



# Ethics of Robotics and AI

*Moral Responsibility and Societal Challenges*

**Mark Coeckelbergh**

Professor of Philosophy of Media and Technology  
University of Vienna

[mark.coeckelbergh@univie.ac.at](mailto:mark.coeckelbergh@univie.ac.at) || [coeckelbergh.wordpress.com](http://coeckelbergh.wordpress.com)

# PHILOSOPHY OF TECHNOLOGY



**Philosophy**

**Interdisciplinarity**

**Policy**

# Robophilosophy 2018





[European Commission](#) > [Strategy](#) > [Digital Single Market](#) > [Policies](#) >

Digital Single Market

POLICY

# High-Level Expert Group on Artificial Intelligence

Following an open selection process, the Commission has appointed 52 experts to a new High-Level Expert Group on Artificial Intelligence, comprising representatives from academia, civil society, as well as industry.

# PHILOSOPHY OF TECHNOLOGY



**Thinking about and for technology, but also using technology to think about philosophical issues**

# PHILOSOPHY OF TECHNOLOGY



**Focus: robots and AI**

# PHILOSOPHY OF TECHNOLOGY

A futuristic humanoid robot with a white and red color scheme is shown in a thinking pose, with its right hand resting on its chin. The robot has a sleek, modern design with visible joints and a helmet-like head. The background is a soft, out-of-focus blue and white, suggesting a high-tech or futuristic environment.

**Ethics**

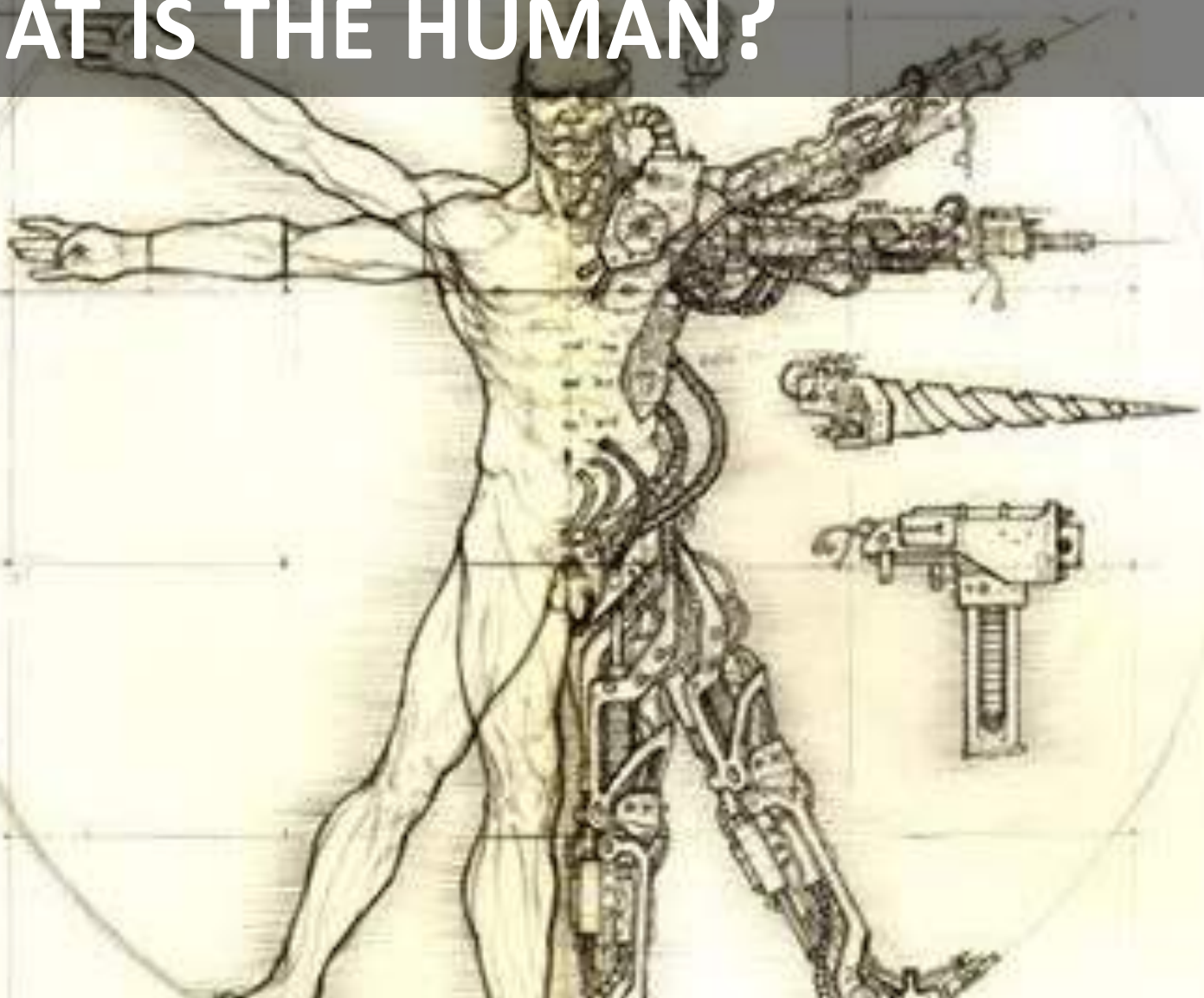
**Philosophical anthropology**

**Epistemology**

**Aesthetics**

...

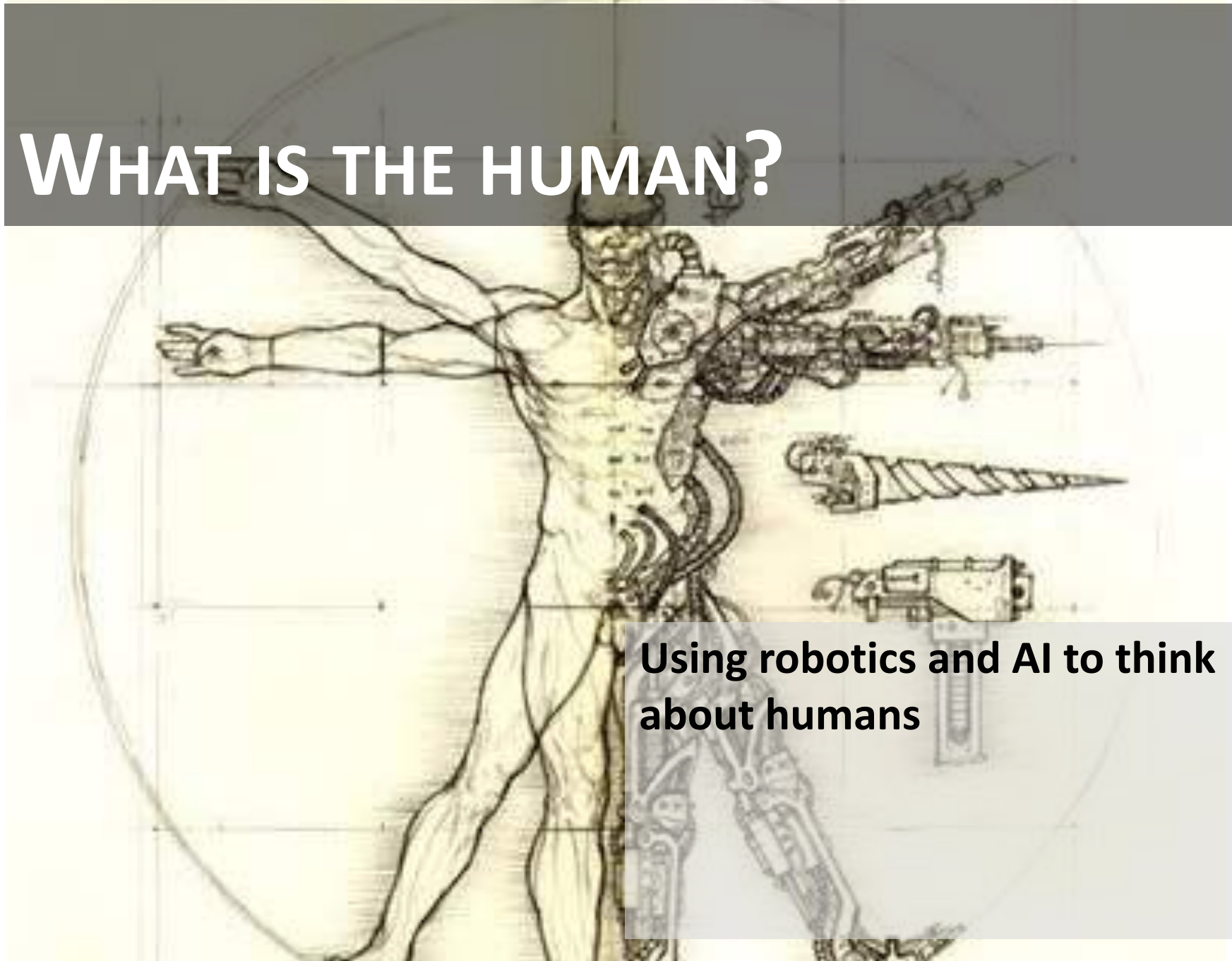
# WHAT IS THE HUMAN?





