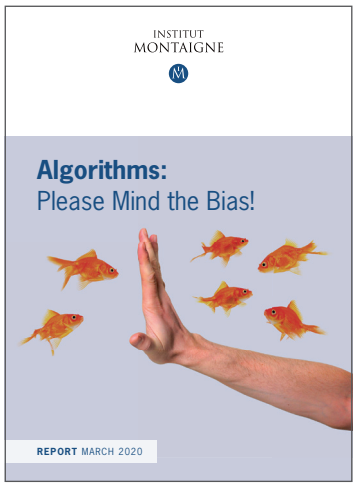

Executive Summary

Findings and challenges

Despite the risk of bias in some cases, algorithms in many ways actually represent an advance where discrimination is concerned. Men and women are often consciously or unconsciously biased, inconsistent in their decisions. **Using an algorithm means formalizing rules that apply to everyone, measuring the results, and ensuring that no bias exists.**

Algorithmic biases leading to discrimination are rarely due to an incorrect code in the algorithm. Incomplete or poor-quality data, or data that reflect biases in society, are much more often at the root of such biases. The fight against algorithmic bias is therefore above all a fight against discrimination that already exists on a daily basis. The challenge is not only to produce algorithms that are fair, but also to reduce discrimination in society.

This battle is difficult for numerous reasons. First of all, **it is not simple to define a bias-free algorithm.** While some biases are voluntary, such as promoting need-based scholarship recipients in the school admissions, others are involuntary or ignored, leading to discrimination against certain groups.

A fair algorithm, i.e. one that treats users fairly, is close to an algorithm without bias. However, this can never be fully guaranteed. Taking equity into account does not help designing unbiased algorithms, since equity can take different forms. Assessing what is fair involves an inherent cultural dimension and depends on each situation. The ethical attitude will not be the same in the case of one algorithm that analyzes lung X-rays and another that recommends political advertisements. In addition, total fairness between individuals and complete fairness between groups are fundamentally incompatible. **There will always be societal and political choices to be made.**

Then, **correcting an algorithm to make it fair often means reducing its performance with respect to its initial design criteria.** When we develop an algorithm, we choose one or many metrics allowing to optimize it and assess whether it reaches its goal. Adding constraints means limiting the capacity to optimize the algorithm vis-à-vis its initial performance criteria. It is always more difficult to pursue many goals at the same time rather than a unique one. It will therefore be difficult and costly for many actors to combat discriminations caused by algorithms.

Finally, **combating algorithmic bias means achieving an equilibrium between protecting citizens against discrimination on the one hand, and giving the possibility to experiment, crucial to the digital economy, on the other.** Restricting the use of algorithms, on suspicion of biases means depriving ourselves of new tools that could make our decisions more objective. It means curbing the growth of the French digital industry and accepting American and Chinese technological superiority in the long term. Adopting a laissez-faire approach would mean ignoring the destructive potential of such innovations for our social fabric.

Recommendations

Faced with these issues, we must be clear: we recommend neither a law against algorithmic bias common to all sectors of activity, nor a systematic check by the State of the absence of bias in algorithms.

Numerous texts that deal with discrimination already exist. They apply to both the physical and digital worlds and are likely to limit the risk of bias insert period. In view of society's limited hindsight in this field, a specific law on algorithmic bias would risk inhibiting innovation without actually solving the underlying problem.

The General Data Protection Regulation (GDPR) has shown that the use of personal data is far too widespread for a public agency to be able to verify all such data before they are used. We believe that the same will be true for algorithms and that it is illusory to expect the State to check each and every algorithm to ensure that they are ethical before they are implemented.

We have endeavored to formulate recommendations that are as realistic as possible in order to allow the rapid development of new technologies within a framework that respects our lifestyles.

Test the presence of bias in algorithms in the same way that the side effects of medication are tested

Like new drugs, it is complicated to understand how all algorithms work, especially those based on artificial intelligence. Furthermore, understanding how they work does not guarantee that they will be bias-free. It is ultimately by testing for the absence of bias that we will create confidence in the fairness of algorithms.

Testing the fairness of an algorithm has a cost and requires test data that specifically include some sensitive information (gender, social origin). Algorithm developers and purchasers will need to incorporate this constraint, and implement functional or performance testing to ensure the absence of bias. In some cases where the creation of these databases is difficult or problematic, the State could be responsible for their compilation.

Promote active fairness, rather than hoping for fairness by ignoring diversity

In order to combat discrimination, France has long chosen to blindfold itself, to see nothing of individuals beyond their status as citizens. As far as algorithms are concerned, this approach is no longer sufficient. An algorithm can introduce biases against women, even if the gender of the variables used has been explicitly excluded: it is easy to "guess" gender from other information such as buying women's products. To combat discrimination, it must therefore first be possible to detect it.

We need to move from an approach that hopes for fairness through unawareness to one of active fairness. We need to accept that the fairness of an algorithm is not achieved by excluding all protected variables such as gender, age or religion. On the contrary, it is obtained by including them, and by testing the independent nature of the result with respect to these variables. To achieve this, it is necessary to have access to this protected information. But if this information is protected, it is precisely because it can be a source of discriminations. The collection and use of this sensitive data must therefore be strictly supervised, and be limited uniquely to the purpose of testing and restricted to a sample of the users concerned. Moreover, such an approach would have to be the subject of an impact study declared beforehand to the CNIL (the French data protection authority).

Require greater stringency for high-risk algorithms (fundamental rights, security, access to essential services)

The sensitivity of an algorithm with respect to society obviously depends on its sector of activity, but above all on its potential impact on citizens. This impact is significant when the algorithm can restrict access to essential services such as a bank account or a job, endanger security (health, police), or violate fundamental human rights. These areas are already subject to strong discrimination obligations. When an algorithm is introduced in these areas, it cannot be at the cost of lowering requirements.

For these algorithms, we recommend an ad hoc framework integrating transparency obligations with regard to the data used and the objectives set for the algorithm, as well as a right of appeal against the decision taken. The creation of such a framework does not require a new law on algorithmic bias, but rather the implementation of good practices in companies and administrations, the use of existing legal provisions, and the addition of provisions in sectoral legislation on a case-by-case basis.

Ensure team diversity within algorithm design and deployment projects

Beyond these four strong recommendations that we are putting forward, we are convinced that a great deal of work remains to be done in terms of training. This concerns researchers and developers, of course, especially in the area of algorithmic bias, but also leaders and citizens within the more general framework of artificial intelligence, so that everyone can take ownership of both the opportunities associated with this technology and its inherent risks.

This vigilance should also be reinforced in organizations implementing charters and best practices. These initiatives, which we noted during our interviews, must be encouraged as, together with technical and operational measures, they would make it possible to generate collective awareness around the dangers of algorithmic bias.

Finally, vigilance must be external and, in the case of high-risk algorithms, it would seem judicious to strengthen controls. This could be done firstly via the issuing of labels, the emergence of which should be supported. Such labels would guarantee the quality of the data used and of the organizations developing the algorithms, the existence of control procedures, and the auditability of these algorithms. The industrial sector would notably need such guarantees in order to take full advantage of the algorithmic revolution. Furthermore, with respect to high-risk algorithms, the ability to audit and monitor certain requirements could be entrusted to a third party or to the State.