

Getting AI ethics wrong could 'annihilate technical progress'

July 30 2019, by Richard Gray, Horizon: The Eu Research & Innovation Magazine



AI can be used to identify potential perpetrators of violent crimes and assess the risk of a criminal re-offending. Credit: Max Pixel, licensed under CC0

An intelligent water gun that uses facial recognition to identify its targets

is helping to highlight some of the emerging human rights issues surrounding artificial intelligence (AI) – an area of research that is on the rise as new technologies become more and more prevalent in our daily lives.

"It's very difficult to be an AI researcher now and not be aware of the [ethical implications](#) these algorithms have," said Professor Bernd Stahl, director of the Centre for Computing and Social Responsibility at De Montfort University in Leicester, UK.

"We have to come to a better understanding of not just what these technologies can do, but how they will play out in society and the world at large."

He leads a project called [SHERPA](#), which is attempting to wrestle with some of the ethical issues surrounding smart information systems that use machine learning, a form of AI, and other algorithms to analyse big data sets.

The intelligent water gun was created with the aim of highlighting how biases in algorithms can lead to discrimination and unfair treatment. Built by an artist for SHERPA, the water gun can be programmed to select its targets.

"Our artist has built a water gun with a face recognition on it so it will only squirt water at women or it can be changed to recognise a single individual or people of a certain age," said Prof. Stahl. "The idea is to get people to think about what this sort of technology can do."

While squirting water at people might seem like harmless fun, the issues are anything but. AI is already used to identify faces on social media, respond to questions on digital home assistants like Alexa and Siri, and suggest products for consumers when they are shopping online.

It is also being used to help make [judgements about criminals' risk of reoffending](#) or even to [identify those who might commit violent crimes](#). Insurers and [tax authorities](#) are employing it to help detect fraud, banks have turned to AI to help process loan applications and it is even being [trialled at border checkpoints](#).

Impacts

Over the past year, Prof. Stahl and his colleagues have compiled 10 case studies where they have empirically analysed the impacts of these technologies across a number of sectors. These include the use of AI in smart cities, its use by the insurance industry, in education, healthcare, agriculture and by governments.

"There are some very high-profile things that cut across sectors, like privacy, data protection and cyber security," said Prof. Stahl. "AI is also creating new challenges for the right to work if algorithms can take people's jobs, or the right to free elections if it can be used to meddle in the democratic process as we saw with [Cambridge Analytica](#)."

Perhaps one of the most contentious emerging uses of AI is in predictive policing, where algorithms are trained on historical sets of data to pick out patterns in offender behaviour and characteristics. This can then be used to predict areas, groups or even individuals that might be involved in crimes in the future. Similar technology is already being trialled in some parts of the US and the UK.

Biases

But these algorithms can also learn the biases that already exist in data sets. If a police database shows that mainly young, black men are arrested for a certain crime, it may not be a fair reflection of the actual offender profile and instead reflect historic racism within a force. Using

AI taught on this kind of data could exacerbate problems such as racism and other forms of discrimination.

"Transparency of these algorithms is also a problem," said Prof. Stahl. "These algorithms do statistical classification of data in a way that makes it almost impossible to see how exactly that happened." This raises important questions about how legal systems, for example, can remain fair and just if they start to rely upon opaque 'black box' AI algorithms to inform sentencing decisions or judgements about a person's guilt.

The next step for the project will be to look at potential interventions that can be used to address some of these issues. It will look at where guidelines can help ensure AI researchers build fairness into their algorithms, where new laws can govern their use and if a regulator can keep negative aspects of the technology in check.

But one of the problems many governments and regulators face is keeping up with the fast pace of change in [new technologies](#) like AI, according to Professor Philip Brey, who studies the philosophy of technology at the University of Twente, in the Netherlands.

"Most people today don't understand the technology because it is very complex, opaque and fast moving," he said. "For that reason it is hard to anticipate and assess the impacts on society, and to have adequate regulatory and legislative responses to that. Policy is usually significantly behind."

Prof. Brey is coordinator of the [SIENNA](#) project, which is developing recommendations and codes of conduct for a range of emerging technologies, including human genomics, human enhancement, AI and robotics.

Mining

"Information Technology has, of course, already had a major impact on privacy through the internet and the mobile devices we use, but [artificial intelligence](#) is capable of combining different types of information and mining them in a way that reveals fundamentally new information and insights about people," said Prof. Brey. "It can do this in a very fast and efficient way."

AI technology is opening the door to real time analysis of people's behaviour, emotions along with the ability to infer details about their mental state or their intentions.

"That's something that wasn't previously possible," said Prof. Brey. "Then what you do with this information raises new kinds of concerns about privacy."

The SIENNA team are conducting workshops, consultations with experts and public opinion surveys that aim to identify the concerns of citizens in 11 countries. They are now preparing to draw up a set of recommendations that those working with AI and other technologies can turn into standards that will ensure ethical and human rights considerations are hardwired in at the design stage.

This wider public understanding of how the technology is likely to impact them could be crucial to AI's survival in the longer term, according to Prof. Stahl.

"If we don't get the ethics right, then people are going to refuse to use it and that will annihilate any technical progress," he said.

Provided by Horizon: The EU Research & Innovation Magazine

Citation: Getting AI ethics wrong could 'annihilate technical progress' (2019, July 30) retrieved

20 March 2024 from

<https://techxplore.com/news/2019-07-ai-ethics-wrong-annihilate-technical.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.