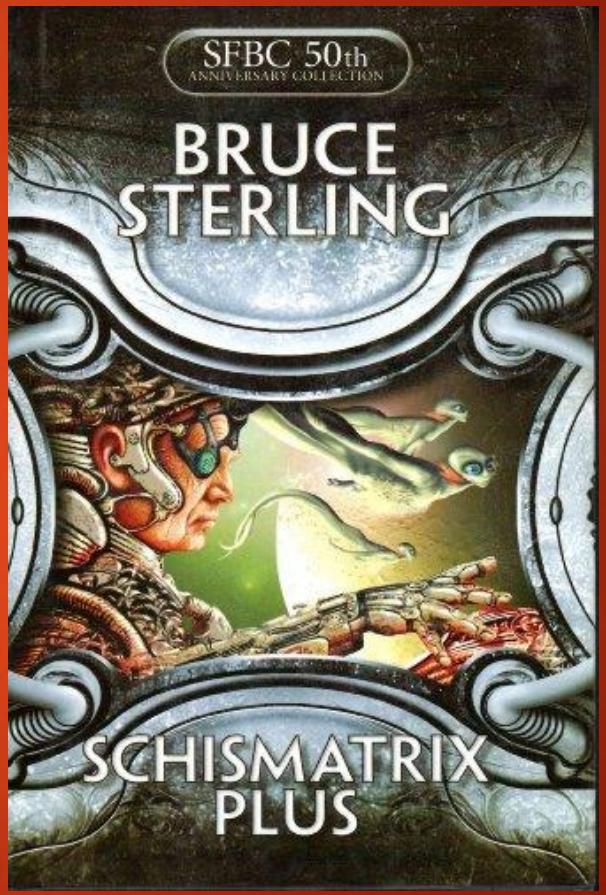
Posthumanism in the Age of AI: Expanding Humanistic Attitude

Sang Wook Yi

Department of Philosophy

Hanyang University



"Beyond Humanism: Artificial Intelligence, Information, and Posthumanism", International Conference, 17 December 2019, KIAS

AI's Challenge to Commonsense Humanism

- Machine-Learning Based AI's Impressive Performance without Awareness
- Challenging the so-called 'only human' area
- Bas Korsten's 'The Next Rembrandt'
- Unfamiliar Nature of AI Intelligence
- Providing an Opportunity for us to re-examine the place of humans in the Universe

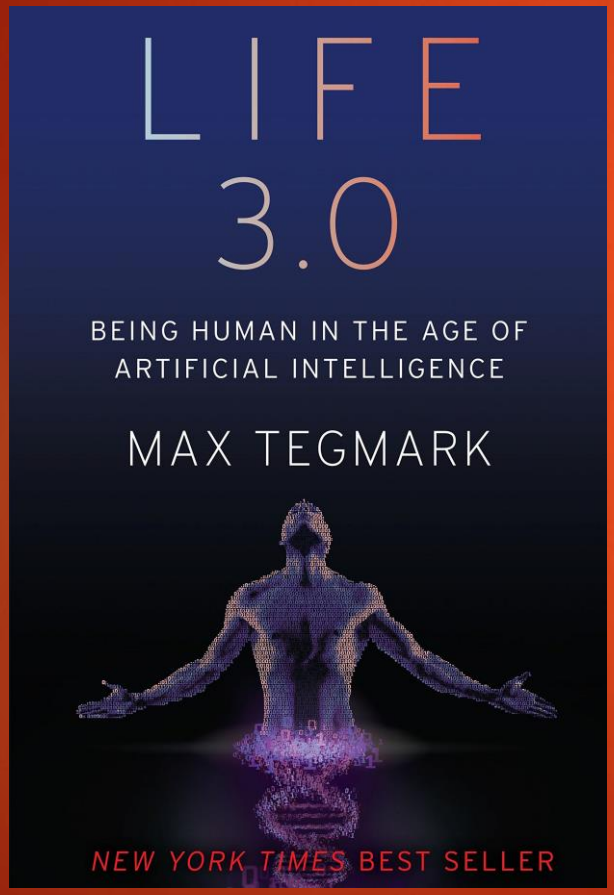
Humans facing strange intelligence

- Homo Neanderthalensis: a cousin species of Homo Sapiens, lived between 400,000 and 40,000 years ago in Eurasia region
- For long time, we never met a 'comparably intelligent' being other than fellow humans
- Naturally we tend to presuppose human-like mind behind human-like performance
- Now AI allows us to recognize this is just a historical contingency!

Artificial Intelligence or Machine Intelligence?

- The name AI highlights that they are human-made.
- Consequently, AI-related discussions are dominated by anthropomorphizing talk of AI and asking whether and when AI will surpass human intelligence.
- (Jerry Kaplan): What if we call artificial bird, not airplane?
- Accepting the plurality of minds, and treat machine intelligence as another kind of intelligence

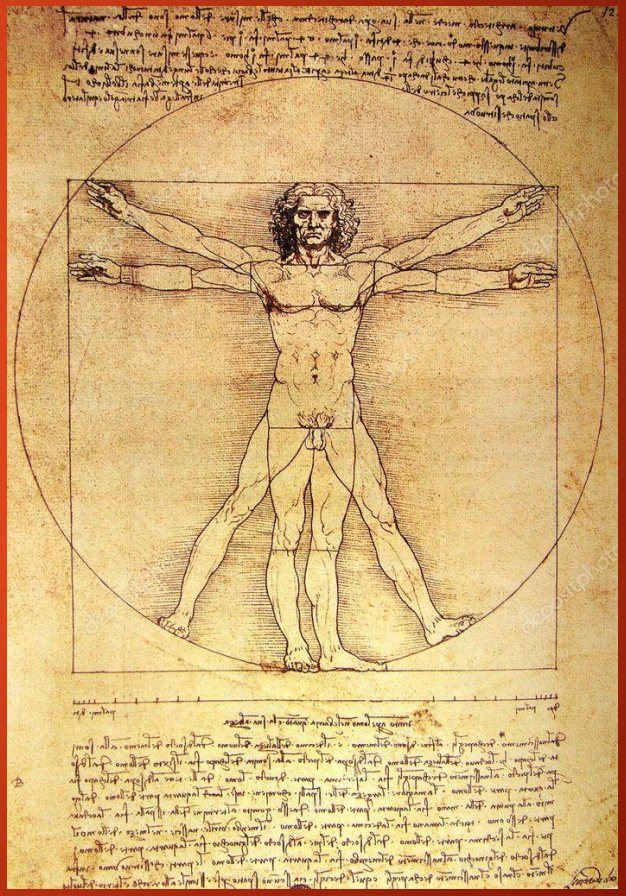
Singularity, Superintelligence, Existential Risk



- Exponentially developing AI >> Singularity shall be reached in near future, and Superintelligence will emerge. (Kurzweil)
- Wise warning of 'existential risk'? (Hawking, Tegmark, Russell, Tallin)
- Many critics including Pinker point out that the discussion relies on very 'thin' (and outdated) conception of intelligence!
- IEEE prefers AI/S instead of AI.

Ever Changing Conception of Humanism

- **Commonsensical Humanism (人文主義/人本主義)**: Respecting human existence, capabilities, characters, hopes, happiness, and examining human culture
- **Renaissance Humanism**: Admiring ancient Greek and Roman civilization >> Affirming human activities such as literature, art and architecture
- **French Humanism**: Declaring universal and inalienable human rights
- **Expanding Moral Circle or Moral Progress**



Posthumanism, Reexamining Humanistic Values

- Two ways of conceptualizing Postmodernism

1) Reexamining modernistic values
2) Radical relativistic destructionism

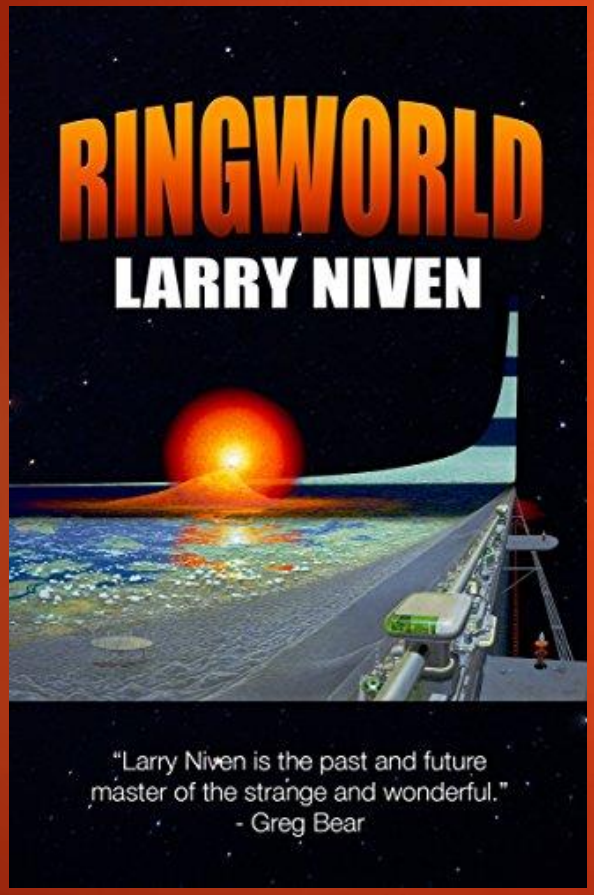
- Similarly, we might understand Posthumanism in two ways,

1) Reexamining taken-for-granted humanistic values
2) Promoting human enhancement (transhumanism)

Posthumanism as Methodological Attitude

- Notice that posthumanism does NOT presuppose that it is justified to treat humans, animals, and machines morally and legally equal.
- Posthumanism does NOT deny ALL humanistic values.
- Rather, Posthumanism pursues reevaluating and reexamining modernistic presuppositions as regards human rights, human dignity, and human uniqueness.

Future of Humanity and Posthumanism

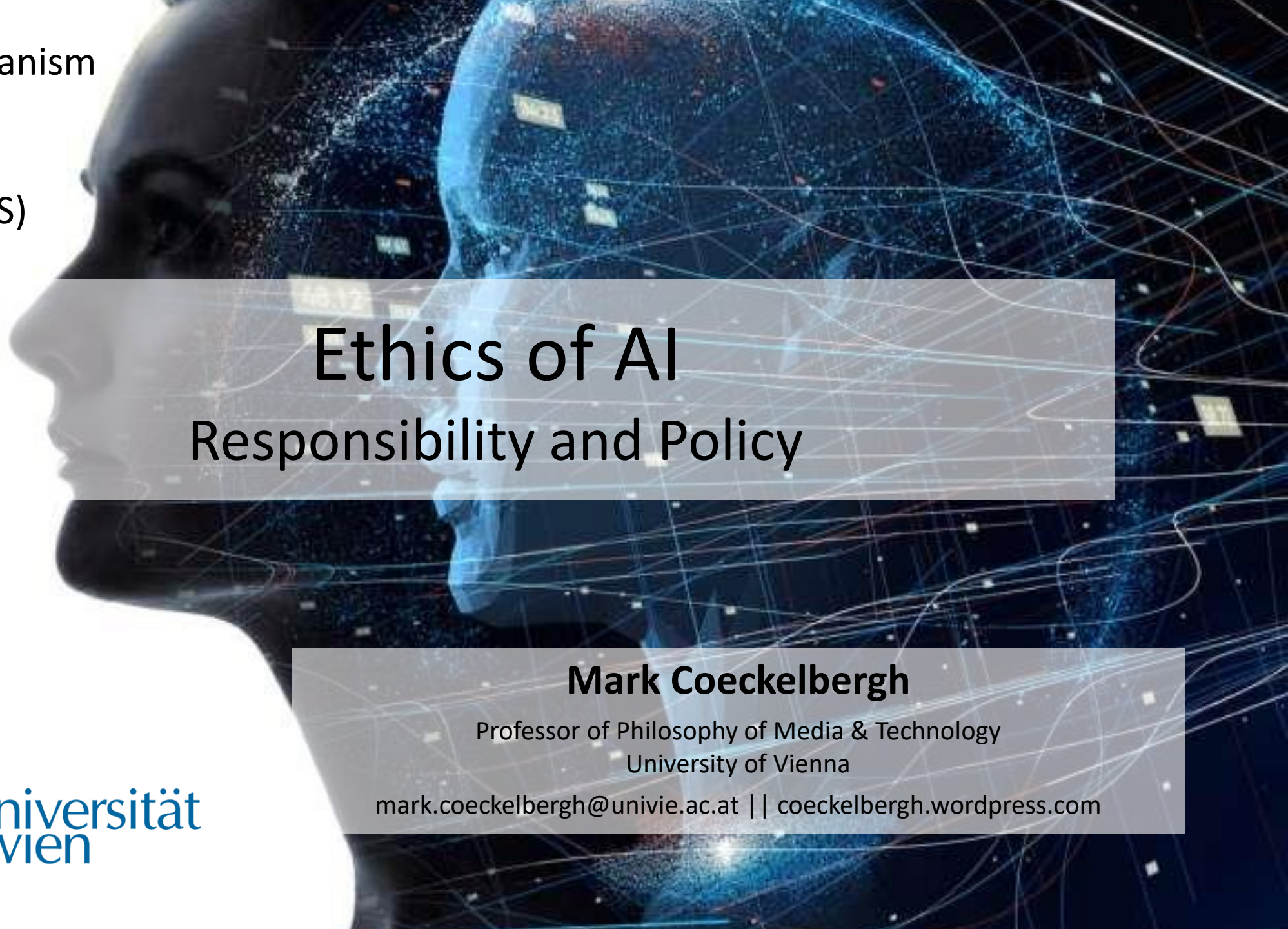


- **Redefining Humanistic Values**: Machine intelligence which is impressive but strange to humans urges us to redefining humanistic values without familiar human-centered viewpoint.
- Universe is likely to be populated by many (and strange to us) intelligent beings. And we need to navigate through the Universe in our future.
- Posthumanism as methodological attitude will be instrumental.

Thank You!


$\Psi_{\text{kitty}} = \frac{1}{\sqrt{2}}\Psi_{\text{alive}} + \frac{1}{\sqrt{2}}\Psi_{\text{dead}}$

Talk for Beyond Humanism conference
Korea Institute for Advanced Study (KIAS)



Ethics of AI

Responsibility and Policy

Mark Coeckelbergh

Professor of Philosophy of Media & Technology
University of Vienna

mark.coeckelbergh@univie.ac.at || coeckelbergh.wordpress.com

universität wien

Board

SPT President

Mark Coeckelbergh

Department of Philosophy
Universitätsstraße 7 (NIG)
1010 Vienna
Austria
https://coeckelbergh.wordpress.com/

Phone +43-1-4277-46466
mark.coeckelbergh(at)univie.ac.at

DETAILS FOR SPT 2019 NOW AVAILABLE!

