
http://ethics4eu.eu

# Ethics in Human-centric Design. A Case of Autonomous Cars

Gordana Dodig Crnkovic, Professor of Computer Science

Chalmers Technical University | University of Gothenburg & Mälardalen University, Sweden

http://gordana.se/

http://www.gordana.se/work/presentations.html

Lecture Plan, Part 1

13.15-14:00

## Ethics in Human-centric Design

- Ethics and morality
- Normative systems
- Classical ethical approaches
- Ethics for design
- Designer's code of Ethics

## The Case of Autonomous Cars

1. Development of AVs
2. Trolley Problem's unsolvability
3. Real-world engineering ethics of AV's
4. Technical challenges to AV ethics
5. Social challenges to AV ethics
6. Ethical guidelines
7. Further reading

14:0-14:15 Break

14:15-14:30 Preparation For Discussions

14:30-14:45 Breakout Room Discussions

14:45- Questionnaire & Conclusion

## Lecture Plan, Part 2

14.15-15:00

# Discussions in Breakout Rooms

# Questionnaire

10 minutes 1-page questionnaire

The aim: to help us develop the course as an element in the program of ethics education for students at technical universities in Europe.

# Ethics in Human-Centric Design

The aim of this lecture – to offer new views

As the topic of Design Ethics, AI ethics and even AV ethics are huge, what this lecture can do is to open the window with a view, giving you just a glimpse of a huge unexplored territory in front of us.

”I invite readers not on a visit to an archaeological museum, but rather on an
adventure in science in making”

Ilya Prigogine. The End of Certainty: Time, Chaos and New Laws of Nature, 1997

https://www.onventanas.com/historia-vidrio/ventana-japonesa/#iLightbox[postimages]/0

# Wicked Problems in Design Thinking

**Design Unbound. Designing for Emergence in a White Water World**

Ann Pendleton-Jullian and John Seely Brown, two volume set, MIT Press 2018

https://mitpress.mit.edu/books/design-unbound-designing-emergence-white-water-world-volume-1

Richard Buchanan (1992) Wicked Problems in Design Thinking. Design Issues, Vol. 8, No. 2, pp. 5-21. The MIT Press

http://www.jstor.org/stable/1511637

# Ethics and Morality - Etymology

Morality and ethics have the same roots, mores which means manner and customs from the Latin and etos which means custom and habits from the Greek.

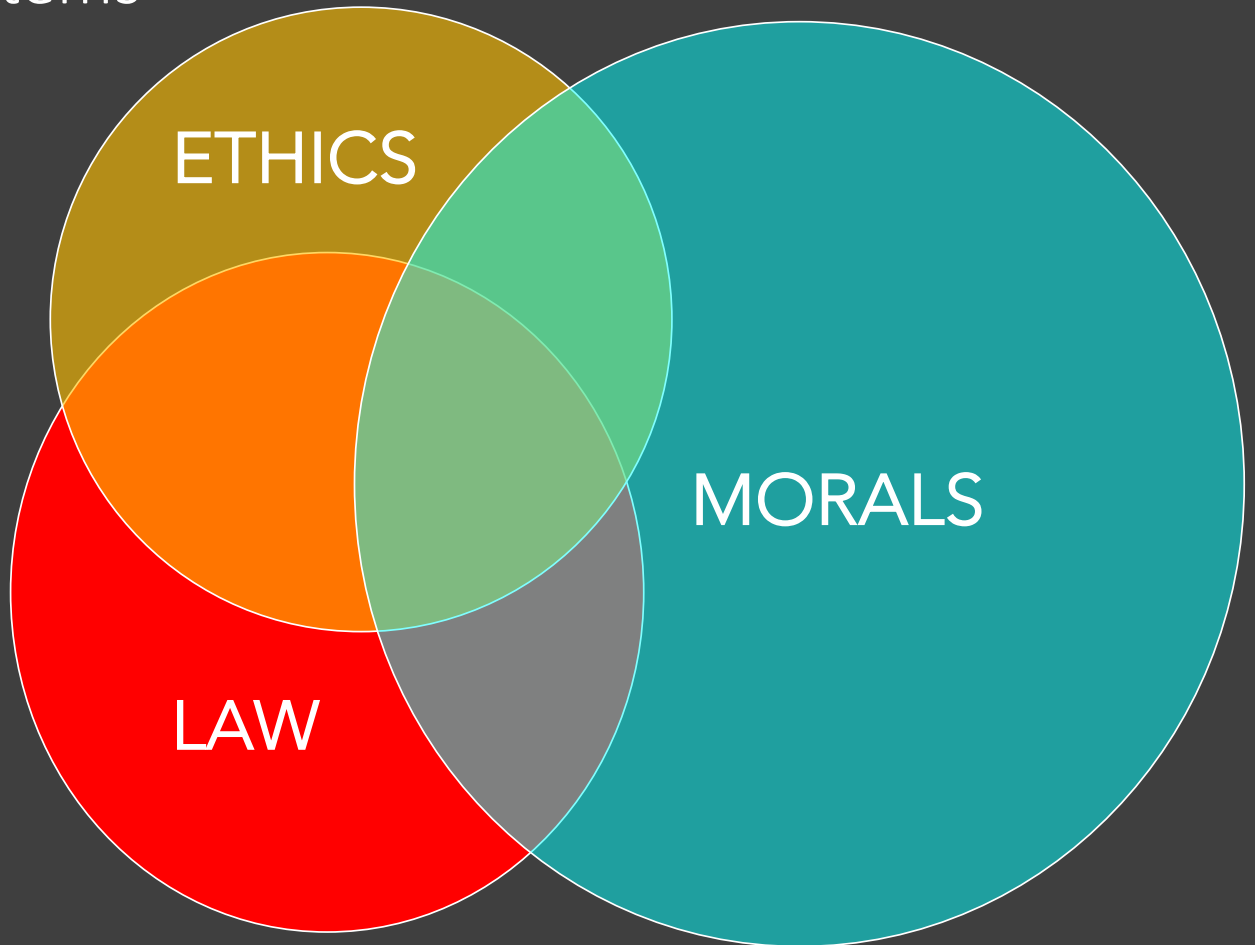
Robert Louden, Morality and Moral Theory

# Ethics and Morality

Morality: first-order set of beliefs and practices of a good life.

Ethics: a second-order, conscious reflection on the adequacy of our moral beliefs.

# Normative Systems



# Values

Values serve as a guide to action and knowledge.

They are relevant to all aspects of scientific and engineering practice, including discovery, analysis, and application.

# Classical Approaches to Ethics

## The Virtue Ethics

Focuses on attitudes, dispositions, or character traits that enable us to act in ways which develop our human potentials. Examples: honesty, courage, faithfulness, trustworthiness, integrity, etc.

The principle is: What is ethical is what develops moral virtues in ourselves and our communities.

# Classical Approaches to Ethics

## The Utilitarian Ethics

Focuses on the consequences that actions or policies have on the well-being ("utility") of persons directly or indirectly affected by the action or policy.

The principle is: Of any two actions, the most ethical one will produce the greatest balance of benefits over harms.

# Classical Approaches to Ethics

## The Rights-Based Ethics

Each person has a fundamental right to be respected and treated as a free and equal rational individual capable of making his or her own decisions.

The principle is: An action or policy is morally right only if those persons affected by the decision are not used merely as instruments for advancing some goal but are fully informed and treated only with their informed consent.

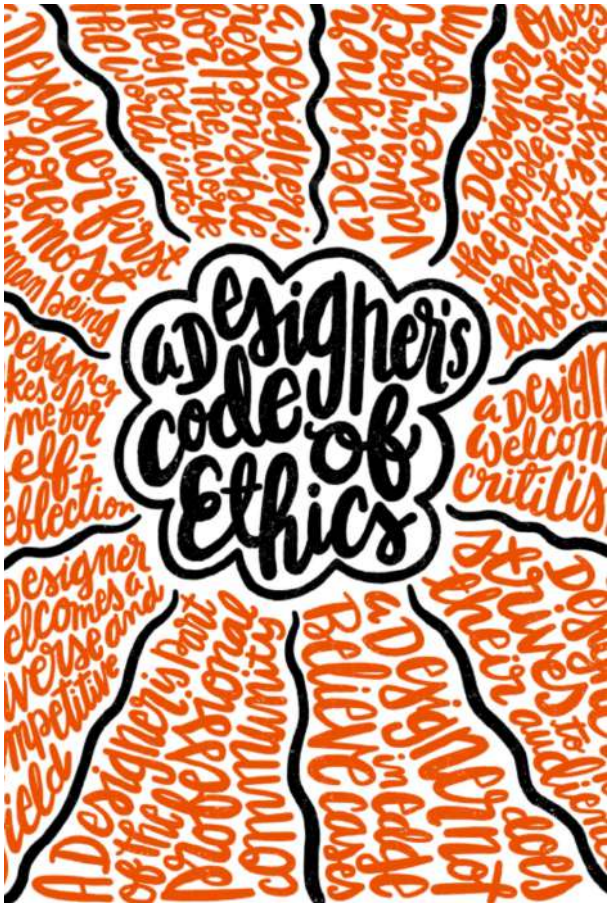

12 designers and researchers from 8 European cities discuss the impact of design on our societies and the paths to follow for designers to work for the good of all.

1. What does a designer do?
2. Is there something wrong with design?
3. Ethics and Morality
4. How designers can do better?

http://ethicsfordesign.com/player?lang=en
http://www.ethicsfordesign.com/player?lang=en

# Designer's Code of Ethics

(one among many)




## A Designer's Code of Ethics

A designer is first and foremost a human being.

is responsible for the work they put into the world.

values impact over form.

owes the people who hire them not just their labor, but their counsel.

welcomes criticism.

strives to know their audience.

does not believe in edge cases.

is part of a professional community.

welcomes a diverse and competitive field.

takes time for self-reflection.