# WHAT IS THE HUMAN?

Using robotics and AI to think about humans



# WHAT IS THE HUMAN?



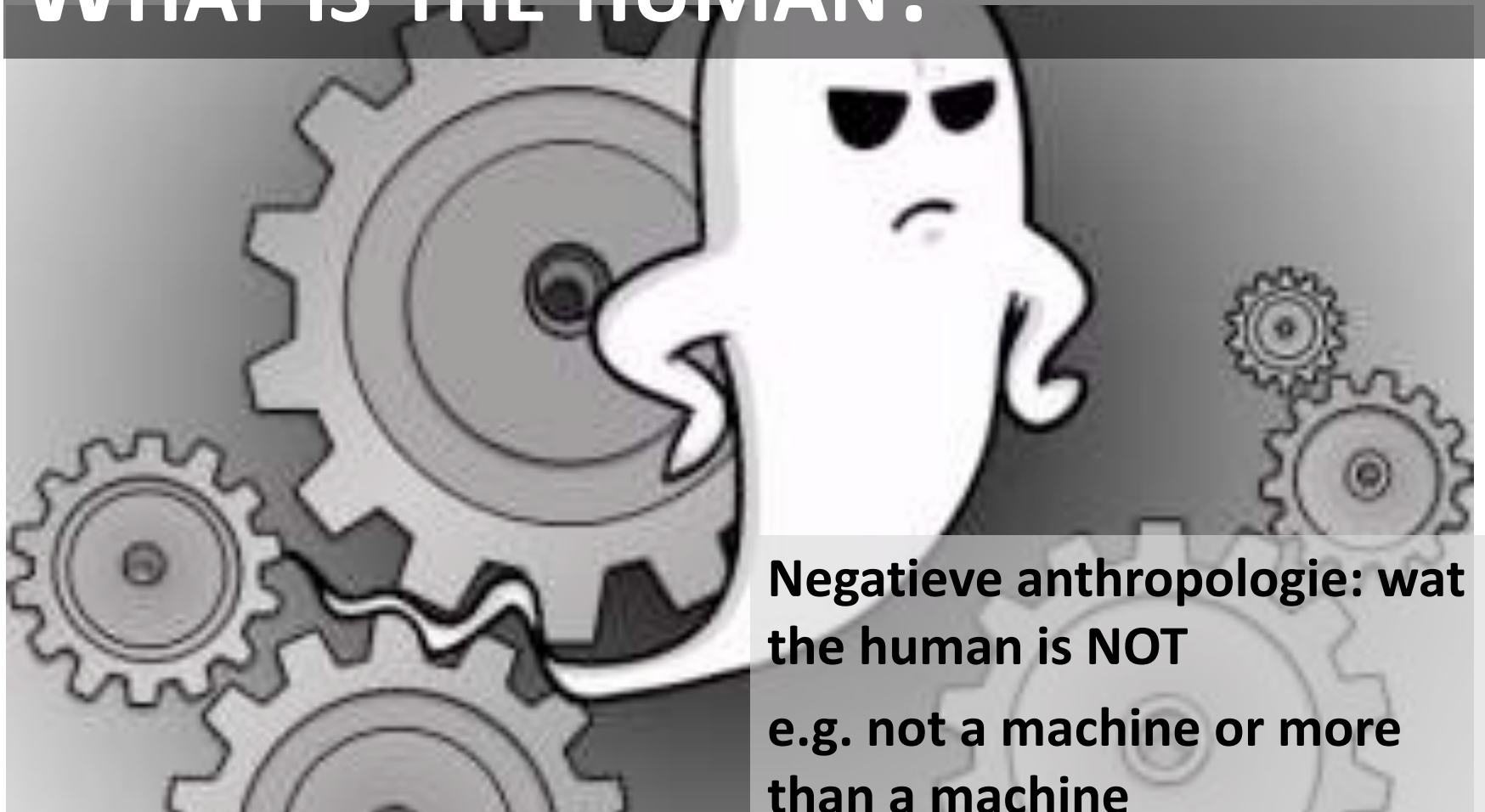
**Negative anthropology: what the human is NOT**

# WHAT IS THE HUMAN?



**Negative anthropologie: wat  
the human is NOT  
e.g. not a chimp**

# WHAT IS THE HUMAN?



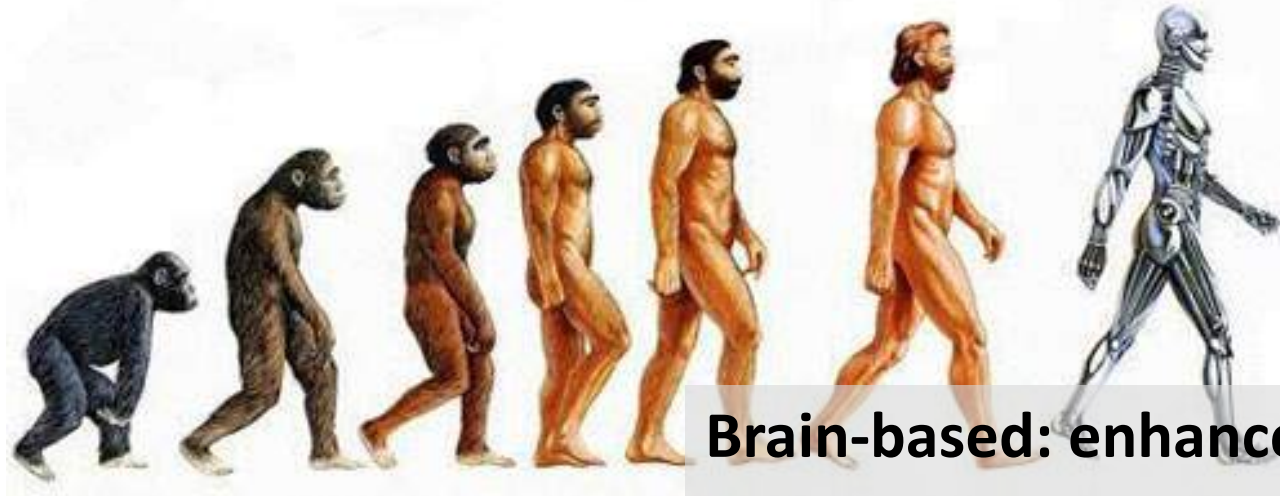
**Negative anthropologie: wat  
the human is NOT  
e.g. not a machine or more  
than a machine**

# WHAT IS THE HUMAN?



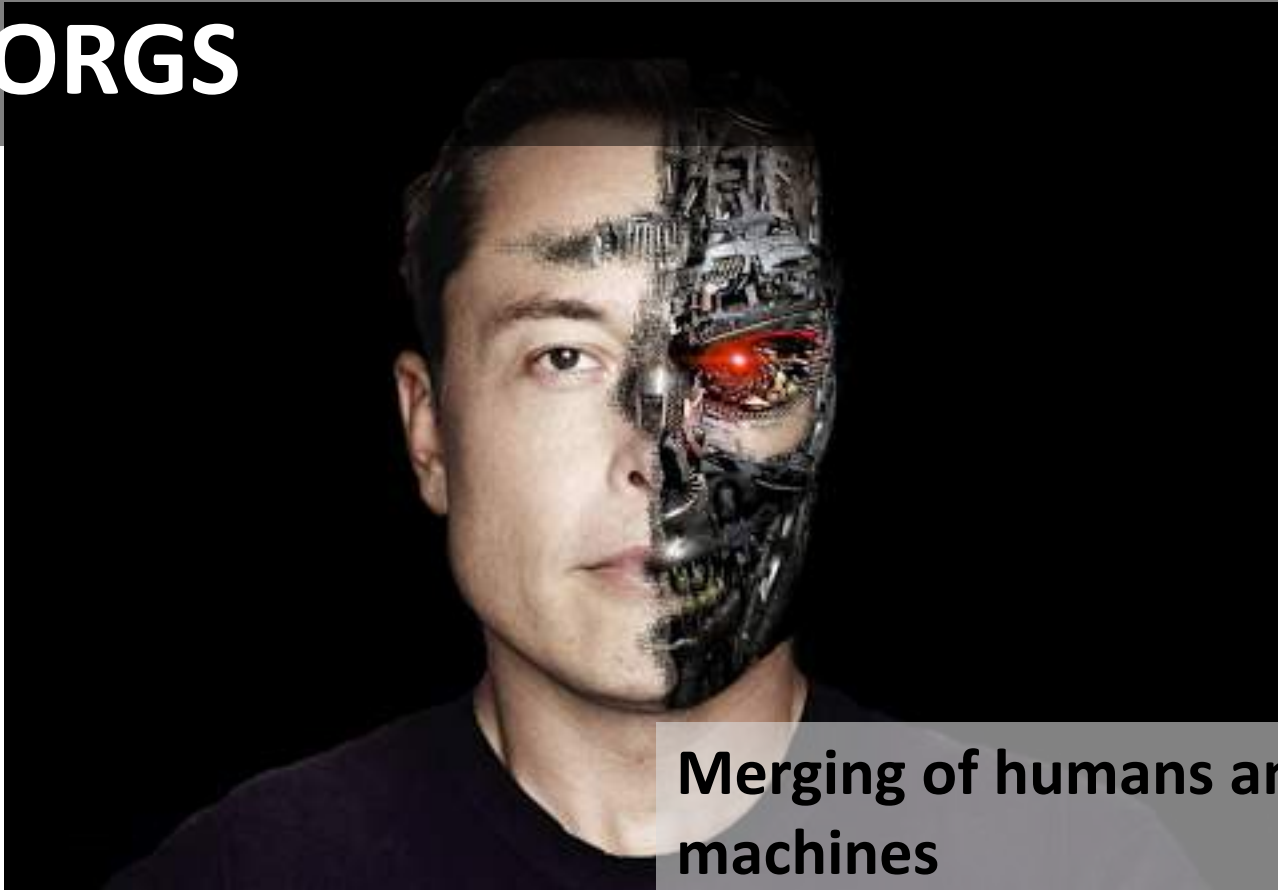
**Positive anthropology: what the human is**  
**e.g. a computational or information being**