SPT's 21st biennial meeting will take place at Texas A&M University. For all relevant information, please visit the official website: https://www.spt2019.org/. We look forward to seeing you all there!

SPT RESEARCH

Member Publications

FOLLOW SPT ON FACEBOOK

ABOUT SPT
BOARD
RESOURCES
CONTACT US

SEARCH

spt
The Society for Philosophy and Technology

HOME | NEWS | BIENNIAL MEETING | RELA

Board

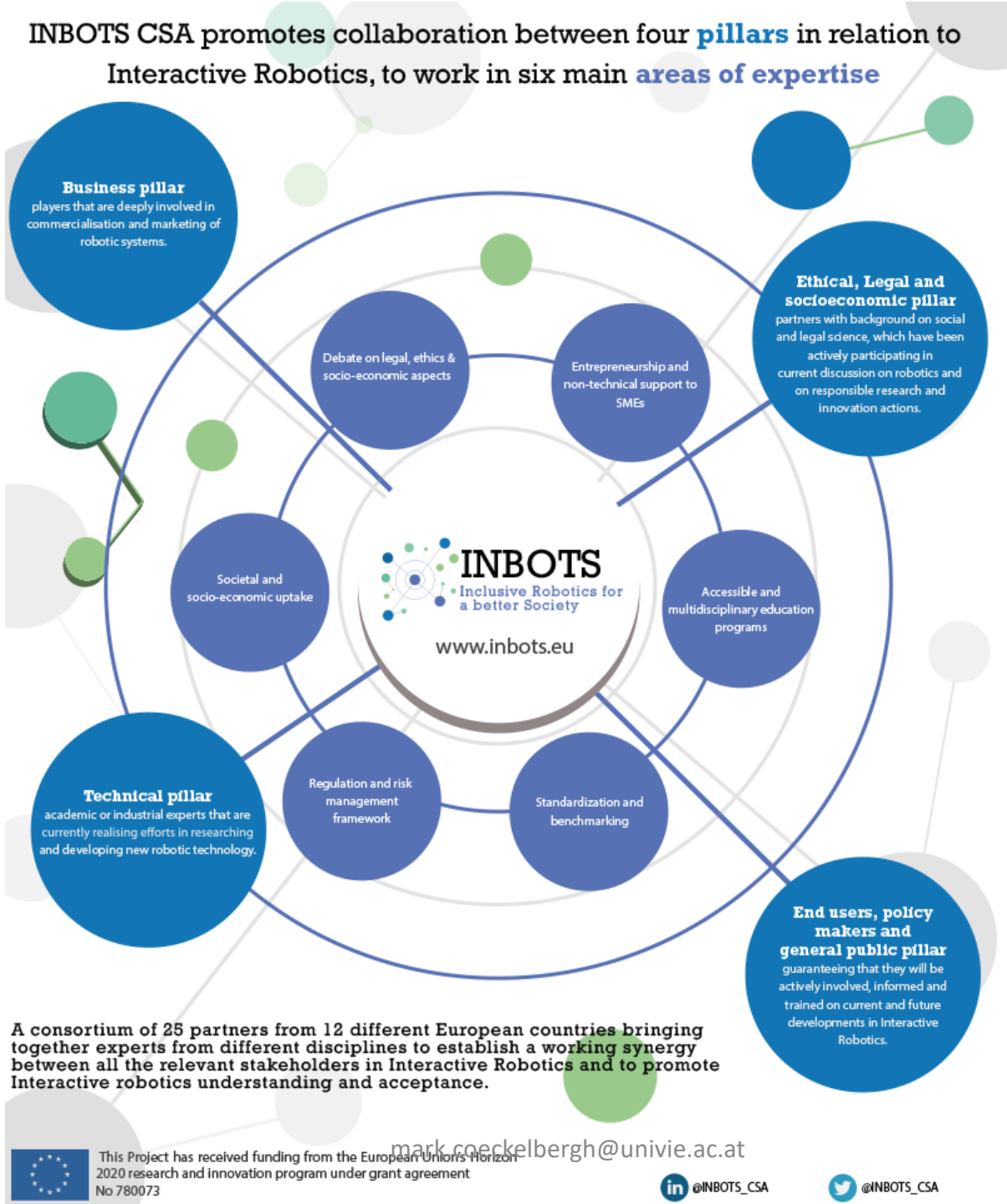
SPT President

Mark Coeckelbergh

Department of Philosophy
Universitätsstraße 7 (NIG)
1010 Vienna
Austria
https://coeckelbergh.wordpress.com/

Phone +43-1-4277-46466
mark.coeckelbergh(at)univie.ac.at

spt
The Society for Philosophy and Technology



This Project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 780073

mark.coeckelbergh@univie.ac.at

@INBOTS_CSA @INBOTS_CSA

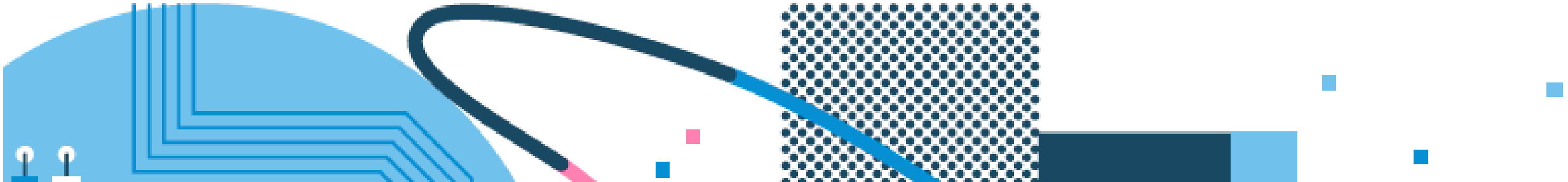

European Commission
INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION
AI
ETHICS GUIDELINES FOR TRUSTWORTHY AI

Wien, November 2018

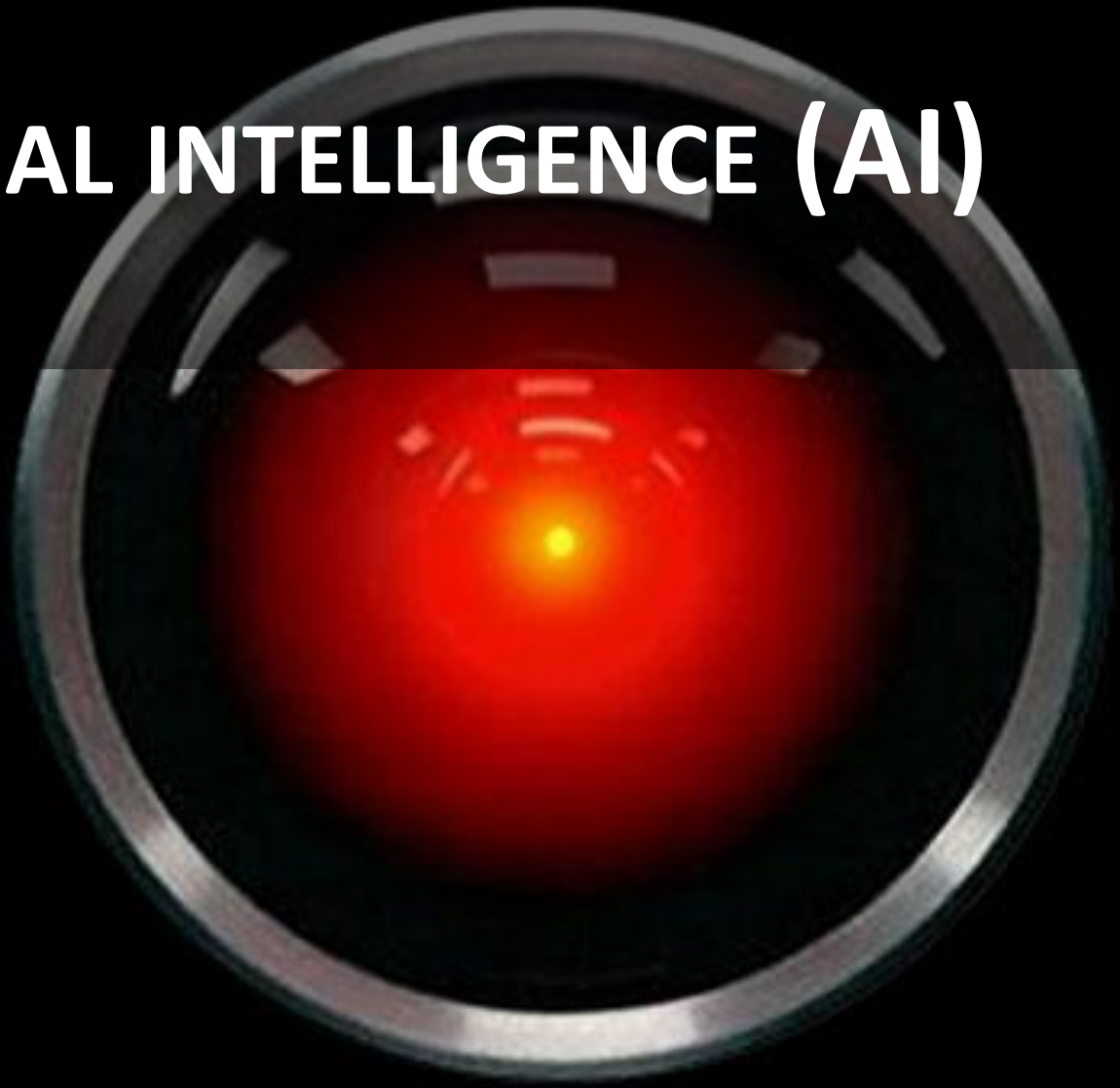
Österreichischer Rat
für Robotik und
Künstliche Intelligenz

Die Zukunft Österreichs mit Robotik und Künstlicher Intelligenz positiv gestalten

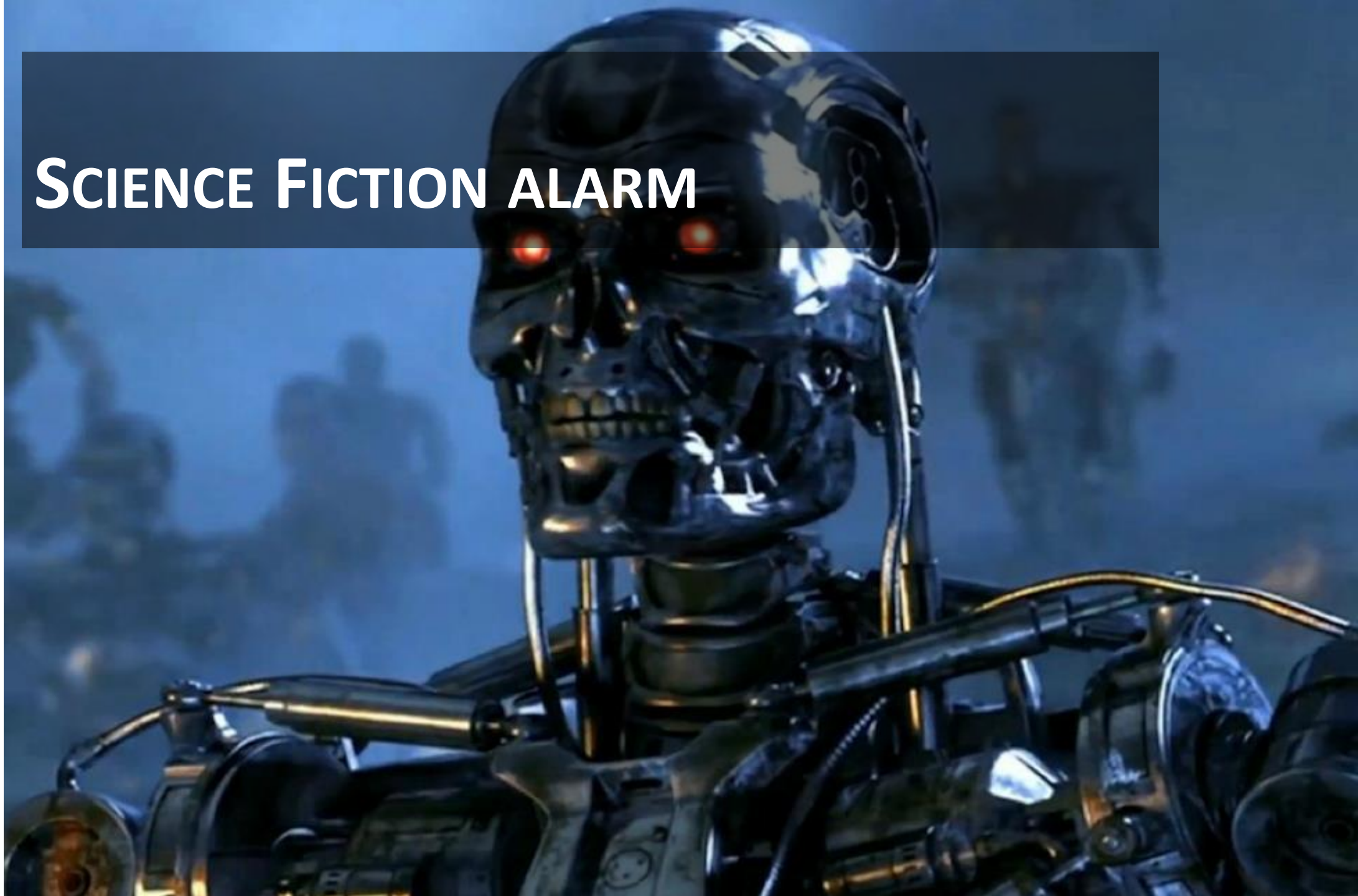
White Paper des Österreichischen Rats für Robotik und Künstliche Intelligenz



Artificial intelligence (AI)



