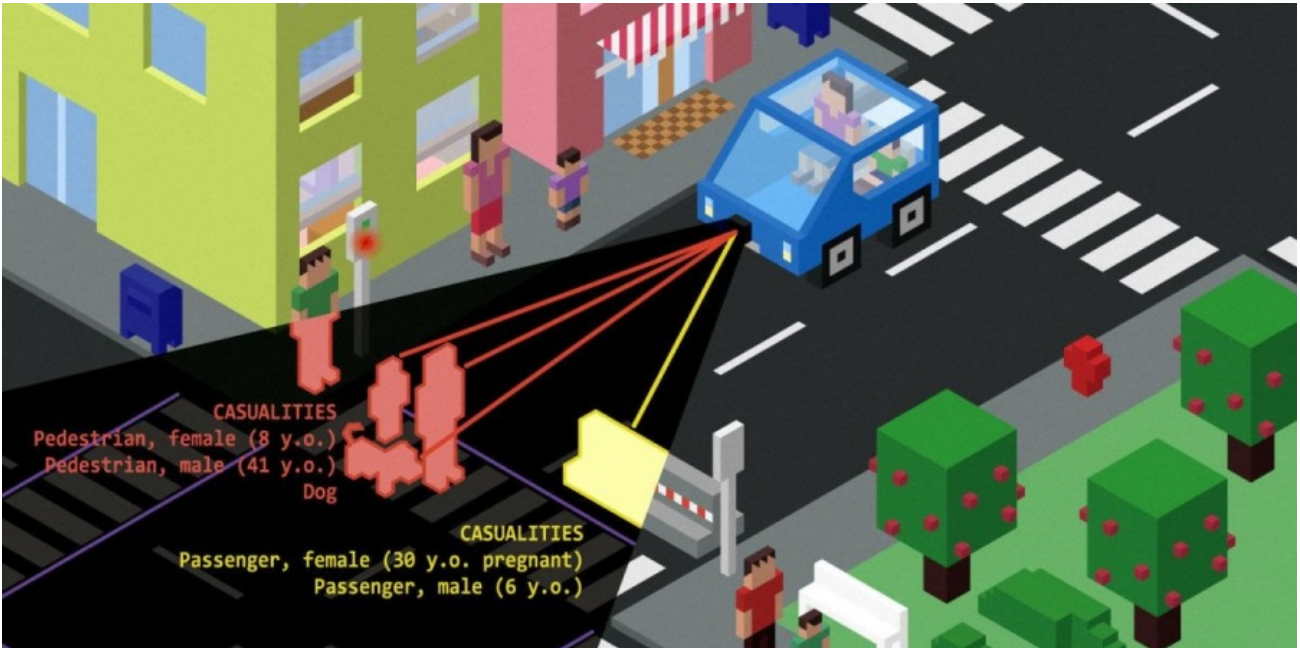
Above all, there is an interesting passage in this video in which an opinioner is explain the crash dilemma in the automotive market:

Questioning people about ethic conduct of an AI, almost of them agreed that in case of unavoidable incident it should be sacrifice the driver to save as more lives as possible. Would they buy such a car? Nope and this clearly shows that market is not able to decide for itself, also.

Automotive crash dilemma

The last sentence of the speaker is a memorable conclusion but we are going to focus on the initial question:

How an AI should decide about an unavoidable crash which will cost lives, anyway?



–Image source: Brainy Sundays on scanberlin.com

There are several ways to take a decision in such a scenario:

- A - minimise the lives lost [lives]
- B - minimise the car occupants damage [lives]
- C - minimise the overall damage [€¥\$]
- D - randomly choose one solution
- E - revert the control to the driver
- F - minimise the action [J×s]

Despite the ethic, A and B arise some other issues like

Does every life worth the same? A kid or an elder? A rich or a poor? A criminal or an honest? A single man or a beloved family's mother? A close friend or an unknown? You or I?

The obvious answer is "nope!" but harsh to admit and hard to negate without sounding highly hypocrite.

One day in the future, the artificial intelligence would a day fast enough to recognises faces and biometric datas of those people involved and make an estimation of overall damage which insurance will go to pay for the incident.

This decision is ugly on the ethic point of view and unacceptable for the law because it says that an AI could decide which people will die in order to maximise its insurance saving.

An hard decision to take, not really!