# A Case of Autonomous Cars

Based on:

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global



Picture: https://www.aarete.com/insights/what-is-the-business-case-for-autonomous-vehicles-in-the-supply-chain/

# Why Self-Driving/Autonomous Cars?

We take Self-driving cars as an example of emerging technology that is combining advances in several underlying emergent technologies such as electric mobility and artificial intelligence (with connected driving, intelligent cities, intelligent infrastructure, etc.)

Technology emerges not in vaccuum but in its social context that today is global technosocial environment

How can we contribute in different roles as stakeholders to the development of good society with help of new powerful technologies. Who are the main actors/stakeholders and how do they affect the development? Autonompus cars have been studied a lot and we can learn from the development so far.

# PART 1
# AUTONOMOUS CARS DEVELOPMENT

Source: Wikipedia

On January 29, **1886,** Carl Benz applied for a patent for his “vehicle powered by a gas engine.” The patent – number 37435 – may be regarded as the birth certificate of the automobile.

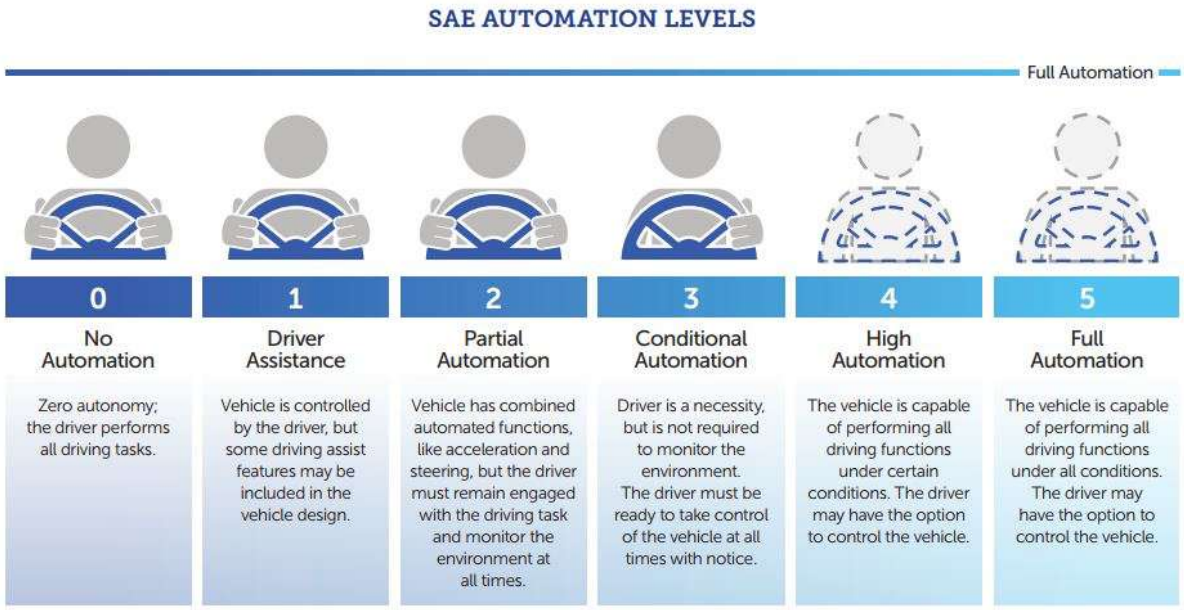
# Car History

Source: Tesla

Complex driving tasks are successively replaced by advanced driving assistant systems

Source: T. Holstein (taken at IAA 2015)

# LEVELS OF AUTOMATION



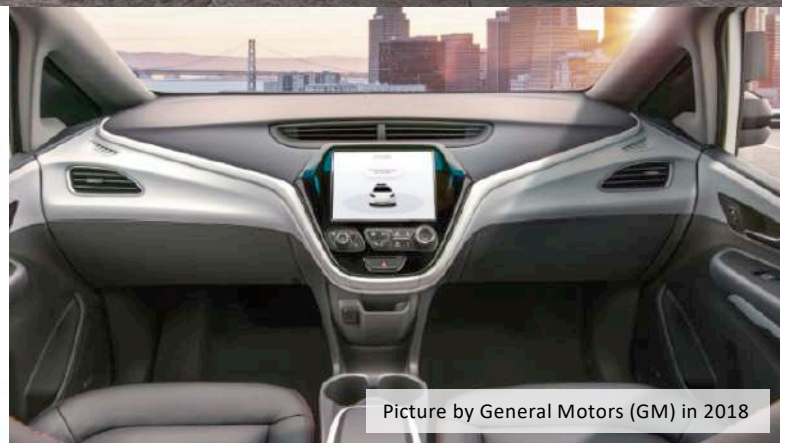
# With Self-Driving Cars into the Future

Picture by Google/Waymo in 2015


Picture by General Motors (GM) in 2018

Google I/O in 2016

Picture Source: http://blog.csloxinfo.com/wp-content/uploads/2017/01/google-self-driving-car-2016-photos-15-420-90.jpg

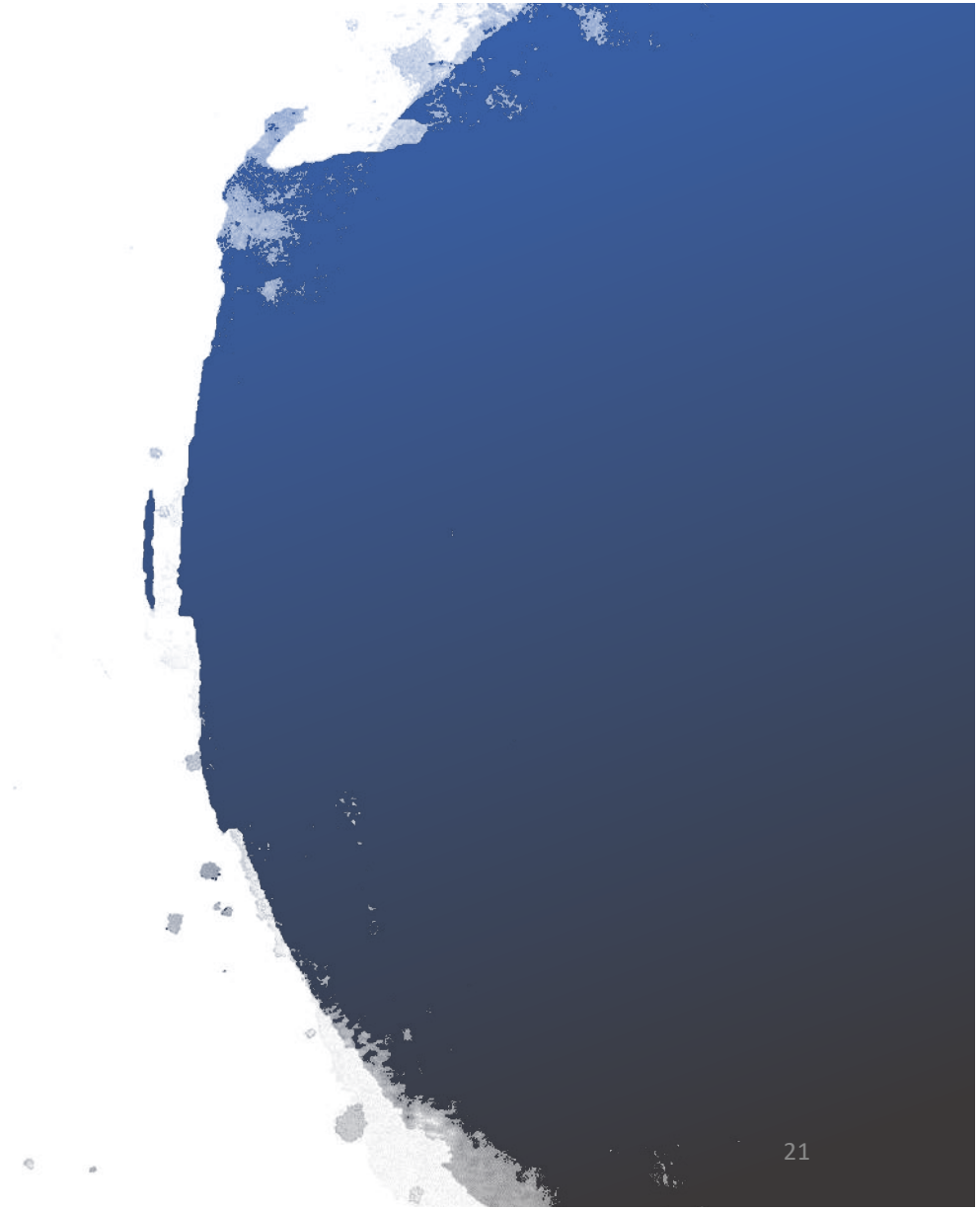


Picture by General Motors (GM) in 2018

Highest level of autonomous driving (Level 5 of 5), where a car can drive from A to B without human supervision.