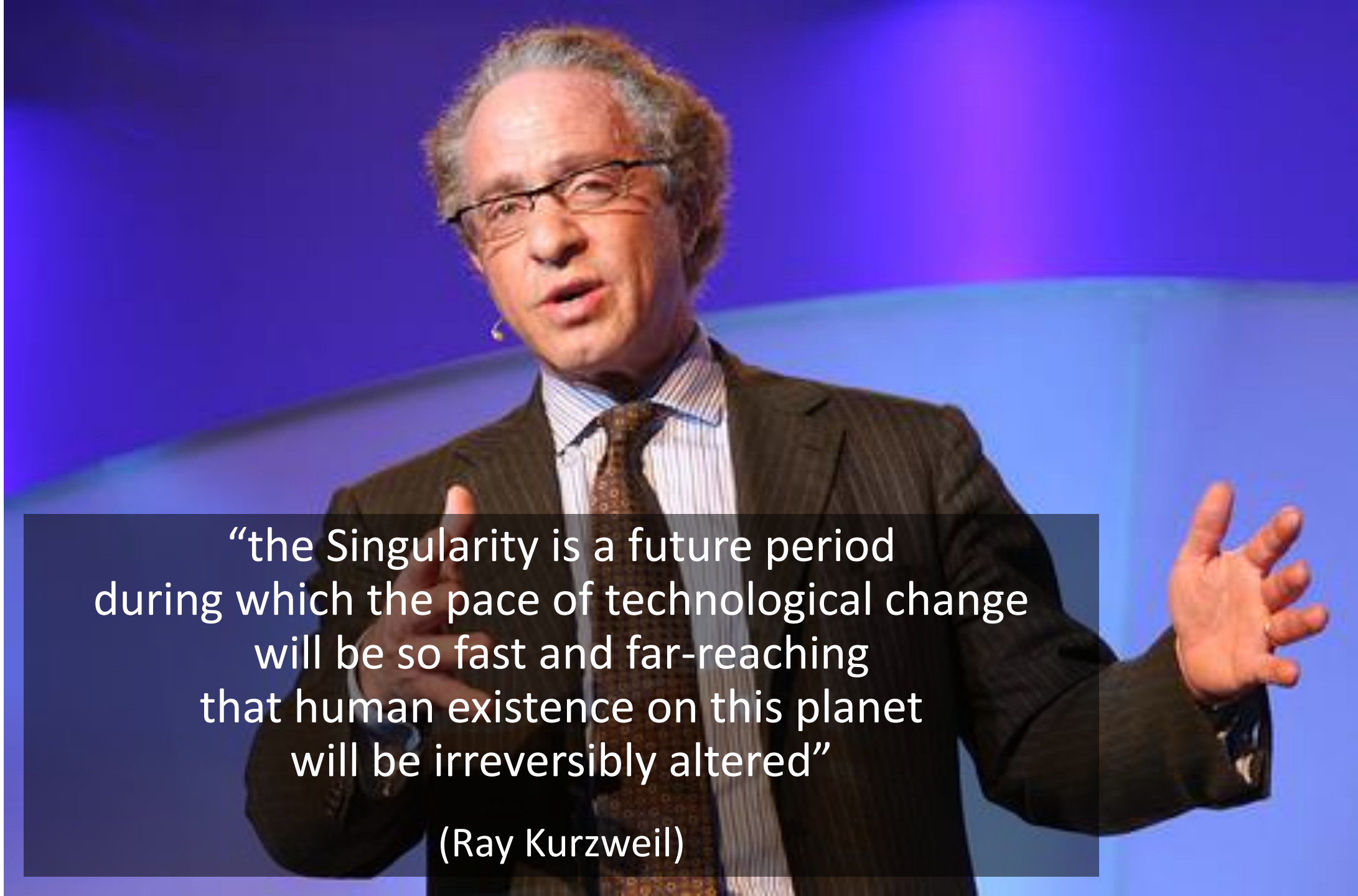
Science Fiction alarm




“AI is a fundamental existential risk for human civilization”
(Elon Musk)

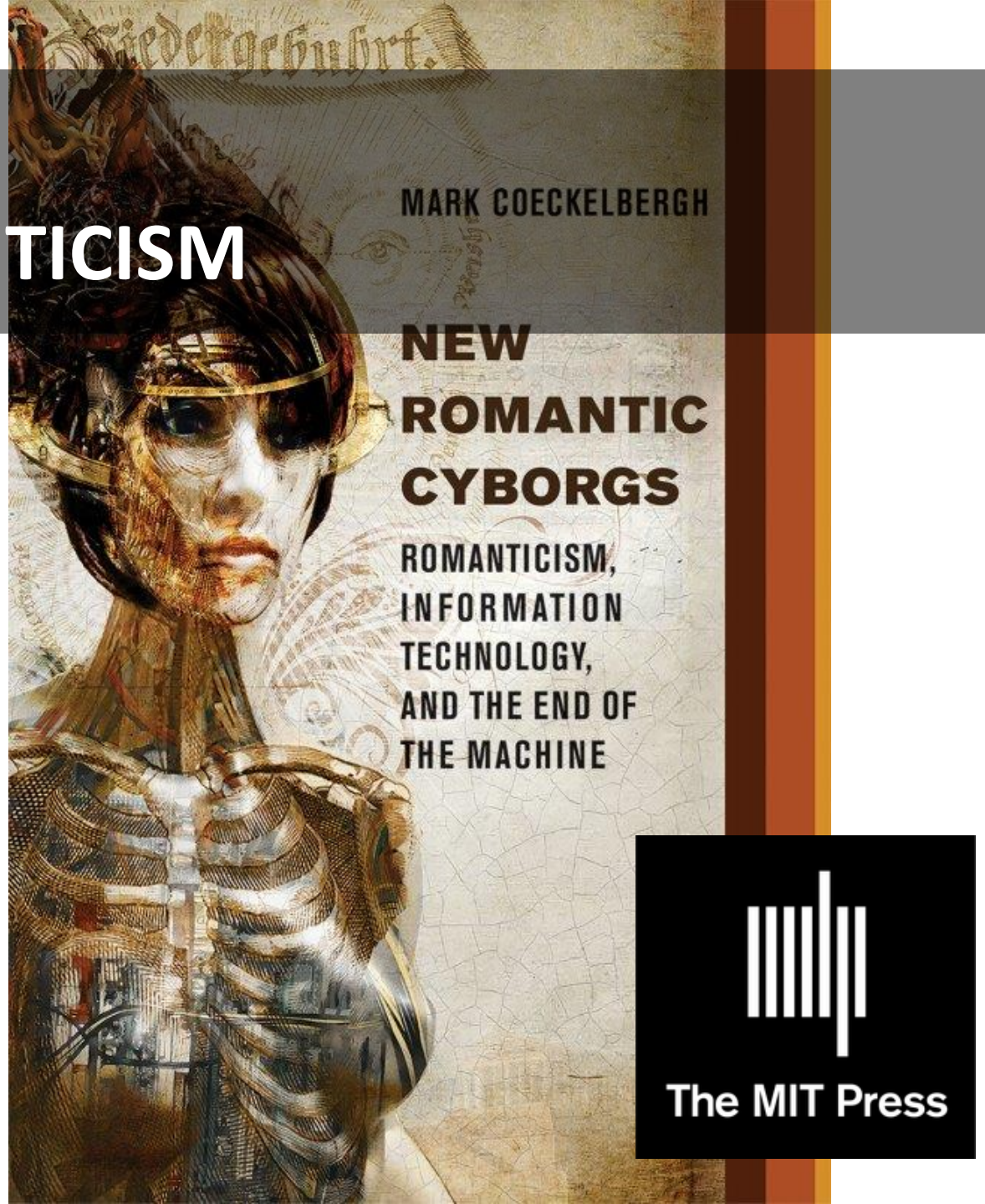
“we humans are like small children playing with a bomb”

(Nick Bostrom)


“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



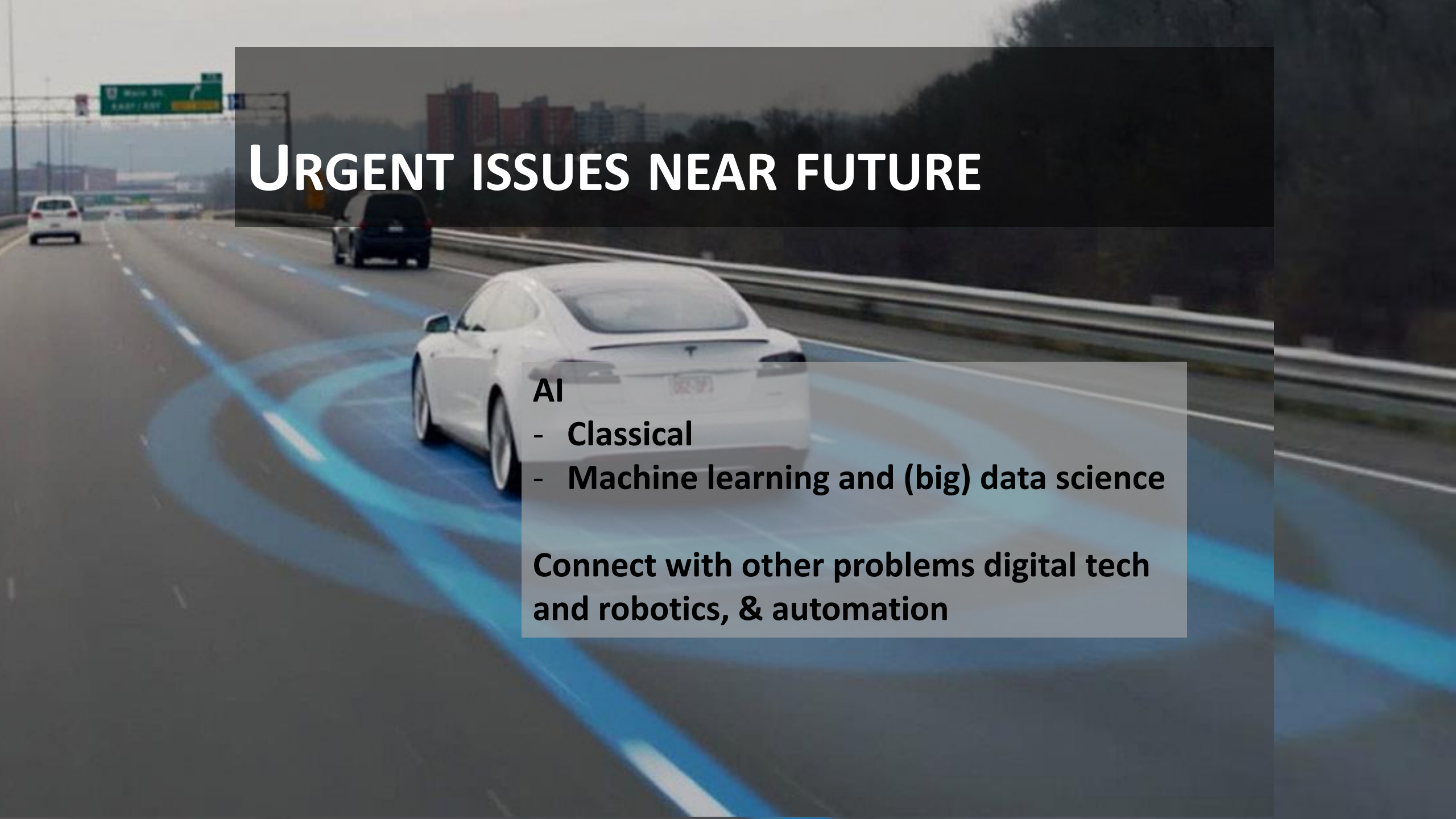
AGAINST ALARMISM

KEEP
CALM
AND
DO YOUR
HOMEWORK

There is no general AI!

AI: In your pocket

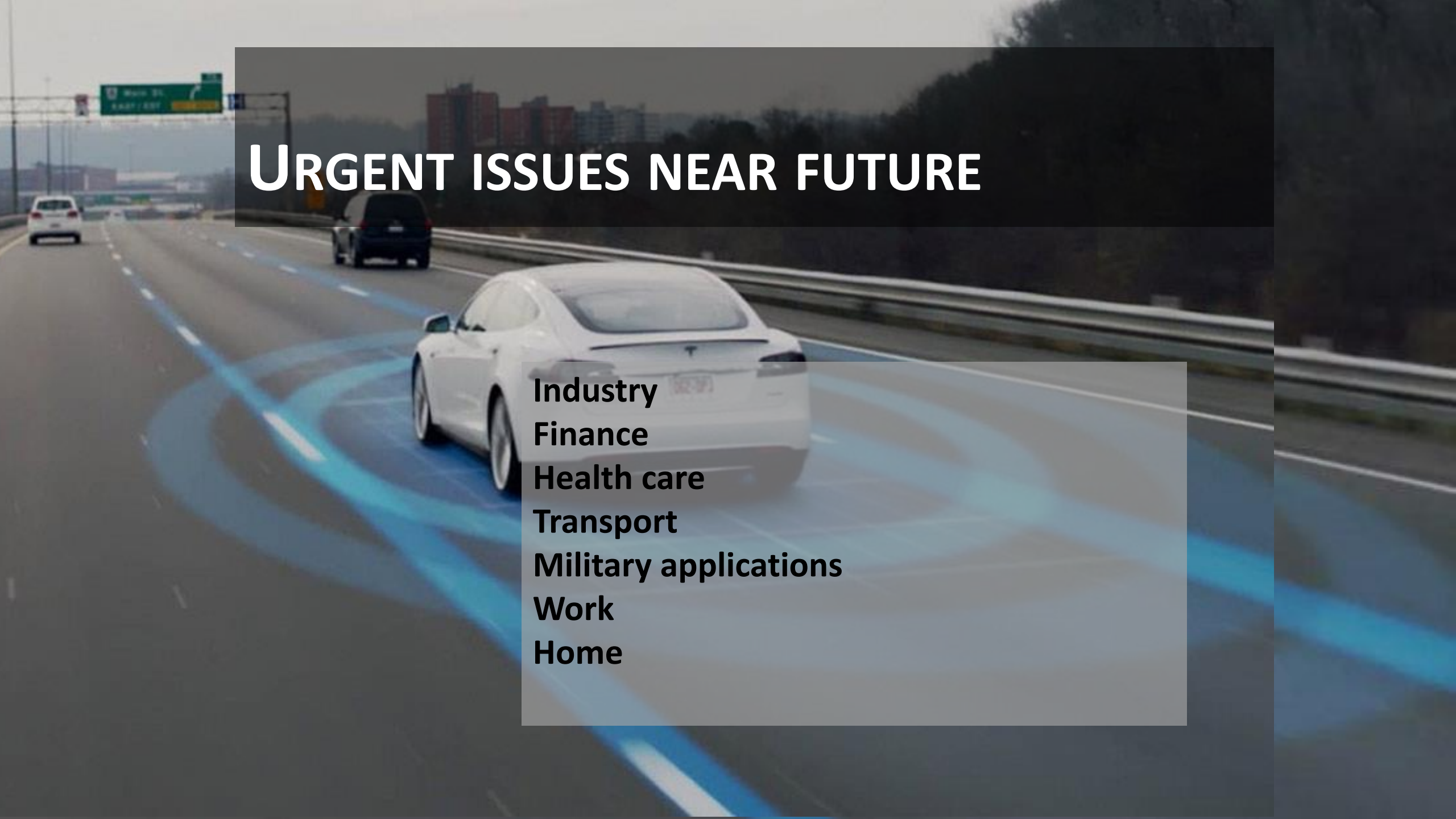
URGENT ISSUES NEAR FUTURE



AI

- **Classical**
- **Machine learning and (big) data science**

Connect with other problems digital tech and robotics, & automation



URGENT ISSUES NEAR FUTURE

- Industry
- Finance
- Health care
- Transport
- Military applications
- Work
- Home

