In: Emre Bayamlıoğlu, Irina Baraluic, Liisa Janssens and Mireille Hildebrandt (eds), BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen, 2018: Amsterdam University Press, 46-50. DOI 10.5117/9789463722124/CH7.

Transparency is the perfect cover-up (if the sun does *not* shine)

*Jaap-Henk Hoepman**

Abstract

In this paper we reflect on Louis Brandeis' famous quote "Sunlight is said to be the best of disinfectants; electric light the most efficient policeman". We critique the current focus on transparency, and discuss some limitations to relying on transparency as a mechanism to counter the ill effects of automated decision making. First, investigating an automated decision making system requires a sufficient number of experts that are motivated to investigate in the first place. It also requires special domain knowledge, and sufficient resources to process and analyse the often huge amount of data that underlie the decision making process. Second, transparency is useless without agency. Without the power to challenge a decision, information underlying that decision is useless. Third, being transparent is hard, and in fact organisations may appear to be transparent while actually obfuscating the actual decision making process. Finally, transparency may conflict with other legitimate business interests. This is not to say that transparency is useless. To the contrary: the mere fact that decision-makers are forced to be transparent will make them behave more diligent most of the time. But this is not enough. We need new, stronger, models of accountability that take the above limitations of transparency into account.

Keywords: privacy, transparency, algorithmic decision making, accountability

Introduction

Even though calls for transparency in a modern form go as far back as the early Age of Enlightenment (Annany & Crawford 2018), perhaps Louis Brandeis can be considered the father of 'transparency theory' because of this famous quote (Brandeis 1914):

> *Publicity is justly commended as a remedy for social and industrial diseases. Sunlight is said to be the best of disinfectants; electric light the most efficient policeman.*

Indeed, transparency is commonly advocated as an important tool to counter the ill effects of automated, data driven, decision-making (Hildebrandt & Gutwirth 2008; Pasquale 2015).

Now Brandeis never used the term 'transparency' itself, but if we read publicity as transparency, I cannot fail to wonder: what if the sun does not shine?.... What if we all lived in glass houses but there is no light to see inside? Wouldn't that render transparency useless? Indeed, wouldn't that turn transparency into a perfect cover-up, allowing organisations to hide in plain sight, pretending not to be engaged in any nefarious activities?

Do many eyeballs make bugs shallow?

It is a common mantra in the open source community: 'many eyeballs make bugs shallow' (Raymond 2000). In fact, it is one of the main arguments why the source code of all software we develop should be open. By publishing the source code of the software, one allows public scrutiny of that code by other, independent, experts. Bugs (i.e. programming mistakes) will be found that would otherwise lay undetected in the source code forever. As a result, systems will become more reliable and more secure (Hoepman & Jacobs 2007). Moreover, fundamental design decisions can be challenged, possibly leading to improved designs.

However...

The mantra assumes three things. First, that an unlimited number of eyeballs, i.e. independent experts, is available to scrutinise the growing pool of open source projects. Second, that these experts have an interest or incentive to spend some of their (valuable) time on this. And third, that every open source project is equally likely to attract the attention of a sufficient number of experts.

All three assumptions are unfounded.

The number of experts is severely limited. These experts may often be inclined to start their own open source project rather than contributing to someone else's project. And many open source projects remain unnoticed. Only a few, high profile projects receive the eyeballs they need. Advocating transparency to balance data driven decision making, suffers from the same set of potential problems. Systems that make automated decisions are complex, and require considerable expertise to understand them (an issue that we will return to further on). Even if all automated decision making by all organisations is done in a transparent way, there will always be only a limited number of experts that can scrutinise and challenge these decisions. Which decisions will actually be challenged depends on the incentives; again, we cannot be sure high-profile cases are likely to attract the attention they deserve.

Transparency by itself is useless without agency

Let's assume transparency works in the sense that 'bugs', i.e. improper data driven decisions, come to light and people want to act. Transparency by itself does not allow them to do so, however. The situation also requires agency, i.e. the ability to address and redress the problem. (Note that for exactly this reason a large class of open source software is in fact *free*, as in free speech. This allows anyone *with the necessary technical capabilities* to change the source code, fix whatever bug they find, and redistribute the solution.)

In many cases you have no agency whatsoever. Computer says no, tells you why, but no matter how you try, you will not be able to successfully challenge that decision. (See Ken Roach's excellent movie "I, Daniel Blake" for a compelling illustration of this point.) This is caused by several factors.

The first, most important one, is the lack of power. A single person, wronged by a decision of a large organisation, is but an itch that is easily scratched. Even if the case involves a larger, powerful, group of subjects that are collectively impacted by the decision, or if the case is taken over by a powerful consumer organisation or a fancy law firm, one would still need laws and regulations that create a (legal) basis on which the decision can be challenged. Finally, the process of appealing a decision may be so cumbersome that the effort to challenge the decision may thwart the benefit of doing so. Individuals easily get stuck into bureaucratic swamps.

The 'house of mirrors' effect

A mirror is made of glass, but it is not transparent. A house of mirrors is a seemingly transparent maze where one easily gets lost. The same problem plagues transparency theory: a decision-maker may be transparent about the decision-making process, but the description may in effect be opaque, hard to understand, hard to access/find, and/or hard to compare with others. For example, many privacy policies are overly legalistic, making them unintelligible by the average user. They are often far to long too, requiring so much reading time that no one ever reads all privacy policies of all sites they visit (McDonald and Cranor 2008).

