Ethics in Algorithms

Term Paper on Critical Algorithm Studies

Martin Asmus - 1126588
martin.asmus@student.tuwien.ac.at
Technical University of Vienna

10. Oktober 2016

Abstract

This paper explains some approaches to ethical analysis of algorithms and computational systems and it highlights some of the areas of conflict, based on previous research in this field.

1 Theories of normative ethics

To understand where the ethical conflics in the design and use of algorithm based systems – and anywhere else – it is necessary to understand how people define something as ethical. The papers on this topic primarily discuss two to three ethical orientations. These three theories of normative ethics are distinguished by the following properties:

The deontological approach says „a fixed set of duties, rules, and policies define actions as ethical. Break these rules and you have behaved unethically.“ [2]

An teleological (consequentialist or utilitarian) orientation derives from the consideration of consequences to a decision and results in considering the option with the best – or least harmful – consequences for a particular group (the person herself, a group that the individual identifies with, the majority, etc) the most ethical one.

Following the personalist or virtue model, people neither care for duties nor for consequences, but for subjective attraction or spiritual guidance to choose what is ethical.

2 Ethics in the design process

When designing a product, the designers have to make decisions about the products properties. Some of these decisions belie on rational requirements or goals, others are a question of the use case or the position of the designers. If there is only one way to design a product and therefore the design process as well as the use of the product leads to a single output – at least for certain criteria -, it consists of genuine value-judgments. This theory is appliable to material products as well as immaterial products, like algorithms, equally. An algorithm depending on the personal position of the designers can be considered a value-lade algorithm. One suggested definition of Value-laden algorithms is as follows: „An algorithm comprises an essential value-judgment if[D?] and only if, everything else being equal, software designers who accept different value-judgments would[D?] have a rational reason to design the algorithm differently (or choose different algorithms for solving the same problem).“ [1]

[1] also shows a very clear and rational approach to define groups for all variables, that have to be set in the course of the design process and therefore identify possible input vectors for personal values and biases.

Another ethical issue deriving from the design process can arise, if the designers and the users have conflicting ethical bases. If the designers think of a certain use case or approach, but do not clearly state the limited usability of their product, users might not be aware, that they rely on judgments based on different values than they would choose. (See A real life example: medical image analysis)

In [1] it is clearly recommended to implement user-defined ethical values wherever possible, so the users are aware of the ethical implications and can choose a value based on their position and/or use case. However it is recognized, that this is not feasible in every case, but in cases where it is not, it should be at least made be transparent to the users, that the results such program delivers are bound to particular values and that the use is therefore not universally ethically harmless.

3 Ethical significance of use and input

Mike Ananny argues, that at least some categories of algorithms cannot be evaluated in respect of any approach of ethical analysis independently from their sociotechnical context and the given input [2]. The output of machine learning algorithms, social network algorithms and predictive algorithms depend on the given input and their use might have a completely different ethical significance, depending on the actions triggered by a certain output. Peter Hanink shows in [3] the problems with biased output of compstat, the implementation of an predictive policing algorithm used by the NYPD, deriving from biased input and leading to police officers stopping and frisking approximately 400,000 innocent Afro-Americans. It is obvious that an algorithm processing statistical data can give biased results, if the data set is biased and that any breach of ethics in the creation or collection of input data is propagated through the algorithm to its output.

The example of US crime statistics as input data reflects this problem. It is well known that US crime statistics take into account which ethnic background a convict has, but no information on the social background or any information on the (non-criminal) development of that individual, therefore only leaving the ethnicity as significant correlation to criminal activity. [6]

This also shows a limit of a purely quantitative approach to scientific research, as some of the factors that are needed for a non-biased input for a predictive policing algorithm – despite other ethical questions about the use of such a system – lie far in the field of qualitative sociological studies. By now many ethically relevant, big data based algorithms are in one way or another reinforcing stereotypes by delivering results based on alleged similarities.