# TOWARDS AN ARTIFICIAL HUMAN?



**Brain-based: enhancement  
and/or  
Robots and AI**

# CYBORGS



**Merging of humans and machines**

# ROBOTS: HUMAN-LIKE

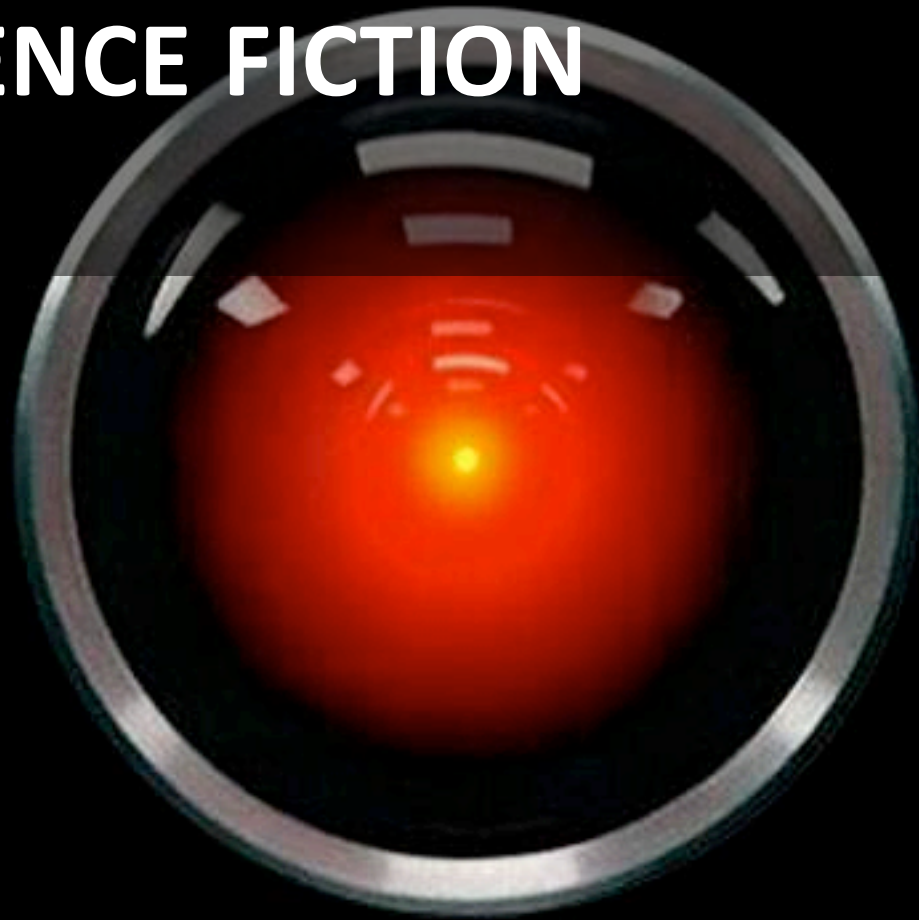




# ROBOTS: NOT NECESSARILY HUMAN- LIKE



# AI: SCIENCE FICTION



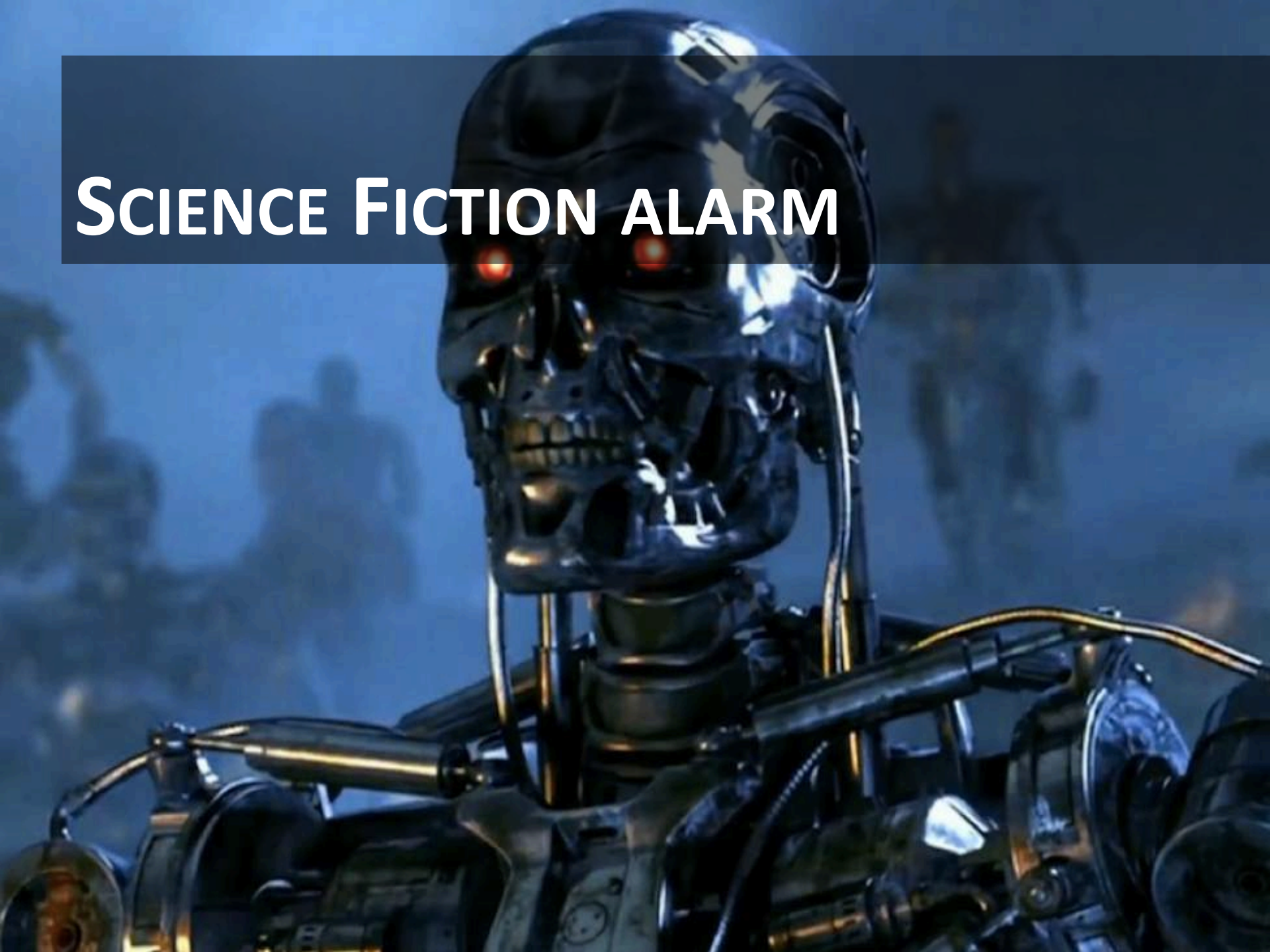
# AI: IN YOUR POCKET



# ETHICS!



# SCIENCE FICTION ALARM



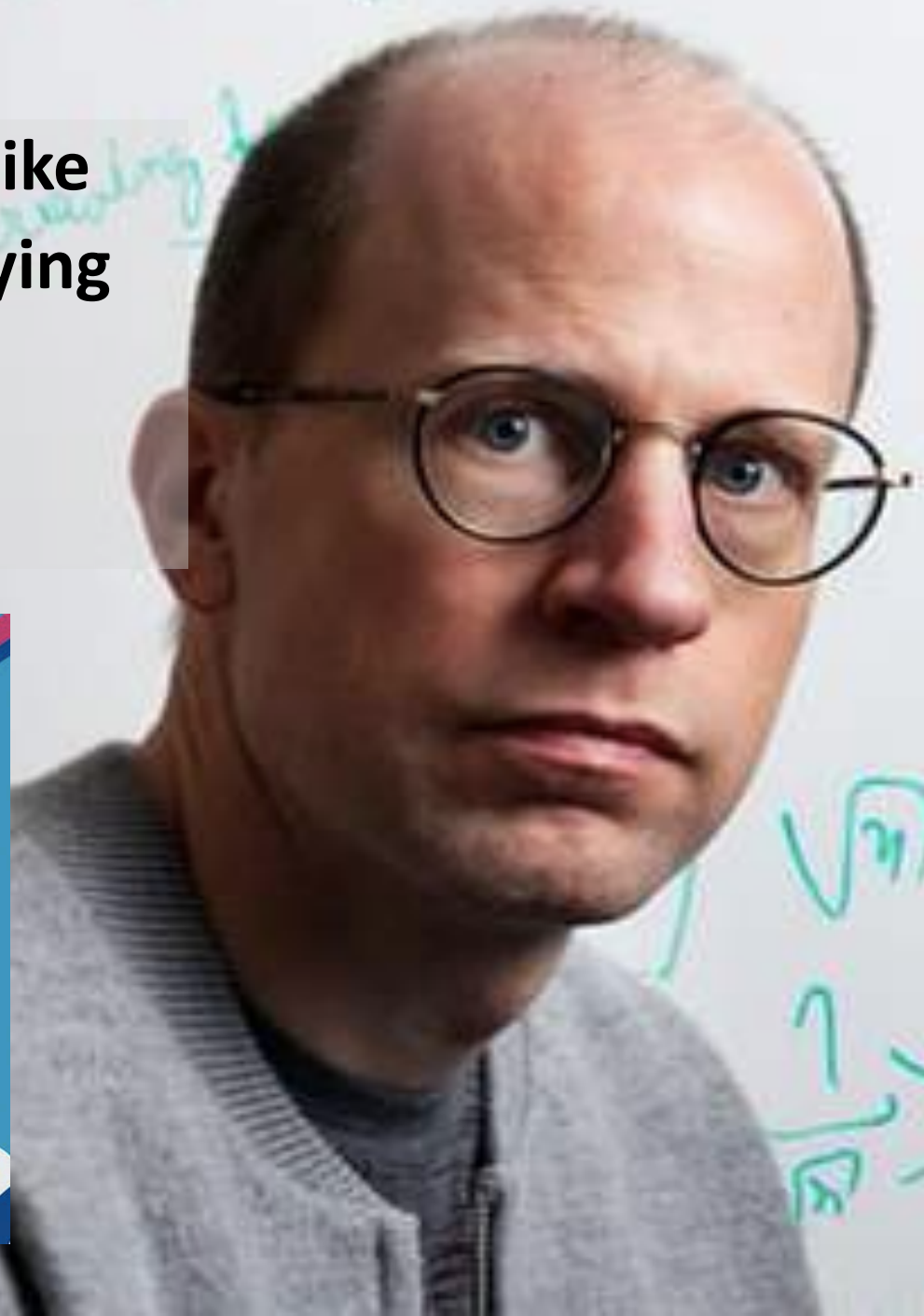


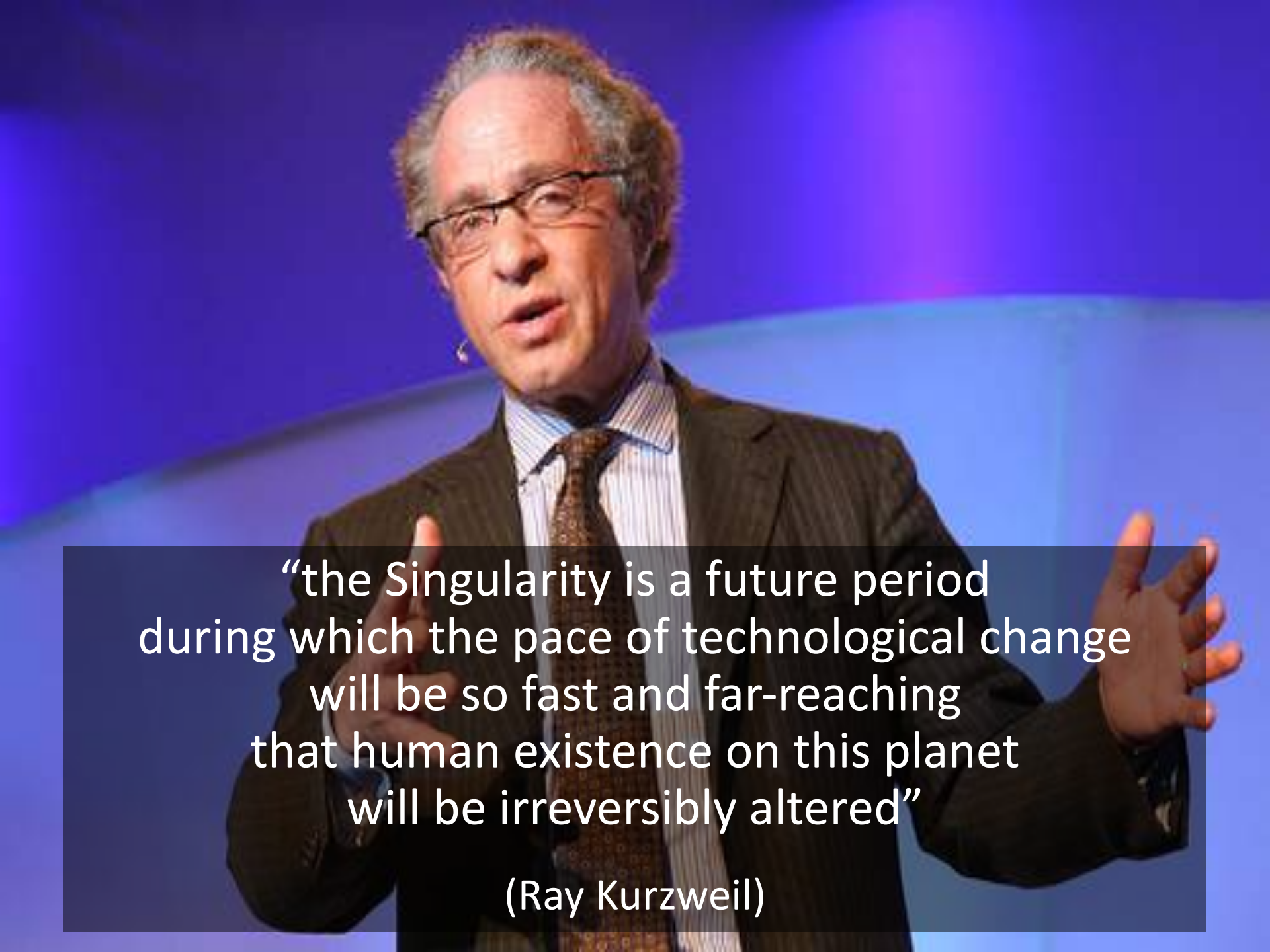
**“AI is a fundamental  
existential risk for  
human civilization”**

