

The fundamental rights challenges of algorithms

Netherlands Quarterly of Human Rights
2019, Vol. 37(3) 205–209
© The Author(s) 2019



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0924051919861773
journals.sagepub.com/home/nqh



Janneke Gerards

Utrecht University, the Netherlands

Abstract

Algorithms form an increasingly important part of our daily lives, even if we are often unaware of it. They are enormously useful in many different ways. They facilitate the sharing economy, help detect diseases, assist government agencies in crime control, and help us choose what series or film to watch. Yet, there is also a darker side to algorithms, and that is that they (and their applications) can easily interfere with our fundamental rights. This column explores some of the main fundamental rights challenges set by the pervasiveness of algorithms, and it presents a brief outlook for the future.

Keywords

Algorithms, Artificial Intelligence (AI), private life, freedom, equality, non-discrimination, procedural fundamental rights

Algorithms are everywhere. They have become instrumental in our everyday lives. They enable us to search the internet effectively, they allow us to find interesting new books, movies and music, and they can help detect certain diseases with great precision. Algorithms assist government bodies in making tax decisions, in crowd control, in police investigations and in the detection of social security fraud. Companies use algorithms in price determination and in making employment decisions. Algorithms, in short, are wonderful tools and in our present-day world, we cannot do without them.

Nevertheless, over time, we have come to recognise that there is a dark side to algorithms. As is often the case, it is this dark side that makes them highly relevant to lawyers – fundamental rights lawyers in particular. Their relevance for the law has to do with three main, interrelated, characteristics of algorithms. Put shortly, algorithms are non-transparent, non-neutral, human constructs.

Algorithms are human constructs in that they are created, programmed, trained and applied by human beings. As such, that is not problematic; it merely means that, in many ways, we have

Corresponding author:

Janneke Gerards, Utrecht University, Newtonlaan 231, 3584 BH Utrecht, the Netherlands.

E-mail: j.h.gerards@uu.nl

control over how they work and how they are used. However, it also implies that flaws in human reasoning and human decision-making may have a bearing on algorithms and their application.

This feeds into the second characteristic of algorithms, which is that they are unavoidably non-neutral. After all, the human beings responsible for designing and applying algorithms all hold particular values or beliefs, and they all have biases. These influence the choices humans make in creating an algorithm, even if this influence is completely unintentional. Similarly, as an effect of the prevalence of bias and stereotypes in human society, the data sets that are fed into algorithms are equally non-neutral. Moreover, the users of the algorithms will have biases too, which means they may easily misinterpret the outcome of an algorithmic application. As a result, it is almost unavoidable that subjective, non-neutral elements slip in at various stages of algorithmic decision-making.

This, then, is closely related to the third characteristic of algorithms, which is their non-transparency. Some algorithms may be relatively simple, in that they are a sort of ‘recipes’ that describe certain ingredients and processes of decision-making, and contain choices that we can see and understand if we try. However, many algorithms are much more complex, in particular if they are based on techniques of deep learning or machine learning. Such techniques allow an algorithm to search and analyse incredible amounts of data to discover patterns and correlations, which, in turn, may be the basis for decision-making. In some cases, this has the effect of decreasing the importance of the human factor; the ‘learning’ of the algorithm takes place ever more independently, in an ever more opaque process. The complexity and independence of these learning processes make it very difficult for outsiders (and even insiders) to understand, let alone explain, what is going on inside. This is why algorithms are often called ‘black boxes’.

Together, these three characteristics make that algorithms and algorithmic applications have an impact on a wide variety of fundamental rights. First, privacy rights may be affected. We use increasing numbers of laptops, tablets and hand-held devices, we have smart heating and electricity systems installed in our houses that are connected to the internet, and sometimes we even own smart fridges or smart toys such as talking dolls. These applications and ‘things’ generate a constant stream of data that can easily be intercepted and – using algorithms – connected, analysed and used, either by the government or by the companies that produce them. On the streets, too, our behaviour is monitored by many cameras and other tools that collect information. Algorithms and other increasingly advanced techniques, such as facial recognition, may help government authorities to detect risky situations and protect security. Clearly, all such devices and applications are useful and they may serve important public values, but they also affect our private lives. Indeed, we may start behaving differently if we are aware of all our preferences and movements being known to companies and the government alike, and our data being constantly compared to those of millions of other people.

