

The Ethics of AI

People, Policy, Politics, Society, Technology



Roberto V. Zicari

With contributions from: Irmhild van Halem, Matthew Eric Bassett, Karsten Tolle, Timo Eichhorn, Todor Ivanov, Jesmin Jahan Tithi (*)

Frankfurt Big Data Lab, Germany

www.bigdata.uni-frankfurt.de

(*) Intel Labs, USA

November 26, 2019

THE AUTOMATORS - Revolution of Workforce
Frankfurt School of Finance & Management

© 2019 by Roberto V. Zicari and his colleagues

The content of this presentation is open access distributed under the terms and conditions of the Creative Commons (**Attribution-NonCommercial-ShareAlike** CC BY-NC-SA) license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

AI holds a mirror up to humanity



Image <https://sites.google.com/site/learnvocabinieltsreading/home/first-step-to-ielts-practice-test/reflecting-on-the-mirror>


Approaching Ethical Boundaries



“But if we just let machines learn ethics by observing and emulating us, they will learn to do lots of unethical things.

So maybe AI will force us to confront what we really mean by ethics before we can decide how we want AIs to be ethical.” ()*

--Pedro Domingos (Professor at University of Washington)

 (*) Source: **On Artificial Intelligence, Machine Learning, and Deep Learning. Interview with Pedro Domingos**, ODBMS Industry Watch, June 18, 2018

The Ethics of Artificial Intelligence



- ❧ AI is becoming a sophisticated tool in the hands of a variety of stakeholders, including political leaders.
- ❧ Some AI applications may raise new **ethical** and **legal** questions, and in general have a significant impact on **society** (for the good or for the bad or for both).
- ❧ **People motivation** plays a key role here.

Examples of AI Projects (Government)



- ❧ **Detecting movement from drone images**
- ❧ **Classifying objects from national archive**
- ❧ **Maintenance planning in national archive**
- ❧ **Deciding emergency level and needs from 911 calls**
- ❧ **Tax evasion and likelihood of return payment**
- ❧ **Traffic accident prediction**
- ❧ **Animal movement and species mapping**
- ❧ **Neurotheater**
- ❧ **Process mapping and redesign on national portal**
- ❧ **Personalised teaching in primary school**
- ❧ **Automating internal work processes related to customer support**
- ❧ **Chatbot**
- ❧ **Monitoring price manipulation**

Ministry of Economic Affairs and Communications (Europe)

Personalised Healthcare

Digital Biomarkers



*“...our application of mobile and sensor technology to monitor symptoms, disease progression and treatment response – the so called “**Digital Biomarkers**”.*

We have our most advanced programmes in Multiple Sclerosis (MS) and Parkinson`s Disease (PD), with several more in development. Using these tools, a longitudinal real-world profile is built that, in these complex syndromes, helps us to identify signals and changes in symptoms or general living factors, which may have several potential benefits.”

– Bryn Roberts

Global Head of Operations for [Roche Pharmaceutical Research & Early Development](#)



Source: On using AI and Data Analytics in Pharmaceutical Research. Interview with Bryn Roberts ODBMS Industry Watch, September 10, 2018



Do no harm.
Self Driving Cars

Ω

Let`s consider an autonomous car that relies entirely on an algorithm that had taught itself to drive by watching a human do it.

What if one day the car crashed into a tree, or even worse killed a pedestrian?

The Uber Case for *False positive* for plastic bags...



2018: „The newsletter "The Information" has reported a leak from Uber about their fatal accident. The relevant quote:

The car's sensors detected the pedestrian, who was crossing the street with a bicycle, but Uber's software decided it didn't need to react right away. **That's a result of how the software was tuned.** Like other autonomous vehicle systems, Uber's software has the **ability to ignore "false positives,"** or objects in its path that wouldn't actually be a problem for the vehicle, such as a plastic bag floating over a road. In this case, Uber executives believe the company's **system was tuned so that it reacted less to such objects.** But the tuning went too far, and the car didn't react fast enough, one of these people said." (*)

(*) *How reliable is this Source?* : <https://ideas.4brad.com/uber-reported-have-made-error-tuning-perception-system>

Story also in *Der Spiegel* Nr. 50/8.12.2018 *Tod durch Algorithms* (Philipp Oehmke)

Who is responsible?



2019: Article which reports the results of the "investigation" done by the US National Transportation Safety Board **after** the Uber crashed last year killing a homeless crossing a highway.

It seems that one of the main problems were how the **AI was designed and how the training data was chosen.**

Moreover, the **human in the loop**, who was supposed to control the AI, was not paying attention (watching a video on her smartphone...)