ETHICAL, LEGAL, AND SOCIETAL PROBLEMS

Derek Bacon

Definition problems

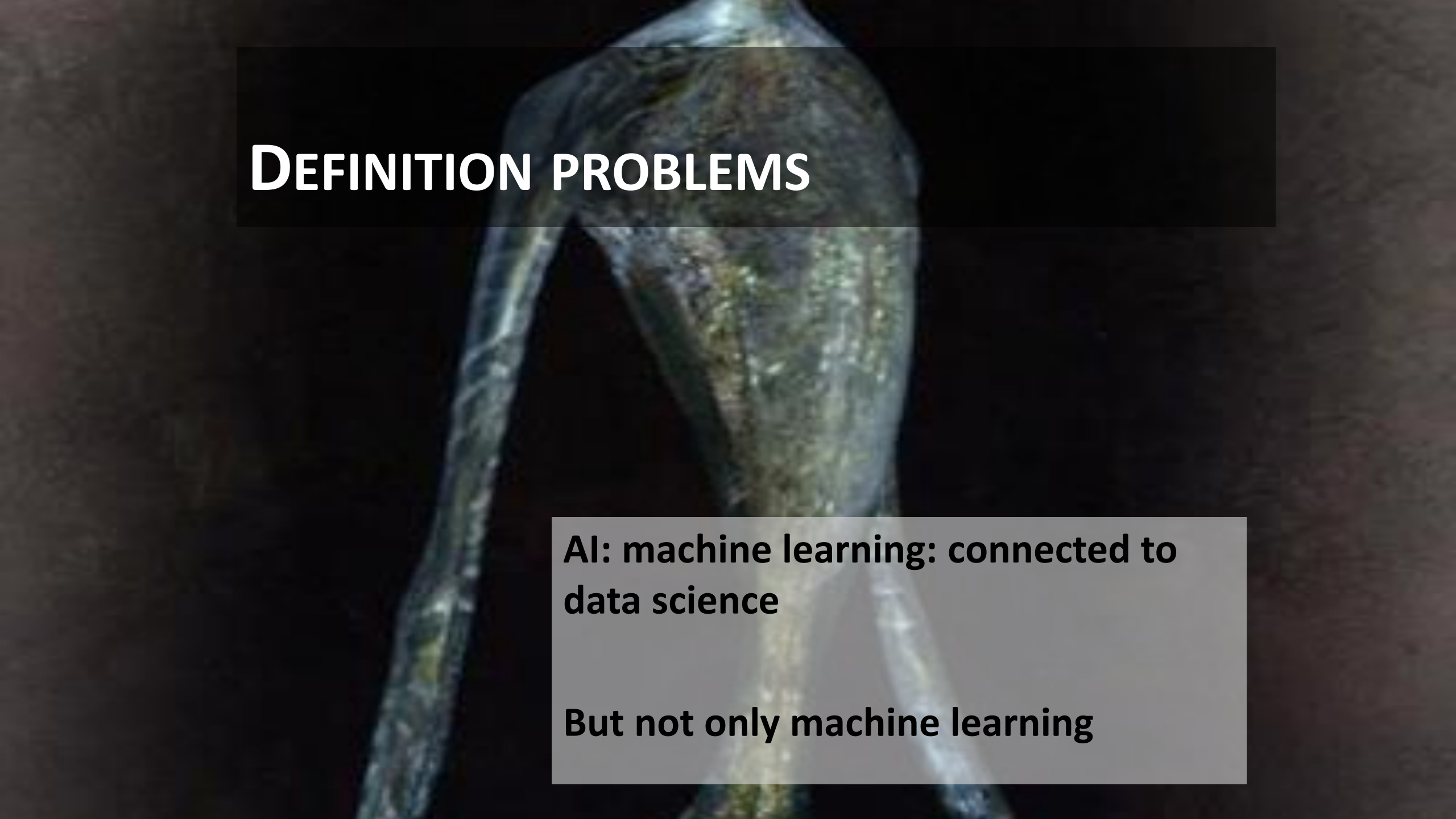
Problem for discussion and regulation:

- **What are we talking about?**
  - **AI, robots, algorithms, code, smart tech, internet of things, 'cyber-physical systems' … ?**

Definition problems

Problem for discussion and regulation:

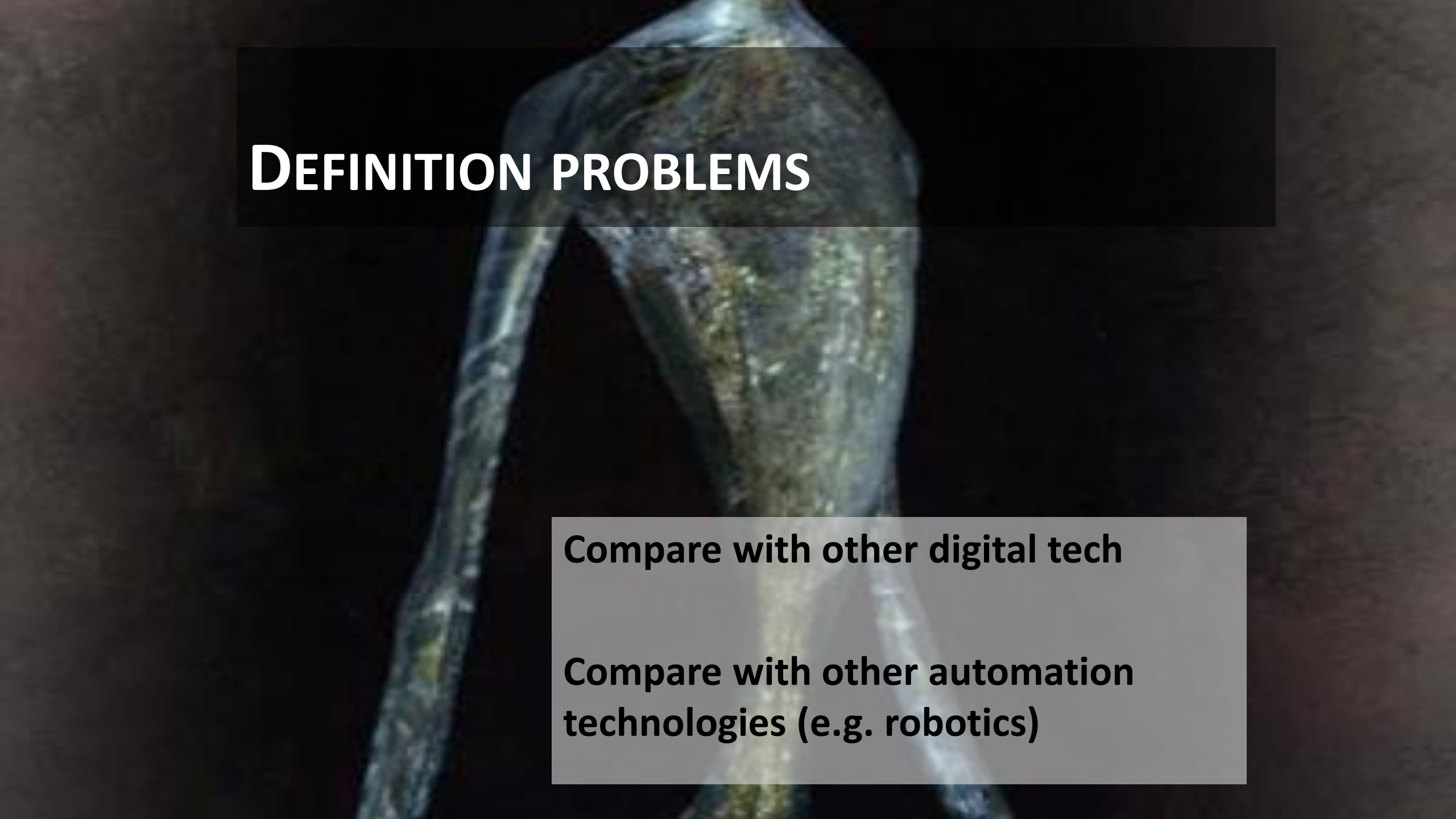
- **“the AI”???**
- **How autonomous, intelligent, etc.?**



Definition problems

AI: machine learning: connected to data science

But not only machine learning



Definition problems

Compare with other digital tech

Compare with other automation technologies (e.g. robotics)

Privacy, Security, Surveillance

- **Privacy and data protection**
  - **The system records what you do and transfers data… to whom? Company? Third Party? (and do you know it?)**
- **Security**
  - **What if your software gets hacked?**

EXPLOITATION AND MANIPULATION

Safety


Artificial Intelligence and Robotics
+ Add to myFT
Worker at Volkswagen plant killed in robot accident
Fatality touches on concerns about spread of automation

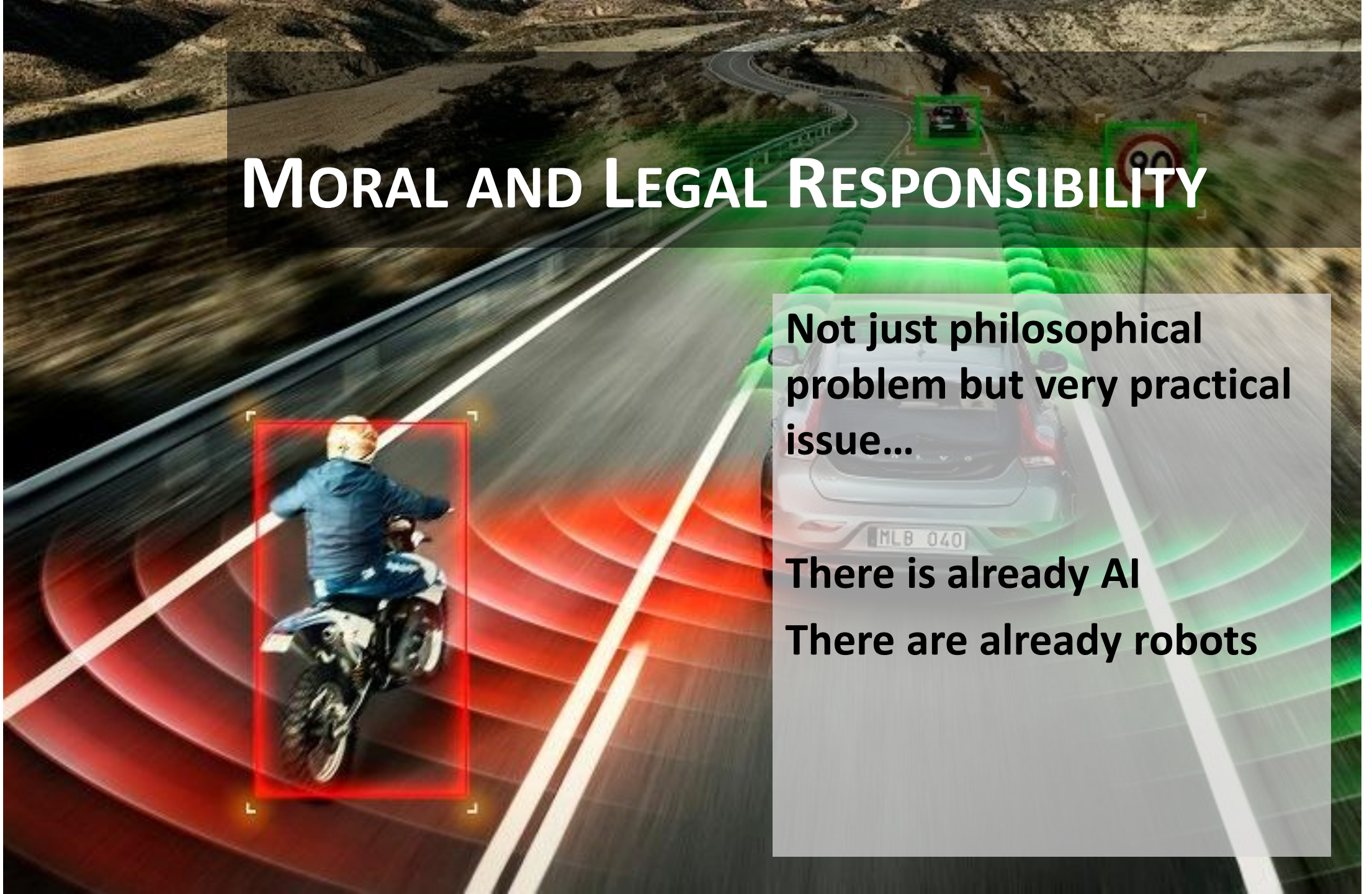
Vulnerable Users, Attachment and Deception

Human Dignity

REPLACEMENT?

- Not just about jobs but also about tasks
- What about human-AI and human-robot collaboration?

Moral and Legal Responsibility



Not just philosophical problem but very practical issue…

There is already AI

There are already robots

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

THE PROBLEM

Given that AI gets more agency, who is responsible?

- **How to attribute responsibility?**
  - **What is required to take or ascribe responsibility?**

Moral and Legal Responsibility

Responsibility conditions since Aristotle:

- control condition (agency)

- knowledge condition

Moral and Legal Responsibility

Responsibility conditions since Aristotle:

- **control condition (agency)**
- **knowledge condition**

Moral and Legal Responsibility

Responsibility

- **The technology cannot be held responsible**
  - **Lacks the required moral agency capacities, e.g. free will**
    - **Does not really "know" what it is doing, is not aware of what it is doing**
- **<u>Humans</u> can**
  - **BUT WHO????**

Self-driving Uber kill s Arizona woman in first fatal crash involving pedestrian

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

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 kill s Arizona woman in first fatal crash involving pedestrian

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

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?**

Moral and Legal Responsibility

Some problems

- How to attribute and distribute responsibility if there are not only many hands but also *many things*?
- how to make sure responsibility traces back to humans? human in control?

- **Measures**
  - insurance?
  - regulating or ban?
  - new legal instruments or not? (e.g. existing liability law enough?)