Second, there may be an impact on rights such as the freedom of expression and access to information. The exercise of these rights is greatly facilitated by the availability of search engines, social media and internet forums. At the same time, our access to the wealth of available information is strongly determined by algorithmic analyses of our viewing, reading and clicking behaviour. These allow us to get access only to the information that we like to have access to – or, in fact, the information that an algorithm decides we would like to have access to. This generates the infamous ‘filter bubbles’, which may have the result that the information we receive is one-sided and coloured, instead of pluralist and balanced. Moreover, the information we get access to can easily be influenced and even manipulated using clickbait or bots, or by formulating messages and posts in a particular manner.

In addition, algorithms may be used to limit the freedom of expression in order to facilitate the protection of conflicting values, such as the protection of morals or reputation, or respect for intellectual property rights. This may have beneficial effects in that it helps fight the spreading of expressions we do not find acceptable, such as child pornography or hate speech. At the same time, the freedom of expression is not only intended to protect inoffensive expressions, but also those that 'offend, shock or disturb'.¹ Because of their combined characteristics, it is far from certain that algorithms can fairly and understandably classify a picture as an offensive, yet satirical or artistic and therefore permissible expression, or an unacceptable glorification of violence. Moreover, if there is a risk that some offensive or shocking expressions will be removed, people might feel discouraged from posting them. However beneficial algorithms are in this respect, they thus may have chilling effects and pose a threat to the freedom of expression.

Third, algorithms may have discriminatory effects. Surely the right to non-discrimination does not require that everyone is treated the same. Indeed, in many cases this would create unfair situations, because all human beings are different. In fact, it would be best to treat everyone in accordance with their own, uniquely personal characteristics, merits, needs and behaviour. Interestingly, algorithms are good at detecting differences between individuals and their behaviour, and at calculating how someone ought to be treated so as to take account of these differences. Accordingly, algorithms constantly discriminate, and often this seems fair enough. This can be seen, for example, in price discrimination on the internet. If a man sees different prices in a web shop than a woman, for the same product, it might just be that men buy different things than women do, and they are willing to pay different prices for this. An algorithm then simply allows the price to reflect those differences.

The point is, however, that the non-discrimination principle is not just about ensuring individualised and differentiated treatment, but it is also about our moral and ethical compass. There is a moral consensus that it is not (or only rarely) acceptable to base decisions on characteristics like gender, ethnicity or sexual preference, even if an algorithm finds they are relevant. In addition to this, many of the correlations algorithms detect are, the result of historical discrimination and of flaws and biases in our thinking and reasoning. Indeed, if an algorithm detects a seemingly relevant correlation, this may be the result of long-standing stereotyped views and prejudice in society. For example, it is well-known that historically, in many fields mainly men were appointed in high functions. If an algorithm is then used as part of a hiring policy to detect potentially successful employees, and 'successfulness' is defined in terms of the likeliness of being promoted, the algorithm will probably select only male candidates. Thus, even if it seems that an algorithm is simply being honest and reflects existing individual differences, the principle of non-discrimination may require that we reject its outcomes.

Finally, algorithms may affect procedural fundamental rights. To generate procedural justice, it is essential that in court proceedings, all necessary information and evidence comes to the table. All parties to a case should have equal access to that information and all arguments and evidence should be carefully considered by an impartial and independent court that can make an objective and unbiased decision. Algorithms can be of great value in this context. For example, the use of algorithmic technologies to help determine a sentence or assess the risk of re-offending might result in better-informed, more consistent, and more objective judgments. There are also some

1. See famously *Handyside v the United Kingdom* App no 5493/72 (ECtHR 7 December 1976) para. 49.

risks here, however, which derive from the three interconnected characteristics of algorithms. Previous judgments (and the sentences provided in them) are based on human interpretations and decisions, which are necessarily subjective and are subject to change over time. Similarly, risk assessments may be based on stereotyped or biased data. If an algorithm is not sufficiently carefully programmed, or it is trained with coloured data, there is a risk that subjective views and decisions are fed into it. Because of the black box effect, the parties to a case or a judge may find it hard to detect any resulting biases. Moreover, even if a judge is aware of potential problems, the algorithmic outcome may involuntarily provide her with an anchor, which makes it psychologically difficult to deviate from it. Algorithmic tools may thus end up having a great influence on a judgment, without there being much possibility to check if that is really fair.

