DATA ETHICS

Data Challenge 2019

Roberto V. Zicari

"*Values*"

How societies approach Tech and the *values* it wants to build into the new technology?

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.

*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.

Do no harm.
Self Driving Cars

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/

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.

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).

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.

- *Source: http://www.bigdata.uni-frankfurt.de/ethics-artificial-intelligence/*

AI Ethical Inspection: Benefits

"If **governments** deploy AI systems on human populations without framework for accountability, they risk losing touch with how decisions have been made, thus making it difficult for them to identify or respond to bias, errors, or other problems. The public will have less insight into how agencies function, and have less power to question or appeal decisions."

An Ethical assessment "would also benefit **vendors (AI developers)** that prioritize fairness, accountability, and transparency in their offering. Companies that are best equipped to help agencies and researchers study their system would have a competitive advantage over others. Cooperation would also help improve public trust, especially at a time when skepticism of the societal benefits of AI is on the rise."

Source: Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability, AI NOW, April 2018.

Ethics in Practice

"*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.

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
Big Data and AI made in USA

The tech giants, such as Amazon, Apple, Facebook, Google, developed their own Ecosystems.

The Politics of AI
Ethics and Democracy

Democracy has substantial ethical content.

~Charner Perry (*)

(*) Source Ethics and Democracy, Charner Perry, *Ethics* Vol. 83, No2, (Jan. 1973), pp-87-107, The University pf Chicago Pres

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 (*)

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

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.. (2017). Scientific American (February 25, 2017).

Conceptual Clusters

- 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 implication**s;
  - avoid concentrations of power,
  - contemplate implications for employment.

Discover potential ethical issues

- Use **Socio-technical scenarios** to describe the *aim of the* system, the *actors and their expectations*, the *goals of actors´ action*, the *technology* and the *context*. (*)
  - What kind of **ethical challenges** the deployment of the AI in the **life of people** raises;
  - Which **ethical principles** are appropriate to follows;
  - What kind of **context-specific values and design principles** should be embedded in the design outcomes.
- **Mark possible ethical issues as** FLAGS!
- **Socio-technical scenarios and the list of FLAGS! are constantly revised and updated.**

(*) source: Ethical Framework for Designing Autonomous Intelligent Systems. J Leikas et al. J. of Open Innovation, 2019, 5, 1

Concept Building

- *Mapping and clarifying ambiguities*
- *Bridging disciplines, sectors, publics and cultures*
- *Building consensus and managing disagreements*

Source: Whittlestone, J et al (2019

Developing an evidence base

- Understand technological capabilities and limitations
- Build a stronger evidence base on the current uses and impacts (*domain specific*)
- Understand the perspective of different members of society

Source: Whittlestone, J et al (2019)

Identify *Tensions*

- ***Identifying Tensions*** (*different ways in which values can be in conflict*), e.g.
  - **Accuracy vs. fairness**

    e.g. *An algorithm which is most accurate on average may systematically discriminate against a specific minority.*

    Using algorithms to make decisions and predictions more accurate versus ensuring fair and equal treatment

  - **Accuracy vs explainability**

    e.g Accurate algorithm (e.g. deep learning) but not explainable (degree of explainability)
  - **Privacy vs. Transparency**
  - **Quality of services vs. Privacy**
  - **Personalisation vs. Solidarity**
  - **Convenience vs. Dignity**
  - **Efficiency vs. Safety and Sustainability**
  - **Satisfaction of Preferences vs. Equality**

Source: Whittlestone, J et al (2019)

Address, Resolve *Tensions*

- ***Resolving Tensions*** (Trade-offs)
  - ***True ethical dilemma*** - the conflict is inherent in the very nature of the values in question and hence cannot be avoided by clever practical solutions.
  - ***Dilemma in practice***- the tension exists not inherently, but due to our current technological capabilities and constraints, including the time and resources we have available for finding a solution.
  - ***False dilemma*** - situations where there exists a third set of options beyond having to choose between two important values.

- ***Trade-offs***: How should trade-off be made?

Source: Whittlestone, J et al (2019)

List of potential ethical issues

- The outcome of the analysis is a list of potential ethical issues, which need to be further deliberated when assessing the design and the system`s goal and outcomes. (*)

(*) source: Ethical Framework for Designing Autonomous Intelligent Systems. J Leikas et al. J. of Open Innovation, 2019, 5, 1

Fairness: *Different definitions*

- Suppose we are concerned with whether an algorithm used to make healthcare decision is *fair* to all patients. *Different definitions*, e.g.
  - *Egalitarian* concept of fairness: *assess if the algorithm produces equal outcomes for all users (or all "relevant" subgroups)*
  - *Minimax* concept of fairness: *ensure the algorithm results in the best outcomes for the worst off user group.*

Source: Whittlestone, J et al (2019)

Example: Fairness/*Bias*

- *Differences between disciplines*, e.g.
  - "Biased sample" (Statistics)
  - "Bias" –negative attitude/ prejudices towards a particular group (Law, Social Psychology)
  - A dataset which is "unbiased" (in the statistical sense) may nonetheless encode common biases (in the social sense) towards certain individuals or social groups (*)

(*) source: Whittlestone J (2019)

Example: Fairness/*Discrimination*

- *Unfairness* due to *discrimination*

- Normative (prescriptive) definition of Fairness (non-discriminatory)

Vs.