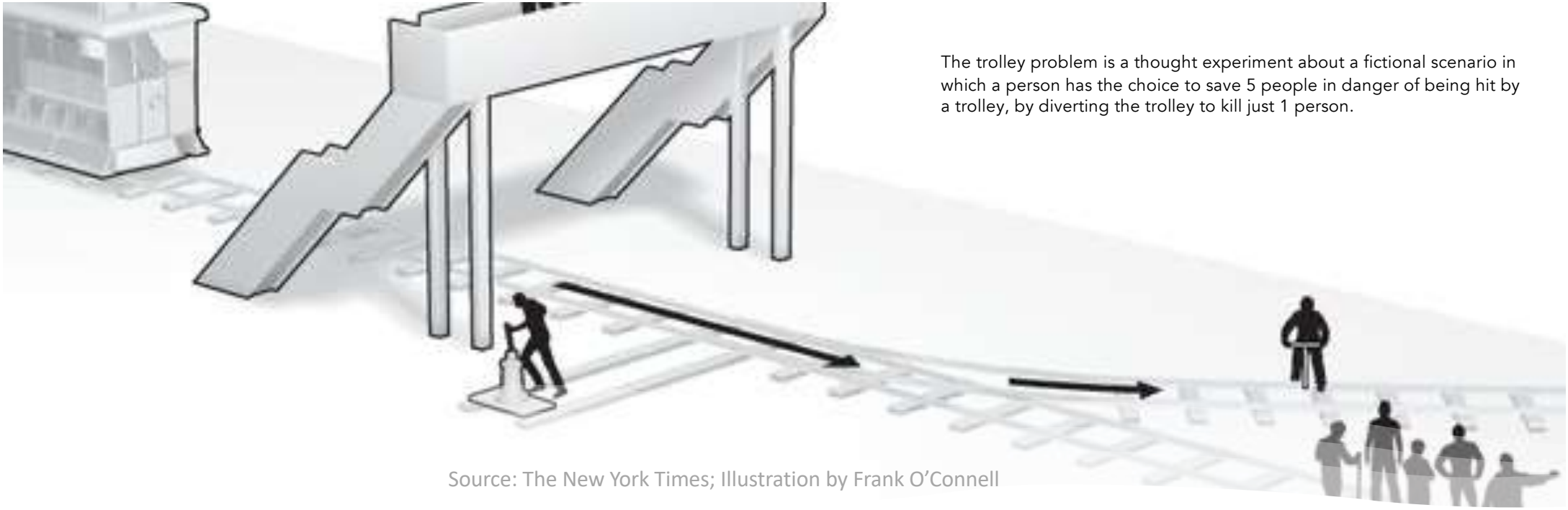
No Steering-Wheel or other primary driving controls => the former driver becomes solely passenger.

# PART 2 CURRENT ETHICS OF AUTONOMOUS CARS – TROLLEY PROBLEM



Current discussions about self-driving cars ethics repeatedly take form of decision-making problem borrowed from philosophy

THE TROLLEY PROBLEM:
Whom will the self-driving car kill when it has to decide?



The trolley problem is a thought experiment about a fictional scenario in which a person has the choice to save 5 people in danger of being hit by a trolley, by diverting the trolley to kill just 1 person.

Source: The New York Times; Illustration by Frank O'Connell

# The Trolley Problem Ethical Dilemma

Ethical thought experiment defined by philosopher Philippa Foot in "The Problem of Abortion and the Doctrine of the Double Effect," pp. 5-15, *Oxford Review,* 5, (1967). Focus on the difference between responsibility of acting vs. non-acting.

Many different variants, such as the use of personas to include an emotional perspective. But there is always a single decision: Whom to kill?

# Typical Approaches to the Trolley Problem…

…are based on the following ethical theories:

- Utilitarianism
- Other forms of consequentialism
- Deontological ethics

For example, utilitarianism would aim to minimize casualties, even if it means to kill the passenger in the car, by following the principle: the moral action is the one that maximizes utility (or minimizes the damage).

Depending on the ethics framework, different arguments can be used to justify the different decisions.

# The Trolley Problem is Unsolvable by Construction!

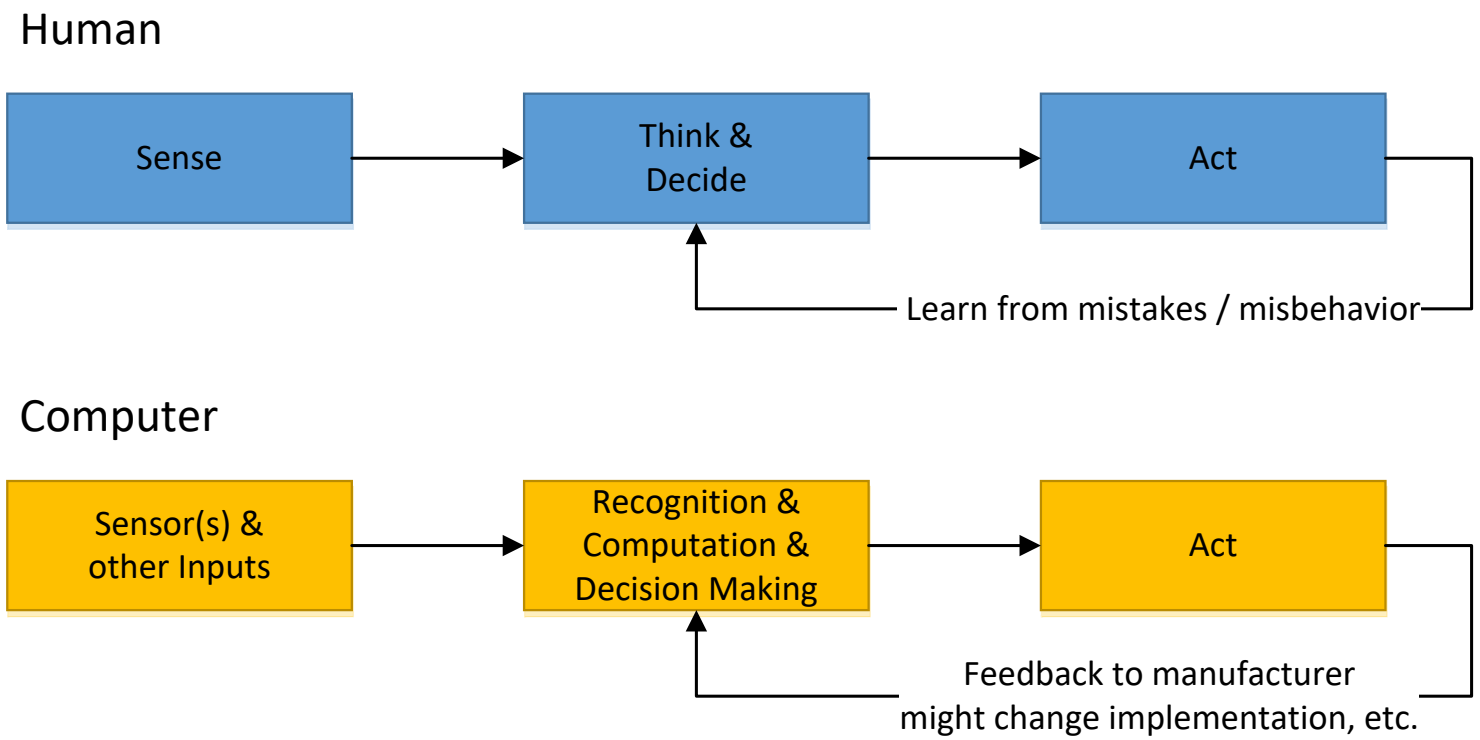
The problem is that for the question “whom to kill?” all answers are bad or wrong.

There is no correct answer to the Trolley Problem and therefore it is not the right kind of problem representative of a real-life situation and even less possible to approach by engineering.

# PART 3 THE REAL-WORLD ENGINEERING PROBLEM IS NOT WHOM TO KILL BUT HOW NOT TO KILL!

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

# Human Decision-making Process versus Self-Driving Car (Computer)



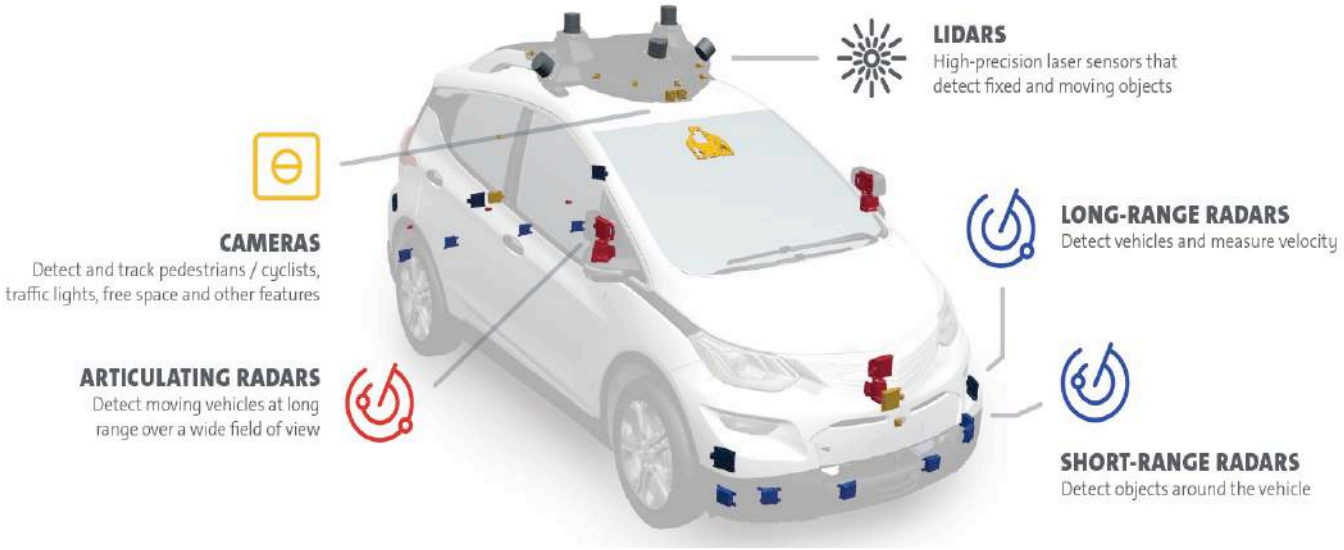
# Decision Making in Self-Driving Cars

Decision making process involves sensors, external sources of information, networks, hardware, software, etc.

Environmental influences, such as weather conditions (rain, bright sun, storm, …)

Complex input must be filtered and only represents an abstraction of the real world.