Source: <https://www.wired.com/story/ubers-self-driving-car-didnt-know-pedestrians-could-jaywalk/>



Do no harm
Can we explain decisions?



What if the *decision* made using AI-driven algorithm *harmed* somebody, and you *cannot explain* how the decision was made?

☞ This poses an ethical and societal problem.

Another kind of Harm



"Big Nudging"

He who has large amounts of data can manipulate people in subtle ways.

But even benevolent decision-makers may do more wrong than right. ()*

(*) Source: *Will Democracy Survive Big Data and Artificial Intelligence?*. Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R. V., & Zwitter, A.. (2017). *Scientific American* (February 25, 2017).

Who is Responsible?



*“Who will decide what is the impact of AI
on Society?”*

Policy Makers and AI



*“**Citizens and businesses** alike need to be able to **trust** the technology they interact with, and have effective safeguards protecting fundamental rights and freedoms.*

*In order to increase **transparency** and **minimise the risk of bias**, AI systems should be developed and deployed in a manner that allows humans to **understand** the basis of their actions.*

***Explainable AI** is an essential factor in the process of strengthening people’s trust in such systems.” (*)*

*-- **Roberto Viola** Director General of DG CONNECT (Directorate General of Communication Networks, Content and Technology) at the **European Commission**.*

(*) Source [On the Future of AI in Europe. Interview with Roberto Viola](#), ODBMS Industry Watch, 2018-10-09

Mindful Use of AI



We are all responsible.

The individual and collective conscience is the existential place where the most significant things happen.

So, my claim is that we are all responsible to look at the Ethical aspects and of AI and that ethics should be (possibly) considered upfront.

Closing the Gap



“Most of the principles proposed for AI ethics are not specific enough to be action-guiding.”

“The real challenge is recognizing and navigating the tension between principles that will arise in practice.”

“ Putting principles into practice and resolving tensions will require us to identify the underlying assumptions and fill knowledge gaps around technological capabilities, the impact of technology on society and public opinion” . ()*

(*)Whittlestone, J et al (2019) Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. London: Nuffield Foundation.

Formulating universal AI principles?



“ Given different cultural traditions, philosophers could spend many lifetimes debating a set of universal AI principles”

-- John Thornhill. (*)

(*) *Formulating AI values is hard when human fail to agree*, John Thornhill, Financial Times, July 22, 2019

The Politics of AI *Ecosystems*



- ❧ The Rise of (Digital) Ecosystems paving the way to disruption.^(*)
- ❧ Different Countries, Different Approaches, Cultures, Political Systems, and Values (e.g. China, the United States, Russia, Europe,...)

^(*) Source: Digital Hospitality, Metro AG-personal communication.

The Politics of AI

AI made in China



In China, questions about ethics unlike in most democracies, are not framed around the individual but instead the collective ()*

(*) China's Techno-Utilitarian Experiments with Artificial Intelligence, Dev Lewis, Digital Asia Hub ,2019

The Politics of AI

Big Data and AI made in China



☞ *Facial recognition makes up 35% of all AI applications in China.*

e.g. Sensetime (商汤科技), Megvii Face++, Yitu (*)

The Politics of AI China



*Spotlight on China: Is this what the Future of
Society looks like?*

*How would behavioural and social control impact
our lives?*

*The concept of a Citizen Score, which is now being
implemented in China, gives an idea (*).*

(*) Source: *Will Democracy Survive Big Data and Artificial Intelligence?*. Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R. V., & Zwitter, A. *Scientific American* (February 25, 2017).

What Practitioners Need



Need for ethical frameworks and case studies



- ☞ “ Several interviewees suggested it would be helpful to have access to domain-specific resources, such as **ethical frameworks and case studies**, to guide their teams’ ongoing efforts around **fairness**”
- ☞ 55% of survey respondents indicated that having access to such resources would be at least “Very” useful (*)
- ☞ (*) **Based on 35 semi-structured interviews and an anonymous survey of 267 ML practitioners in USA.** Source: Improving Fairness in Machine Learning Systems: What Practitioners Need? K. Holstein et al. CHI 2019; May 4-0, 2019

Need for More Holistic Auditing Methods



“Interviewers working on applications involving richer, complex interaction between the user and the system bought up needs for more *holistic*, system-level **auditing methods**.” (*)

(*) source: Improving Fairness in Machine Learning Systems: What Practitioners Need? K. Holstein et al. CHI 2019; May 4-0, 2019

Need for Metrics, Processes and Tools



☞ “Given that *fairness* can be highly context and application dependent, there is an **urgent need for domain-specific educational resources, metrics, processes and tools** to help practitioners navigate the unique challenges that can arise in their specific application domains” (*)

☞ (*) source: Improving Fairness in Machine Learning Systems: What Practitioners Need? K. Holstein et al. CHI 2019; May 4-0, 2019

Z-inspection

A process to assess Ethical AI



Why doing an AI Ethical Inspection?



