Universität Basel

Research Project

Trusting Black-Box Algorithms? Ethical Challenges for Biomedical Machine Learning

Project funded by own resources

Project title Trusting Black-Box Algorithms? Ethical Challenges for Biomedical Machine Learning
Principal Investigator(s) Elger, Bernice Simone ;
Co-Investigator(s) De Clercq, Eva ; Roth, Volker ;
Project Members Starke, Georg ;
Organisation / Research unit
Ethik / Bio- und Medizinethik (Elger)
Project start 01.02.2019
Probable end 01.10.2021
Status Completed

The integration of Artificial Intelligence (AI) into everyday life will deeply reshape many areas of society. Also within the medical domain, applying machine learning (ML) techniques on health-related data promises paradigm-shifting advances. More accurate and efficient diagnostic tools, personalised therapeutic regimes as well as prognostic or predictive measures are bound to improve the treatments of patients. Yet, while many authors have voiced general ethical concerns, in-depth ethical analysis of biomedical machine learning is still nascent.

ă

Intricate ethical questions arise when the very design of a program renders it opaque to human understanding. Artificial neural networks employed for Deep Learning (DL) can serve as a classic example for this. In DL, programs can find their own, multi-layered representations based on vast training data, allowing the program to find novel patterns in the data. In practice, this can be of great value if it renders a program's predictions and decisions more accurate or allows for new and improved scientific descriptions of a phenomenon. However, such novel forms of representations, recognizing patterns yet undetected and potentially undetectable by human agents, often make it impossible to fully explain and understand these so-called "black boxes".

ă

In clinical contexts, such opacity poses particular ethical challenges. How can we address so-called responsibility gaps, created by complex interactions between human agents and black-box algorithms, if a program's recommendation is erroneous and endangers patients? How can informed consent be obtained to use a program if it is by principle incomprehensible to both patients and health care professionals? How can we avoid discrimination against socially salient groups and protect vulnerable populations from systematic bias without understanding the underlying computational processes?

ă

A popular strategy to tackle these challenges is to call for trust in AI. After all, trust can be a means to deal pragmatically with uncertainty and incomplete knowledge in complex societies - and is generally considered to be of vital importance in healthcare settings. Taking trust as a starting point, this project hence asks if and under which conditions we can and should trust medical black boxes, bringing together theoretical considerations and empirical analysis informed by semi-structured qualitative interviews. The results of both parts shall be integrated in the sense of critical applied ethics to evaluate social practices

concerning biomedical ML, improve bioethical theory addressing this field and provide guidance for ethics committees and regulatory bodies.

Financed by
University funds

Add publication

Published results

4618283, Starke, Georg; De Clercq, Eva; Elger, Bernice S., Towards a pragmatist dealing with algorithmic bias in medical machine learning, 1386-7423 ; 1572-8633, Medicine, Health Care and Philosophy, Publication: JournalArticle (Originalarbeit in einer wissenschaftlichen Zeitschrift)

4610505, Starke, Georg; De Clercq, Eva; Borgwardt, Stefan; Elger, Bernice Simone, Why educating for clinical machine learning still requires attention to history: a rejoinder to Gauld; et al, 0033-2917 ; 1469-8978, Psychological medicine, JournalItem (Kommentare, Editorials, Rezensionen, Urteilsanmerk., etc. in einer wissensch. Zeitschr.

4621675, Starke, Georg; van den Brule, Rik; Elger, Bernice Simone; Haselager, Pim, Intentional machines: A defence of trust in medical artificial intelligence, 1467-8519, Bioethics, Publication: JournalArticle (Originalarbeit in einer wissenschaftlichen Zeitschrift)

Add documents

Specify cooperation partners