Natascha Machner 1027745 31.7.2016

Critical Algorithm Studies

Final Assignment

Algorithms that judge?

Algorithms are already part of our daily life. Usually, they work silently and unnoticed to the user in the background, for example when they define what advertisement is shown to you on online shopping websites, what videos and music is suggested on streaming platforms, or how your personal social network news feed looks like.[1]

But not only did they find their way into social and personalized shopping platforms, they also became part of global business e.g. in the form of “algorithmic trading”[2] at the stock market, or of political issues like security and surveillance[3], areas where they can cause great impact and also great harm.

Studies already have shown that algorithms can influence the human mood[4] and human behaviour.[5] And they are already used as a means to judge human work and human behaviour as well. Turnitin for example is a tool which is used to check academic writing on plagiarism. It compares the written text to a huge database of other papers and academic publications and calculates a ratio on how much of the text can also be found elsewhere, what could be labelled as how much of it was only copied.[6] It could be a simple, easy method to judge the amount of plagiarism in an academic text.

Not only in the case of Turnitin it is often said that algorithms are more accurate, more trustworthy and more impartial than humans. For a teacher, it would be impossible to check the huge amount of papers and academic texts and compare them to a student's writing, but the algorithm in Turnitin can handle this huge amount of data in short time. Also, it does the same check for every student in the same way and therefore should not be biased by sympathy or other aspects. The ratio it calculates seems like a matter of fact on how much in the text can be found in other text. That said, why not entrust the algorithm with the check of plagiarism and simply taking the calculated ratio as a measurement?

One problem is that the comparison method doesn't recognize correct quoted passages, so the ratio doesn't necessarily only contain plagiarism. Students can also check their writings with the tool, and rearrange phrases so that they aren't recognized anymore by Turnitin, but still remain copied, what is especially easier for people who are native speakers in the language the text is written in.
So how much truth remains in that simple ratio, how unbiased when knowing that native speakers can deceive the tool easier, and that it can be deceived at all? Can teachers trust it, and can the students trust a teacher that bases his or her decision on a tool like that?

And this example is only a small one, a – you could say – not exactly life threatening one. What about algorithms that are bigger, much bigger, so big that not even the developers, let alone the users can tell how exactly it comes to its decision and where some problems could lie hidden? How can you recognize it might be (unintentionally) biased by something very unexpected like the ability to rearrange sentences in a specific language?

What if the algorithms is meant to make a decision on a medical diagnosis? Or on the threat of a person to public safety? What if its decision not only may influence your academic career, but your entire life?

And this is not a question that we should ask ourselves in the far future. Those kinds of algorithms already exist.
The US military has already made use of algorithms to identify possible terrorists in Pakistan in order to have them killed by drones.[7] Even though it was stated that the algorithm only selected possible targets and that it was confirmed if "justified" or not before the actual strike happened, a part of the decision who should be killed was made by an algorithm. There are similar programs in development in order to identify possible criminals even before they actually commit the crime, especially in the case of terrorism[8].
At the same time, drones and also robots for killing people are already in use, and in the USA recently the Dallas shooting suspect was killed by a robot that was equipped with an explosive device.[9]

Taking those recent examples and developments, how far should technical systems be allowed to go? Although especially the crime prevention algorithms are known as not very accurate at the moment, the development is going fast. They scan through the communication of people, how they are connected with known criminals etc. and calculate how high the possibility for a crime of a person would be. Let alone the huge privacy issues aside, another big and dangerous problem of many algorithms nowadays, those algorithms have to be so complex and have to be fed with so much data that probably no human could reconstruct the decision making process. What if the algorithms has found a "possible threat"? What if there is not much time to check this person in detail, because it is very urgent? Is it fine to trust the judgement an algorithm has made, and – in the worst case – eliminate the target?

What does it mean for a human being to be judged by – easily said – a machine? To be arrested or killed for unknown reasons that somehow resulted in the decision that one person was a threat?