In addition, there is an issue of equality of arms. The right to a fair trial requires that both parties to a court case should be able to participate in the proceedings on an equal footing. This may turn out to be difficult if a case concerns an algorithm-based decision, such as a decision to refuse a tax return or an employment decision. In such cases, one of the parties usually knows much better how the algorithm works than the other does. The resulting imbalance to some extent may be compensated for by inviting technology experts into the courtroom to explain the algorithm, but this is difficult if the algorithm is protected by a trade secret or a non-disclosure policy. Perhaps this could be solved if the court would play an active role in determining the fairness and quality of the algorithm-based decision-making. Still, however, the black box problem then persists, and even if that can be solved, there will be a strong need to increase judicial expertise in understanding and dealing with algorithms.

This short *tour d'horizon* demonstrates that there is indeed a dark side to algorithms. Policy-makers, legislators and courts are increasingly confronted with the fundamental rights challenges of algorithmic decision-making. In addition, some cross-cutting issues can be identified that affect the protection of all such rights. In particular, most fundamental rights instruments are still designed to deal with so-called vertical relationships, in that they aim to protect individuals against harm done by the State or by State actors. The examples discussed above show that, indeed, there continues to be a need for that protection, even more so because the availability of algorithms has exacerbated some risks of government interference. The example of dataveillance comes to mind, just like that of discrimination in policy-making and legislation in the sphere of criminal prosecution or the detection of fraud. Combined, this makes that there is a clear need to be aware of algorithmic risks and ensure that our legal systems contain sufficient checks and balances to deal with them.

However, the examples provided also show that protecting against government behaviour is not enough. There is an increasing need for protection of fundamental rights against private parties, ranging from discriminating employers and web shops to internet service providers and social media that determine our access to information and collect our personal data. Many legal mechanisms are not designed to deal with fundamental rights violations in such 'horizontal' relations. Although there are important exceptions, most systems of protection of fundamental rights are still very much based on the traditional distinction between the public and the private, or between vertical and horizontal relations. To deal with the fundamental rights impact of algorithms effectively, we either should get rid of that distinction completely, or we should at least make sure that also in horizontal relations, effective protection of fundamental rights is offered to a much stronger degree.

So what, then, should be our outlook? Clearly, there is a need to adapt our legislation and our legal mechanisms. They should be able to deal more effectively with the typical characteristics of

algorithms, they should better reflect and accommodate the increasing intertwinement of the public and the private, and they should find solutions for the newly arising conflicts between fundamental rights and other public values and interests. We are only at the beginning of finding out how this can be done, and it needs a joint effort of many people to take the next step. In designing practical solutions to the procedural inequalities in court proceedings, it may be crucial to team up legal, ethical, socio-psychological and technical experts. Similarly, we are in clear need of interdisciplinary training and education so as to ensure that lawyers understand how technology works, and technical experts comprehend the intricacies of the law. The many studies currently being conducted into such topics and an increasing number of fundamental right scholars that obtain an interest in the effects of artificial intelligence, big data, algorithms and the like are a step in the right direction. In the end, it is their combined effort that will allow us to make the most of the many benefits of algorithmic decision-making, while also finding a satisfactory answer to the fundamental rights challenges involved.

Author contribution

Janneke Gerards is Professor of fundamental rights law at Utrecht University, the Netherlands (see www.uu.nl/staff/jhgerards). This contribution is loosely based on the study she conducted into the impact of algorithms on fundamental rights together with Max Vetzo and Remco Nehmelman, which also contains numerous references to the sources of the examples provided in this short piece. The original (Dutch language) version of this study can be found at <https://www.rijksoverheid.nl/documenten/rapporten/2018/03/01/algoritmes-en-grondrechten>.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.