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 || coeckelbergh.wordpress.com
PHILOSOPHY OF TECHNOLOGY

Philosophy
Interdisciplinarity
Policy
Robophilosophy 2018
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.
Thinking about and for technology, but also using technology to think about philosophical issues
PHILOSOPHY OF TECHNOLOGY

Focus: robots and AI
PHILOSOPHY OF TECHNOLOGY

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?

Negatieve anthropology: wat the human is NOT e.g. not a chimp
What is the human?

Negatieve 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)
“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

NEW ROMANTIC CYBORGS

ROMANTICISM, INFORMATION TECHNOLOGY, AND THE END OF THE MACHINE

MARK COECKELBERGH
AGAINST ALARMISM

KEEP CALM AND DO YOUR HOMEWORK
Urgent issues near future
INDUSTRY
IN THE OFFICE
FINANCE
TRANSPORT
HEALTH CARE
MILITARY APPLICATIONS
ALL THINGS - EVERYWHERE
Changes to our Daily Lifes
ETHICAL AND LEGAL 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?

- Robot/AI - human teams
- Degrees of autonomy
- Distributed agency
MORAL AND LEGAL RESPONSIBILITY

• 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?
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?
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

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.
Case: Fatal accident

• Uber self-driving car in autonomous mode causes accident in Arizona: pedestrian dies (March 2018)

• See also 2016 Tesla accident
Case: Fatal accident


- 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?
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?

Self-driving Uber kills first fatal crash involving pedestrian...
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 machines
Moral Status of AIs/Robots

Philosophically interesting, but also practical issue?

Robot Citizen
Technology changes Morality

• Privacy today
• How will AI and robotics change our values?
VULNERABLE USERS, ATTACHMENT AND DECEPTION
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
Drones, information technology, and distance: mapping the moral epistemology of remote fighting

Mark Cresswell

In Introduction

When on August 6, 1945 at 8:15 AM Hiroshima Frieda Gay dropped an atomic bomb on the city of Hiroshima, the carnage was instantaneous. Seeing little and a mushroom-shaped cloud, someone on the street exclaimed, “It looks like the blackest day of the world.” What they didn’t see or feel was, in fact, and how far the explosion killed, approximately 70,000 people and some in the following years. They didn’t see the skin of other victims was burning and burning. They didn’t see people that looked “like walking ghosts” in a manner described here. They didn’t see the suffering of death, of pain, sorrow, and children. They must have felt how their bodies were lit by the shock of the explosion, for sure, but their experience of the effects on them was most likely not very different from the of Captain Yonosuke Shimada when he was using the bomb. I know it’s quite a bit to ask, but I’m certain that it is possible to imagine some ways of technology-mediated fighting.

Keywords: Military ethics - Drones - Ethics - Distance - Information technology - Nuclear weapons
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 computernisation? 

Carl Benedikt Frey, Michael A. Osborne

ARTICLE INFO

Abstract

We examine whether there is going to be a large scale displacement of workers that can be explained by computernisation. To do this, we begin by estimating the impact of computernisation on the probability of computernisation for 702 detailed occupations, using a logistic model. Then, using a weighted sum of these estimates we examine whether computernisation is likely to affect occupational employment levels. We find that computernisation will result in a dislocation of workers from less computernisable occupations to more computernisable ones. The main conclusion is that computernisation is likely to result in a significant reduction in employment levels for many occupations and an increase in employment levels for others.

INTRODUCTION

The rise of artificial intelligence (AI) has raised concerns about the potential displacement of workers. AI technologies are increasingly being used in a wide range of industries, from manufacturing to healthcare. This raises questions about the future of employment and the potential impact on workers.

METHODS

To estimate the impact of AI on employment, we used a logistic regression model to predict the probability of an occupation being computernisable. We then used these estimates to calculate the potential impact on employment levels.

RESULTS

Our results indicate that AI is likely to result in a significant reduction in employment levels for many occupations. This finding is consistent with previous research on the potential impact of automation on employment.

CONCLUSIONS

The rise of AI is likely to result in a significant dislocation of workers from less computernisable occupations to more computernisable ones. This has important implications for policymakers, who will need to develop strategies to support workers during this transition.

DATA

The data used in this study was sourced from various online sources, including the World Economic Forum's occupational classification system.

ACKNOWLEDGEMENTS

We would like to thank the European Economic and Social Committee for their support in conducting this research.

REFERENCES

[List of references]

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!)
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
GENDER ISSUES AND HUMAN RELATIONSHIPS

Harmony, The First AI Sex Robot
FEMINIST PHILOSOPHY OF TECHNOLOGY

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

KEYNOTE SPEAKERS

CORINNA BATH
RICK DOLPHIJN
NINA LYKKE
KATHLEEN RICHARDSON
LUCY SUCHMAN
LUDY WAIGMAN
ETHICS: APPROACH

- Bottom up
- Pro-active
- Global
- Positive
Ethical & legal theory and principles

Experience – Practices
‘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.
Ethical & legal theory and principles

Experience – Practices
ETHICS AND REGULATION: LET’S TRY TO BE PRO-ACTIVE
ETHICS: HOW NOT TO DO IT
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!
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 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
Beyond fear

Ethical

Interdisciplinary, incl. humanities

Connected to wider society

Europe: expertise in tech ethics
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 || coeckelbergh.wordpress.com