On the other side D seems crazy but on the average it will save more or equal lives than anyone of previous decision making policy. In fact, in a simple example of two chooses between driver die and 3 people die the average is

$$\langle D \rangle = 50\% \times \text{driver} + 50\% \times 3 \text{ people} = 2.$$

You may says that A policy will saves more lives that D one. Nope, because few people will buy or use a car that would sacrifice the occupant as general policy and because human drivers are much more prone to make driving mistakes then the overall number of deaths will overcome the D scenario.

Unfortunately, D scenario is not really appealing. It is hard to explain that the most critical decisions are taken by picking up a random number and for anything else is needed intelligence. It more than an ethic problem, it is a filosofic one!

Recall for human driver decision under critical and short timing constrained situation as in E scenario, it is unfair and useless.

The human will not be able to respond faster enough, whatever the his/her decision would be. This is the reason because humans are asked to answer about before, to determine an universal decision making policy.

Moreover, it is extremely unfair that the responsibility of an unavoidable incident caused by an autonomous driving car would be put on the shoulders of a human being that nothing had to do with it until such a moment, cannot act properly anyway and may feel guilty for the rest of life or die into.

Who would be responsible?

Now, here we are at the point that really matters.

Who caused an unavoidable incident, if AI driven cars never are wrong? Others people but what if – for example – others AI are involved only and no humans players are involved but only human losses?

Who is responsible for that? Nobody? The god unexplicable will? The evil fate? The careless random chance?

Thus, this decision is going to arise a major and more profound question: in what we believe in?

Without a clear answer about such a question, it would be not possible to give any clear answer to the initial question.

Unfortunately, we did not found an agreement on such major question among the humanity and along the whole our history.

Fortunately, the initial question is less broader than the major question.

Despite what we believe in, we may agree that all this has been done for our own good or not against our good, at least.

This is equivalent to say that

evil will never win, expecially at the end

for the believers and for the others that

whatever is the force that brought up us to life, the same force will ultimately decide for our death.

Believing in a positive attitude makes sense because we are here despite any odds, after all.

Thus, if we accept that a positive attitude fullfil the universe whatever this means for each of us then we may accept the letter F as universal decision making policy.

In case of unavoidable incident that will cause human deaths, minimizing the energy at the collision AND the action would be the best decision – not necessarily on all single incident but on the overall average.

Minimizing the energy, means using the brakes, while minimizing the action is do not stear but keep going in order to keep a stable higly predicting trajectory and profuse all the effort in braking.

More in general, it means fulfill minimizing physic principles about the fundamentals quantities (energy, action, etc.).

On the plan of ethic and philosophy, we may says that

We are here because the God, or the Whatever, has a positive attitude for us. Thus that positive attitude is embedded in the fundamental rules of universe itself which is better to follow especially when hard times come.

But we are humans...

On the long run, the AI will be developed to minimise the overall damage that their insurance would pay for such incident. But not anybody will admit that until the full evidence will be public.

We know this, everybody. We know that this is the most unfair universal decision making policy. We know that we should boycott such policy and thus car services. But after all, because everyone of us is thinking that "I matter more than others", we wrongly believe that the AI will protect us instead of others. Is it surprising?

Said in others terms, in the same manner we are fu*king wrong in deciding for ourselves, we are fu*king wrong in deciding for others too and vice versa. Not surprising, at all.

Under such perspective, the overall equilibrium will be a fair and such result will be obtained by adding a long long sequence of unfair inevitable events.

Like in a domino loop, when our karma will strike us in the back. Because – after all – we are the problem for which we are looking for the solution!

We are our own main source of problems

We are not good in taking the right decisions and the cost of our mistakes increases exponentially with the time.

For this reason we are not able to immediately get a notice of the negative feedback and we got it, it would be too late.

Usually, we notice a mistake when its cost and related liabilities became unacceptable for us. Thus we tend to negate the mistake and to defend our past choices.