**(Elon Musk)**

**“we humans are like  
small children playing  
with a bomb”**

**(Nick Bostrom)**



A photograph of Ray Kurzweil, an older man with glasses, wearing a dark suit, a light blue striped shirt, and a brown patterned tie. He is gesturing with both hands while speaking. The background is a blue gradient with a faint globe-like shape.

“the Singularity is a future period during which the pace of technological change will be so fast and far-reaching that human existence on this planet will be irreversibly altered”

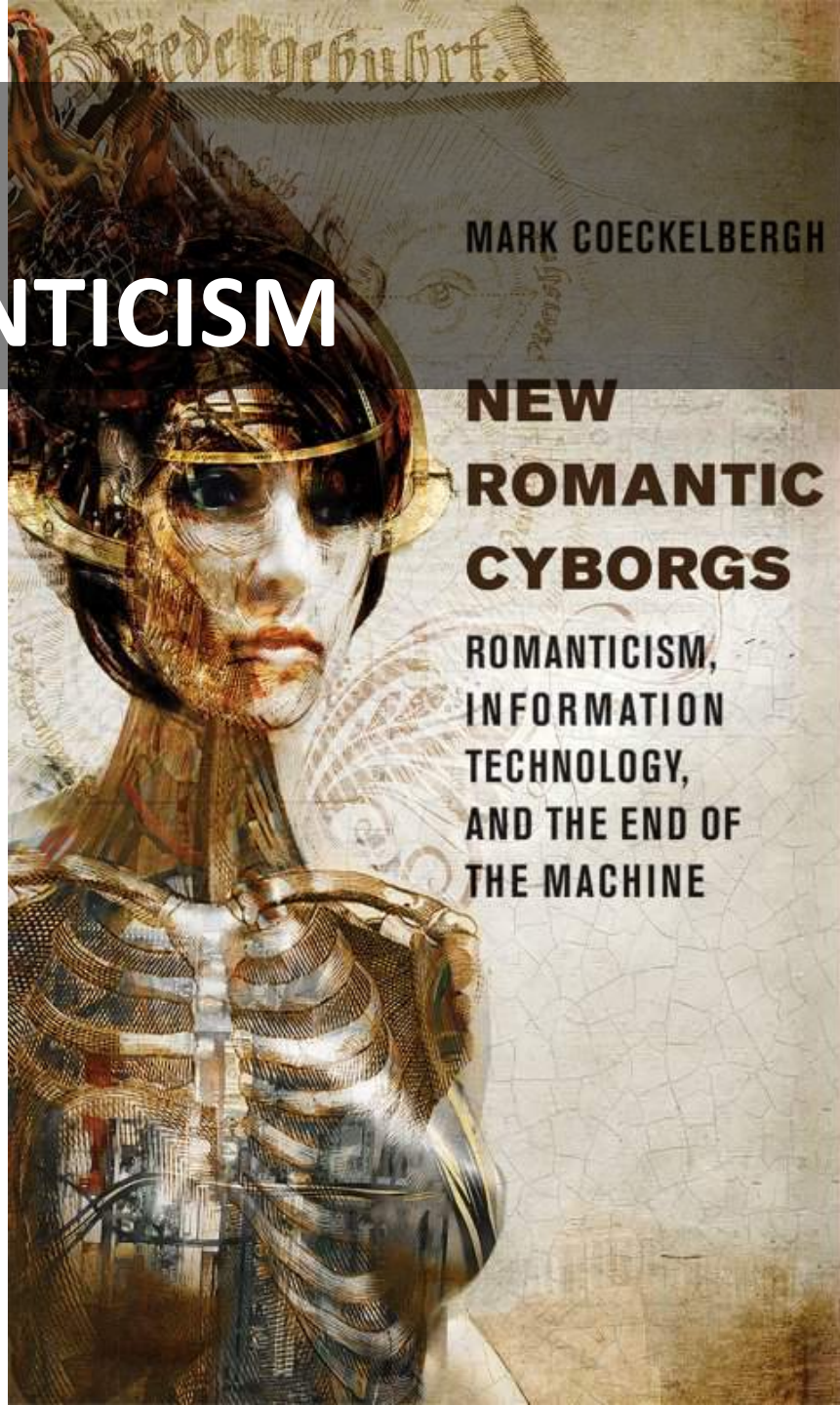
(Ray Kurzweil)



