The quality, fairness, transparency and accountability of algorithmic decisions

Serge Abiteboul,
Autorité de Régulation des Communications Électroniques et des Postes (ARCEP),
Institut National de Recherche en Informatique et en Automatique (INRIA)

Abstract:
We consider aspects, especially technical, of the quality, fairness, transparency, and accountability of algorithmic decisions.

Computer science is accelerating scientific discoveries and innovations while transforming our lives and society.[1] Software programs are increasingly involved in making decisions that affect our lives in the fields of health, justice, banking, etc. (ABITEBOUL & DOWEK 2017). They choose the information to which the social media and search engines expose us. Algorithms have assumed a form of authority. By assuming (or not) the responsibility that should come along with this authority, they are putting users in the position of either trusting or distrusting them. The authority-responsibility-confidence triangle is inherent in algorithmic decision-making. In the social media for example:

- AUTHORITY: To cope with the toxic comportment of some users or even governments, some social networks have blocked contents, closed access to webpages or groups, or even banished users from their platforms. Should we let them the responsibility of defining our society's values? What legitimacy do they have for doing this?
- RESPONSIBILITY: Although the social media provide us with information and allow us to enter into exchanges and weave new bonds, the list of abuses is long: pedopornography, fake news, hate messages, harassment…. As participants in civic affairs, should we not require a more responsible comportment?
- CONFIDENCE: Citizens are ever more objecting to the social media. What to do so that they can enjoy the benefits without having to put up with the nuisance? so that they place, once again, confidence in these networks?

This triangle leads to questions about the quality, fairness, accountability and explicability of the decisions made by algorithms, the questions dwelled on herein. The problem with algorithms is that we do not know what they actually are, nor how they work. A minimum of education is needed to peek inside these black boxes. Although this is a necessary condition for placing confidence in algorithmic decisions, this article does not deal with this essential aspect.

[1] This article has been translated from French by Noal Mellott (Omaha Beach, France). The translation into English has, with the editor's approval, completed a few bibliographical references. All websites were consulted in June 2021.

The quality of decisions

We often enough face errors made by software. How could it be otherwise? A professional proofreader skips spelling errors in a novel of a few thousand words. How could we expect perfection from a piece of software, which (like Windows XP) contains 40 million lines of code and has been coauthored by a battalion of programmers? Besides, even though a software program's results derive from the application of logical rationality, the latter might make a mistake because its computing (algorithm) is incorrect, or the data used are incomplete, erroneous or skewed, the software has been poorly used, or the problem itself is too complex or has not been adequately analyzed.

When a software program makes an important decision in the place of human beings, serious questions arise about quality. We are tempted to set high standards. For example, some people think it inadmissible for a driverless vehicle to take to the road if it risks causing a fatal accident. Since software programs are simpler to analyze than human beings, and simpler to correct, we legitimately require that the aforementioned risk related to driverless vehicles be statistically very low. But would it not be disproportionate to require perfection from such vehicles while we accept to share the road with drivers who are under the influence of alcohol or who have poor eyesight?

Furthermore, software is increasingly being put to use in fields (typically related to the social or human sciences) where concepts are complex and hard to formulate precisely, where the best human experts hesitate or disagree. In these fields, machine learning and (big) data analytics are often the only way we now have of coming up with answers (ABITEBOUL & PEUGEOT 2017). In this case, the quality of decisions depends no longer on computer code alone but also on the data that guide this code. For instance, the results of the analyses made of COMPAS — a case management program that assists US courts in making decisions about conditional discharges — prove that this program is of very poor quality (YONG 2018).

Fairness

When proposing decisions, software programs must comply with the law, of course. But we expect something more. Since they are now part of life in society, they should also be in line with our ethical values. They should be fair and unbiased. In *Weapons of Math Destruction*, Cathy O'Neil (2016) has fully explained how biases can be introduced when data are analyzed.

For those who control the design of algorithms, the temptation to silently introduce biases is strong (DUCOURTIEUX & PIQUARD 2019). Besides, biases might not be introduced deliberately. On Stapples.com, staplers were priced higher in poor neighborhoods without the firm even being aware of this, because the software being used correlated prices with the distance from a store selling staplers and there are few such stores in poor neighborhoods (VALENTINO-DEVRIES 2012). Machine learning might also bias decision-making. Some such biases come from the inadequacy of the data used to "train" the software, as has been reported for facial recognition (SIGNORET 2018). Biases might also crop up because the digital realm is satisfied with holding up a mirror to certain aspects of reality. For instance, COMPAS's decisions about conditional discharges, by reproducing the biases of certain judges, discriminated against ethnic minorities (LARSON *et al*. 2016). Mention should also be made of the algorithms for rankings based on popularity (YANG *et al*. 2019), which are used by dating websites (*e.g.*, Match.com) or for crowdfunding (*e.g.*, Kickstarter). The bias introduced by such a ranking might result in less diversity, in discrimination and even the exclusion of certain users: "The rich get richer, the poor get poorer."