- Descriptive (comparative*): Human Perception* of Fairness

Q. Is it fair to use a feature in a given decision making scenario?

Fairness Disagreements

Source: Human Perception of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction, N. Grgic-Hlaca,et al. WWW2018

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

Clinical Medical Ethics in the context of Ecosystems

The four classical principles of *Western* clinical medical ethics (*):

- Justice
- Autonomy
- Beneficence
- Nonmaleficence

Where "*Western*" define a set of implicit *ecosystems*...

(*) Source. Alvin Rajkomar et al. (2018)

ML and *Fairness* criteria in healthcare

(domain specific)

Using for example ***Distributive justice*** **(from philosophy and social sciences) options for machine learning (*)**

Possible Mitigation (*Fairness* criteria)

Equal Outcomes

Performance

Equal Allocation

Could we use other another criteria? e.g **Kaldor–Hicks criterion:**

This criterion is used in welfare economics and managerial economics

to argue that it is justifiable for society as a whole to make some worse off if this means a greater gain for others.

(*) Source. Alvin Rajkomar et al. Ensuring, Fairness in Machine Learning to Advance Health, Equity, Annals of Internal Medicine (2018). DOI: 10.7326/M18-1990

Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6594166/

ML Bias in healthcare

(domain specific)

- **Biases in model design**
  - *Labels bias, Cohort bias*
- **Biases in training data**
  - *Minority bias*
  - *Missing Data bias*
  - *Informativeness bias*
  - *Training-serving skew*
- **Biases in interactions with clinicians** *(domain specific)*
  - *Automation bias*
  - *Feedback Lops*
  - *Dismissal bias*
  - *Allocation discrepancy*
- **Biases in interactions with patients** *(domain specific)*
  - *Privilege bias*
  - *Informed mistrust*
  - *Agency bias*

Source. Alvin Rajkomar et al. Ensuring, Fairness in Machine Learning to Advance Health, Equity, Annals of Internal Medicine (2018). DOI: 10.7326/M18-1990

Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6594166/

From Domain Specific to ML metrics

- Different interpretations/definitions of *fairness* pose different requirements and challenges to Machine Learning (metrics) !

Mapping Domain specific "Fairness" to Machine Learning metrics

Several Approaches: **Individual fairness , Group fairness, Calibration, Multiple sensitive attributes, casuality**.(*).

In Models : Adversarial training, constrained optimization. regularization techniques,....(*)

Resulting Metrics	**Formal "non-discrimination" criteria**
Statistical parity	Independence
Demographic parity (DemParity) (average prediction for each group should be equal)	Independence
Equal coverage	Separation
No loss benefits	
Accurate coverage	
No worse off	
Equal of opportunity (EqOpt) (comparing the false positive rate from each group)	Separation
Equality of odds (comparing the false negative rate from each group)	Separation
Minimum accuracy	
Conditional equality,	Sufficiency
Maximum utility (MaxUtil)	

(*) Source *Putting Fairness Principles into Practice: Challenges, Metrics, and Improvements*

Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Allison Woodruff, Christine Luu, Pierre Kreitmann, Jonathan Bischof, Ed H. Chi (Submitted on 14 Jan 2019)

Machine Learning "Fairness" metrics

Some of the ML metrics depend on the training labels (*):

- When is the *training data trusted*?
- When do we have *negative legacy*?
- When *labels are unbiased*? (Human raters)

Predictions in conjunction with other "signals"

These questions are highly related to *the context* (e.g. ecosystems) in which the AI is designed/ deployed.

They cannot always be answered technically...

(*Trust in the ecosystem*)

(*) Source *Putting Fairness Principles into Practice: Challenges, Metrics, and Improvements*

Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Allison Woodruff, Christine Luu, Pierre Kreitmann, Jonathan Bischof, Ed H. Chi (Submitted on 14 Jan 2019)

Applying ML and *Fairness* criteria

in healthcare (domain specific)

Do we have protected groups? If yes:

- **Does the Model produces Equal Outcomes?**
  - Do both the protected group and non protected group benefit similarly from the model **(equal benefit)**?
  - Is there any outcome disparity lessened **(equalized outcomes)**?

- **Does the Model produces Equal Performance?**
  - Is the model equally accurate for patients in the protected and non protected groups?
    - 1**. equal sensitivity (equal opportunity)**

      A higher false-positive rate may be harmful leading to unnecessary invasive interventions (angiography)
    - 2**. equal sensitivity and specificity (equalized odds)**

      Lower positive predictive value in the protected group than in the non protected group, may lead to clinicians to consider such predictions less informative for them and act on them less **(alert fatigue)**
    - 3**. equal positive predictive value (predictive parity)**

- **Does the Model produces Equal Allocation (demographic parity)?**
  - Are resources proportionally allocated to patients in the protected group?

