THEME 02

Can algorithms be neutral?

ALGORITHMS INEQUALITY BIG DATA ANALYTICS

Human bias is often ascribed to the fact that we have a subjective point of view on reality, caused by our personal history, circumstances, emotions, desires, political agenda, etc. This is problematic when we wish to make fair decisions in which a neutral or objective point of view is considered highly important. AI and algorithmic decision-making seem to offer a promising solution to this problem, because they lack human subjectivity. Unfortunately, bias appears to have resurfaced once again in AI and algorithmic decision-making. Can this be attributed to human programming with subjective values or the input of biased data, or is there more to it?

Our observations

- Earlier this year, MIT technology Review published the article This is how AI bias really happens—and why it's so hard to fix on the various ways algorithms can become biased. These variations are mainly ascribed to three causes: 1) Framing the problem: the aim toward which an algorithm is designed can create a biased view on the data it collects, since it is gathered in order to achieve something. 2) Collecting the data: deep learning programs are either fed data that is unrepresentative of reality, which creates a false definition of certain phenomena, or data that reflects existing prejudices. 3) Preparing the data: instilling in an algorithm which characteristics (e.g. age, income, or number of paid-off loans) are important to consider and which need to be ignored, can result in a one-sided view on reality that may, for example, exclude certain groups.
- In the recent study Discrimination, artificial intelligence, and algorithmic decision-making, the Council of Europe (CoE) recommended additional regulations for AI decision-making that escapes current non-discrimination laws. However, they also agree that it is still unclear what kind of additional laws would be sufficient and that more research and debate, such as in computer sciences, are required in order to determine how to proceed in this matter.
- Considering the long list of publications that recently came out, such as: Hello World : How to be Human in the Age of the Machine by Hannah Fry, Algoritmisering, wen er maar aan! by Jim Stolze, Sensemaking: The Power of the Humanities in the Age of the Algorithm by Christian Madsbjerg or World Without Mind: The Existential Threat of Big Tech by Franklin Foer, it appears that concerns about algorithms are omnipresent. The main concern is that algorithms are being used by powerful big tech companies to gather our data and make personal profiles that can be used to influence our behavior. Another big concern is the (unjustified) trust we put in the outcome of big data analytics.
- The danger of discrimination and inequality caused by the use of algorithms has already become apparent in several cases, such as in Amazon's AI recruiting tool that appeared to discriminate women or its biased facial-recognition software. In order to still be able to use algorithms but to avoid discrimination or bias, different parties are searching for solutions. The city of Amsterdam, for example, recently hired KPMG to help them eliminate bias in the algorithms they use in their city (e.g. to automatically handle complaints concerning public spaces).

>> see next page

Connecting the dots

The tendency of people to believe in the validity of recommendations made by algorithms over human advice (so-called automation bias), is a common phenomenon, according to the CoE. It can be motivated in several ways. First, recommendations that are generated through an algorithm have an air of rationality, mainly caused by the algorithm's superior calculation power and the absence of human subjectivity. Second, automation bias can be caused by a lack of skills, context or time to evaluate whether the computed recommendation has followed a valid path of reasoning. Finally, human decision-makers may try to minimize responsibility by following the advice provided by AI. One of the biggest challenges AI and algorithmic decision-making face, however, concerns discrimination or a bias policy when operating.

In many of the studies, books, articles and reports on bias and discrimination in algorithms, the blame is assigned to the developers of an algorithm, who (whether consciously or not) either program it in a biased manner or feed it data that misrepresents reality, is one-sided or is simply biased. Considering the problem in this manner gives the impression that, although these problems are very hard to solve, they can be solved at some point nonetheless. This is caused by the hidden premise that there is an objective truth about reality to which we have access and which can objectively be represented in (a) language. For example, sentences such as "The distance between the sun and earth is 149,600,000 km", "Since January 20th, 2017, Donald Trump has been president of the U.S.A." or, "Water freezes below 0 degrees and boils at 100 degrees Celsius" are considered to express objective truths about reality. However, it can be said that in order to gather and express the objective truth about reality, we need to interpret reality. For example, in order to express the distance between the sun and the earth, we have to agree on measurement principles and from where to where we will measure, after which we interpret the distance we come across. Studies on language have already shown that an objective representation of reality in language is not unproblematic, to say the least. A prominent voice in this context is Wittgenstein, who argued that the world cannot simply be represented in a series of (language) expressions, but can only be expressed in a series of interpretations and communal understandings in which meaning is in constant change and always dependent on the participants' conception of a certain definition. In other words, there is no such thing as a fixed definition and therefore, reality cannot be represented by language in a neutral/objective manner. This idea was recently enforced by a scientific experiment in which two quantum scientists made contradictory observations of the same phenomenon. The possibility that reality cannot be observed or expressed in a neutral manner is problematic for the use of algorithms, since they are developed to help us grasp the objective truth about reality. This problem is briefly touched upon by, for example, the study of the CoE and the MIT article on bias in algorithms. At some point, both mention that language itself always carries a certain degree of ambiguity and sometimes even plain contradictions when it comes to the definition of a concept. This would imply that bias and discrimination in algorithms cannot be solved entirely, since they too work with definitions. This of course doesn't mean that algorithms can't be improved or still be of great value in advancing all sorts of processes in which data analytics are required. We should, however, consider the fact that the recommendations of algorithms will be biased by definition, as are those of humans, for that matter. In order to estimate the value of the recommendations made by algorithms, we do not only need to improve algorithms, but also our own ability to evaluate their contribution to understanding the world.

Implications

- The reason that algorithms are not simply dismissed in all cases in which it is important to avoid bias and discrimination, is because they offer possibilities and potential solutions that transcend human capabilities. It is a technology that can cause economic growth and generate better (scientific) understanding of all sorts of important aspects of our lives. However, the idea that algorithms make decisions on important matters without human intervention might become a no-go in the future.
- Automation bias might be hard to change, because in many cases, we are used to trusting the outcome arrived at by computers. When we ask Alexa what kind of weather it will be, when we use a calculator, when we want to know the time, in so many cases, we trust the information provided by computers without thinking of the process that precedes it.
- If we agree that definitions of concepts are in constant change and very much dependent on (human) interpretations, the demand for human programmers of AI to reduce bias will remain a constant factor in the development and use of algorithms.
- In many aspects of our lives, we have accepted the fact that total neutrality is not possible (e.g. a human judge, no matter how experienced and educated, will always carry some of his subjectivity into his ruling). However, it is questionable if, when we indeed consider AI as a bias tool, we will ever accept their recommendations when it comes to sensitive matters in which our personal or common faith is at stake.

>> see next page