F1TENTH Autonomous Racing

Moral Decision Making for Autonomous Systems

Rahul Mangharam
University of Pennsylvania
rahulm@seas.upenn.edu

Johannes Betz
University of Pennsylvania
joebetz@seas.upenn.edu

F1 TENTH

Safe Autonomy Lab
University of Pennsylvania

Penn Engineering

I. Autonomous Driving

Autonomous Driving: Engineering Goals

As an engineer when designing autonomous Systems we need to take care of the following:

- **Functionality:** Drive everywhere, anytime, under every condition
- **Safety:** Do not crash, safe pedestrian, safe passengers
- **Security:** No access to the car from outside for hackers
- **Reliability:** No failures, no downtimes, no outages

What is your opinion?

Do you think an engineer of an autonomous system needs to consider ethics?


PREFECT
Winterthur
Ohringen
Zentrum
Polizei
71
2016-05-19 07:04:02 --- km/h
DR650GW-2CH/FHD-HD

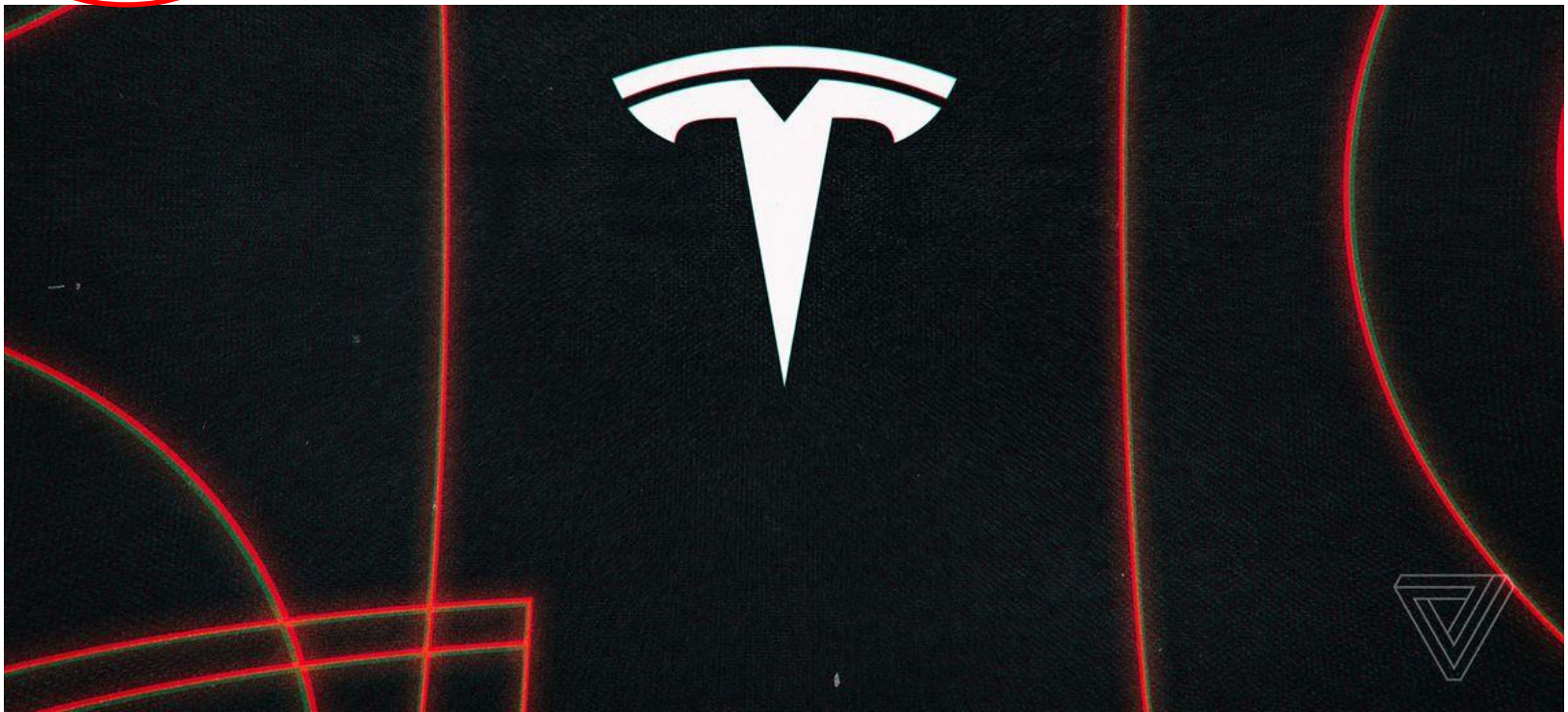

gb
1 CCTV


2018/11/17 12:17:04
2018 Tesla crash – these failures are real!

Tesla ‘recalls’ more than 285,000 vehicles in China over cruise control safety concerns

Most of the affected vehicles were made in China at Tesla’s Shanghai plant

By Kim Lyons | Jun 26, 2021, 9:02am EDT



Autonomous Driving: Responsibility?