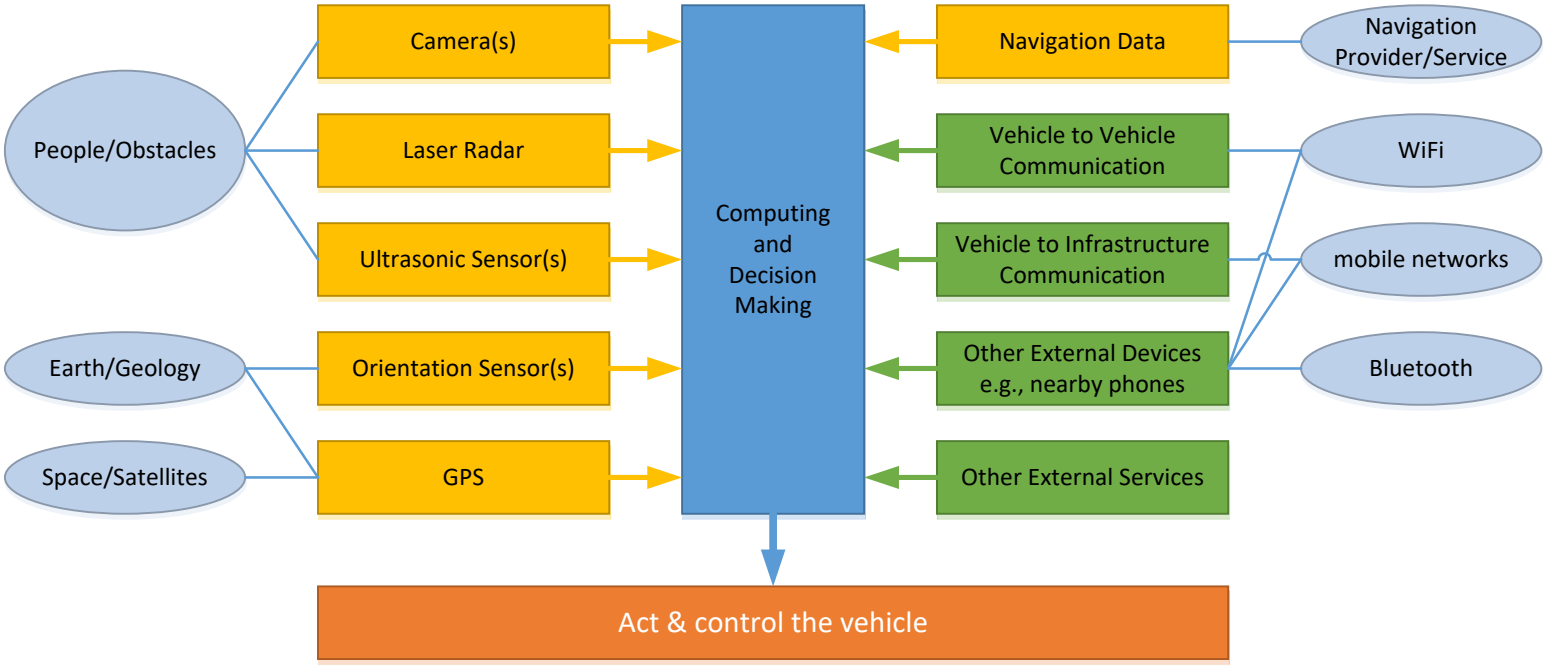
# Technical Components




Picture Source: General Motors Safety Report 2018

# What does a Self-Driving Car “see”…

# Abstract Decision Making Process




This is an outline of what a decision making process might include.
It is based on a literature review and official press releases (Tesla, Google, GM).

# PART 4 TECHNICAL CHALLENGES WITH ETHICAL CONSEQUENCES IN AUTONOMOUS CARS

# Safety

| | |
|---|---|
| **How can we test self-driving cars?** | and when is testing sufficient? |
| **Real world vs Abstract World** | Training of Neural Networks |

# Security

Attacks against car systems and sensors

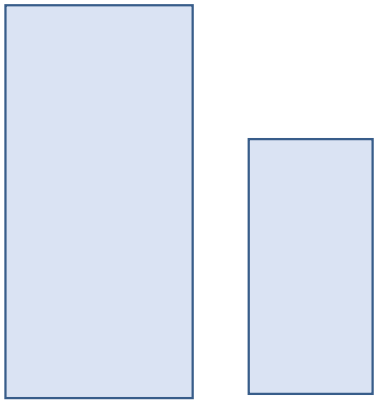
System & security updates

Do we need a “black box” in self-driving cars like in aircrafts?

# Privacy

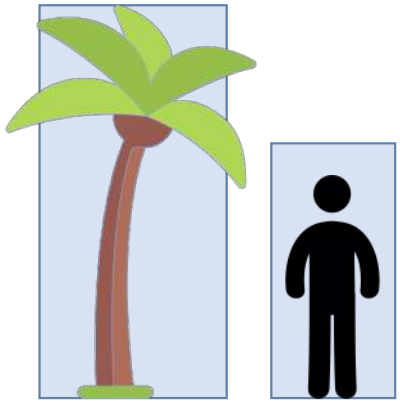
- What data should the car have access to?
  - Who will have access to that data?
  - How will the data be used?
- How is data collected/stored/communicated?
- Privacy in connected driving

# Privacy
# What does the car “recognize”?

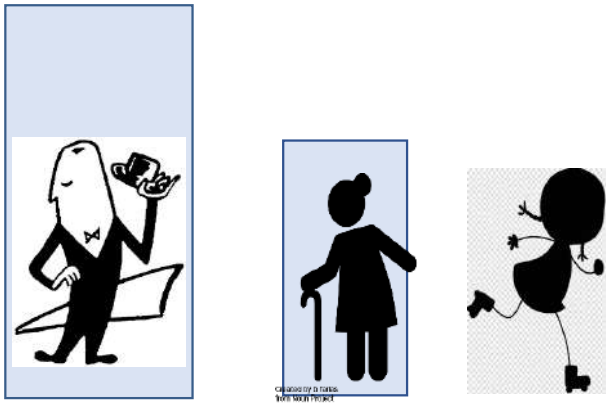


Objects, different size,

Position, moving or stationary



Objects vs Person(s)



„Everything“ including human identity – connected to data-bases

# Trust

**How trustworthy are data sources?**

E.g., GPS, map data, external services

Trust between self-driving car and services

**How trustworthy is the self-driving car?**

E.g., Trust between user and car

# Transparency

Multi-disciplinary challenge to ensure transparency, while respecting intellectual property rights, corporate secrets, security concerns, etc.

How much should be disclosed, and disclosed to whom?

# Reliability

What do we have to rely on?

- What if sensor(s) fail?
- What if networks fail?

Redundancy for everything?

# Responsibility and Accountability

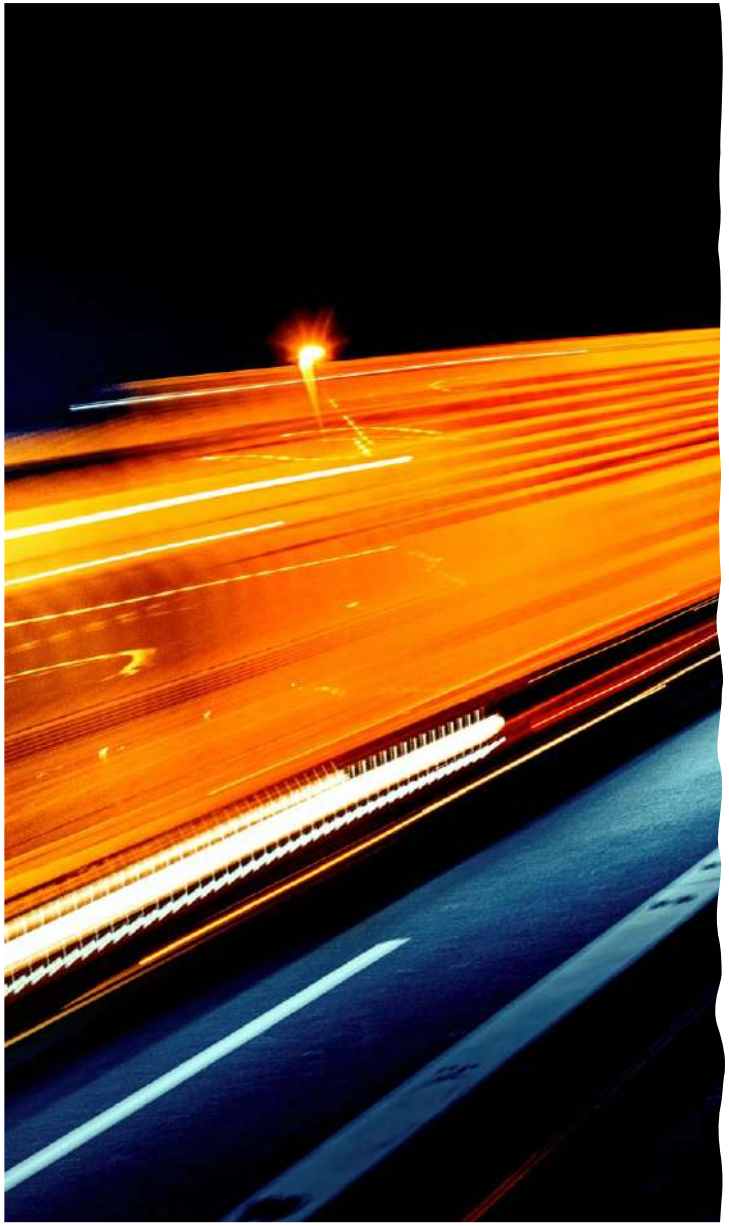
- Who is responsible and for what?
- Who is accountable and for what?
- How is responsibility distributed among:
  - Developers
  - Car manufacturers
  - Safety inspectorates
  - Governmental institutions
  - Involved participants in the traffic
  - Other stakeholders

# Quality Assurance Process

Lifetime of components

Maintenance

Ethics-aware decision making in all processes will help to make ethically justified decisions.



# PART 5 SOCIAL CHALLENGES OF AUTONOMOUS CARS WITH ETHICAL CONSEQUENCES

# Stakeholders Interests

- Loss of jobs (for cabs/taxi/truck/heavy industrial vehicles drivers)
- Humans in the loop
- Impact on Society

# Stakeholders Interests

Freedom of movement

Will the car go, where I want it to go?