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
    - …

Moral and Legal Responsibility

Example: Classificaton 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

Moral and Legal Responsibility

ANOTHER EXAMPLE:

Learning algorithms and data science

-> responsibility for data collection, selection, bringing datasets together, etc. – responsibility relevant at all stages of the process

Responsibility attribution difficult because of long causal history with many hands and many things

Moral and Legal Responsibility

Responsibility

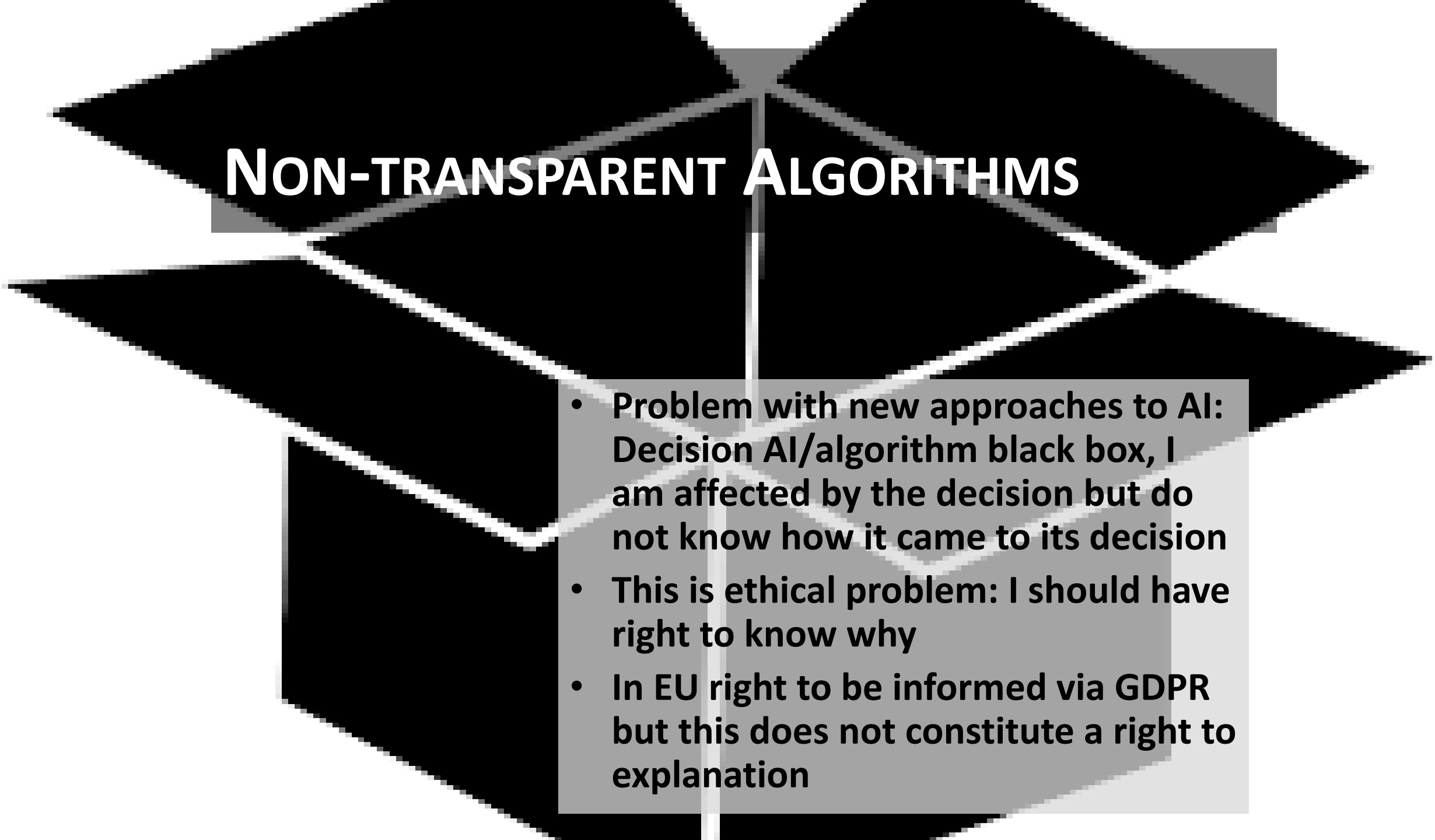
- control condition

- knowledge condition

Moral and Legal Responsibility

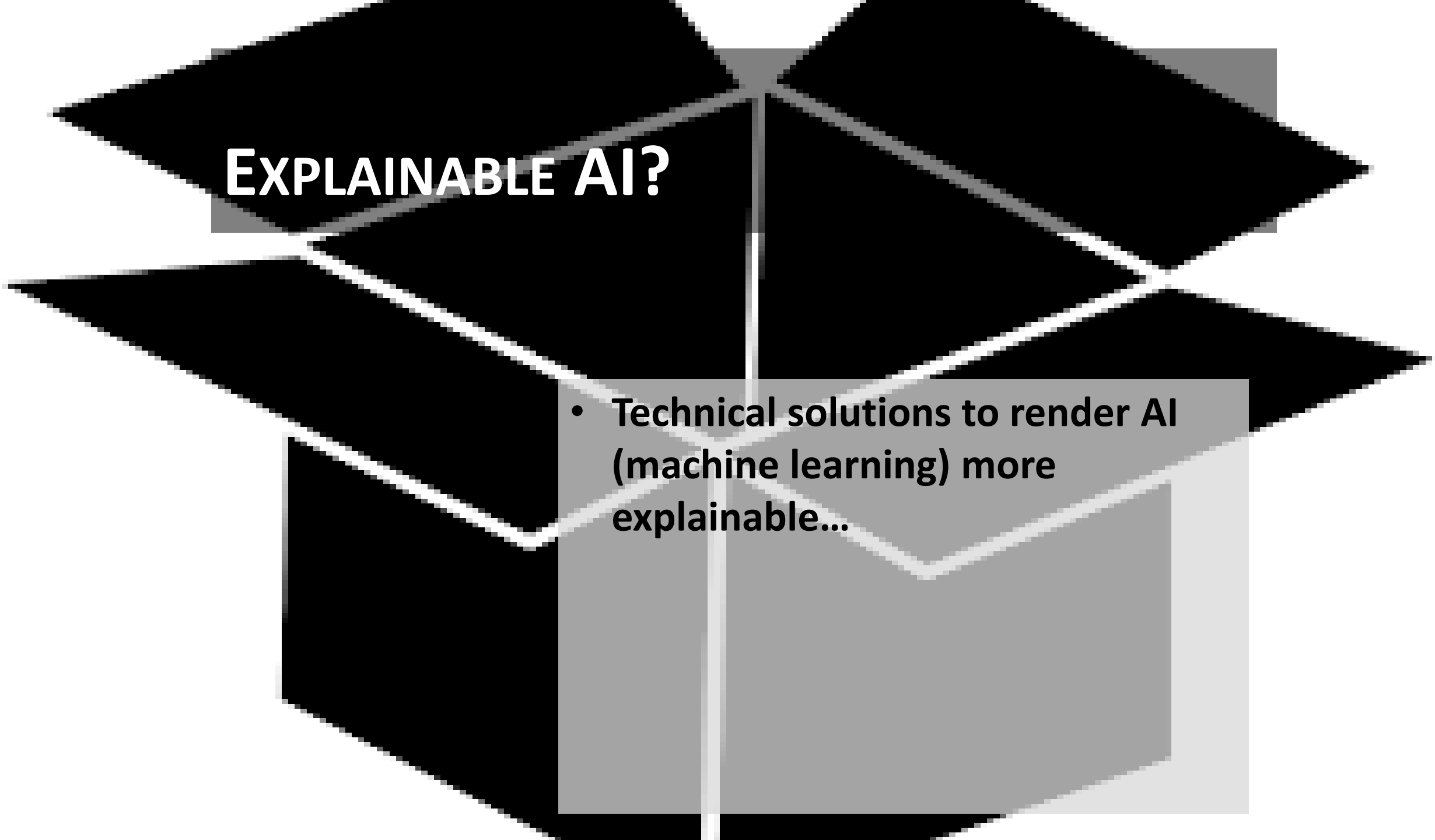
Responsibility

- **Many humans don’t know what they are doing when they use AI**
  - **Don’t know limitations of the system**
  - **Don’t know the potential ethical consequences**
  - **The technical system itself may not be transparent (not even to technical people)**



Non-transparent Algorithms

- Problem with new approaches to AI: Decision AI/algorithm black box, I am affected by the decision but do not know how it came to its decision
- This is ethical problem: I should have right to know why
- In EU right to be informed via GDPR but this does not constitute a right to explanation



EXPLAINABLE AI?

- **Technical solutions to render AI (machine learning) more explainable…**

Responsibility for decisions

- The AI does not “decide” but makes recommendations
- In the end the human decides and remains responsible for the decision
  - Includes duty to be able to explain decision to those affected (not just a technical matter!) >>>

Responsibility to whom?

- To whom are we responsible?
  - Responsibility is not only about agents and their knowledge; it is also about the ”responsibility patients”: to whom are we and should we be responsible?
  - Responsibility as answerability
  - Relational approach
- This is extra reason for explainability and transparency: we owe an explanation to those affected by AI (see new article >>>)

Science and Engineering Ethics
https://doi.org/10.1007/s11948-019-00146-8

ORIGINAL RESEARCH/SCHOLARSHIP

Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability

Mark Coeckelbergh[1]

Received: 22 March 2019 / Accepted: 9 October 2019

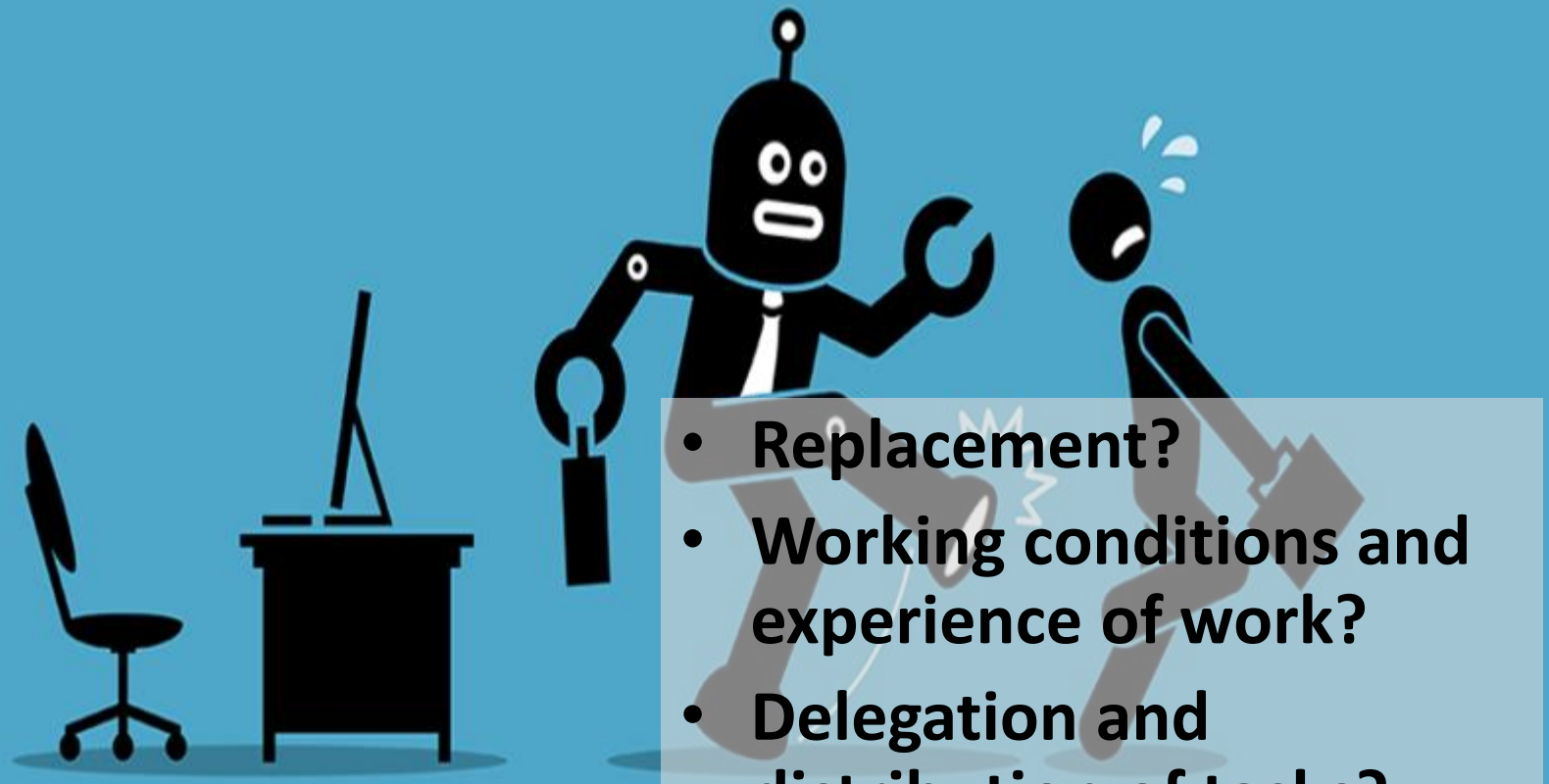
Abstract

This paper discusses the problem of responsibility attribution raised by the use of artificial intelligence (AI) technologies. It is assumed that only humans can be responsible agents; yet this alone already raises many issues, which are discussed starting from two Aristotelian conditions for responsibility. Next to the well-known problem of many hands, the issue of "many things" is identified and the temporal dimension is emphasized when it comes to the control condition. Special attention is given to the epistemic condition, which draws attention to the issues of transparency

mark.coeckelbergh@univie.ac.at

Societal Implications

The Future of Work



- Replacement?
- Working conditions and experience of work?
- Delegation and distribution of tasks?

Technological Forecasting & Social Change 114 (2017) 254–280

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

The future of employment: How susceptible are jobs to computerisation?☆

Carl Benedikt Frey[a,*], Michael A. Osborne[b]

[a]Oxford Martin School, University of Oxford, Oxford OX1 1PT, United Kingdom
[b]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
J62
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.