# FRANKENSTEIN



# ROMANTICISM



MARK COECKELBERGH

## **NEW ROMANTIC CYBORGS**

**ROMANTICISM,  
INFORMATION  
TECHNOLOGY,  
AND THE END OF  
THE MACHINE**



**AGAINST ALARMISM**

**KEEP**

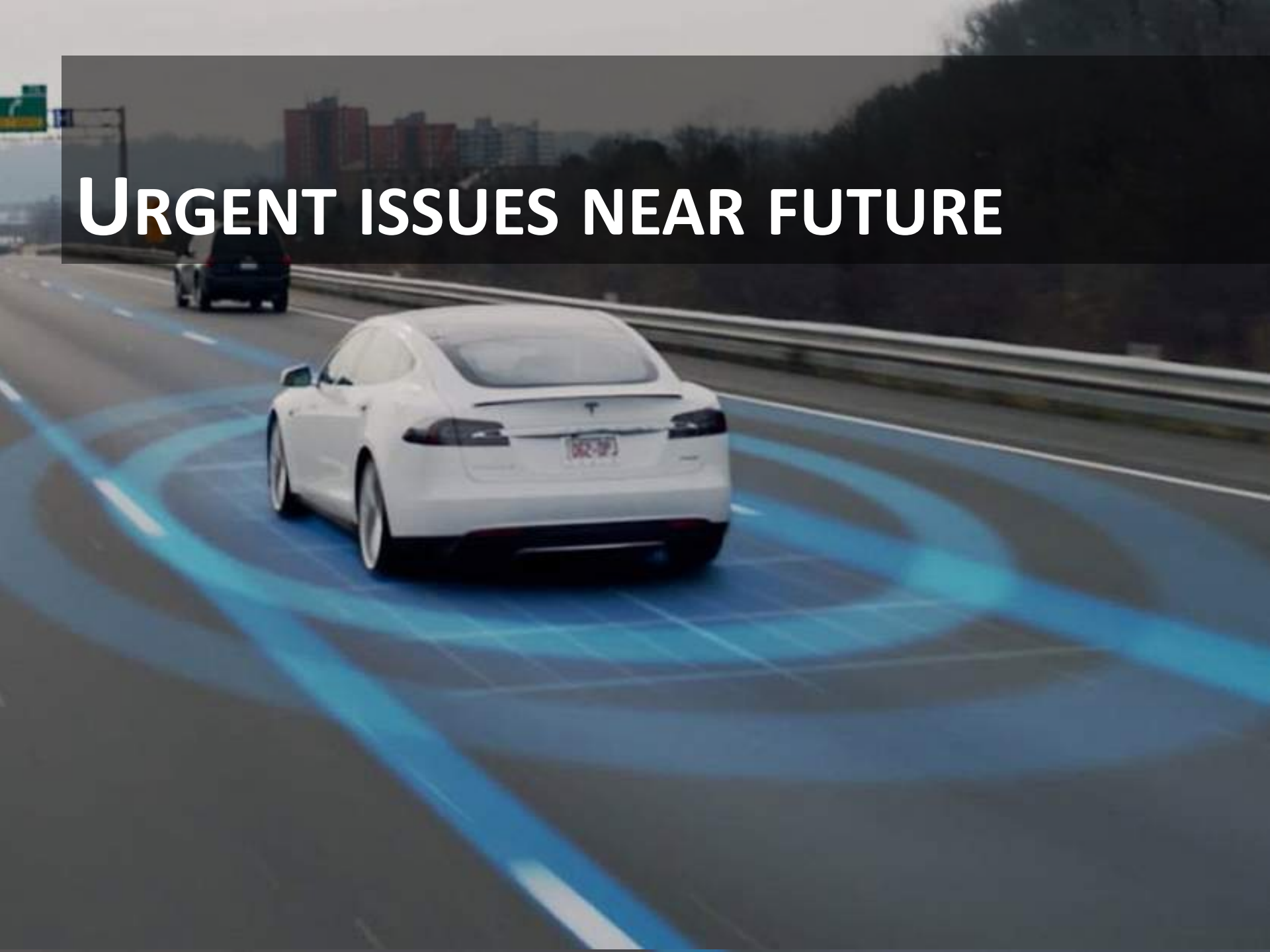
**CALM**

**AND**

**DO YOUR**

**HOMework**

# URGENT ISSUES NEAR FUTURE



INDUSTRY



# DAILY LIFE



# IN THE OFFICE

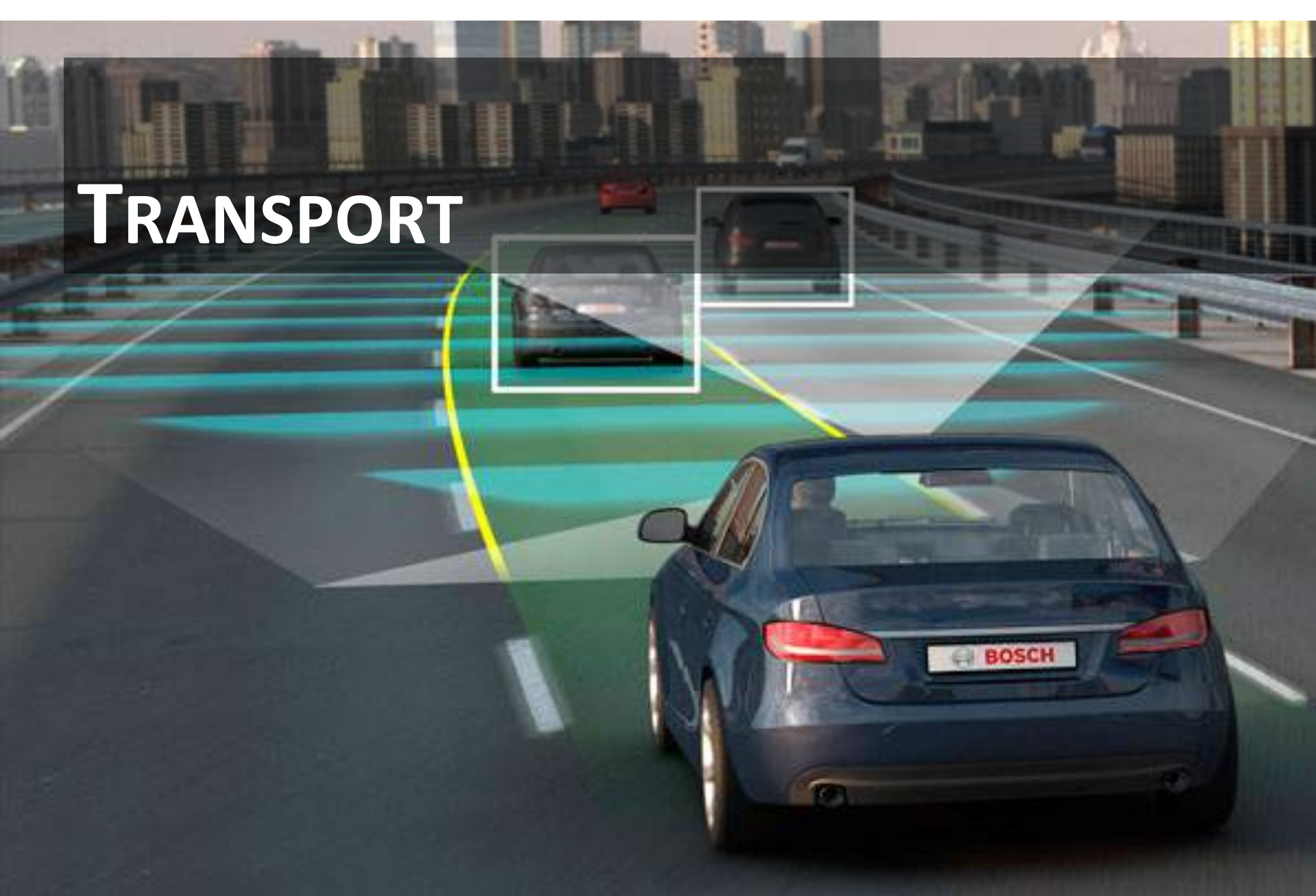


# FINANCE





# TRANSPORT



# HEALTH CARE



# MILITARY APPLICATIONS



# DATA





# CHANGES TO OUR DAILY LIVES

HOW WAS YOUR DAY?

GOOD, YOUR'S



# ETHICAL AND LEGAL PROBLEMS



# DEFINITION PROBLEMS



## Problem for regulation:

- Due to nature of new technologies: robots, AI, algorithms, code, smart tech, internet of things, 'cyber-physical systems' ... ?
- How autonomous, intelligent, etc.?



# PRIVACY, SECURITY, SURVEILLANCE

- The AI records what you do and transfers data... to whom? Company? Third Party?
- What if your robot gets hacked?

# HEALTH



**ADDICTION**



# REPLACEMENT, AUTONOMY, LOSS OF AGENCY?

A woman with short brown hair, wearing a light blue hospital gown, is sitting up in a hospital bed. She is looking towards a white humanoid robot that is leaning over her. The robot has a bear-like head with black eyes and a small black nose. It is wearing a white long-sleeved shirt with blue accents and white pants with a black belt. The robot's right arm is extended towards the woman. In the background, another person in a white lab coat is partially visible, and the setting appears to be a modern hospital room with large windows and a staircase.

- Robot/AI - human teams
- Degrees of autonomy
- Distributed agency

# MORAL AND LEGAL RESPONSIBILITY

The image shows a top-down view of a winding road. A silver Volvo car is in the center, moving away from the viewer. It is surrounded by green concentric sensor waves. To the left, a motorcycle is moving in the same direction, surrounded by red concentric sensor waves. In the distance, another car is visible on the road, enclosed in a green bounding box. A speed limit sign with the number 90 is also visible on the right side of the road. The background consists of a hilly, arid landscape.

- Who?
- AI/robot as moral agent?
- Legal questions

# MORAL AND LEGAL RESPONSIBILITY

## Examples

- AI causes crash on financial markets
- Machines harms worker in factory
- Autonomous car drives into group of children
- Care robot gives the wrong medication
- Killer robot kills civilian
- Child gets too attached to educational robot



# MORAL AND LEGAL RESPONSIBILITY

## Some problems

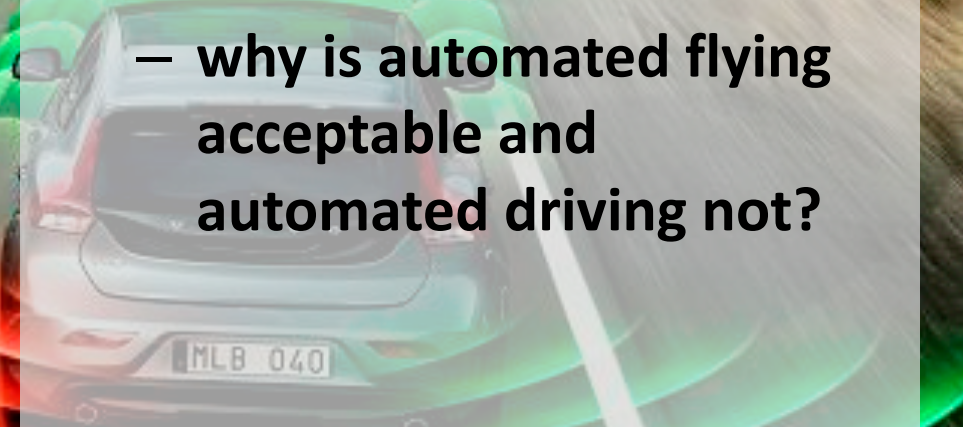
- what about distributed responsibility?
- how to make sure responsibility traces back to humans? human in control?
- insurance?
- regulating or ban?
- new legal instruments or not? (e.g. debate in European context about legal personhood robots versus using existing liability law)



# MORAL AND LEGAL RESPONSIBILITY

## Some problems

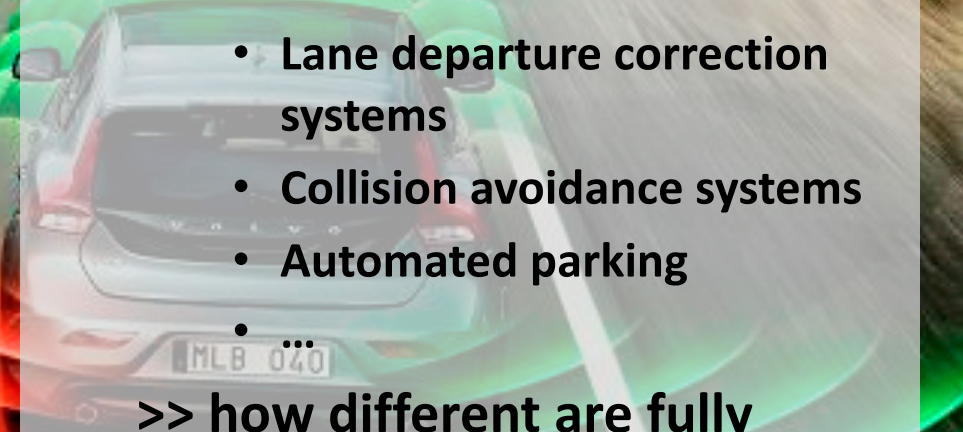
- acceptance:
  - accident and death more acceptable if human agent, e.g. human driver
  - why is automated flying acceptable and automated driving not?





# MORAL AND LEGAL RESPONSIBILITY

- gradations of automation
  - E.g. gradations of autonomous driving; there is already automation in existing cars:
    - Cruise control
    - Lane departure correction systems
    - Collision avoidance systems
    - Automated parking
    - ...
  - >> how different are fully autonomous technologies, e.g. autonomous cars?
  - >> new legal framework needed?



# MORAL AND LEGAL RESPONSIBILITY

Example: Classification  
Society of Automotive  
Engineers (SAE)

5 levels of self-driving:

- Level 0: monitoring, warnings
- Level 1: adaptive cruise control, automated parking
- Level 2: automated driving, but driver must be alert and be able to take over any time
- ...
- Level 5: no human intervention needed



# MORAL AND LEGAL RESPONSIBILITY

**Information and knowledge**

**- Do users and operators understand the system and its limitations?**

**- (Mis)information by manufacturers?**

**Important for discussions about liability and negligence**

**Difference with aviation, which is highly regulated and relatively safe**



# Self-driving Uber kills first fatal crash invol

Tempe police said car was in autonomous mode at the time of the crash and that the vehicle hit a woman who

## Case: Fatal accident

- Uber self-driving car in autonomous mode causes accident in Arizona: pedestrian dies (March 2018)
- See also 2016 Tesla accident



# Self-driving Uber kills first fatal crash invol

Tempe police said car was in autonomous mode at the time of the crash and that the vehicle hit a woman wh



## Case: Fatal accident

- **Who is responsible? Volvo? Uber? Vehicle operator/driver? Pedestrian? State of Arizona? Problem of “many hands”**
  - Draw on tort law: Uber/driver failed to exercise reasonable care
  - Draw on product liability law: Volvo and Uber
  - Conduct pedestrian: accident avoidable?
  - State of Arizona: sufficient regulation? E.g. one could require someone to be in driver seat – but enough?