Fairness refers to the relation between an individual and a social group, but another aspect of interest is the relation between an individual and the software itself. Software programs are offered with features, each of which represents a promise to users. We expect a piece of software to obey the rules it has set. If a platform for recommending restaurants announces that its ratings are based exclusively on its users' opinions, it would be treacherous for the ratings to be skewed in favor of the restaurants that purchase its services.

This is but a sketch of the concept of responsibility in relation to the decisions made by algorithms. This responsibility is essential for confidence-building, and must go in hand with transparency and verification.

Transparency and verification

Confidence in algorithms can be built by understanding what they do and thanks to their transparency. When forced to make algorithms transparent, firms or governments have to make their choices visible, and maybe even reveal their errors or mistakes. Let us take the case of how students in France are allowed to enroll in establishments of higher education as a function of their preferences and the places available. For a long time, this decision was made without any transparency by persons unknown. Replacing these people with software programs (at first APB and now Parcoursup) has led to fairer decision-making. Once the software code was placed in open access, the procedure became transparent. A considerable advantage is that it is now possible to study the rules that guide the decisions made and to object to them. Transparency opens the door for debate.

The transparency of algorithms is gaining ground in Europe. The coming Digital Service and Digital Market acts should enshrine this principle. It is worthwhile pointing out the contrast between the transparency increasingly expected of the Internet giants or of governments and the relative opacity advocated for citizen privacy in regulations such as the EU's General Data Protection Regulation (GDPR). This different treatment is intended to correct, at least in part, the current asymmetry in information between these players and to restore citizen confidence.

Transparency alone does not suffice however. For instance, Debian OpenSSL package (for SSL certificates) contained an error that left its security keys vulnerable. The transparency of this software (its code in open, free access) left open the possibility that someone might examine the code and find the error; but it took two years before someone actually did this and found the "bug". Transparency facilitates but does not replace verification. The makers and sellers of algorithms can always declare their good intentions; but only the verification of their algorithms can convince us that the decisions made are of good quality, that they uphold ethical values and are trustworthy.

In many cases, the formal specification of a software program's properties is often an impediment, even though this action involves human interventions. The aforementioned COMPAS software was demonstrated to be fair for one specification but unfair for another even though both these (apparently convincing) specifications could not be satisfied at the same time (KLEINBERG 2018).

To verify the actions of algorithms, we can either analyze their code and the data they use, which is similar to proving a mathematical theorem, or else examine their effects, which is more like studying physical or biological phenomena, such as the climate or the human heart. The first approach means having access to the code and data (*e.g.*, the training data) that guide the code. This can be done if they are in open access. Otherwise, we have to turn toward more cumbersome methods for auditing the software. Let me insist on the complexity of this verification.

While several results have already been obtained from verifications of software in matters of safety (such as guaranteeing that the software for automatically piloting subway lines will not falter), very few studies have yet focused on verifying the properties related to the fairness of the decisions made by algorithms. Examining the effects relies on observations (often statistical) of the software's actions. In this approach, the algorithms are seen as a black box. This does not make matters easier. The automated tool AdFisher detected that, for Google ads, "*setting the gender to female results in getting fewer instances of an ad related to high paying jobs than setting it to male*" (DATTA *et al*. 2015). To detect this bias, AdFisher made identical profiles that differed only by gender and examined the job ads reported.

Explicability and disputed decisions

While the transparency of algorithms tells us something about their actions in general, an individual, when faced with a decision made by an algorithm, might want an explanation of his "particular case".

In 2016, both the French Digital Republic Act and the EU's GDPR introduced requirements about explicability.[2] This seems natural: when decisions affect us, we want to know why. When human beings make a decision, we have to be satisfied with a few, often unverifiable, explanations. When algorithms make the decision, it is technically possible to require much more of an explanation. If a doctor relies on an information system to make a general diagnosis of a patient, he cannot be satisfied with the response "appendicitis". He needs a justification, the reasons that led to this diagnosis, the statistics underlying it, etc. He needs these explanations to accept the diagnosis, inform the patient, and eventually propose other examinations or ask a colleague for an opinion. For many decisions made by algorithms, the same need exists for making a decision intelligible and, thereby, acceptable. Obtaining such explanations is not always simple. In particular, the algorithms produced through machine learning, which come out of an enormous volume of operations and the fine-tuning of numerous parameters, are often hard to explain. This technique is used in medicine to detect tumors. But for general diagnoses, techniques are preferred that are based on rules that propose explanations.

Just as we can object to a judge's decision and appeal it, we ought to be able to object to a decision made by an algorithm, since it, too, can make mistakes. On social media such as Facebook, the user may now object if a moderator takes down or blocks contents. Since it was criticized because this process was handled internally, Facebook has even set up an independent "oversight board" that may change its policy for moderating posts. The fact that this board is like a "supreme court" tends to reduce a little more the difference between social networks and governments.