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)

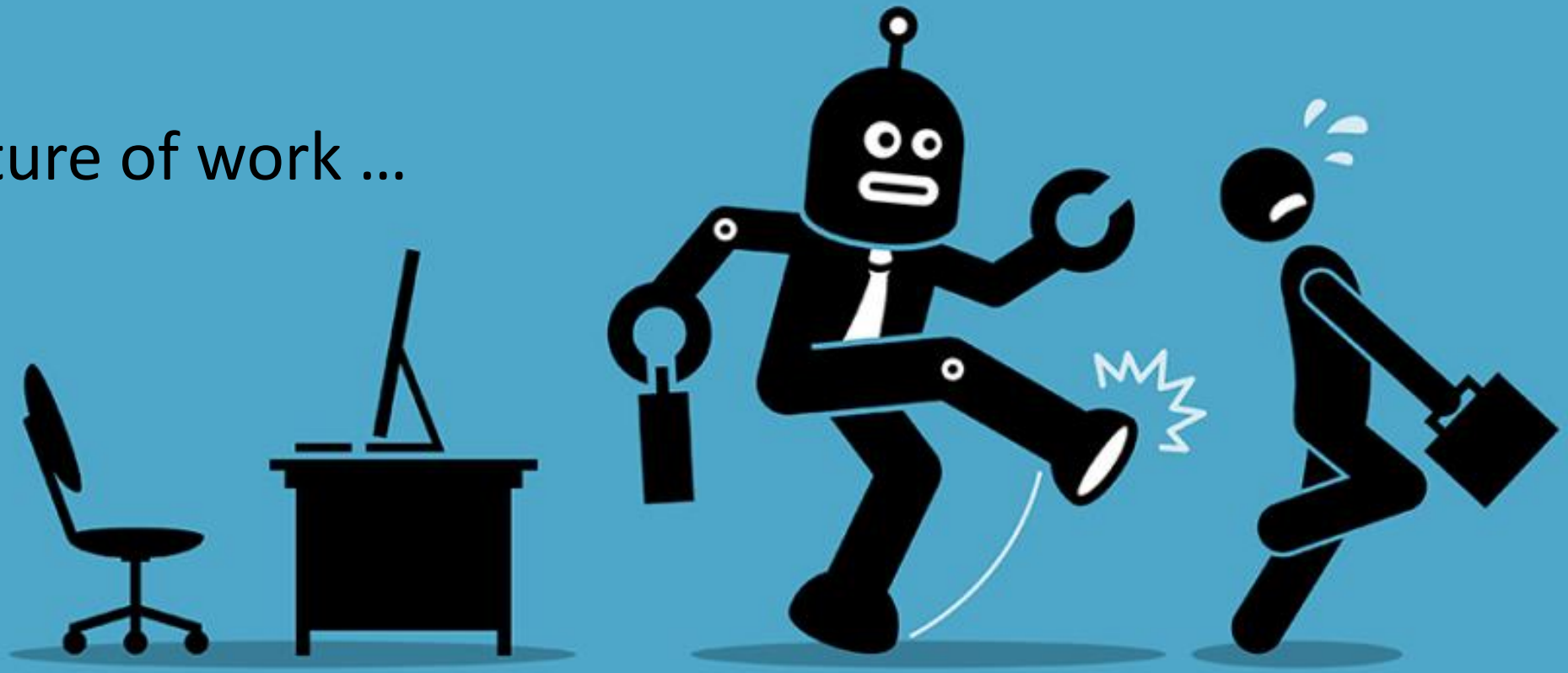
Rapporteur: **Catelijne MULLER**

THE SECOND MACHINE AGE

WORK, PROGRESS, AND PROSPERITY IN A TIME OF BRILLIANT TECHNOLOGIES

ERIK BRYNJOLFSSON
ANDREW McAFEE

The future of work …



… and the meaning of life

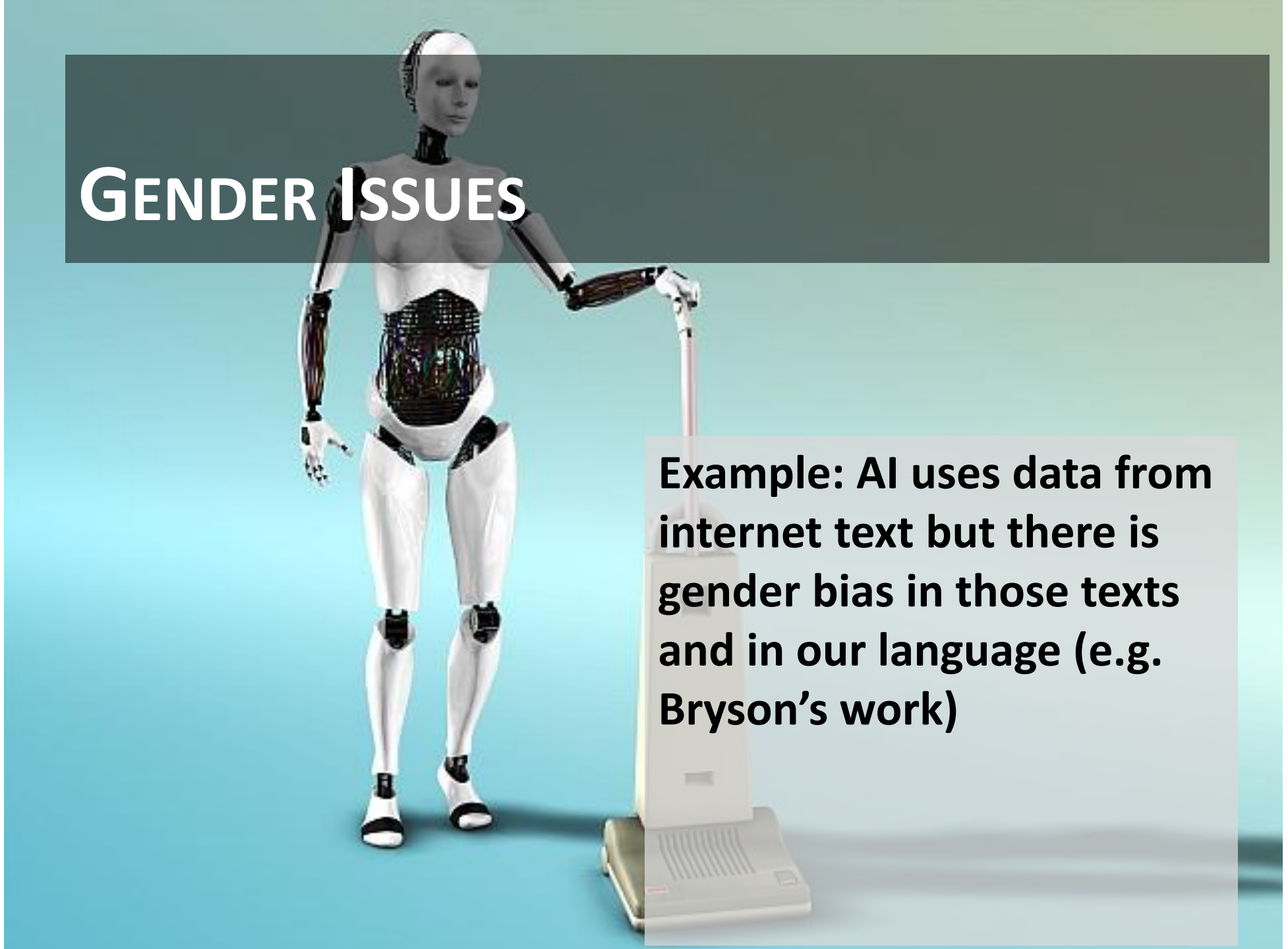
Biased Algorithms

- **Problem in machine learning: bias in algorithm or dataset**
  - **Bias can arise at all stages (data collection, cleaning, algorithm, training data versus implementation)**
  - **Problem of algorithm or society, or both? How to deal with this?**

Biased Algorithms

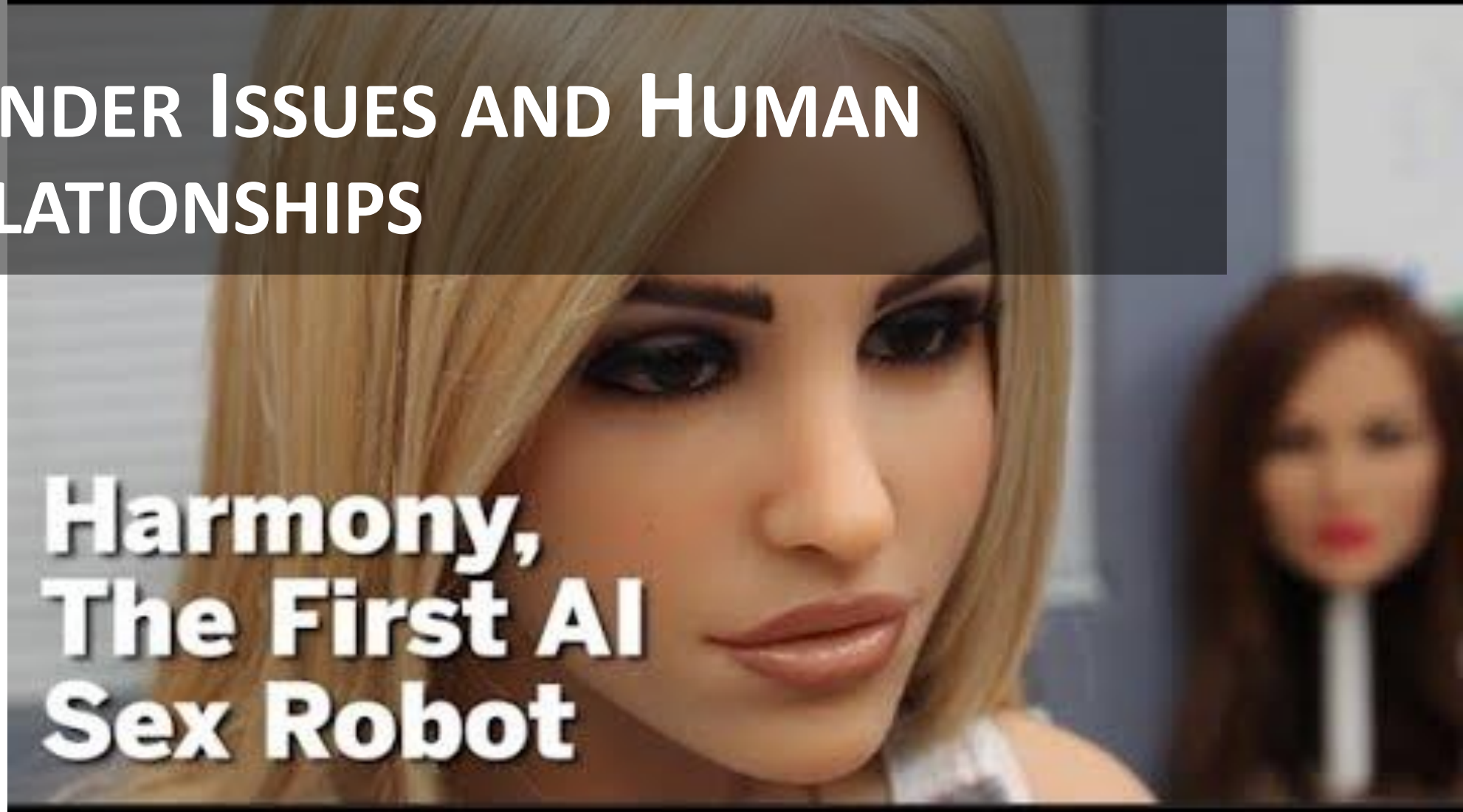
- Is bias avoidable? No, but we can explicitly discuss, analyze, and intervene (kind of bias, degree of bias)
- Ethical & political question is whether bias or discrimination is just/fair or not
- Algorithms teach us something about our societies (see also digital humanities: use AI!)

Gender Issues



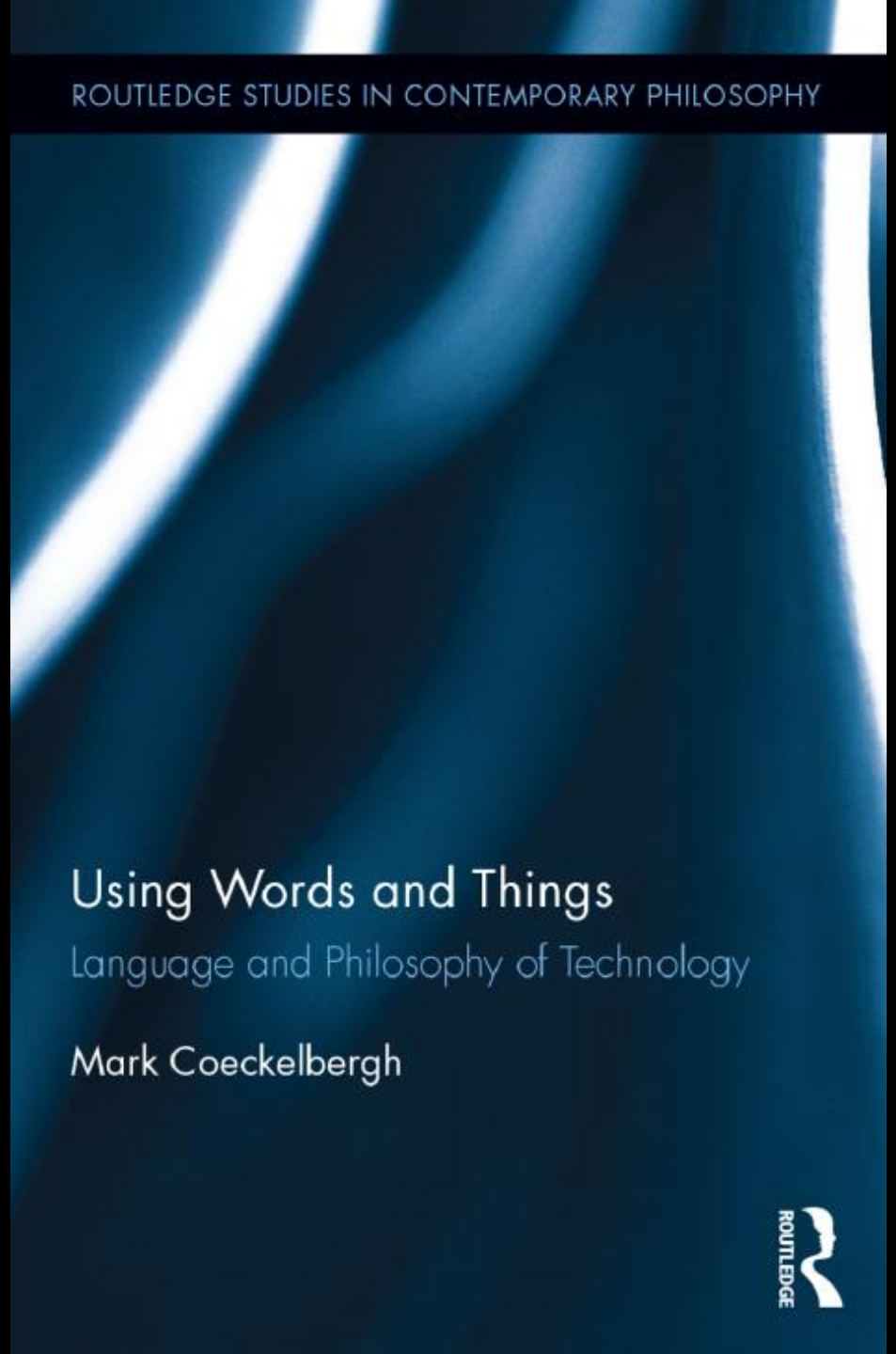
Example: AI uses data from internet text but there is gender bias in those texts and in our language (e.g. Bryson's work)