Even if the decision-maker honestly tries to be transparent about the decision-making process and honestly aims to explain a particular decision to the subject of that decision, this explanation may still be too complex to understand. The explanation may use jargon, may depend on complex rules (if rule-based at all), and may depend on so many variables that data subjects will easily lose track. These properties of transparency may also be put into use disingenuously, to make the explanation unintelligible on purpose, while claiming to be transparent. One can observe a similar effect in the telecommunications market where mobile phone subscription plans are complex, and where different operators use incomparable tariff plans. As a result, ordinary users have a hard time figuring out which offer suits them best (and a whole market of comparison services was born, not only for the telecommunications market, but also for the health insurance market for example).

Being transparent is hard

It very much depends on the decision-making process whether it is easy to supply a proper explanation for every decision made. In classical rule based expert systems this is certainly possible (by disclosing the rules applied and the facts/data/propositions on which they were applied), but in modern machine learning settings this is much less clear (Burell 2016). In many cases the machine learning system constructs an internal representation 'explaining' the example cases presented to it during the learning phase. But this internal representation, the model of the type of cases the algorithm is supposed to be applied to, is not necessarily close to how humans understand these types of cases and the logic they apply to decide them. A complex vector of weighing factors that represent a neural network does nothing to explain the decision made with that neural network, at least not in how humans understand an 'explanation'.

Challenging a decision is hard

Challenging a decision is hard. Even when given the explanation of the decision and the data underlying the decision, it may be hard to verify that the decision is valid. This is caused by several factors.

First of all, you need the necessary domain knowledge to understand the explanation, and to spot potential problems or inconsistencies in it. For example, to understand whether a decision in, say, environmental law is correct you need to be an expert in environmental law yourself. (This partially overlaps the first argument of the difficulty of finding and incentivising experts to challenge a decision.) Secondly, the validity of a decision depends both on the interpretation of the data on which it is based, and the interpretation of the rules used to arrive at the decision. Moreover, the selection of the rules matters a lot: it may very well be that applying a different set of rules would have led to an entirely different set of decisions. (And all this assumes that the decision-making is in fact rule based to begin with, allowing such a clear interpretation.) Thirdly, the data set may be so large and the model used to 'compute' the decision so complex, that even a basic verification of the consistency of the decision itself (let alone any complex 'what-if' scenario analysis) cannot be done 'by hand' and thus requires access to sufficiently powerful data processing resources. In the worst case the problem is so complex that only the decision-maker itself has enough resources to perform such an analysis. This totally undermines the principle of independent oversight. Lastly, the explanation of the decision may be valid and reasonable, but may not be the *actual* reason for the decision. A common example is the (inadvertent) use of proxies (like home address or neighbourhood) for sensitive personal data categories like race or religion. Sometimes this happens on purpose, sometimes this is a mistake.

Transparency may conflict with other legitimate interests

Even if the system used allows for the proper explanation of all decisions made, publishing these explanations may reveal too much information about the underlying model used to arrive at the decision. Of course, that is the whole point of requiring transparency. However, certain organisations may wish to keep their decision-making logic a secret, and may have a legitimate interest for this. For example, law enforcement or intelligence agencies have every reason *not* to reveal the models they use to identify potential terrorists (for fear that terrorists will change their modus operandi to evade detection). Similar arguments apply to fraud detection algorithms for example. Business, like credit scoring agencies, may not want to reveal their models as these algorithms, these models, may be the only true asset, the crown jewels, of the company.

Conclusion

We have discussed six arguments to show that transparency *by itself* is insufficient to counterbalance the ill effects of automated, data driven, decision making. This is not to say that transparency is useless. To the contrary: the mere fact that decision-makers are forced to be transparent will make them behave more diligently most of the time. But this is not enough. We need new, stronger, models of accountability that take the above limitations of transparency into account (Annany and Crawford 2018). For transparency to work, agency is a prerequisite. We need suitably incentivised experts that can help challenge decisions. Proper enforcement of transparency requirements is necessary, to ensure that the information provided is accessible and intelligible. Using decision-making processes that are hard to explain should be made illegal. And independent verification platforms that make it possible to verify and analyse decisions based on complex models and data sets must be made available. Finally, where transparency conflicts with other legitimate interests, a clear set of principles are necessary to decide when an explanation is not required.

Because without sun, transparency is the perfect cover, hiding in plain sight what everyone fails to see.

Notes

* Jaap-Henk Hoepman is associate professor at the Digital Security group of the Radboud University, Nijmegen, the Netherlands. He is also an associate professor in the IT Law section of the Transboundary Legal Studies department of the Faculty of Law of the University of Groningen.

[1] This research is partially funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement no 732546 (DECODE) and by the Netherlands Organization for Scientific Research (NWO) as project 'Patterns for Privacy' (CYBSEC.14.030).

References

Annany, Mike, and Kate Crawford. 2018. “Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability.” New Media & Society, 20(3): 973–89.

Brandeis, Louis. 1914. Other people's money and how the bankers use it. New York: Frederick A. Stokes.

Burell, Jenna. 2016 “How the machine ‘thinks’: Understanding opacity in machine learning algorithms”. Big Data & Society 3(1): 1-12.

Hildebrandt, Mireille, and Serge Gutwirth. eds. 2008. Profiling the European citizen: Cross-disciplinary perspectives. Dordrecht: Springer.

Hoepman, Jaap-Henk, and Bart Jacobs. 2007. “Increased security through open source.” Communications of the ACM, 50(1): 79-83.

McDonald, Aleecia M., and Lorrie Faith Cranor. 2008. “The cost of reading privacy policies.” I/S: A Journal of Law and Policy for the Information Society. 4(3): 540-65.

Pasquale, Frank. 2015. The black box society: The secret algorithms that control money and information. Boston: Harvard University Press.

Raymond, Eric S. 2001. The cathedral and the bazaar. Sebastopol CA: O’Reilly.