Implementation of restrictions/preferences?

Route to destination

Can the passenger define the route, or is it determined by the system?

Road trips (similar to GPS choices of route)?

# Addressing Ethical Issues of AI

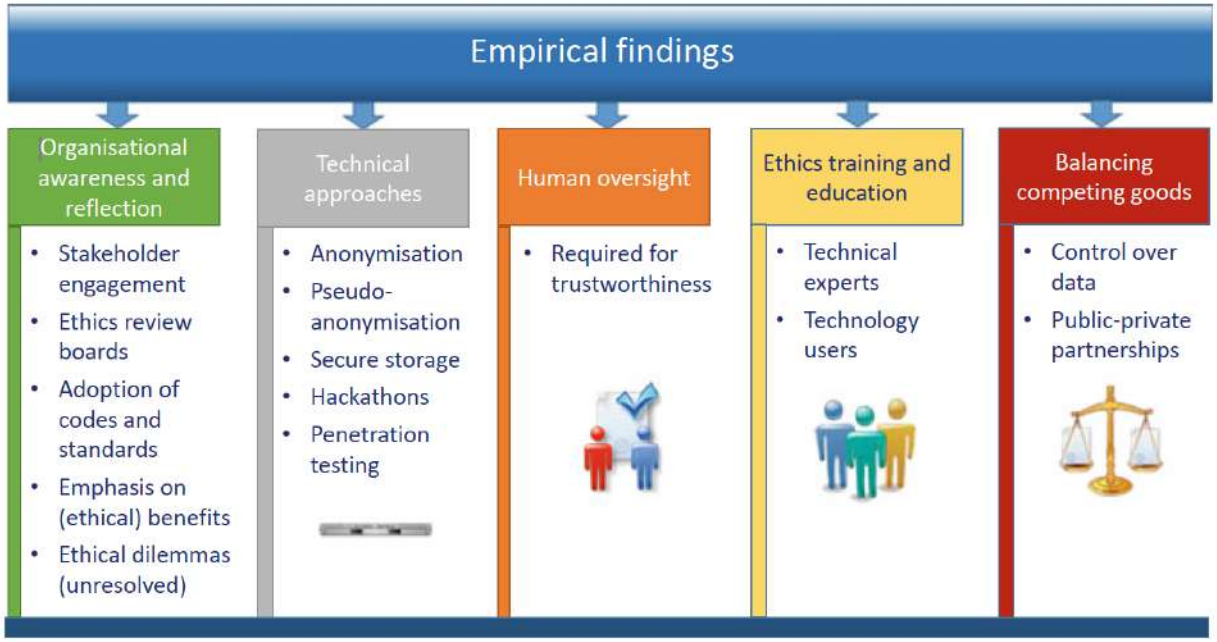
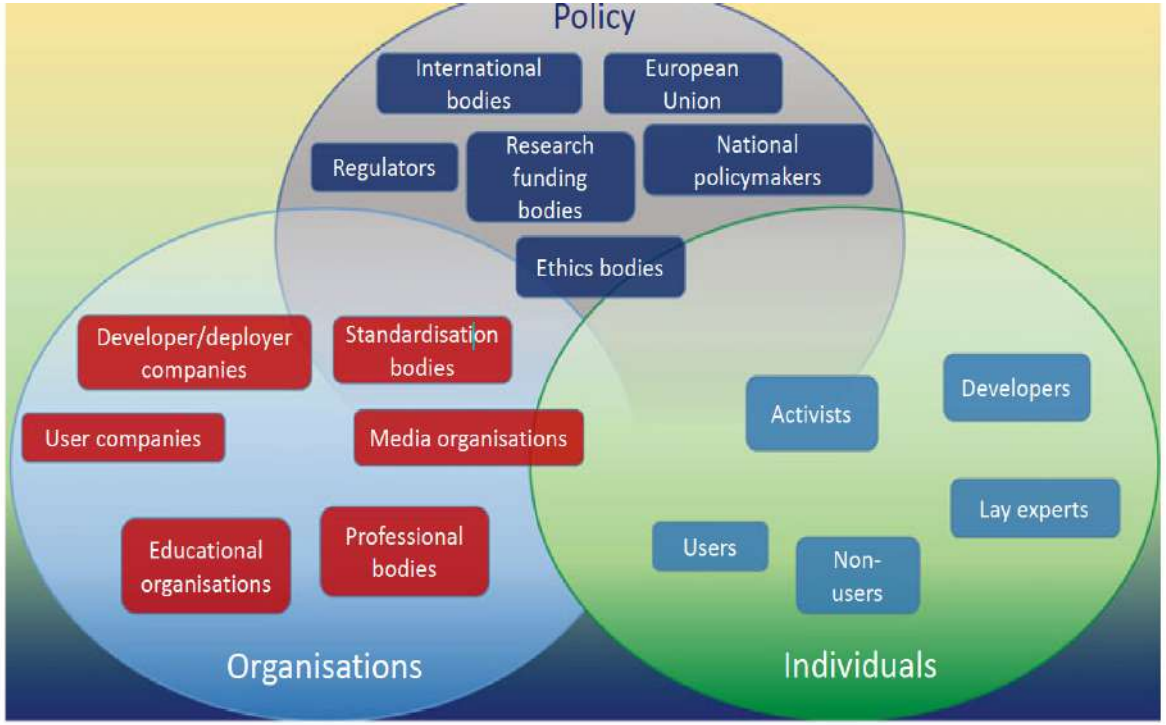



**Fig. 5.2** How case study organisations address ethical issues of AI: empirical findings

Bernd Carsten Stahl (2021) Artificial Intelligence for a Better Future, https://link.springer.com/book/10.1007%2F978-3-030-69978-9

# Overview of AI stakeholders, Artificial Intelligence for a Better Future




Bernd Carsten Stahl (2021) Artificial Intelligence for a Better Future, https://link.springer.com/book/10.1007%2F978-3-030-69978-9

# KEY CHALLENGES OF ETHICAL GOVERNANCE OF AI SYSTEMS

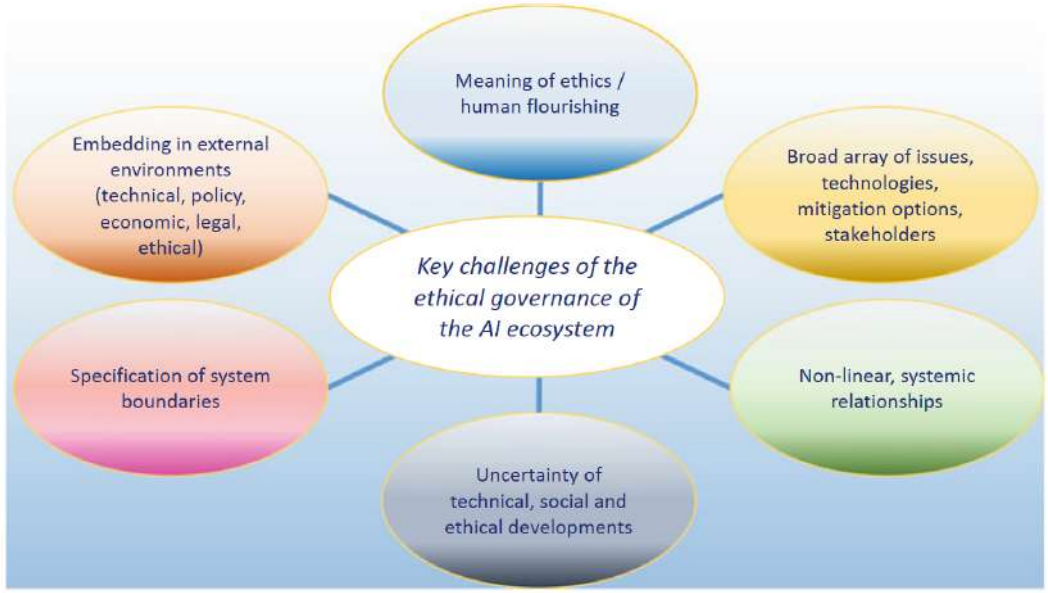



**Fig. 7.1** Key challenges of ethical governance of AI ecosystems

Bernd Carsten Stahl (2021) Artificial Intelligence for a Better Future, https://link.springer.com/book/10.1007%2F978-3-030-69978-9