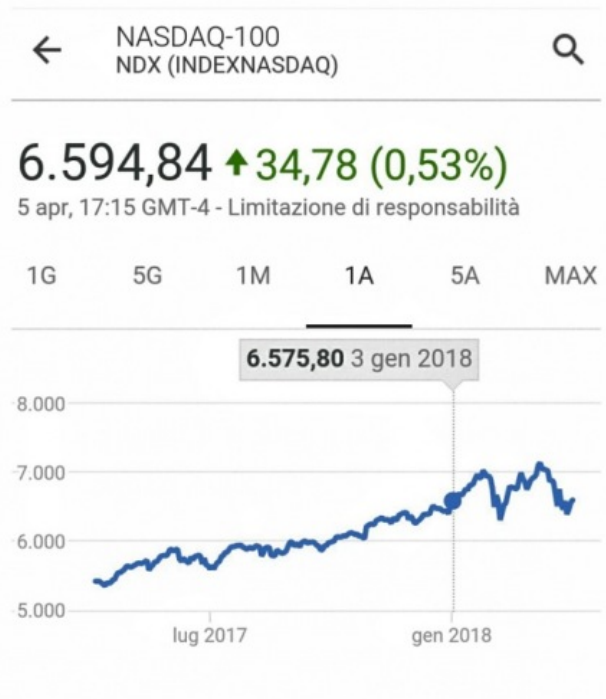
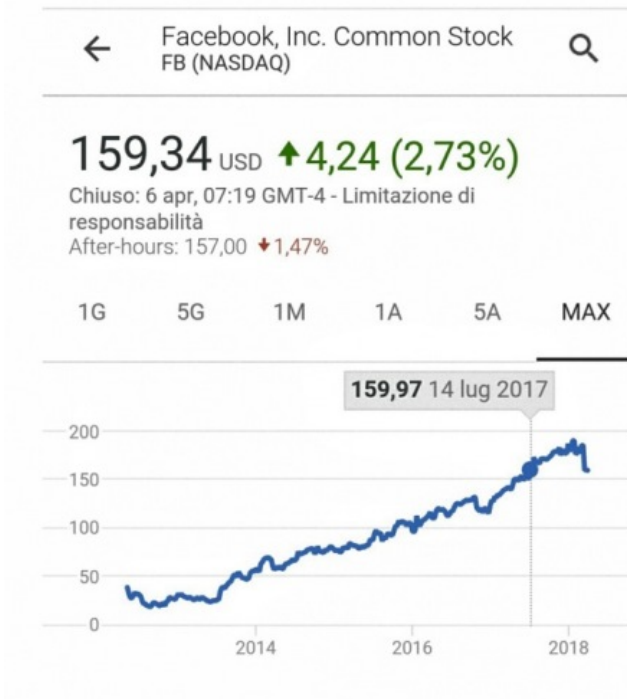
A present-day mistake is about Facebook:

Facebook drops a bombshell and says most of its 2 billion users may have had their personal data scraped

We knew, they knew that they were selling our data but we let them do and we continue to use Facebook.

Claiming the opposite would be like admitting that we still believe in Father Christmas.

Despite the last blasting news about Mark Zuckerberg admissions, the Facebook stock fell down a little compared to the whole capitalisation gained since its start. Moreover, compared to the composite index Nasdaq-100 which lost five trimesters, the Facebook stock lost just three trimesters only.



So far, Facebook had generated a lot of value for the stockholders and not only for them.

All such data have been extracted from Facebook will stay around for much time after. Such data could be used in the future to teach AI driven cars how to react to an unavoidable incident. In particular how to ponderate the weight of a life loss.

There is no solution for this problem because

| we are the problem for which we are looking for a solution!

All winners in a losing game

| With very rare exceptions, automakers are famously coy about crash dilemmas. They don't want to answer questions about how their self-driving cars would respond to weird, no-win emergencies. This is understandable, since any answer can be criticized—there's no obvious solution to a true dilemma, so why play that losing game?

We will buy the autonomous driving car services from those would be able to lie to us in a more convincing and convenient way.

This is appalling but unavoidable because no one would feel responsible or involved in such mess but just move safely, quickly and in a convenient way to their destinations.

However, because the AI driven cars will greatly reduce the number of incidents and fatalities on the road – and there is no doubts that AI will be able to drive in safer way than humans on average – everybody will win.

The MIT moral machine

A platform for gathering a human perspective on moral decisions made by machine intelligence, such as self-driving cars.