# Self-driving Uber kills first fatal crash involving

Tempe police said car was in autonomous mode at the time of the crash and that the vehicle hit a woman who



**Case: Fatal accident**

- **Civil proceedings versus criminal law (but robots/AI cannot be charged with a crime)**
- **Need for better technology and more regulation (or ban? Or self-regulation by private companies (laissez-faire)? Too early or too late?**



# MORAL STATUS OF AIs/ ROBOTS

## Moral agents?

- What capacities needed for moral judgment? Also emotions?
- Rules enough?
- Too anthropocentric?



# MORAL STATUS OF AIs/ ROBOTS

**Moral patients?**

- **Thing or more than that?**
- **Machine as (quasi)other?**
- **Vulnerability humans versus ma**



palgrave  
macmillan

# Growing Moral Relations

Critique of Moral Status Ascription

Mark Coeckelbergh



# THE MACHINE QUESTION

CRITICAL PERSPECTIVES ON AI,  
ROBOTS, AND ETHICS

DAVID J. GUNKEL



# MORAL STATUS OF AIs/ ROBOTS

A close-up photograph of a female humanoid robot's face. The robot has a realistic human-like appearance with brown eyes, a slightly open mouth showing teeth, and a surprised or concerned expression. The background is a plain, light blue color.

Philosophically interesting,  
but also practical issue?

## ROBOT CITIZEN

The background of the slide features the Facebook logo in a light blue color against a darker blue background. In the foreground, there are black silhouettes of two people sitting at a desk, each using a laptop. The overall theme is technology and its impact on society.

# TECHNOLOGY CHANGES MORALITY

- **Privacy today**
- **How will AI and robotics change our values?**

# VULNERABLE USERS, ATTACHMENT AND DECEPTION



**SAFETY**

Artificial Intelligence and Robotics

+ Add to myFT

# Worker at Volkswagen plant killed in robot accident

Fatality touches on concerns about spread of automation

# HUMAN DIGNITY AND AUTONOMY



# ADAPTING TOO MUCH?



**Do we want to adapt to robots or should robots adapt to us?**

# MORAL DISTANCE



Money Machines

Coeckelbergh

## Money Machines

Electronic Financial Technologies,  
Distancing, and Responsibility  
in Global Finance



Mark Coeckelbergh





# MORAL DISTANCE

Ethics Inf Technol (2013) 15:87–98  
DOI 10.1007/s10676-013-9313-6

ORIGINAL PAPER

## Drones, information technology, and distance: mapping the moral epistemology of remote fighting

Mark Coeckelbergh

Published online: 8 March 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** Ethical reflection on drone fighting suggests that this practice does not only create physical distance, but also moral distance: far removed from one's opponent, it becomes easier to kill. This paper discusses this thesis, frames it as a moral-epistemological problem, and explores the role of information technology in bridging and creating distance. Inspired by a broad range of robotics, psychology, phenomenology, and media reports, it is first argued that drone fighting, like other long-range fighting, creates epistemic and moral distance in so far as 'screenfighting' implies the disappearance of the vulnerable face and body of the opponent and thus removes moral-psychological barriers to killing. However, the paper also shows that this influence is at least weakened by current surveillance technologies, which make possible an opponent on the ground is re-humanized, re-faced, and re-embodied. This 'mutation' or unintended 'hacking' of the practice is a problem for drone pilots and for those who order them to kill, but revealing its moral-epistemic possibilities opens up new avenues for imagining morally better ways of technology-mediated fighting.

**Keywords** Military robotics · Drones · Ethics · Distance · Information technology · Phenomenology

### Introduction

When on August 6, 1945 at 8:15 AM B-29 bomber Enola Gay dropped an atomic bomb on the city of Hiroshima, the crew soon witnessed blinding light and a mushroom-shaped cloud, covering the entire city in smoke and fire. They also felt the shockwave of the explosion of "Little Boy". What they didn't see or feel, however, was that and how this explosion killed approximately 70,000 people (and more in the following years). They didn't see how the skin of their victims was bleeding and burning. They didn't see people that looked "like walking ghosts", as a survivor described them. They didn't see the suffering and death of described them. They didn't see the effects of the explosion, for sure, but their victims were men, women, and children. They must have been fascinated by the looks of their victims was most likely experience of its effects on that of Captain William S. Parsons not very different from that of Captain William S. Parsons before the event, when he was arming the bomb: "I knew the Japs were in for it, but I felt no particular emotion about it." (Parsons quoted in Takaki 1995, p. 43).

In the beginning of the twenty-first century, a different bombing practice is becoming increasingly common: one that involves vehicles without a crew on board. Unmanned aerial vehicles (UAVs), also known as 'drones', are aircrafts that are controlled by computers or by pilots on the ground. Here I will restrict my discussion to aircrafts that are controlled by humans at all times, and use the terms "UAV" and "drone" interchangeably<sup>1</sup> to refer to such human-controlled aircraft.

The military can and do use UAVs for surveillance, but also for bombing targets on the ground. Today many

# MORAL DISTANCE

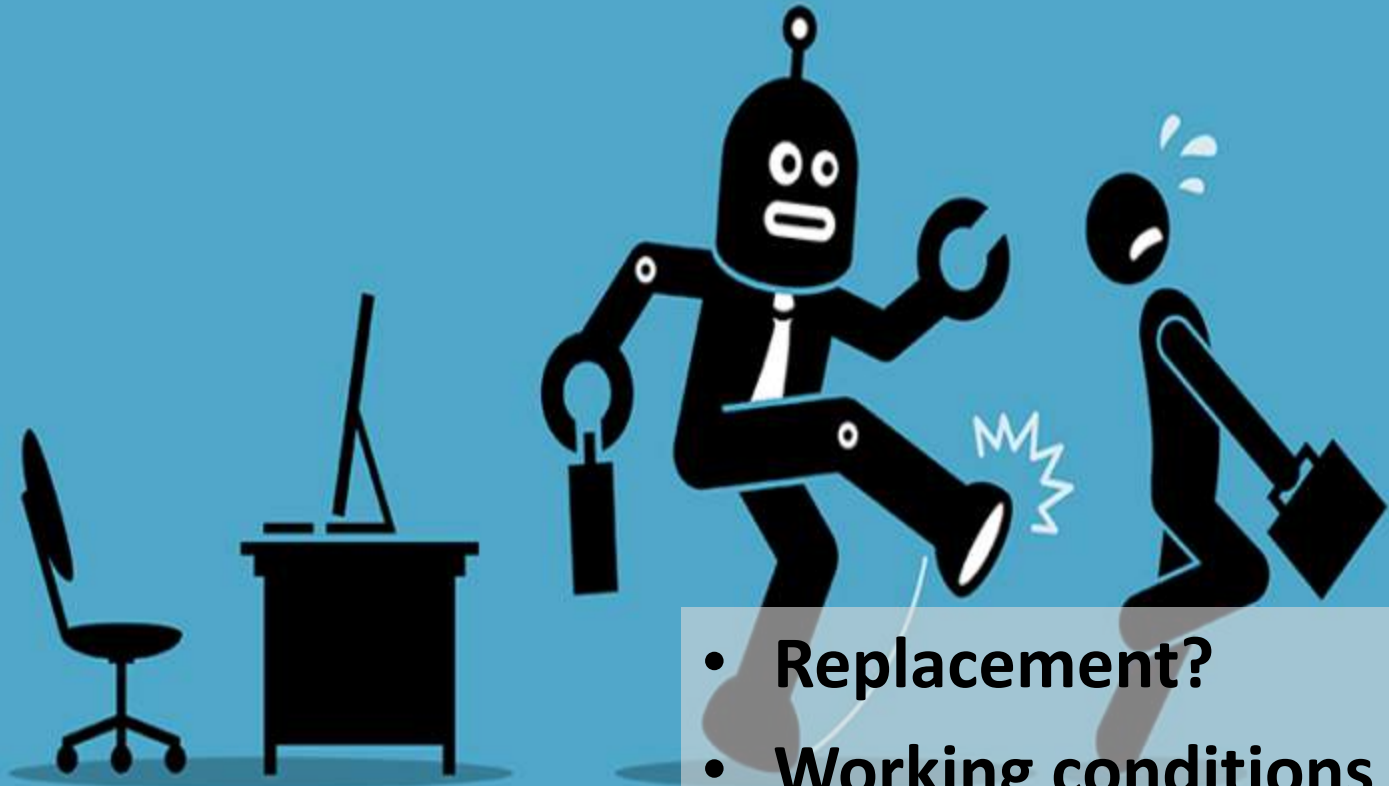


# SOCIETAL IMPLICATIONS



- Justice, fairness, power
- Inclusive society?
- Biased and non-transparent algorithms  
>>
- Social relations, e.g. intimate relations
- Sustainable economy?
- Future of work >>

# THE FUTURE OF WORK



- **Replacement?**
- **Working conditions and experience of work?**
- **Delegation and**



# The future of employment: How susceptible are jobs to computerisation?<sup>☆</sup>

Carl Benedikt Frey<sup>a,\*</sup>, Michael A. Osborne<sup>b</sup>

<sup>a</sup>Oxford Martin School, University of Oxford, Oxford OX1 1PF, United Kingdom

<sup>b</sup>Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, United Kingdom



## ARTICLE INFO

Article history:  
Received 24 September 2015  
Accepted 19 August 2016  
Available online 29 September 2016

JEL classification:  
E24  
J24  
J31  
J52  
O33

Keywords:  
Occupational choice  
Technological change  
Wage inequality  
Employment  
Skill demand

## ABSTRACT

We examine how susceptible jobs are to computerisation. To assess this, we begin by implementing a novel methodology to estimate the probability of computerisation for 702 detailed occupations, using a Gaussian process classifier. Based on these estimates, we examine expected impacts of future computerisation on US labour market outcomes, with the primary objective of analysing the number of jobs at risk and the relationship between an occupations probability of computerisation, wages and educational attainment.  
© 2016 Published by Elsevier Inc.