Another example for this field of conflict is unexpected user input, manipulating the calculations of a formally correct, non-biased algorithm. For example search engine algorithms are designed to display a consensus on the relevance of particular sites to a particular keyword or search phrase and if everything goes as expected, they do. But if a user or a group of users unravel the inner workings of a search engine and create information the search engine uses to calculate the results, the rationality of the results is anulled. Adam Mathes re-

searched this effect and created the term “Googlebombing” for setting up a big amount of links, associating a site with a search phrase or keyword, as incoming links are one of the main parameters for Google’s judgment on the relevance of a site to a keyword and its overall popularity. The most popular example for this technique was the association of George W. Bush’s biography with the search term “miserable failure”. [5]
This more hacktivistic approach, where users embed this data on “real” websites, is closely related to the operation of so-called link farms, where a – usually commercial - proponent operate sites with the sole purpose of publishing links that should produce a higher listing rank. Google gives penalties on ther PageRank, when they suspect a site of such behaviour, like happened to SearchKing, an Oklahoma local commerce search engine. [5]
These incidents show that while search engines are only repesenting biases, that exist in the information flow and network they analyse and can be considered a consensus of the people that publish content on the Internet, this consensus can be constructed on purpose as well and by that create a ethically questionable output of an otherwise correctly and rationally working algorithm.

4 Ethical implications of results

Algorithms are nowadays used for a number of ethically sensitive matters. Be it the calculation of medical diagnoses, predictive policing or even weaponry systems. If the results of such systems lead to direct action instead of involving manual judgment of a human being, the use at least poses some serious ethical questions, no matter how they are designed.

Even search engine algorithms can be entangled in ethical issues in this context. In [5] it is shown, that search terms like “jew”, which gives results including antisemitic sites and has set an ethical predicament for Google [5], rise the question, if an algorithm that processes possible unethical input data should filter it, mark it with a disclaimer, or allow it as output without any difference to other less sensitive data. This quickly leads to a debate about censorship, free speech and hate content.

5 Censorship or unethically manipulating allegedly unethical data

Especially nation states’ governments often have a distinct agenda on what information should reach the local users and are able to create a legal foundation to enforce these. Two more popular and very diverse examples, where these regulations were enforced on computational output was on the one hand the censoring of search results on tiananmen from Chinese users and on the other hand the results on the mentioned search for „jew“ in Germany and Austria, where laws forbid antisemitic content due to the history in the nazi regime.

The first case seems as a clear confinement due to uncomfortable facts that might endanger people's loyalty toward the state itself. But there are people in China, that prefer to see the more glorious aspects of their history and feel the international, uncensored results to „tiananmen“ as American, anti-communist propaganda.

The second case is even more difficult to analyse in respect of ethics, because the censored content as well as censorship itself can be considered unethical.

6 Accountability

Which actions are derived by an algorithm's results also leads to the question of accountability. Who is responsible for the output of an autonomous system, and who is responsible for subsequent actions that rely on those results? Since algorithmic calculations produce decisions or just recommendations, the people making use of such algorithm try to evade any charges by pointing to the allegedly rational, deterministic system that aided an questionable decision or created the reason for a unethical or even criminal action. There is no consensus on the question, if the designers or the users should be accountable for such a result.

7 Can algorithms be agonistic?

A quite different approach is to ask, what effects has an algorithm on perception and the world view of the public. Can it be agonistic and neutrally accompany and support human societies through a way of development through discourse? Where and in what ways may the introduction of algorithms into human knowledge practices have political - according to Chantal Mouffe, politically means in this context, any process that involved a confilict and discourse - ramifications? In fact, most of the large-scale analysis and prediction systems are designed within a clear agenda and are therefore rather "governing agents" [7] than mirrors or neutral spaces. Either politics or economics allways have an interest in not only analysing, but also influencing peoples behavior. What information is shown to a user by a seemingly autocratic system is seldom without intention, like Amazon's "...also bought"-list has the clear goal to sell more books and is embedded in a complex ecosystem of authors, publishers, reviewers, potential customers and, of course, Amazon. [7] And even if it would be a system that is designed according to the agonistic principle, it still can be levered by others, like the 4chan community manipulated online votes. [7] The contesting algorithm of reddit even was involved in a huge witch hunt against an innocent man, who was believed to be one of the boston bombers from 2013. The gravitational effects of strong, already supported positions in big online communities outline a rather plutocratic than autocratic working of these systems in context with their interaction with humans.