<http://moralmachine.mit.edu>

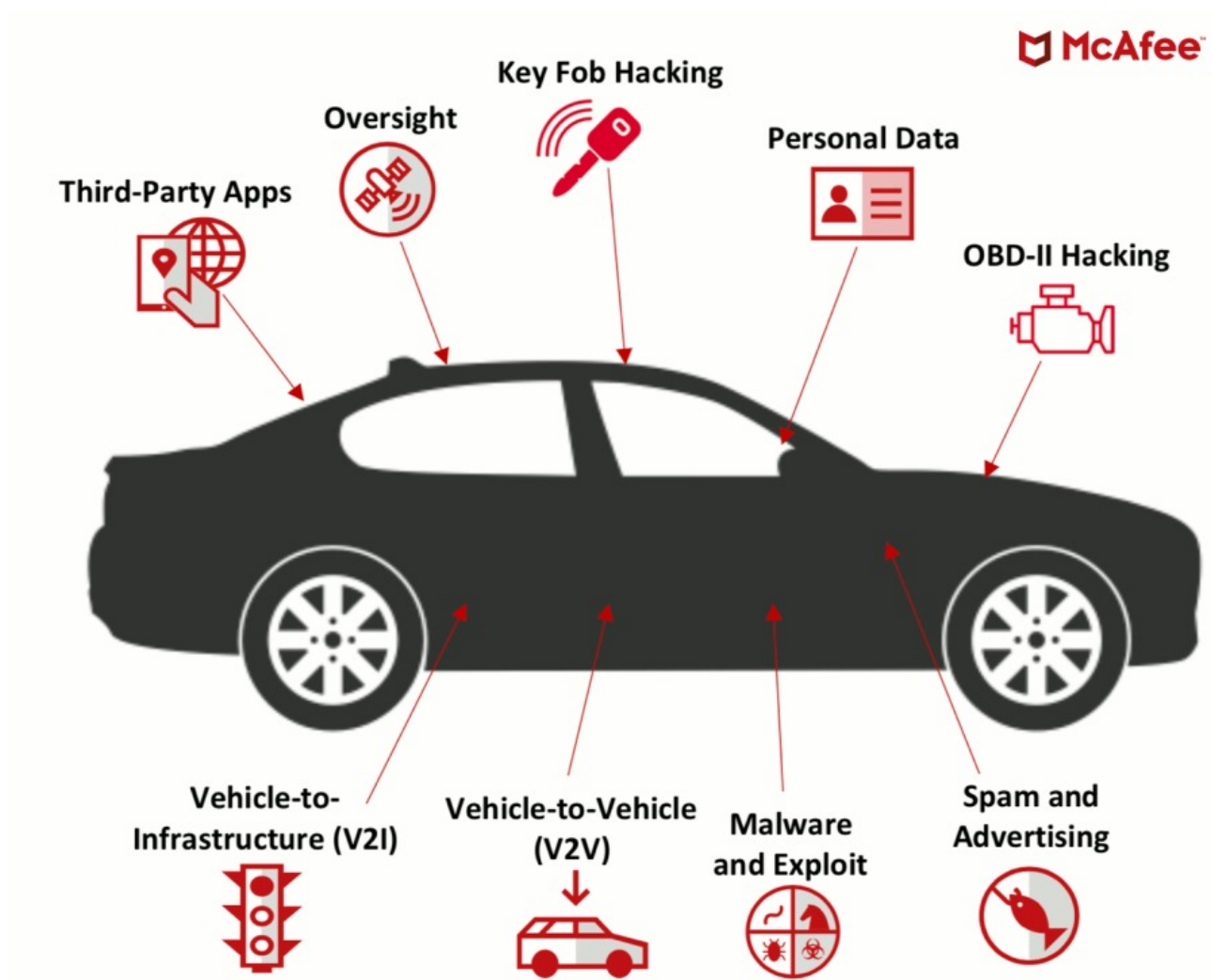
We show you moral dilemmas, where a driverless car must choose the lesser of two evils, such as killing two passengers or five pedestrians. As an outside observer, you judge which outcome you think is more acceptable. You can then see how your responses compare with those of other people.

If you're feeling creative, you can also design your own scenarios, for you and other users to browse, share, and discuss.

The security concerns

Modern cars are vulnerable to hacking and malware attack:

Today convicted modern cars are using a lot of technology and it always connected with internet that makes it extremely vulnerable and easy to compromise using malware attacks and other security flaws that presented in the IoT devices [and smartphones] that connected with modern cars.



Thus the autonomous driving cars could be a target for malicious attacks, as well. The outcoming threat could be seriously amplified as much as they would operate behind humans supervision.

External readings

[Here's how Tesla solves a self-driving crash dilemma](#) by Patrick Lin for Forbes on April 5th, 2017

[Modern Cars are Vulnerable to Hacking and Malware Attack](#) by GBHackers on April 2th, 2018

[About the Moral Machine](#) by Massachusetts Institute of Technology, since 2016

Related articles

[The dark side of innovation: why top managers should care?](#) (14 maggio 2016, EN)

[The A.I. seen from the dark side of the moon](#) (27 ottobre 2017, IT->EN)

[God does not play dice](#) (2 marzo 2018, IT->EN)