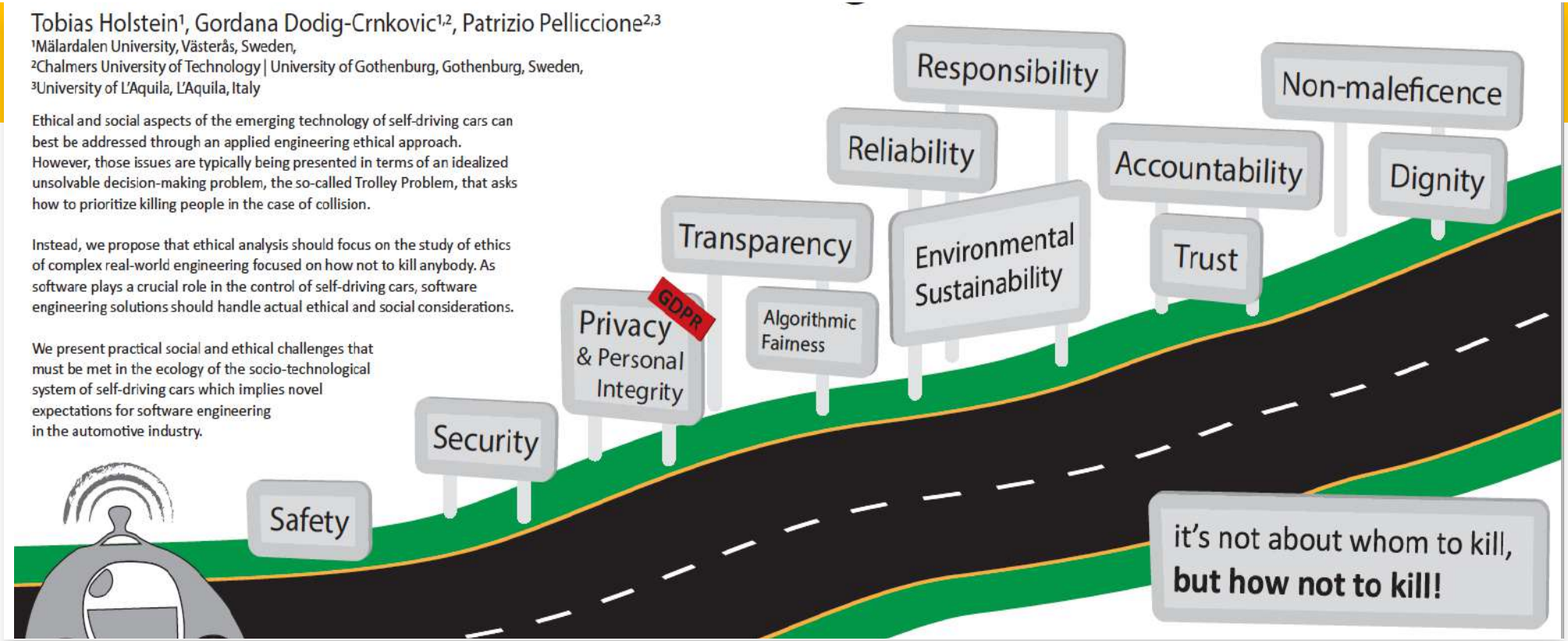
# Ethical Issues of AI

Table 4.1 Three categories of ethical issues of artificial intelligence

| | |
|---|---|
| **1. Issues arising from machine learning** | |
| **Privacy and data protection** | Lack of privacy |
| | Misuse of personal data |
| | Security problems |
| **Reliability** | Lack of quality data |
| | Lack of accuracy of data |
| | Problems of integrity |
| **Transparency** | Lack of accountability and liability |
| | Lack of transparency |
| | Bias and discrimination |
| | Lack of accuracy of predictive recommendations |
| | Lack of accuracy of non-individual recommendations |
| **Safety** | Harm to physical integrity |
| **2. Living in a digital world** | |
| **Economic issues** | Disappearance of jobs |
| | Concentration of economic power |
| | Cost to innovation |
| **Justice and fairness** | Contested ownership of data |
| | Negative impact on justice system |
| | Lack of access to public services |
| | Violation of fundamental human rights of end users |
| | Violation of fundamental human rights in supply chain |
| | Negative impact on vulnerable groups |
| | Unfairness |
| **Freedom** | Lack of access to and freedom of information |
| | Loss of human decision-making |
| | Loss of freedom and individual autonomy |
| **Broader societal issues** | Unequal power relations |
| | Power asymmetries |
| | Negative impact on democracy |
| | Problems of control and use of data and systems |
| | Lack of informed consent |
| | Lack of trust |
| | Potential for military use |
| | Negative impact on health |
| | Reduction of human contact |
| | Negative impact on environment |
| **Uncertainty issues** | Unintended, unforeseeable adverse impacts |
| | Prioritisation of the "wrong" problems |
| | Potential for criminal and malicious use |
| **3. Metaphysical issues** | |
| | Machine consciousness |
| | "Awakening" of AI |
| | Autonomous moral agents |
| | Super-intelligence |
| | Singularity |
| | Changes to human nature |

https://link.springer.com/book/10.1007%2F978-3-030-69978-9 Artificial Intelligence for a Better Future (book)

# PART 6 Ethical Guidelines for Self-Driving Cars




# Ehics Of Self-Driving Cars

Presented at major SE conference ICSE2020 as poster

Extended version in a book chapter:

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

# Ehics Of Self-driving Cars Paper* Method

1. A list of ethical values (requirements) for autonomous cars was compiled from an extensive literature study and connected to challenges together with approaches for their fulfilment.
2. The proposed analysis with values, challenges and approaches for technical and social aspects was presented to different stakeholders and discussed via seminars.
3. Finally we number of experts was invited to review the results.

* Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

# Practical Use of the Proposed Ethical Framework

Ethical requirements must be fulfilled in all phases in the life-cycle of a product (autonomous car) The context of:

1. Conceptualization/Design/Prototyping/ Construction/Development/Testing/Production
2. Deployment/Application/
3. Maintenance/Support
4. Oversight/Regulation


https://www.indiamart.com/kaynes-tecnology-india-pvt-ltd/product-life-cycle-management.html

# The First Ethical Guidelines For Automated Driving

German Ethics Commission's report comprises 20 propositions. The key elements are:

Automated and connected driving is an ethical imperative if the systems cause fewer accidents than human drivers (positive balance of risk).

Damage to property must take precedence over personal injury. In hazardous situations, the protection of human life must always have top priority.

https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

# The First Ethical Guidelines For Automated Driving

In the event of unavoidable accident situations, any distinction between individuals based on personal features (age, gender, physical or mental constitution) is impermissible.

In every driving situation, it must be clearly regulated and apparent who is responsible for the driving task: the human or the computer.

It must be documented and stored who is driving (to resolve possible issues of liability, among other things).

Drivers must always be able to decide themselves whether their vehicle data are to be forwarded and used (data sovereignty)

https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

# “Learning By Experience” And “Proven In Use” Concepts

“Learning by experience” (recording data from autonomous cars) presupposes a functioning socio-technological system that provides strong coupling among legislation, guidelines, standards and use, and promptly adapts to lessons learned.

H. Schäbe and J. Braband. Basic requirements for proven-in-use arguments. CoRR, abs/1511.01839, 2015.

# Challenges

| Legislation | Global framework | Guidelines | Ethics in the whole techno-social ecology |
|---|---|---|---|
| Keeping legislation up-to-date with current level of automated driving, meeting emergence of self-driving cars. | Creating and defining global legislation frameworks for the interoperable implementation and development of increasingly automated vehicles. | Defining the (globally-compatible) guidelines for self-driving cars. | Including ethical guidelines in the design and development processes, in the entire life-cycle and AV techno-social ecology. |

Holstein, Dodig-Crnkovic, Pellizzione: Ethical and Social Aspects of Self-Driving Cars, ArXives arXiv:1802.04103v1 [cs.CY] 5 Feb 2018

# Recommendations

Car producers supporting and collaborating with legislators in their task to keep up-to-date with the current level of automated driving.

Legislative support and contribution to global frameworks to ensure a smooth enrollment of the emerging technology.

Inclusion of ethics in the overall process of design, development and implementation of self-driving cars. Ensure adequate ethics training for involved engineers.

Establishing and maintenance of a functioning socio-technological system in addition to existing functional safety and design standards.

# Conclusions

It is time to stop discussing unsolvable ethical dilemmas that obfuscate much bigger actual ethical challenges in gthe development of autonomous cars.

Discuss the real-world ethical challenges of increasingly automatized, autonomous and driverless vehicles.

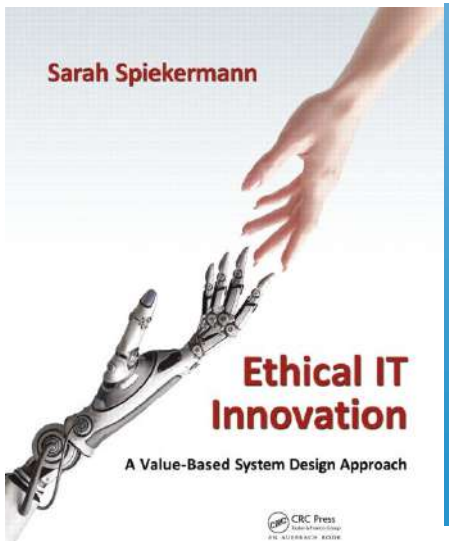
Define what is technically possible and ethically justifiable.

Create transparency to support evaluations by independent organisations/experts.

We propose Ethicality as non-functional property.
(Ethicality: the state, quality, or manner of being ethical.)

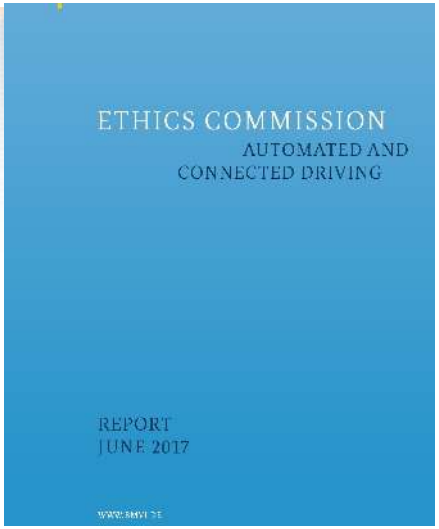
There is already a body of normative documents supporting ethicality of AV's and they continue to develop, to match technology development.