Gender Issues and Human Relationships




OCTOBER 25 - 26 2018
FEMINIST PHILOSOPHY OF TECHNOLOGY
https://philtech.univie.ac.at/
KEYNOTE SPEAKERS
CORINNA BATH
RICK DOLPHIJN
NINA LYKKE
KATHLEEN RICHARDSON
LUCY SUCHMAN
JUDY WAJCMAN


AI TECH IS EMBEDDED IN RELATED TO HUMAN ACTIVITIES AND EMBEDDED IN A WIDER SOCIAL AND CULTURAL CONTEXT, IN “TECHNOLOGY GAMES” AND A FORM OF LIFE


ROUTLEDGE STUDIES IN CONTEMPORARY PHILOSOPHY
Using Words and Things
Language and Philosophy of Technology
Mark Coeckelbergh
ROUTLEDGE

Ethics of AI: Approach

- Bottom up
- Pro-active
- Global
- Positive

Ethical & legal theory and principles

Experience – Practices


Ethical & legal theory and principles
Experience – Practices

Ethics and Regulation: Let's try to be Pro-Active

ETHICS: HOW NOT TO DO IT

REGULATION

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

Responsible Research and Innovation and Ethically aligned design

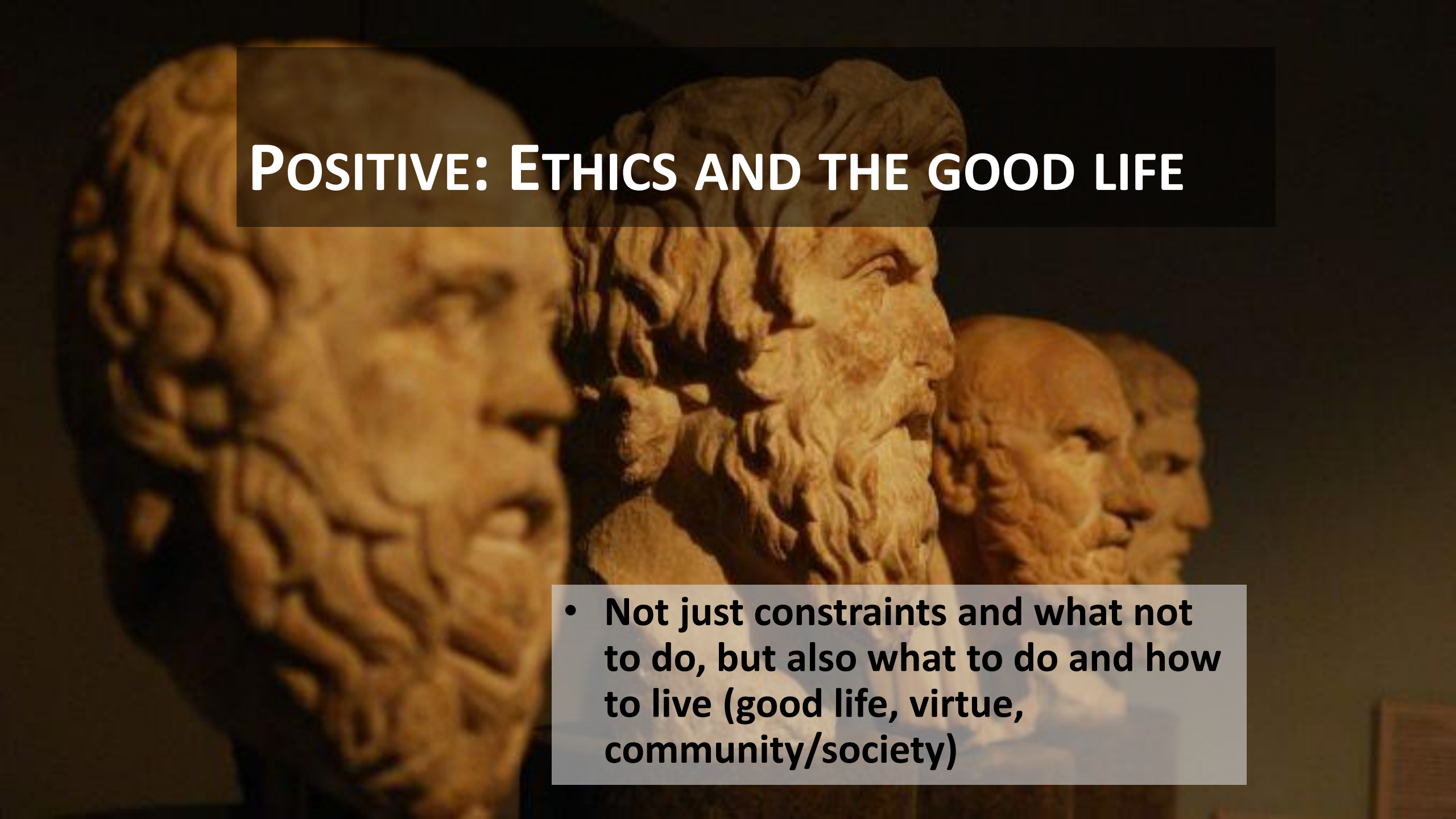
- Pro-active
- Stakeholders
- But: problems

SOME PROBLEMS

- How to translate from principles to practice? Problem of method and operationalization
- Power differences; democracy?
- No possibility to stop the technology

EUROPEAN BUT ALSO GLOBAL ACTION NEEDED

- Due to nature of new technologies
- Do we have suitable institutions for this? Or only big corporations who decide?



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)

EXPLORE HOW OTHER COUNTRIES AND CULTURES DEAL WITH AI ETHICS

- Cultural differences
- Different conceptions of the good life and the good society

INNOVATION, DESIGN, ART

- **Imagination needed, art *can* provide a different perspective**

POLICY

INDEPENDENT

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

SET UP BY THE EUROPEAN COMMISSION

AI ethics guidelines

- Aim is to promote trustworthy AI
  - Lawful
  - Ethical
  - Robust

AI ethics guidelines

- Approach
  - Human-centred
  - Prevent and minimize risks
  - Fundamental rights as a basis

AI ethics guidelines

- Ethical principles
  - respect for human autonomy
  - prevention of harm
  - fairness and explicability

AI ethics guidelines

- Trustworthy AI: 7 requirements:
  1. human agency and oversight
  2. technical robustness and safety
  3. privacy and data governance
  4. transparency
  5. diversity, non-discrimination and fairness
  6. environmental and societal well-being
  7. accountability

AI ethics guidelines

- Zoom in on non-discrimination and fairness:
  - Fairness:

    Avoidance of unfair bias. Data sets used by AI systems (both for training and operation) may suffer from the inclusion of inadvertent historic bias, incompleteness and bad governance models. The continuation of such biases could lead to unintended (in)direct prejudice and discrimination against certain groups or people, potentially exacerbating prejudice and marginalisation.(…)

    Identifiable and discriminatory bias should be removed in the collection phase where possible. The way in which AI systems are developed (e.g. algorithms' programming) may also suffer from unfair bias. This could be counteracted by putting in place oversight processes to analyse and address the system's purpose, constraints, requirements and decisions in a clear and transparent manner. Moreover, hiring from diverse backgrounds, cultures and disciplines can ensure diversity of opinions and should be encouraged.

AI ethics guidelines

- Methods to ensure implementation
  - Technical (architectures for trustworthy AI, testing, explanation methods, …)
  - Non-technical
    - Regulation
    - Codes of conduct
    - Standardization
    - Certification
    - Accountability via governance frameworks
    - Education and awareness
    - Stakeholder participation
    - Diverse and inclusive design teams

AI ethics guidelines

- Trustworthy AI Assessment List
  - Human agency and oversight
  - Technical robustness and safety
  - Privacy and data governance
  - Transparency (traceability, explainability , communication)
  - Diversity, non-discrimination, and fairness
  - Societal and environmental well-being
  - Accountability

AI ethics guidelines

- Trustworthy AI Assessment List
  - Example Explainability:

Did you assess: to what extent the decisions and hence the outcome made by the AI system can be understood? to what degree the system's decision influences the organisation's decision-making processes? why this particular system was deployed in this specific area? what the system's business model is (for example, how does it create value for the orga nisation)?

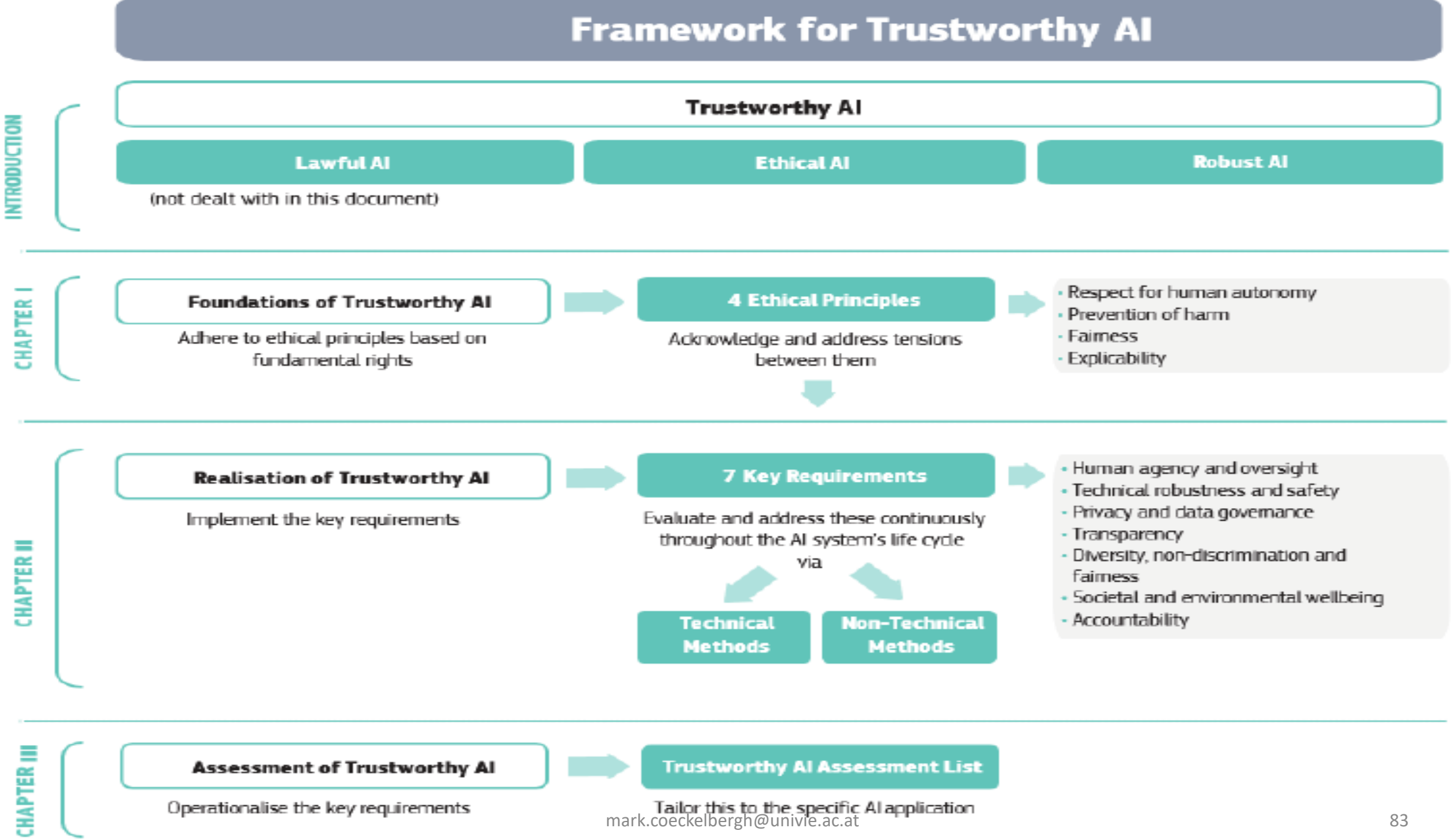
Did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand?

Did you design the AI system with interpretability in mind from the start?

Did you research and try to use the simplest and most interpretable model possible for the application in question?

Did you assess whether you can analyse your training and testing data? Can you change and update this over time?

Did you assess whether you can examine interpretability after the model's training and development, or whether you have access to the internal workflow of the model?


Framework for Trustworthy AI
INTRODUCTION
Trustworthy AI
Lawful AI
(not dealt with in this document)
Ethical AI
Robust AI
CHAPTER I
Foundations of Trustworthy AI
Adhere to ethical principles based on fundamental rights
4 Ethical Principles
Acknowledge and address tensions between them
Respect for human autonomy
Prevention of harm
Fairness
Explicability
CHAPTER II
Realisation of Trustworthy AI
Implement the key requirements
7 Key Requirements
Evaluate and address these continuously throughout the AI system's life cycle via
Technical Methods
Non-Technical Methods
Human agency and oversight
Technical robustness and safety
Privacy and data governance
Transparency
Diversity, non-discrimination and fairness
Societal and environmental wellbeing
Accountability
CHAPTER III
Assessment of Trustworthy AI
Operationalise the key requirements
Trustworthy AI Assessment List
Tailor this to the specific AI application

What’s going on now?

- Re-design assessment list
  - For this purpose asking feedback from stakeholders
    - Interviews
      - Quantitative
      - Deep dives selected organizations

Questions for policy makers

- What do do?
- How to do it?
- Who should do it?

What to do?