Since the early days when it became apparent that self-driving cars are a real possibility, engineers, regulators and jurists started asking who would be liable when the car injures or kills someone through a mistake of its own?
For instance, take the case of the Uber incident in Arizona, where an SUV in Uber autonomous mode struck and killed 49-year-old Elaine Herzberg on Sunday March 18, 2018[1].

Question: *Who is responsible for her death?*

- The car's passenger, though they were not in physical control of the car at the time
- The car's autonomy manufacturer (in this case, Uber)
- The designer of the computer vision algorithm that failed to identify the woman with sufficient
- The company that provided the dataset on which the computer vision algorithm (a deep neural net) was trained
- The city and state regulators who allowed Uber to test on public roads at such an early stage

Autonomous Driving:

Moral and Legal Responsibility?

- We will explore this issue in our discussion, but in a different context, which brings out the urgency and salient traits of this issue in sharp relief: namely, in the context of autonomous weapons, that make the kill decision and implement it independently of humans.
- This is not because we think autonomous weapons are fine and dandy - we don't.
- Rather, it is because thinking about autonomous weapons makes it very clear what is at stake in terms of responsibility assignment.
- Who is morally responsible when an autonomous robot takes a decision to kill a human being?

II. Killer Robots

Reading Assignment: Killer Robots

Journal of Applied Philosophy, Vol. 24, No. 1, 2007

Killer Robots

ROBERT SPARROW

ABSTRACT *The United States Army's Future Combat Systems Project, which aims to manufacture a 'robot army' to be ready for deployment by 2012, is only the latest and most dramatic example of military interest in the use of artificially intelligent systems in modern warfare. This paper considers the ethics of the decision to send artificially intelligent robots into war, by asking who we should hold responsible when an autonomous weapon system is involved in an atrocity of the sort that would normally be described as a war crime. A number of possible loci of responsibility for robot war crimes are canvassed: the persons who designed or programmed the system, the commanding officer who ordered its use, the machine itself. I argue that in fact none of these are ultimately satisfactory. Yet it is a necessary condition for fighting a just war, under the principle of* jus in bellum, *that someone can be justly held responsible for deaths that occur in the course of the war. As this condition cannot be met in relation to deaths caused by an autonomous weapon system it would therefore be unethical to deploy such systems in warfare.*

Robert Sparrow:

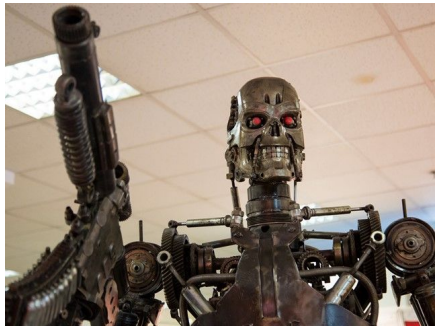
- Professor of Philosophy
- Research:
  - Political Philosophy
  - Applied Ethics
- Robotics and AI, Bioethics

Penn Engineering F1 TENTH

Killer Robots

Question 1:

What is the thesis that Sparrow developed in his paper?
Give a short summary what you think, Sparrow wants us to know about autonomous weapons and killer Robots.



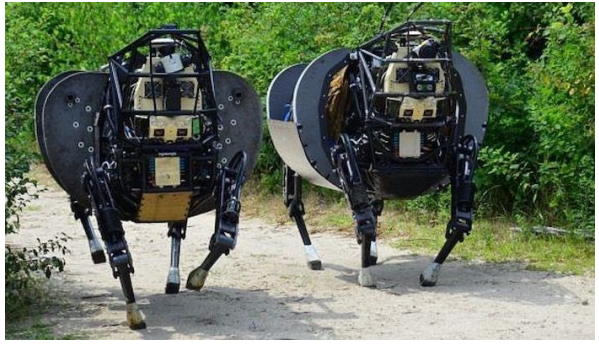
Penn Engineering F1 TENTH

Killer Robots

Question 2:

At what point would you consider a system to be autonomous?

E.g. if the system is only told "Find the enemy infrastructure and destroy it **vs.**
Having a human guide the weapon to within 1 mile of a pre-specified target **vs.**
Guiding it to the final target and impact point.



Killer Robots

Question 3:

Sparrow says that "Military interest in UCAVs stems from two main sources":

1. Uninhabited systems offer the prospect of achieving military objectives without risking the politically unacceptable cost of friendly casualties.

2. They are expected to be substantially cheaper than the piloted systems they are intended to replace

What is the implicit value system that allows this calculation of political cost?
Whose lives must be protected according to this calculation?