8 Narrowing down human experience

While it has some practical use to compute the “best” option, acting accordingly to an algorithmic result is “categorically narrowing the set of socially acceptable answers to the question of what ought to be done”. [2] Especially predictive systems, that make recommendations based on similarities give as result “a narrowly construed set that comes from successfully fitting other people, past actions, and inanimate objects into categories”. [2] Recommending similar things to people might feel convenient to some, but it clearly decreases the probability that an individual is spontaneously impressed and attracted by something completely new instead of remaining in the habit. If these systems have unreflected influence on a long term, this might lead to people converging to uniform, stereotypical groups.

9 A real life example: medical image analysis

Algorithms developed in scientific context tend to prefer false negatives over false positives to keep the yielded results valid for further research. In e.g. medical use, this approach leads to an ethically questionable use case. In [1] the authors propose the use of the precautionary principle for medical purposes, rather than the conservative scientific approach, to prevent doctors from falsely judging an ill person as healthy.

10 Connections to other topics of Critcal Algorithm Studies

This topic is closely related to values and biases in algorithms and the discussion about accountability, as these directly influence or even set the ethical meaning of an algorithm. Also visibility and transparency are directly connected, because the users knowledge of the existance and configuration of algorithmic processes alters the way a result is perceived and handled. Therefore it is not possible to summarize this aspect isolated, without at least mentioning the interfaces to and overlapping with these other fields.

11 Conclusions

There are several formalizable vectors for ethical values in the design process, the use of an algorithm as well as in the post-processing and the use of its results. Basically for all of them exist recommendations and best practices, but the responsibility is divided between designers, operators, end-users and other parties involved in providing input. In most cases these groups are not connected in a single organization, but might not even know each other. In a setting, where these groups support different ethical values or one of them assumes an antagonistic role, it is very difficult to ensure ethical behavior of any system,

that has some influence on human lives or values.
It seems clear, that the desires to understand and even influence people in a large scale are very dominant and make it difficult to create a agonistic system and keep it that way.

References

[1]: Kraemer, Felicitas, Kees Overveld, and Martin Peterson. 2010. ''Is There an Ethics of Algorithms?'' Ethics and Information Technology 13 (3): 251-60.
https://link.springer.com/article/10.1007%2Fs10676-010-9233-7

[2]: Mike Ananny. 2015. „Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness“ Science, Technology, & Human Values 1-25.
http://sth.sagepub.com/content/early/2015/09/23/0162243915606523.full.pdf+html

[3]: Peter Hanink. 2013. „Don't Trust the Police: Stop Question Frisk, Compstat, and the High Cost of Statistical over-Reliance in the NYPD“ 13 JIJIS 99.
http://heinonline.org/HOL/LandingPage?handle=hein.journals/jijis13&div=12&id=&page=

[4]: Merrill, J. C. 2011. ''Theoretical Foundations for Media Ethics.'' In Controversies in Media Ethics, 3rd ed., edited by A. D. Gordon, J. M. Kittross, J. C. C. Merrill, W. Babcock, and M. Dorsher, 3-32. New York: Routledge.
https://www.routledge.com/Controversies-in-Media-Ethics/Gordon-Kittross-Merrill-Babcock-Dorsher/p/book/9780415963329

[5] Grimmelmann, James. 2008. "The Google Dilemma." New York Law School Law Review, 53: 939.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1160320

[6] Lee Ellis, Kevin M. Beaver, John Wright. 2009. „Handbook of Crime Correlates“ Academic Press
https://www.elsevier.com/books/handbook-of-crime-correlates/ellis/978-0-12-373612-3
[7] Kate Crawford. 2015. "Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics" Science, Technology, & Human Values 2016, Vol. 41(1) 77-92