Source. Alvin Rajkomar et al. Ensuring, Fairness in Machine Learning to Advance Health, Equity, Annals of Internal Medicine (2018). DOI: 10.7326/M18-1990

Known Trade Offs
(Incompatible types of fairness)

Known Trade Offs (Incompatible types of fairness)

Equal positive and negative predictive value vs. equalized odds

Equalized odds vs. equal allocation

Equal allocation vs. equal positive and negative prediction value

Which type of fairness is appropriate for the given application and what level of it is satisfactory?

It requires not only Machine Learning specialists, but also clinical and ethical reasoning.

Source. Alvin Rajkomar et al. Ensuring, Fairness in Machine Learning to Advance Health, Equity, Annals of Internal Medicine (2018). DOI: 10.7326/M18-1990

Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6594166/

Which Tools to Use for what?

Open Source Tools (non-exhaustive list)

Tool	*Purpose*	*Map to Ethical Values*	*Limitations*

AI Fairness 360 (IBM)

AI ***Explainability*** 360 Open Source Toolkit ***(IBM)***

What-if Tool, Facets, Model and Data Cards (Google)

Aequitas *(Univ. Chicago)* https://dsapp.uchicago.edu/projects/aequitas/

Lime *(Univ. Washington)* https://github.com/marcotcr/lime

FairML https://github.com/adebayoj/fairml

Contribution from

- **Jesmin Jahan Tithi**
  Research Scientist,
  Parallel Computing Labs, Intel.

Privacy, Ownership, Rights, Accountability

- **Questions:**

- **▪ Who can sell your data?**
  ▪ Shouldn't you be able to see all data collected and remove them completely?

- **▪ Can apps secretly monitor your online activity – Crime/Disease/Productivity pre-screening? ▪ Should the Gov. Regulate?**

- **o Remove customized advertising culture – risk for human health and society**
  • Addiction to harmful entertainment (porn, violence), narcissism, ideological extremist

- **o Apps should have Opt-out default vs Opt-in**
  o Anonymize data completely and collect & retain no more than needed with clear disclosure

Ethics in “Techniques for large-scale data”

- **By Graham J.L. Kemp**

Chalmers Universisty of Technology

"JUSTICE: WHAT'S THE RIGHT THING TO DO?"

- If you want a moral philosophy perspective on ethical decision making:
- Professor Michael Sandel's lectures from a course at Harvard introducing moral and political philosophy
- First lecture includes some "trolley car" scenarios
- https://www.youtube.com/watch?v=kBdfcR-8hEY&list=PL30C13C91CFFEFEA6

DIFFERENT ETHICAL FRAMEWORKS EXIST

- and can lead to different recommendations
  e.g. Kantian ethics vs. Utilitarianism
  https://www.youtube.com/watch?v=7FR-FuhN2HM

- But it is assumed that students have a sense of right and wrong, and are capable of discussing a scenario and taking a view on whether an action is ethical, even if they haven't taken a course in moral philosophy and are not familiar with the theories and principles that define different approaches to ethics.

DISCUSSING ETHICAL ISSUES

- **Identify stakeholders**
- **Identify benefits and possible harm for each stakeholder Weigh benefits against possible harm**
- Get input from others
- Would you want an internal ethical assessment to be seen outside the organisation?
- Would you publish an internal ethical assessment on the organisation's web site?

- Recognize **that privacy is more than a binary value**
- “creepy”: when social values and technical capabilities are not aligned
- ”Understand that your attitude towards acceptable use and privacy may not correspond with those whose data you are using, as privacy preferences differ across and within societies.”

- Debate the tough, ethical choices
- “Discussion and debate of ethical issues is an essential part of professional development—both within and between disciplines—as it can establish a mature community of responsible practitioners.”
- ”Why might one set of scholars see [some ethical case] as a relatively benign approach while other groups see significant ethical shortcomings? Where do researchers differ in drawing the line between responsible and irresponsible research and why?”

- Know when to break these rules
- “For example, in times of natural disaster or a public health emergency, it may be important to **temporarily put aside questions of individual privacy in order to serve a larger public good**. Likewise, the use of genetic or other biological data collected without informed consent might be vital in managing an emerging disease epidemic.”
- ”Ethics is often about finding a good or better, but not perfect, answer, and it is important to ask (and try to answer) the challenging questions.”

- **References**

- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
  Raymond, E. (1999). *The cathedral and the bazaar*. Knowledge, Technology & Policy, 12(3), 23-49.
  Sandel, M. J. (2010). *Justice: What's the right thing to do?* Macmillan.
  Zook, M., Barocas, S., boyd, d., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B.A., Metcalf, J., Narayanan, A., Nelson, A. & Pasquale, F. (2017). *Ten simple rules for responsible big data research*. PLOS computational biology, 13(3), e1005399.