[2] Respectively: Act n°2016-1321 of 7 October 2016 for a "digital republic" available at https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746; & The GDPR (General Data Protection Regulation): "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data" available via: http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1478961410763&uri=CELEX:32016R0679.

Conclusion

Hesitancy before the decisions made by algorithms is sometimes set down to the lack of robustness and transparency of the algorithms, biases in the software, or several other reasons, often related to the relative recency of these techniques and their uses. As pointed out, transparency, verification and explicability help build up confidence in algorithms. Let us not be naive however: the interests at stake are so huge that we can hardly imagine that citizens alone will be able to compel big tech to see to the ethical behavior of their software programs. Government must contribute to this process through laws and regulations. We can see possibilities for rules of this sort by examining those proposed by the Facebook Mission (ABITEBOUL *et al*. 2019, ABITEBOUL & CATTAN 2020).

For both individuals and society, the control of algorithms is a concrete question, specific to the context, about the authority to be granted to them and about responsibility and liability. However no degree of transparency or perfection of algorithms can by itself overcome hesitancy.

Questions also crop up about how much decision-making human beings want to keep for their own. We will soon have driverless cars. Some people reject them outright since they consider that no longer driving means that we have lost freedom (as we become dependent on machines) or that we are being treated like children (who do not know how to drive). For people who are unable to drive safely (owing to their age or for other reasons) however, driverless vehicles will represent progress, a completely independent form of mobility. Opinions differ!

Do we want robot nurses, robot judges, a robot police? In each case, there is no simple answer, since our sense of humanity is at stake. Furthermore, the answer changes over time as we learn to live with ever more sophisticated software programs. It is worthwhile repeating that we cannot hand decision-making over to algorithms out of laziness or convenience. This action should ensue from a collectively made choice.

References

ABITEBOUL S. & CATTAN J. (2020) "Nos réseaux sociaux, notre régulation", *Revue européenne de droit*, 1, September, pp.36-44, available at https://legrandcontinent.eu/fr/2020/04/07/nos-reseaux-sociaux-notre-regulation/.

ABITEBOUL S. & DOWEK G. (2017) *Le temps des algorithmes* (Paris: Le Pommier).

ABITEBOUL S. & PEUGEOT V. (2017) *Terra data. Qu'allons-nous faire des données numériques?* (Paris: Le Pommier).

ABITEBOUL S., POTIER F., BERBAIN C., GOURDIN J.B., MARTINON J., SCHWOERER G. & SIGNOUREL A. (2019) "Créer un cadre français de responsabilisation des réseaux sociaux. Agir en France avec une ambition européenne", May, a report to the Secrétaire d'État en charge du Numérique, 34p., available via https://www.economie.gouv.fr/files/files/2019/Mission_Regulation_des_reseaux_sociaux-Rapport_public.pdf.

DATTA A., TSCHANTZ M.C. & DATTA A. (2015) "Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination", *Proceedings on Privacy Enhancing Technologies*, 16p., available via https://www.cs.cmu.edu/~mtschant/publications/1408.6491v1.pdf.

DUCOURTIEUX C. & PIQUARD A. (2019) "Concurrence: L'Europe inflige à Google une troisième amende, d'un montant de 1,49 milliard d'euros", *Le Monde*, 20 March, available at https://www.lemonde.fr/economie/article/2019/03/20/concurrence-l-europe-inflige-a-google-une-troisieme-amende-d-un-montant-d-1-49-milliard-d-euros_5438780_3234.html.

KLEINBERG J. (2018) "On algorithms and fairness", Collège de France, available at https://www.college-de-france.fr/media/claire-mathieu/UPL8987154333826643969_KleinbergColle__geDeFrance.pdf.

LARSON J., MATTU S., KIRCHNER L. & ANGWIN J. (2016),"How we analyzed the COMPAS recidivism algorithm", *Pro Publica*, 23 May, available at https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

O'NEIL C. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown Books).

SIGNORET P. (2018) "Une étude démontre les biais de la reconnaissance faciale, plus efficace sur les hommes blancs", *Le Monde*, 12 February, available at https://www.lemonde.fr/pixels/article/2018/02/12/une-etude-demontre-les-biais-de-la-reconnaissance-faciale-plus-efficace-sur-les-hommes-blancs_5255663_4408996.html.

VALENTINO-DEVRIES J. (2012) "Websites vary prices, deal based on users' information", *The Wall Street Journal*, 24 December, available at https://www.wsj.com/articles/SB10001424127887323777204578189391813881534.

YANG K, GKATZELIS V. & STOYANOVICH J. (2019) "Balanced ranking with diversity constraints", *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, available at https://par.nsf.gov/biblio/10111428-balanced-ranking-diversity-constraints.

YONG E. (2018) "A popular algorithm is no better at predicting crimes than random people", *The Atlantic*, 17 January, available at https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/.