Morality: constraints, red lines, sactions

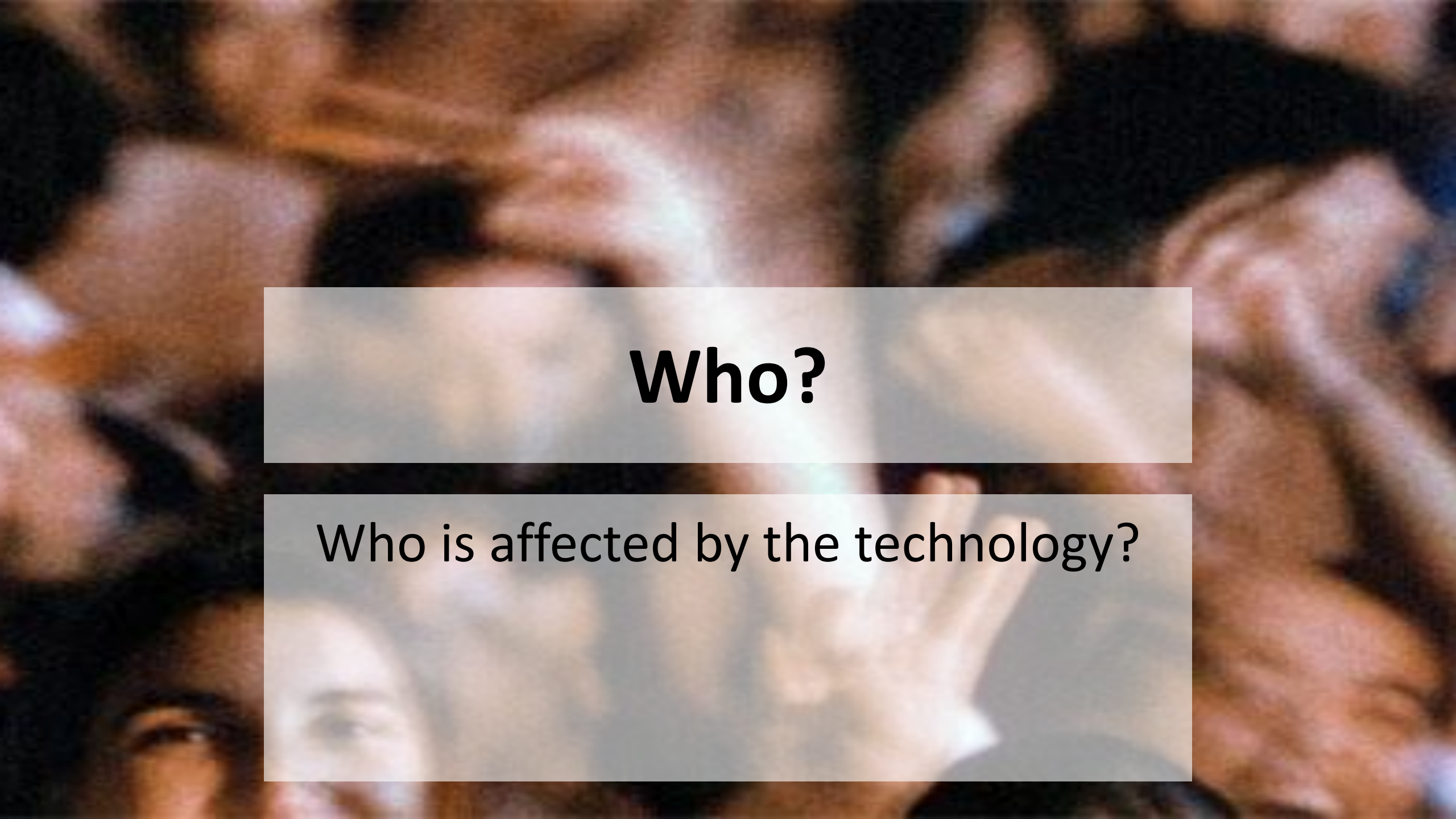
Ethics: the good life, the best life

+ the good society! (see article)

How?

HOW can we reach these goals? Also think about PROCESS

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



Who?

Who is affected by the technology?

Who?

Who should make the rules?

Who?

Experts, citizens, and mediators needed

Who?

Cultural differences (global, Europe)

Who?

Power differences (e.g. big companies)

Who?

What about non-humans?

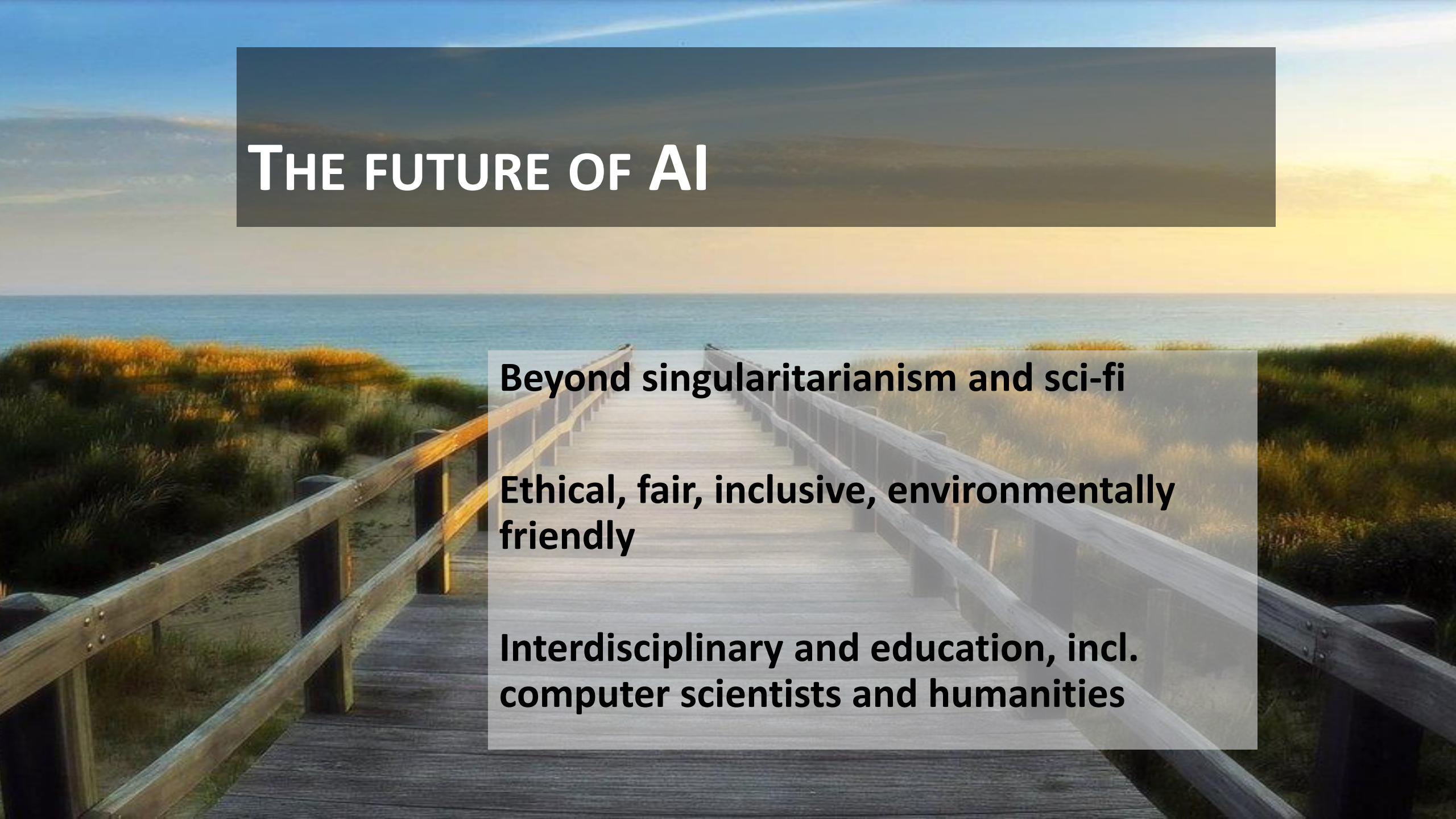
What about the environment?

Can AI "save the planet"?

- Policy is also about priorities: AI or climate change?
- Can AI help to deal with climate change? Or does it make things worse?
  - AI can help us to deal with complex problems
  - But may also reflect a problematic attitude towards the earth and the planet (see also discussion about the Anthropocene)

Conclusion: some barriers to good policy making

- Too much focused on principles, too little work on methods and operationalization
- How democratic is the decision-making really?
- Lack of sufficient interdisciplinary and transdisciplinary expertise
  - Importance of education!
- Lack of discussion about priorities
- Sufficient global action?



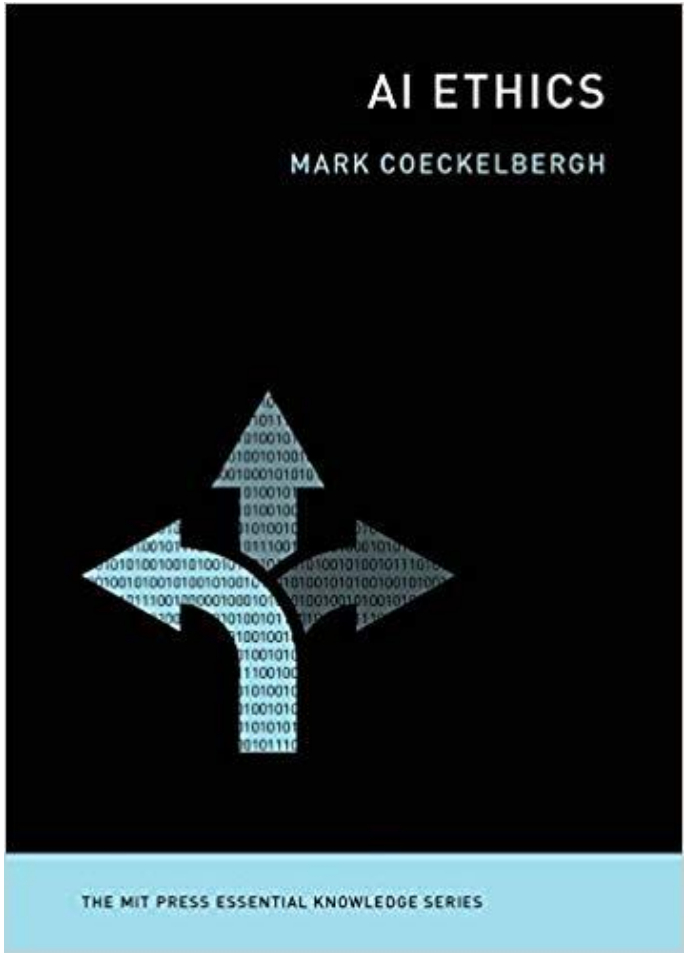
The future of AI

Beyond singularitarianism and sci-fi

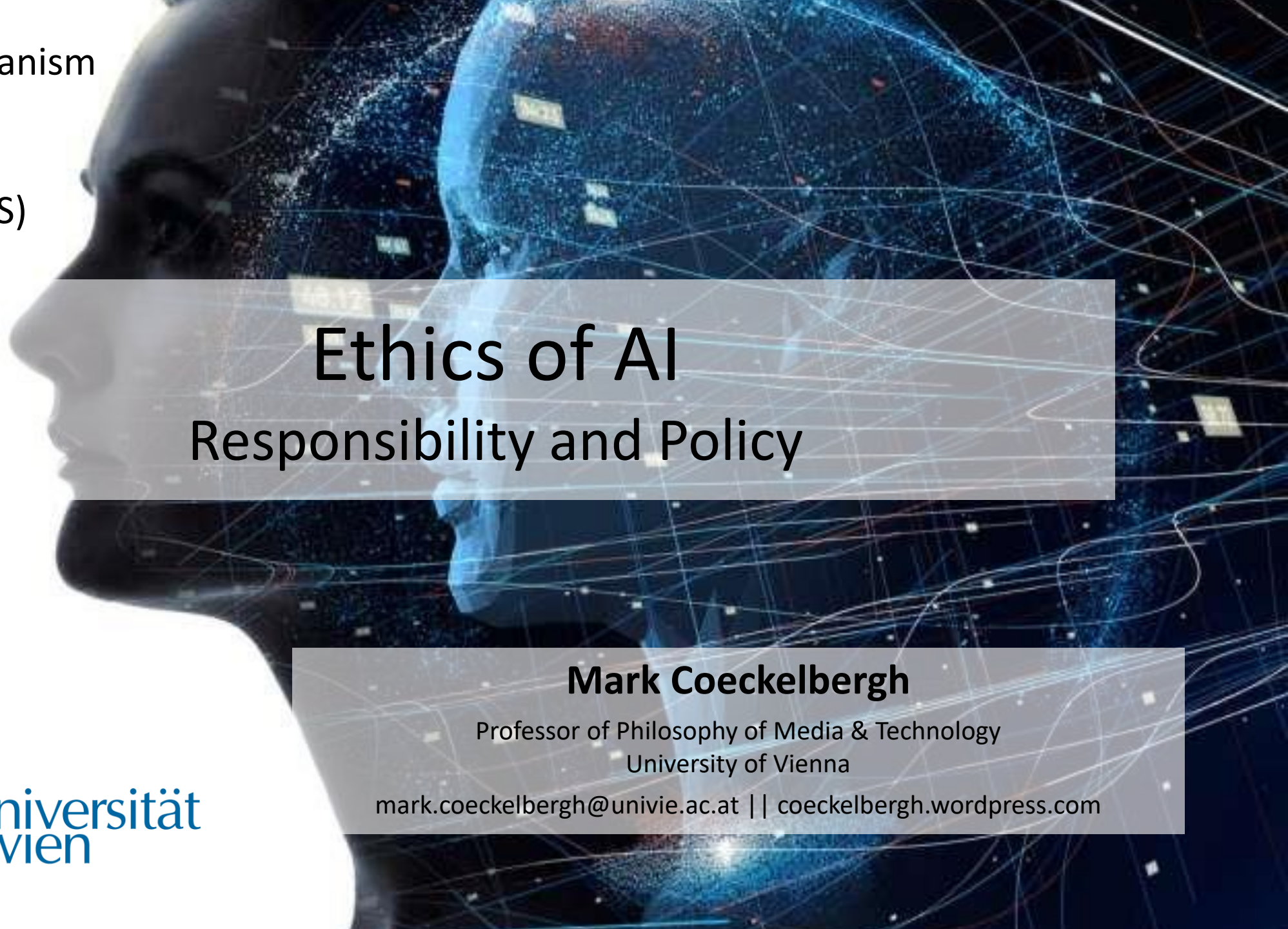
Ethical, fair, inclusive, environmentally friendly

Interdisciplinary and education, incl. computer scientists and humanities

Forthcoming: *AI Ethics* (MIT Press)



Talk for Beyond Humanism conference
Korea Institute for Advanced Study (KIAS)



Ethics of AI

Responsibility and Policy

Mark Coeckelbergh

Professor of Philosophy of Media & Technology
University of Vienna

mark.coeckelbergh@univie.ac.at || coeckelbergh.wordpress.com

universität wien