



---

# On the Ethics of Algorithmic Decision-Making in Healthcare

Dr. Thomas Grote

Ethics and Philosophy Lab, Cluster of Excellence “Machine Learning: New Perspectives for Science”; [thomas.grote@uni-tuebingen.de](mailto:thomas.grote@uni-tuebingen.de)

**07.05.2020**



---

## My aims in this talk

- I try to develop a coherent framework for the different ethical concerns which result from the impending automation of healthcare.
- I will not advocate for a particular normative position, but try to give an outline of what is morally at stake with regards to machine learning in healthcare.



---

## Structure

- (i) Machine learning in medicine – between hype and hope
- (ii) Ethical issues at the individual level
- (iii) Ethical issues at the institutional level
- (iv) The ethics of medical machine learning – a glimpse into the future



---

# Machine Learning in Medicine I

- Recent breakthroughs in deep learning have spiked interest in the application of machine learning within clinical medicine.
- There is a steady influx of high-profile studies about machine learning algorithms exceeding the abilities of medical professionals in the detection of disease or the prediction of critical health conditions.



---

## Machine Learning in Medicine II

- In domains such as ophthalmology, oncology or dermatology, machine learning algorithms have surpassed medical experts in terms of accuracy, when detecting diseases based on clinical images.
- In a similar vein, the relevant algorithms have shown great promise in the prediction of acute health threats, from cardiovascular disease, to kidney deterioration or circulatory failure.



---

## Machine Learning in Medicine III

- The potential benefits of implementing machine learning algorithms into clinical settings are manifold.
- They might:
  - (i) improve the reliability of medical diagnosis,
  - (ii) act as monitoring systems in intensive care,
  - (iii) help in the discovery of new biomarkers,
  - (iv) and at the hospital level, render the allocation of resources more efficient.



---

## Machine Learning in Medicine IV

- Some glaring problems:
  - All these studies have been conducted in environments which do not mirror realistic settings.
  - The metric of performance is somewhat skewed in favour of the algorithm.
- Whether machine learning algorithms are indeed beneficial to clinical medicine still needs to be established.



---

# Ethical Concerns

- Three levels of analysis:
  - (i) Individual
  - (ii) Institutional
  - (iii) Population-level/public health





---

## Ethical Concerns at the Individual Level

- The trade-off between accuracy and opacity:
  - Whereas a machine learning algorithm might have a higher diagnostic accuracy than a clinician, the complexity of its model makes it difficult to interpret why it made a given diagnostic decision.
  - The opacity of machine learning algorithms lies at the heart of many ethical problems



---

## Case-Study: Medical Diagnosis

- Medical diagnosis is an inherent source of uncertainty:
  - (i) A clinician will observe a bundle of symptoms.
  - (ii) She makes various conjectures.
  - (iii) She makes an inference to the best explanation.
  - (iv) She decides on a given medical intervention.
  - (v) Follow-up.



---

## The Problem of Defensive Medicine

- An expert has to choose between two options,  $a$  and  $b$ .
- She thinks  $a$  is superior to  $b$ .
- Yet, she chooses  $b$ .



---

## Peer Disagreement

- Peer Disagreement: Disagreement between two equally competent experts who are in a similar epistemic position.
- A clinician examines a patient who spotted a macule on her back and suspects that she has disease  $x$ , whereby her credences are 0.8.
- The algorithm examines an image of the macule and diagnoses disease  $y$ , with a credence of 0.8
- Who gets to decide, how to decide?



---

## The Problem of Automated Deference

- *Psychological reasons*: It is compelling to defer to the algorithm.
- *Epistemic reasons*: Knowing that an expert disagrees, lowers one's confidence in a given diagnosis.
- *Normative reasons*: By deferring to the algorithm, the clinician avoids being held accountable.
- *Structural reasons*: Institutional pressure to stick to algorithmic decisions.



---

## Patient Autonomy

- Patients need to give their informed consent to treatment decisions.
- Without knowing why treatment decisions are made, the possibility of informed consent is threatened.
- How can the patient assess, whether the algorithm`s diagnosis is trustworthy?



---

## Overcoming Opacity

- What constitutes satisfactory explanations, both for the requirements of the clinician and the patient?



---

## Ethical Concerns at the Institutional Level

- I focus on three issues in particular:
  - Accountability (gaps)
  - Fairness
  - Normative Alignment





---

## Accountability Gap(s)

- Without being able to evaluate algorithmic decisions, who is to blame for diagnostic errors?
- Algorithms do not qualify as moral agents, hence blaming them misses the mark.
- The clinician is destined to become a scapegoat.



---

## How to Close the Gap?

- Is distributed responsibility a feasible concept?
- Can something be learned from the legal handling of questions of responsibility for (other) medical devices?



---

## Fairness I

- The risks of utilising machine learning algorithms will be split unfairly across different populations.
- We have less data on ‘minority groups’ which is why there is a threat that algorithms will generate more false findings for them.



---

## Fairness II

- Obermeyer et al. (2020) scrutinized an algorithm that is widely used in the United States to predict health care needs of patients.
- As evidence shows, the relevant algorithm was miscalibrated for Black patients.
- They were, at a given risk score, considerably sicker than White patients.
- One of the key-factors which the researchers attributed to this miscalibration has been that the health-costs of individual patients were used as the main predictor.
- Due to various socio-economic factors, the access to healthcare is restricted for Black patients.



---

## Fairness III

- A trade-off between fairness and privacy?
  - Depending on where the medical data has been collected, some populations will almost inevitably be underrepresented. An academic hospital somewhere in Central Europe will most likely have lesser Black patients than Whites.
  - By disproportionately collecting data from underrepresented groups, we expose members of said groups to higher risks (as the medical data contains sensitive information on the individual patients)



---

## Normative Alignment I

- By incorporating machine learning into clinical medicine, we defer normative decisions to algorithms.
- Thereby, said algorithms will reshape the values of clinical medicine.



---

## Normative Alignment II

- According to which values do we determine the choice of treatment?
  - (i) Highest cure rate
  - (ii) Lowest risks
  - (iii) The least suffering



---

# A glimpse into the future





---

## The Political Economy of Medical AI/ML

- Much of the cutting-edge research is currently conducted by the big tech companies.
- As these companies enter the healthcare sector, what is it that we should be worried about, if any?



---

# Machine Learning in Medicine – Just a Black Box Among Many?

- Even in the absence of machine learning, clinical medicine is pervaded by black boxes.
- We oftentimes lack mechanistic evidence of a drug`s working.
- Highly complex risk models or technological devices are used already in clinical practice.
- Is there something distinctively opaque about machine learning?



---

# Thanks!

(DFG: (BE5601/4-1; Cluster of Excellence “Machine Learning—New Perspectives for Science”, EXC 2064, project number 390727645).

---



---

If you want to get in touch:

Thomas.grote.@uni-tuebingen.de

---