European Economic and Social Committee

INT/806  
Artificial intelligence

## OPINION

Section for the Single Market, Production and Consumption

Artificial intelligence – The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society  
(own-initiative opinion)

Rapporteur: Catelijne MULLER



# BIASED ALGORITHMS

- Problem in machine learning: AI trains on dataset that may contain a bias (e.g. favors young white men)
- Problem of algorithm or society, or both? How to deal with this?
- Right non-discrimination

# BIASED ALGORITHMS

- **Is bias avoidable? No, but we can explicitly discuss, analyze, and intervene (kind of bias, degree of bias)**
- **Algorithms teach us something about our societies (see also digital humanities: use AI!)**

# NON-TRANSPARENT ALGORITHMS

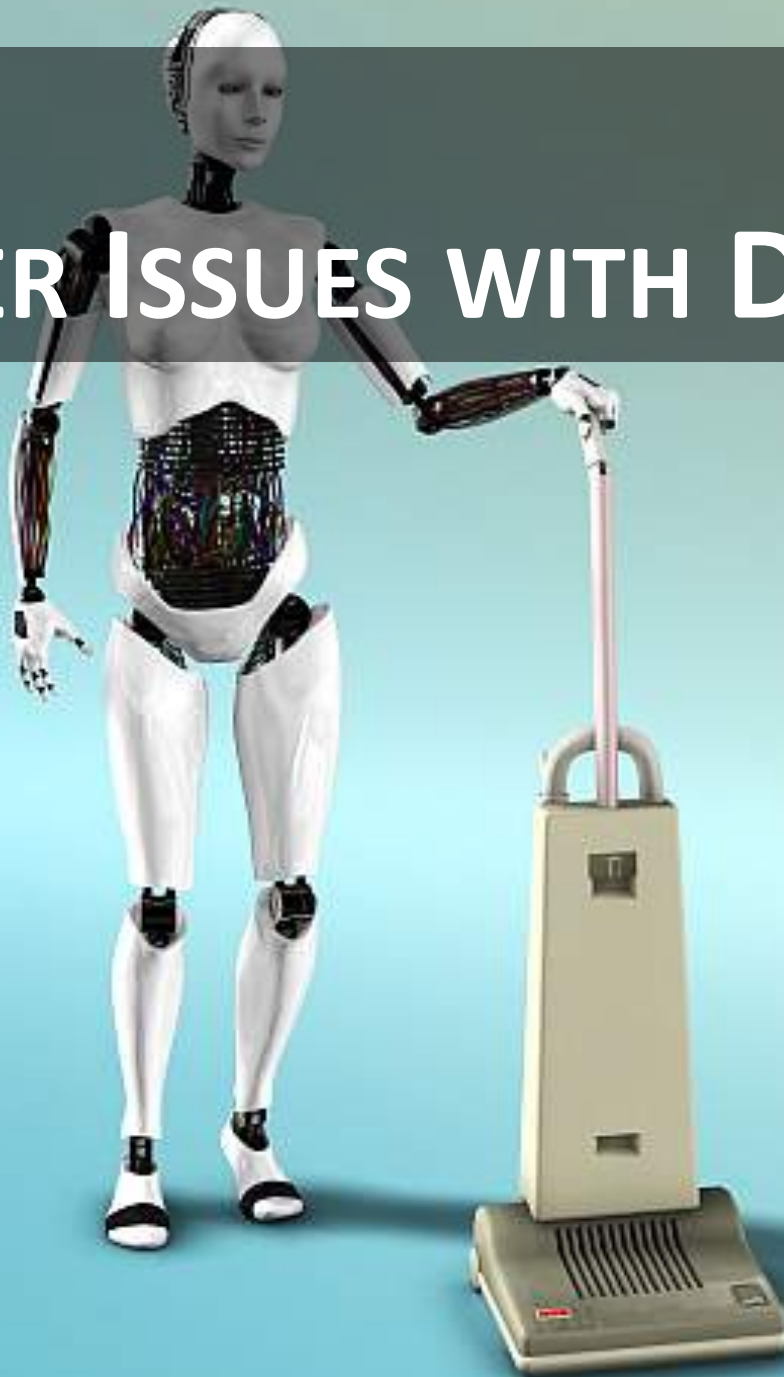
- Problem with new approaches to AI: Decision AI/algorithm black box, I am affected by its decision but do not know how it came to its decision
- Right to be informed, “Right to Explanation of Automated Decision Making” (Wachter et al. 2017) but is that possible?



# TRUST AND TRANSPARENCY

- **Trust in system (technology: reliability) vs trust in people (also emotions)**
- **Transparency of data, process, organisation: again, it depends on people**

# GENDER ISSUES WITH DESIGN



ROUTLEDGE STUDIES IN CONTEMPORARY PHILOSOPHY

# Using Words and Things

Language and Philosophy of Technology

Mark Coeckelbergh



# GENDER ISSUES AND HUMAN RELATIONSHIPS



**Harmony,  
The First AI  
Sex Robot**



OCTOBER 25-26 2018

# FEMINIST PHILOSOPHY OF TECHNOLOGY

<https://philtech.univie.ac.at/>

KEYNOTE SPEAKERS

**CORINNA BATH**  
**RICK DOLPHIJN**  
**NINA LYKKE**  
**KATHLEEN RICHARDSON**

**LUCT SUCHMAN**

[mark.coeckelbergh@univie.ac.at](mailto:mark.coeckelbergh@univie.ac.at)

**JUDY WAJCMAN**



# ETHICS: APPROACH

- **Bottom up**
- **Pro-active**
- **Global**
- **Positive**

**Ethical & legal theory and principles**



**Experience – Practices**

## Artificial intelligence (AI) Shortcuts

# 'Thou shalt not always beat us at chess': an alternative 10 commandments for robots

The lord bishop of Oxford has handed a new list of laws for AI to a select committee. But, if we are to live in harmony with our robotic companions, here are a few more he might wish to include

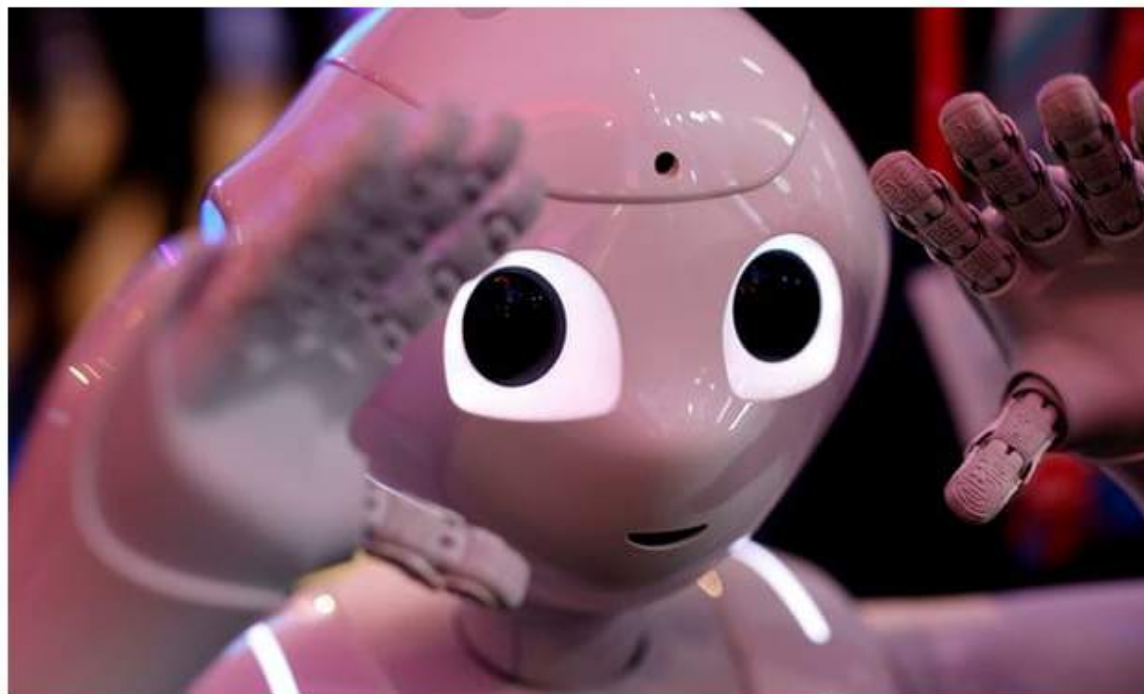


**Stuart Heritage**

@stusheritage  
Mon 5 Mar 2018 12.58 GMT



149 325



▲ A Pepper robot by SoftBank Robotics. Photograph: AFP/Getty



**Ethical & legal theory and principles**



**Experience – Practices**

A row of European Union flags on tall poles in front of a modern building with a curved facade. The building has a grid-like pattern of windows. The sky is overcast.

# **ETHICS AND REGULATION: LET'S TRY TO BE PRO-ACTIVE**

# ETHICS: HOW NOT TO DO IT



theguardian

## Thousands of drivers suffer loss of power following VW emissions 'fix'

41,000 owners are bringing a class action against the manufacturer citing poor performance, worse fuel consumption – and no compensation

THEGUARDIAN.COM



## Volkswagen executive pleads guilty in emissions scandal

A German Volkswagen executive pleaded guilty Friday to conspiracy and fraud charges in Detroit in a scheme to cheat emission rules on nearly 600,000 diesel vehicles.

LATIMES.COM



## Volkswagen: The scandal explained - BBC News

The scandal over VW cheating pollution emissions tests in the US is casting a cloud over the whole car industry.

BBC.COM | BY BBC NEWS

# ETHICS: PRO-ACTIVE IN RESEARCH AND INNOVATION



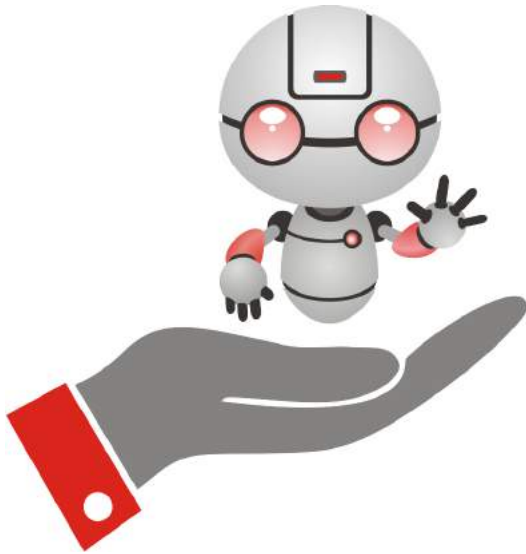
- **Regulation: needed, but always too late?**
- **Work also through standards, see IEEE**
- **Certification**



# GLOBAL ACTION NEEDED

- **Due to nature of new technologies**
- **Do we have suitable institutions for this?**

ALSO NON-GOVERNMENTAL ACTORS!