# A Value-Based Design Approach



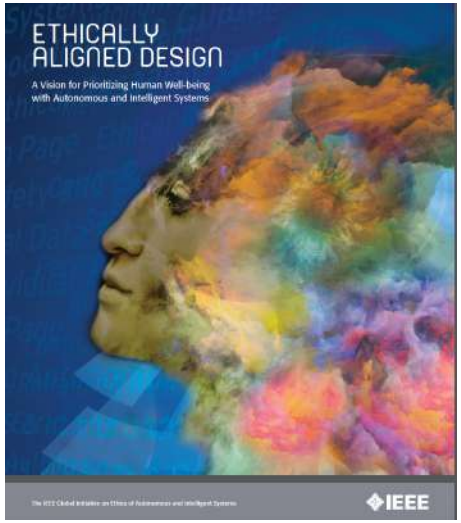

Sarah Spiekermann

Ethical IT Innovation:
A Value-Based System Design Approach



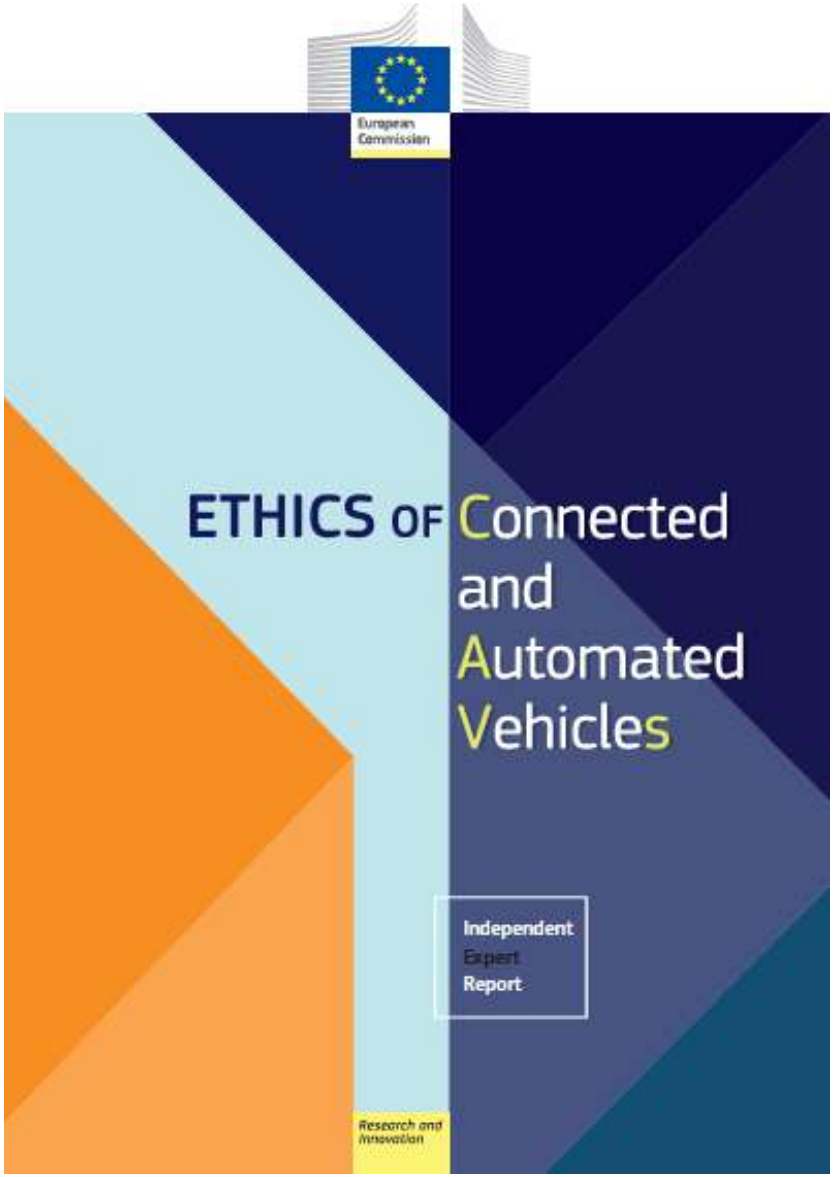

Ethics Commission: Automated and connected driving (Report by Federal Ministry of Transport and Digital Infrastructure of Germany [BMVI])

BMVI = Bundesministerium für Verkehr und digitale Infrastruktur




https://ethicsinaction.ieee.org/

# PART 7 FOR FURTHER READING




# Ethics of Connected and Automated Vehicles
## Recommendations on road safety, privacy, fairness, explainability and responsibility.
## European Commission

"New recommendations for a safe and ethical transition towards driverless mobility" (2020)

https://ec.europa.eu/info/news/new-recommendations-for-a-safe-and-ethical-transition-towards-driverless-mobility-2020-sep-18_en

Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659). Ethics of Connected and Automated Vehicles. 2020. Publication Office of the European Union, Luxembourg. European Commission. Directorate-General for Research and Innovation. https://europa.eu/!VV67my

**New technologies do not appear out of nowhere: they are imagined by people and built with purpose.**

EU values and principles need to be integrated at the core of these new technologies to ensure their ethical use and positive impact. Our ability to reach a just, sustainable and inclusive society depends on them.

# 20 RECOMMENDATIONS to support researchers, policymakers, manufacturers and deployers in the safe and responsible transition towards CAVs

1. Ensure that CAVs reduce physical harm to persons.
2. Prevent unsafe use by inherently safe design.
3. Define clear standards for responsible open road testing.
4. Consider revision of traffic rules to promote safety of CAVs and investigate exceptions to non-compliance with existing rules by CAVs.
5. Redress inequalities in vulnerability among road users.
6. Manage dilemmas by principles of risk distribution and shared ethical principles.
7. Safeguard informational privacy and informed consent.
8. Enable user choice, seek informed consent options and develop related best practice industry standards.
9. Develop measures to foster protection of individuals at group level.
10. Develop transparency strategies to inform users and pedestrians about data collection and associated rights.
11. Prevent discriminatory differential service provision.
12. Audit CAV algorithms.
13. Identify and protect CAV relevant high-value datasets as public and open infrastructural resources.
14. Reduce opacity in algorithmic decisions.
15. Promote data, algorithmic, AI literacy and public participation.
16. Identify the obligations of different agents involved in CAVs.
17. Promote a culture of responsibility with respect to the obligations associated with CAVs.
18. Ensure accountability for the behaviour of CAVs (duty to explain).
19. Promote a fair system for the attribution of moral and legal culpability for the behaviour of CAVs.
20. Create fair and effective mechanisms for granting compensation to victims of crashes or other accidents involving CAVs.

# ETHICS of Connected and Automated Vehicles

## ROAD SAFETY, RISK, DILEMMAS:

Safety benefits and improvements of CAVs should comply with basic ethical and legal principles: they should be publicly demonstrable, monitored and updated through solid and shared scientific research, and continuously adjusted to the needs of all road users.

## DATA AND ALGORITHM ETHICS: PRIVACY, FAIRNESS, EXPLAINABILITY:

Artificial Intelligence (AI) and automated systems used in CAVs should be explainable and transparent to empower users and to protect their data.

This should be reflected through rules and regulations that take into account the fast-changing nature of CAV technologies (especially AI and big data) and favour inclusive deliberation at all levels.

## RESPONSIBILITY:

Responsibilities should be clearly attributed and shared, going beyond blame and compensation in case of a collision. No single person or system can be held solely accountable.

From inception to use, best practices promoting ethical responsibility

# Ethics & Law

Ryan Jenkins (2016) Autonomous Vehicles Ethics & Law: Towards an Overlapping Consensus

https://www.academia.edu/29332066/Autonomous_Vehicles_Ethics_and_Law_Towards_an_Overlapping_Consensus

Patrick Lin (2015) Why Ethics Matters for Autonomous Cars.

In: Autonomes Fahren Technische, rechtlische und gesellschaftliche Aspekte

https://www.springerprofessional.de/en/why-ethics-matters-for-autonomous-cars/4397684

# Ethically Aligned Design

A Vision for Prioritizing Human Well-being With Autonomous and Intelligent Systems

https://ethicsinaction.ieee.org/

Embedding Values into Autonomous Intelligent Systems - The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

https://standards.ieee.org/develop/indconn/ec/ead_embedding_values.pdf

An example of ethical guidelines thinking one step further is described in the book:

Sarah Spiekermann. Ethical IT Innovation: A Value-Based System Design Approach. Taylor & Francis, 2015.

# Policy Concerning Automated Vehicles (US DOT )

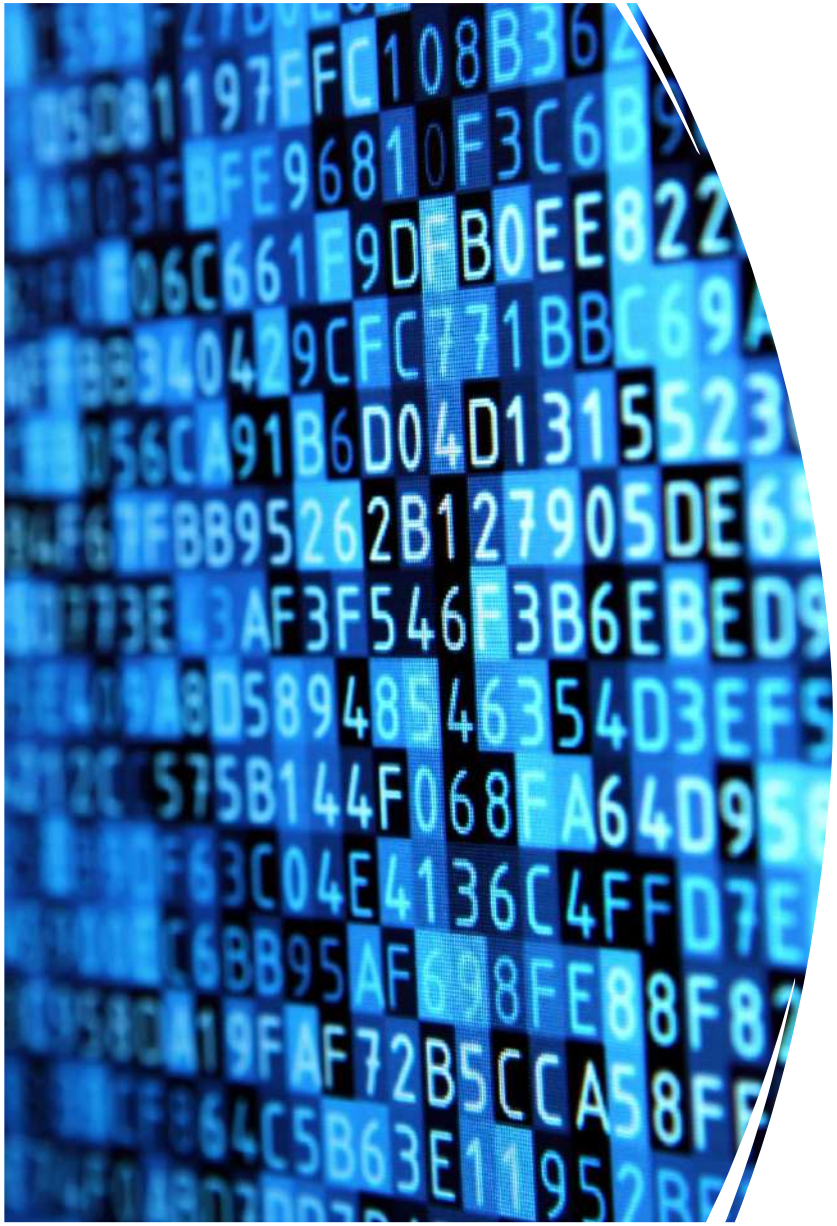
"DOT/NHTSA Policy statement concerning Automated Vehicles" 2016 update to "Preliminary statement of policy concerning automated vehicles".