There are several reasons to do an AI Ethical Inspection:

- ❧ *Minimize Risks* associated with AI
- ❧ *Help establishing “TRUST”* in AI
- ❧ *Improve the AI*
- ❧ *Foster ethical values and ethical actions*
(stimulate new kinds of innovation)

Help contribute to closing the gap between “*principles*” (the “what” of AI ethics) and “*practices*” (the “how”).

Z-Inspection: *Areas of investigations*



We use *Conceptual clusters* of:

Bias /*Fairness*/ Discrimination

Transparencies /*Explainability*/ Intelligibility/Interpretability

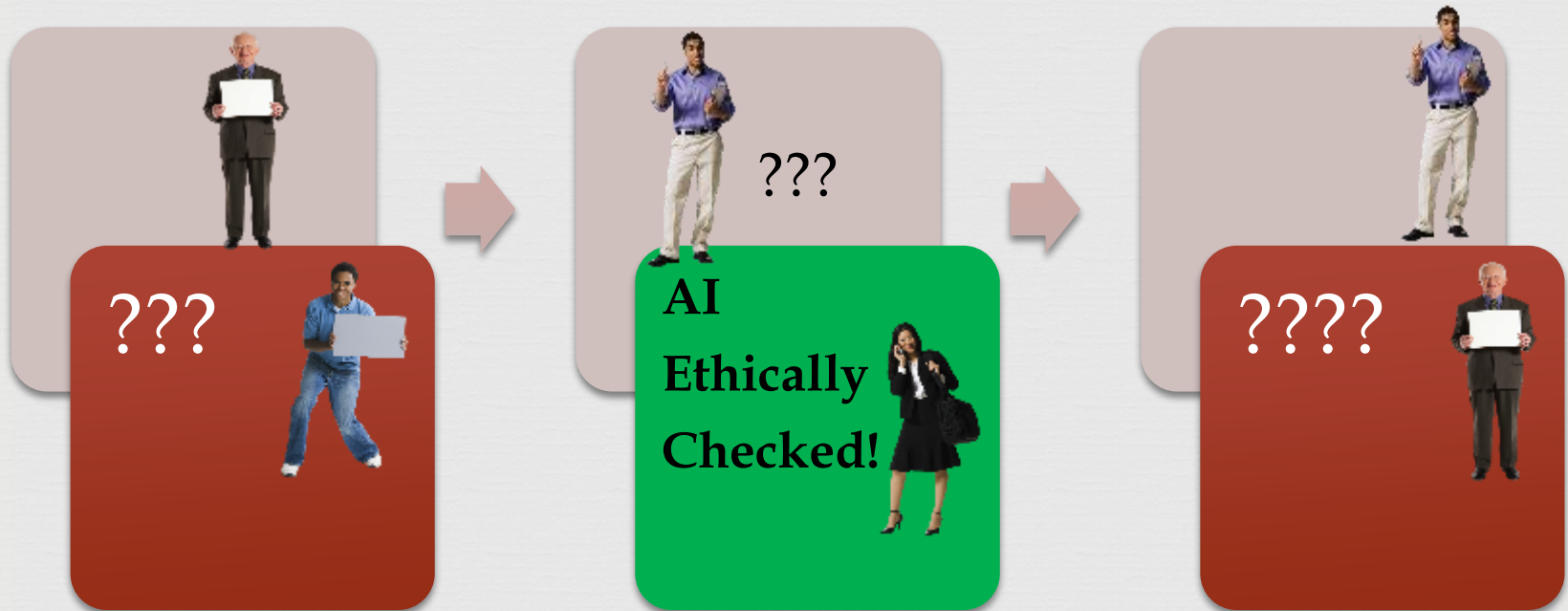
Privacy/ Responsibility/*Accountability*

Safety

Human-AI

- Other (for example chosen from this list):
 - uphold human rights and values;
 - promote collaboration;
 - **acknowledge legal and policy implications;**
 - avoid concentrations of power,
 - contemplate implications for employment.

Micro-validation does not imply Macro-validation




Assessing fairness (Bias / Discrimination)



“Clarifying what kind of algorithmic “fairness” is most important is an important first step towards deciding if this is achievable by technical means” ()*

Identify Gaps/Mapping conceptual concepts between:

1. *Context-relevant Ethical values,*

2. *Domain-specific metrics,*

3. *Machine Learning fairness metrics.*

(*) Source: Whittlestone, J et al (2019) *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research.* London: Nuffield Foundation.