RESPONSIBLE  
**ROBOTICS**

*Accountable Innovation for the Humans Behind the Robots*

The background of the slide features a row of marble busts of ancient philosophers, likely from the Stoic or Epicurean schools, set against a dark background. The lighting is dramatic, highlighting the textures of the stone and the profiles of the faces.

# **POSITIVE: ETHICS AND THE GOOD LIFE**

- **Not just constraints and what not to do, but also what to do and how to live (good life, virtue, community/society)**

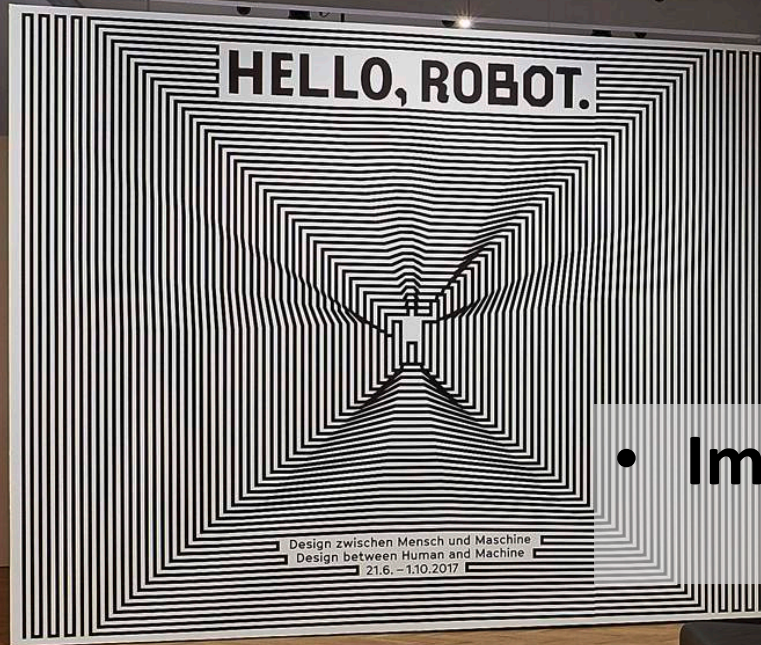
A scientist in a light blue lab coat is working in a laboratory. He is focused on adjusting a piece of complex scientific equipment that features a large metal frame with several curved tubes. In the background, there is a computer monitor displaying a colorful image, and various lab supplies and equipment are visible on the workbench. The overall scene is a professional and technical laboratory environment.

# EXPLORE NEW POSSIBILITIES

- **New experiential and action possibilities**
- **Not only in the West**



# INNOVATION, DESIGN, ART



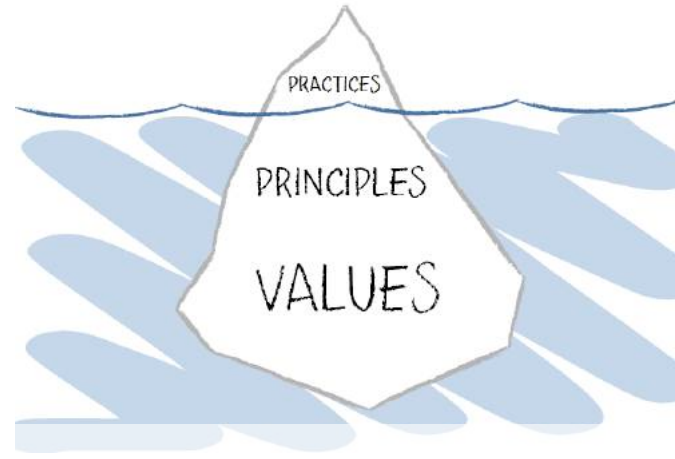
- Imagination needed



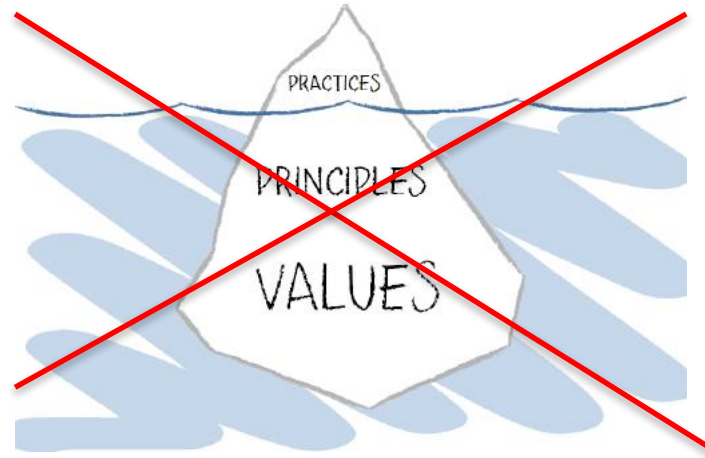


# **Policy needed**

**Everyone affected, need for vision and policy  
NOW**



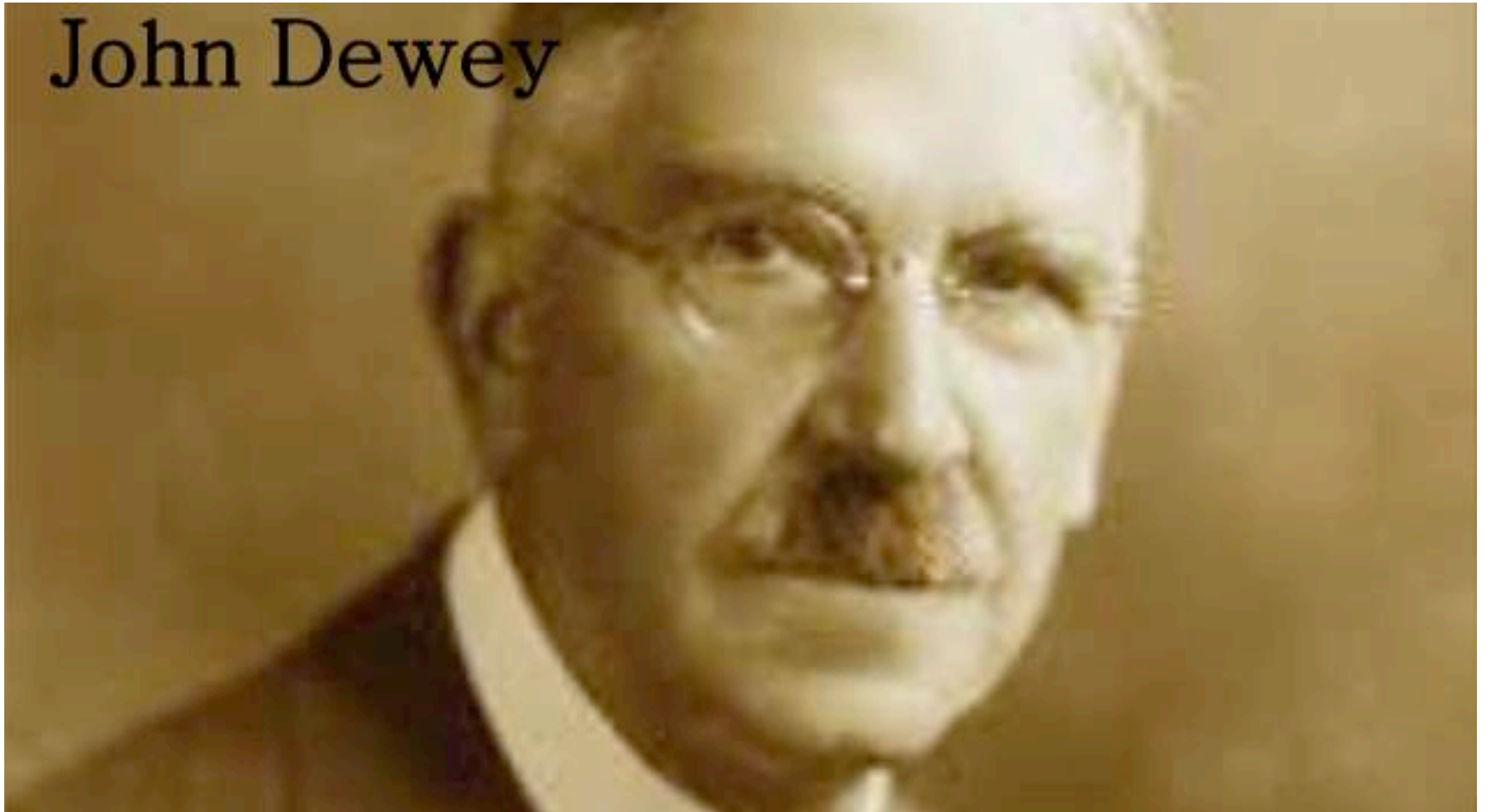
**“It’s the principles, stupid”**



**No, it's not only about principles, values, norms, theory, etc. Challenge is to change technological practices (design, innovation and use) and principles, theory, etc. are instruments to do that**

reflecting on experience

John Dewey



# What to do?

Usually ethics focuses on what (not) to do, but often we agree on what (not) to do; there are also other questions:

- Who does what?
- How to do things (best)?  
>> practical wisdom



# What to do?

Morality: constraints, red lines, sanctions

Ethics: the good life, the best life



# Who and how?

How can we work together to ensure that AI and robotics will contribute to a future we want? Also think about PROCESS

Experts, citizens, and mediators needed



# Who and how?

Role researchers, governmental, intergovernmental, and non-governmental organisations/civil society includes: raise awareness and bring people together, initiate new processes: HOW can we reach these goals?



# Who and how?

Power differences (e.g. big companies versus individual citizens)

Cultural differences (global, Europe)

# SOME BARRIERS

- **Lack of sufficient transdisciplinary expertise**
- **Lack of connections academia – policy makers and short-term views**
- **Insufficient institutional support for more participatory decision making**
- **Not taking into account lessons learnt, re-inventing the wheel**

# ADDRESS PROBLEMS

- **More support for transdisciplinary research**
- **Further institutionalize links academia – policy makers and make room for development of long-term vision**
- **Collaborate with other, non-governmental and non-academic actors in society**
- **More studies taking into account work already done, including work in the areas of philosophy of technology and robot ethics**

# THE FUTURE OF AI (& INFORMATICS)

**Beyond fear**

**Ethical**

**Interdisciplinary, incl. humanities**

**Connected to wider society**

**Europe: expertise in tech ethics**

# THE FUTURE OF AI (& INFORMATICS)

**The future of AI will be ethical or it will not be.**



**Thanks!**

**Mark Coeckelbergh**

Professor of Philosophy of Media and Technology  
University of Vienna

[mark.coeckelbergh@univie.ac.at](mailto:mark.coeckelbergh@univie.ac.at) || [coeckelbergh.wordpress.com](http://coeckelbergh.wordpress.com)