Technical report, National Highway Traffic Safety Administration (NHTSA).

http://www.nhtsa. gov/staticfiles/rulemaking/pdf/Autonomous-Vehicles-Policy-Update-2016.pdf

# References


- Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global
- Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2020). Real-world Ethics for Self-Driving Cars. In Proceedings of the 42nd International Conference on Software Engineering (ICSE '20) Poster Track. https://ethics.se
- Holstein, T. Dodig-Crnkovic G. Avoiding the Intrinsic Unfairness of the Trolley Problem. Accepted for the Proceedings of FairWare workshop at ICSE2018, to be published by ACM.
- Holstein, T. Dodig-Crnkovic G. and Pelliccione P. Ethical and Social Aspects of Self-Driving Cars, http://arxiv.org/abs/1802.04103
- Dodig Crnkovic, G. and B. Çürüklü. Robots: ethical by design. Ethics and Information Technology, 14(1):61–71, Mar 2012.
- Dodig Crnkovic, G. and B. Çürüklü. Robots: ethical by design. Ethics and Information Technology, 14(1):61–71, Mar 2012.
- Dodig-Crnkovic, G. and D. Persson. Sharing moral responsibility with robots: A pragmatic approach. In Proceedings of the 2008 Conference on Tenth Scandinavian Conference on Artificial Intelligence: SCAI 2008, pages 165–168, Amsterdam, The Netherlands, IOS Press. 2008.
- Dodig-Crnkovic, G. and D. Persson. Sharing moral responsibility with robots: A pragmatic approach. In Proceedings of the 2008 Conference on Tenth Scandinavian Conference on Artificial Intelligence: SCAI 2008, pages 165–168, Amsterdam, The Netherlands, IOS Press. 2008.
- Johnsen A., G. Dodig- Crnkovic, K. Lundqvist, K. Hänninen, and P. Pettersson. Risk- based decision-making fallacies: Why present functional safety standards are not enough. In 2017 IEEE International Conference on Software Architecture Workshops (ICSAW), pages 153–160, April 2017.
- Sapienza, G., Dodig-Crnkovic, G. and I. Crnkovic. Inclusion of ethical aspects in multi-criteria decision analysis. In 2016 1st International Workshop on Decision Making in Software ARCHitecture (MARCH), pages 1–8, April 2016.
- Thekkilakattil A. and G. Dodig-Crnkovic. Ethics aspects of embedded and cyber-physical systems. In 2015 IEEE 39th Annual Computer Software and Applications Conference, volume 2, pages 39–44, July 2015.



# ADDITIONAL INFORMATION …

… that you can use in your group discussions

# Discussions in Breakout rooms

1. In what way does intelligence affect design of artifacts (such as cars)?
   Compare classical cars with intelligent cars. What are the new ethical questions that did not exist before intelligence?
2. High-risk AI systems according to the European Parliament are: Unmanned Aircraft, Autonomous traffic management systems, Autonomous robots and Autonomous Vehicles (level 4 and 5). Discuss and compare risks and benefits of intelligent technology. How can design contribute to the "AI for good"?
3. Real-world engineering ethics and its connection to your field of study. How do you see the role of ethics and its possible inclusion into your profession?
4. How do we meet Technical challenges to AV ethics (see p.67)
5. How do we meet Social challenges to AV ethics (see p.67)
6. Ethical guidelines, laws, guidelines, best practices
7. Sustainability aspects (environmental and social sustainability) of intelligent cars

| Requirements | Technical Challenges | Approaches |
|---|---|---|
| Safety | Hardware and software adequacy. Vulnerabilities of machine-learning algorithms. Trade-offs between safety and other factors (like economic). Possibility of intervention in self-driving cars (including for the Police). Systemic solutions to guarantee safety in organizations (regulations, authorities, safety culture). | Setting safety as the first priority. Learning from the history of automation. Learning from driving experience - perception and input interpretation processes. Specification of how a self-driving car will behave in cases when the car is not able to operate autonomously. Clarification of the role of the police. Regulations, guidelines, standards being developed as the technology develops. |
| Security | Minimal necessary security requirements for deployment of self-driving cars. Security in systems and connections. Deployment of software updates. Storing and using received and generated data in a secure way. | Technical solutions that will guarantee minimum security under all foreseeable circumstances. Anticipation and prevention of the worst-case scenarios regarding security breaches. Provide active security. Accessibility of all data, even in the case of accidents, has to be provided, so that it can be analysed to foster knowledge and to provide facts for next generation developments. |
| Privacy | Trade-offs between privacy and data collection/recording and storage/sharing. | Following/applying legal frameworks to protect personal data, such as GDPR. |
| Transparency | Information disclosure, what and to whom. Transparency of algorithmic decision making. Transparency in the techno-social ecosystem. | Assurance of transparency and insight into decision making. Active sharing of knowledge to ensure the interoperability of systems and services. |
| Algorithmic Fairness | Algorithmic decision making is required to be fair and not to discriminate on the grounds of race, gender, age, wealth, social status etc. | This requirement is related to transparency of decision making and expectation of explainability of the ground for decision making. |
| Reliability | Reliability of sensors and software and need for redundancy. Reliability of required networks and solution for the case when the network is unavailable. | Definition of different levels for reliability, such as diagnostics, vehicle input sensors, software, and external services, set the ground for reliability measures of the car as a system and its components. Standardized process required to shift from fail-safe to fail-operational architecture. |
| Environmental Sustainability | Environmental sustainability ethics refers to new ways of production, use, and recycling for autonomous vehicles. | Production, use, and disposal/recycling of technology rises sustainability issues (batteries, car sharing) that must be addressed. |
| Intelligent behavior control | Intelligent behaviour may lead to unpredictable situations resulting from learning and autonomous decision making. | Development of self-explaining capability and other features ensuring desired behavior in intelligent software. |
| Transdisciplinarity -Systemic approach | Ethics in design, requirements engineering, software-hardware development, learning, legal and social aspects, software-hardware interplay. | Adoption of transdisciplinarity and system approaches is increasing and should be given even more prominent role. |
| Quality | Quality of components. Quality of decision making. Lifetime and maintenance. QA process. Adherence to ethical principles/guidelines | Ethical deliberations included in the whole process starting with design and development. Ethics-aware decision making to ensure ethically justified decisions. |

*Table 1 Summary of the technical challenges and approaches, grouped by requirement*

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In S. J. Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

| Requirements | Social Challenges | Approaches |
|---|---|---|
| Non-maleficence | Technology not causing harm.<br>Disruptive changes on the labor market.<br>Change of related markets and business models (e.g., insurances, manufacturers). | Partly covered by technical solutions.<br>Preparation of strategic solutions for people losing jobs.<br>Learning from historic parallels to industrialization and automatization. |
| Stakeholders involvement | In this field different stakeholders are involved – from professionals designing, developing, maintaining cars, to their users, and general public. | Active involvement of stakeholders in the process of design and requirements specification as well as decisions of their use. |
| Beneficence | Values and priorities: Ensure that general public values will be embodied in the technology, with interests of minorities taken into account. | Initiatives as "AI for good" exemplify this expectation that new technology not only do not cause harm, but actively do good for its stakeholders. |
| Responsibility and Accountability | Assignment and distribution of responsibility and accountability are among central regulative mechanisms for the development of new technology. They should follow ethical principles. | The Accountability, Responsibility and Transparency (ART) principle based on a Design for Values approach includes human values and ethical principles in the design processes (Dignum, 2019). |
| Freedom and Autonomy | Freedom of choice hindered by the system (e.g. it may not allow to drive into a certain area) | The freedom of choice determined by regulations.<br>Determination and communication of the amount of control a human has in context of the self-driving car |
| Social Sustainability | In the domain of business, social sustainability is about identifying and managing business impacts on people | Pursuing social equity, community development, social support, human rights, labour rights, social responsibility, social justice, etc. |
| Social Fairness | Ascertaining fairness of the socio-technological system. | Fairness of the decision-making. Related to transparency and explainability. |
| Dignity and Solidarity | This requirement refers to the entire socio-technological system. | Challenges come from the lack of common wholistic view. |
| Social Trust | Establishing trust between humans and highly automated vehicles as well as within the social system. | Further research on how to implement trust across multiple systems.<br>Provision of trusted connections between components as well as external services |
| Justice: legislation, standards, norms, policies and guidelines | Keeping legislation up to date with current level of automated driving, and emergence of self-driving cars.<br>Creating and defining global legislation frameworks. Including ethical guidelines in design and development processes | Legislative support and contribution to global frameworks. Ethics training for involved engineers. Establishment and maintenance of a functioning socio-technological system in addition to functional safety standards |

*Table 2 Summary of the social challenges and approaches, grouped by requirement*

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In S. J. Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

# Questionnaire

10 minutes 1-page questionnaire

The aim: to help us develop the course as an element in the program of ethics education for students at technical universities in Europe.

https://forms.gle/cAT7jopbK232iuzP7