Z-inspection: Trade offs



- ❧ **Appropriate use:** Assess if the data and algorithm are appropriate to use for the purpose anticipated and perception of use.
 - ❧ Suppose we assess that the AI is technically *unbiased* and *fair* –this does not imply that it is acceptable to deploy it.
- ❧ **Remedies:** If risks are identified, define ways to mitigate risks (when possible)
- ❧ **Ability to redress**

Assessing Ethical AI Use case Health Care



Assessing



“The first highly accurate and non-invasive test to determine a risk factor for coronary heart disease.

Easy to use. Anytime. Anywhere.” ()*



(*) Source: <https://cardis.io>



Cardisio: Socio-technical scenario

The Domain



- ❧ *Coronary angiography* is the reference standard for the detection of **stable coronary artery disease (CAD)** at rest (invasive diagnostic 100% accurate)
- ❧ **Conventional non-invasive diagnostic** modalities for the detection of stable coronary artery disease (CAD) at rest are subject to significant limitations: low sensitivity, local availability and personal expertise.
- ❧ Latest experience demonstrated that **modified vector analysis** possesses the potential to overcome the limitations of conventional diagnostic modalities in the screening of stable CAD.

Cardisio: Socio-technical scenario

Cardisiography



- ❧ *Cardisiography (CSG)* is a denovo development in the field of applied vectorcardiography (introduced by Sanz et al. in 1983) using Machine Learning algorithms.
- ❧ **Design:** By applying standard electrodes to the chest and connecting them to the Cardisiograph, CSG recording can be achieved.
- ❧ **Hypothesis:** „By utilizing computer-assisted analysis of the **electrical forces** that are generated by the heart by means of a continuous series of vectors, abnormalities resulting from impaired repolarization of the heart due to impaired myocardial perfusion, it is **hypothesized that CSG is an user-friendly screening tool for the detection of stable coronary artery disease (CAD).**”

Cardisio: Socio-technical scenario

Actions taken based on model`s prediction



- ❧ Patients received “Green” score (*continuous prediction: dark to light Green*). Doctor agree. Patient does nothing;
- ❧ Patients received “Green” (*continuous prediction*). Patient and/or Doctor do not trust, asked for further invasive test;
- ❧ Patient received “Red” (*continuous prediction: dark to light Red*). Doctor agree. Patient does nothing;
- ❧ Patient received “Red” (*continuous prediction*). Doctor agree. Patient asks for further invasive test;
- ❧

In any of the above cases, Patient and/or Doctor may ask for an *explanation*.

Cardisio: Socio-technical scenario

Discover potential ethical issues



Overall, from an ethical point of view the chances that more people with an undetected serious CAD problem will be diagnosed in an early stage need to be weighted against the risks and cost of using the CSG app.

Cardisio: Socio-technical scenario

Discover potential ethical issues



Diagnostic Trust and Competence - ethical issues:

- ❧ When CSG is being used in screening un-symptomatic patients who are “*notified*” by Cardisio with a “minor” CAD problem that might not impact their lives, **they might get worried- change their lifestyles after the *notification* even though this would not be necessary**
- ❧ If due to the CSG test more patients with minor CAD problems are being “notified” and sent to cardiologists, **this might result in significant increase of health care costs, due to further diagnostics tests.**

Cardisio: Socio-technical scenario

Discover potential ethical issues



Diagnostic Trust and Competence - ethical issues:

- ❧ Using a black-box algorithm **might impair the trust of the doctor in the diagnostic app**, especially if the functioning of the app / algorithm has not been verified by independent studies.
- ❧ Using an AI assisted diagnostic app **could in the long-term impair the diagnostic competence of the medical personal** and also the quality of the diagnostic process when more “physician assistance” instead of medical doctors do the diagnostic “ground work”.
- ❧ **The doctor’s diagnostic decision might become biased** by the assumed “competence” of AI - especially when the doctor’s and the AI’s diagnosis differ.
- ❧ **How high is the risk that an application/diagnostic error happens** with the traditional diagnostic instruments compared to using the CSG app?

What if the Z-inspection happens to be false or inaccurate?



- ❧ There is a danger that a *false* or *inaccurate* inspection will create natural skepticism by the recipient, or even harm them and, eventually, backfire on the inspection method.
- ❧ This is a well-known problem for all quality processes. It could be alleviated by an open development and incremental improvement to establish a process and brand (like “*Z Inspected*”).

AI, Ethics, Democracy



Do we want to assess if the *Ecosystem(s)* where the AI has been designed/produced/used is *Democratic*?

Is it Ethical?

Is it part of an AI Ethical Inspection or not?

For more information



✂ The Ethics of Artificial Intelligence

www.bigdata.uni-frankfurt.de/ethics-artificial-intelligence/