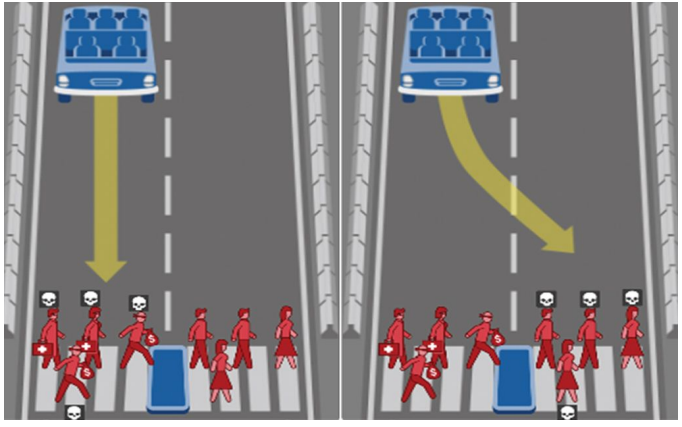
Penn Engineering F1 TENTH

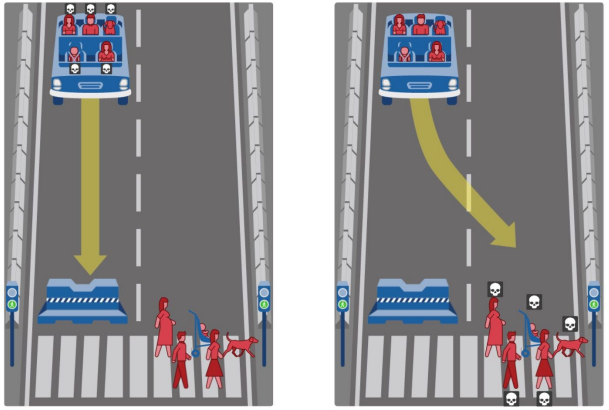
Killer Robots

Question 4:

Let us think about Questions 3 again and try now to adapt this to self-driving cars:

Is there a similar dynamic in self-driving cars? Whose lives must be protected according to this calculation? Is one group's interest advanced over another's?





Killer Robots

Question 5: Read the opening example of p.5. Hypotheticals are a very useful way of analyzing situations and distinguishing the important factors at play.

"Let us imagine that an airborne AWS, directed by a sophisticated artificial intelligence, deliberately bombs a column of enemy soldiers who have clearly indicated their desire to surrender. These soldiers have laid down their weapons and pose no immediate threat to friendly forces or non-combatants. Let us also stipulate that this bombing was not a mistake; there was no targeting error, no confusion in the machine's orders, etc. It was a decision taken by the AWS with full knowledge of the situation and the likely consequences. Indeed, let us include in the description of the case, that the AWS had reasons for what it did; perhaps it killed them because it calculated that the military costs of watching over them and keeping them prisoner were too high, perhaps to strike fear into the hearts of onlooking combatants, perhaps to test its weapon systems, or because the robot was seeking to revenge the 'deaths' of robot comrades recently destroyed in battle."

Killer Robots

Question 6:

Why do you think it is important to assign responsibility, especially when autonomous weapons are used?

You can argue with Sparrow or with your own thoughts.

Killer Robots

Question 7:

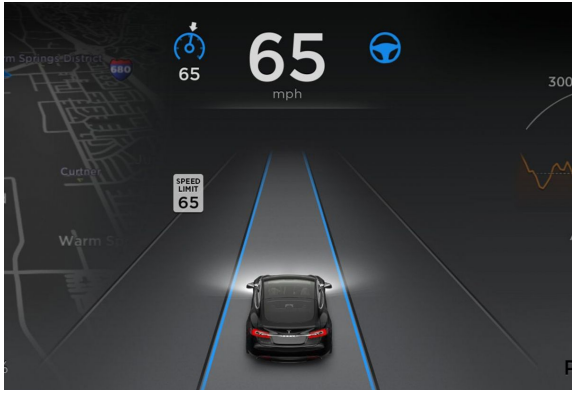
Now let us transfer this arguments to the field of autonomous driving.

Do we see parallel obligation in autonomous vehicles to be able to assign responsibility for the willful killing of pedestrians?

Killer Robots

Question 8:
Sparrow says: "There is an obvious tension involved in holding that there are good military reasons for developing autonomous weapon systems but then not allowing them to fully exercise their 'autonomy'."

Let us try to transfer that to an AV example: *Is there a similar tension in Tesla's so-called AutoPilot between reasons for using autonomy, and promises to limit its use?*



Killer Robots

Question 9:

Sparrow lists three candidates for the locus of responsibility that he consider exhaustive:

1. The programmer
2. The commanding officer
3. The machine itself.

Do you agree with this list or do you think there are more candidates?

Can the programmer be taken as meaning 'the company', and the commanding officer as meaning 'the government'?

Killer Robots

Question 11:

Recall the Commanding Officer sub-section of the "Killer Robots" paper (pages 9-10).

Who is the equivalent of the commanding officer in the case of autonomous car?
I.e., who 'orders' the car onto public streets?
What does this imply for responsibility assignment?

Killer Robots

Question 12:

Recall the argument against holding the machine responsible.

This is a *reducio ad absurdum*: take the reasoning to its logical conclusions and reach a contradiction.

If the machine can be said to genuinely suffer, then we should be concerned about its pain in war, and should be as reluctant to send it into war as a human. This defeats the purpose behind building it!

Similarly, a truly suffering machine would also feel empathy and would not be a cold killing machine.

Killer Robots

Question 13:

Once we hold someone responsible for an action, do we then always follow that by reward or punishment?

What are other purposes of assigning responsibility?

EAS 203 - Guest Lecture

Moral Decision Making for Autonomous Systems

Thank you for your participation!
Questions?

F1 TENTH

Penn Engineering