And – since there is no program without mistakes – what if one person happens to be mistakenly accused? Do you blame an algorithm that might be even 99% accurate, just not exactly in this case? Can the developers be held responsible? Especially when it comes to big projects, where no single developer knows the whole algorithm? When the algorithm can learn and change itself? Probably not. What about the people who trusted the judgement of the algorithm? When usually its decisions are right, and they have no possibility to reconstruct the judgement process? Why could you blame them?

But, after all, you can't hold a machine or an algorithm responsible. At least, it probably wouldn't result in anything, because they don't know guilt, and they can't be convicted.

Algorithms that judge humans are therefore – in my opinion - a very delicate topic.
The transparency of the decision making process, the real accuracy, the mistake rate, the unintentional and often hidden biases and the problem of responsibility are issues many algorithms have.
But when it comes to important decisions when even human lives depend on, they have to be emphasized even more.
Of course, you can say that also humans often can't lay open how exactly they came to a judgement, what unconscious thoughts influenced them. Humans make mistakes as well, probably even more than algorithms do, like drone pilots who work for hours and inevitably lack concentration at some point. Humans are biased, and they aren't as fast as algorithms to react to changing situations, maybe even less than a second of time where it might be important to stop an attack when children suddenly appear in the area.

But, if a human does the judgement, he or she will be responsible for it.
Now, what to prefer? Let humans judge and let them make more mistakes than an algorithm in their place would do? Is that really the case? Can algorithms judge better on plagiarism, can they prevent crime better, can they do “better” drone attacks resulting in less civil casualties? How could someone possibility prove or disprove that?

And even if algorithms would be better judges, what world would that be where people are judged by machines? Where an algorithm decides that they should be arrested, that they are a threat, or even that their life should end. And where maybe even a machine “handles the killing”.

Isn't it the duty of humans to do that themselves? Isn't it the right of humans to be only judged by other humans? Wouldn't it be inhuman to be sentenced to death by an algorithm and killed by a robot, even more, if that is actually possible, than that being done by another human? What about the families of people who were mistakenly killed by drones, what do you tell them if no human was actually a deciding part of this judgement, and that all they can blame would be a set of code and a lifeless weapon?

So in conclusion, I think that algorithms should not judge people. They can suggest their results, in an as transparent as possible way so that it gets comprehensible how this result was created and can be checked in more detail if necessary. They are great tools to support people, to complement them, but in the end, the judgement should be done by humans. Humans who see more than just the plain data, and who can react to situations that couldn't be foreseen.
And who will be responsible for all the consequences and mistakes to come.

1 Berg, Martin. 2014. Participatory trouble: Towards an understanding of algorithmic structures son Facebook. Cyberpsychology: Journal of Psychosocial research on Cyberspace, 8(3).
2 Lenglet, Marc. 2011. “Conflicting codes and codings: how algorithmic trading is reshaping financial regulation.” Theory, Culture & Society, 28(6), 44-66.
3 Introna, Lucas D., and David Wood. 2004. ‘‘Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems.’’ Surveillance & Society 2 (2/3): 177-98
4 https://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds (visited on 31.7.2016)
5 Grosser, Benjamin. 2014. “What Do Metrics Want? How Quantification Prescribes Social Interaction on Facebook.” Computational Culture.
6 Algorithms, Governance, and Governmentality: On Governing Academic Writing
7 https://www.theguardian.com/commentisfree/2016/feb/21/death-from-above-nia-csa-skynet-algorithm-drones-pakistan (visited on 31.7. 2016)
and
Schuppli, Susan. 2014. “Deadly Algorithms: Can legal codes hold software accountable for code that kills?” Radical Philosophy, 187: 1-8.
8 N. F. Johnson, M. Zheng, Y. Vorobyeva, A. Gabriel, H. Qi, N. Velasquez, P. Manrique, D. Johnson, E. Restrepo, C. Song, S. Wuchty. 2016 “New online ecology of adversarial aggregates: ISIS and beyond” Science 17 Jun 2016: Vol. 352, Issue 6292, pp. 1459-1463
9 https://techcrunch.com/2016/07/08/police-use-bomb-robot-to-kill-suspect/ (visited on 31.7.2016)