



Universiteit Utrecht



MONTAIGNE
CENTRUM

VOOR RECHTSSTAAT
EN RECHTSPLEGING

S. Kulk & S. van Deursen, *Legal aspects of algorithmic decision-making. An exploratory study*, The Hague: WODC 2020

Summary

1. Motivation and research question

In our daily lives we all encounter decisions that are made with the help of algorithms. The use of algorithms creates opportunities for the fulfilment of public values and interests. Algorithms, for example, can make decision-making processes more efficient and contribute to finding solutions to various challenges in society. At the same time, the use of algorithms can involve risks and give rise to the question whether the applicable legal frameworks are future proof and able to safeguard public values and interests. Considering the above, the research question of this study is defined as follows:

What opportunities and risks exist in algorithmic decision-making in relation to the protection and realization of public values and interests, and are the current legal frameworks sufficiently future proof to realize opportunities and prevent the occurrence of identified risks or mitigate the consequences of such risks?

The current use of algorithms in decision-making processes and anticipated developments in the next five to ten years constitute the main focus of the study. To help answer the research question, the use of algorithms was examined in case studies in four areas selected by the Research and Documentation Centre (WODC) and the Legislation and Legal Affairs Department of the Ministry of Justice and Security in the Netherlands. These areas are content moderation, self-driving cars, the judicial system and the collection of traffic fines.

2. Algorithmic decision-making and public interests

Algorithmic decision-making as set out in this study, encompasses all processes in which an algorithm is used to take decisions that affect the legal position of persons or decisions that affect their interests in some other way. This can occur when an algorithm takes a decision autonomously, or when an algorithm is included as part of a human decision-making process. Types of algorithmic decision-making vary depending on the objective of the algorithm. From a technical perspective, a distinction can be made between rule-based algorithms and self-learning algorithms. A final important factor is the organizational and social context in which algorithms are used.

Public values and interests serve as a normative framework for identifying opportunities and risks as well as for assessing whether the legal frameworks are future proof. In this study, public values and interests are operationalized via three fundamental rights that are almost always at stake in the various uses of algorithms: the right to data protection, the right to non-discrimination and the right to legal protection. In the different case studies, more specific values such as sustainability or freedom of expression also play a role.

3. Opportunities and risks from algorithmic decision-making

In this study, opportunities are defined as the possibility to fulfil public values and interests, while risks are defined as the possibility that public values and interests are not fulfilled, or are actually harmed. In relation to the question of whether the legal frameworks are future proof, the central issue is to what extent they facilitate the realization of opportunities and avoid or mitigate the occurrence of risks.

Opportunities and risks for public values and interests are related to the type of algorithm that is used, but to an important extent also to the domain and the organizational and social context in which algorithms are used. It is of course possible that both opportunities and risks arise when an algorithm is used. In algorithmic decision-making processes, a balance must always be found between the needs in the area in question, and the values and interests that apply there. The legal frameworks provide the parameters within which such a balancing exercise should take place.

3.1 Opportunities from algorithmic decision-making

The use of algorithms can yield efficiency gains because algorithms are able to handle decision-making processes faster, better or more accurately. Self-learning algorithms, in particular, can discover relations in large amounts of data. The efficiency gains that are achieved in this way have mainly business and economic value, but can also contribute considerably to the realization of public values.

More particularly, the use of algorithms creates clear opportunities with regard to legal protection and non-discrimination. **Opportunities for realizing legal protection** are mainly connected to the efficiency gains that can be achieved by using algorithms, as is most evident from the case study on their use in the judicial system.

The use of algorithms also creates **opportunities for realizing the right to non-discrimination**. Since algorithms are capable of processing large amounts of information, they can include many individual characteristics of people in decision-making processes. In addition, algorithms can help to better tailor decisions to the persons in question, so that the use of algorithms can contribute to creating substantive equality. Algorithms can also contribute to equality by achieving consistency in decision-making. Moreover, well-programmed and validated algorithms are in principle better than humans in taking decisions without discriminating. Lastly, algorithms can also certainly be used in detecting discrimination in decision-making processes.

No potential opportunities for realizing the right to personal data protection were identified within the scope of this study.

3.2 Risks from algorithmic decision-making

Risks as a result of using algorithmic decision-making can arise for all three general public values and interests. The **right to personal data protection** is at stake when algorithms are used to collect or otherwise process personal data on a large scale. In addition, the use of algorithms enables the connection of information to discover (even) more about individuals. This leads to people not only losing control of their personal data, but also to the blurring of the line between what is considered to be personal data, and what is not. Not only that, if algorithms establish links that are incorrect, this can have an impact on the identity and reputation of the individuals concerned.

Risks in relation to the **right to non-discrimination** arise if algorithms incorrectly make a certain distinction, and also if they incorrectly do not make that distinction. This can, for example, happen if algorithms are overinclusive or underinclusive with regard to specific groups of people. Specific risks related to discrimination can also arise if bias or unacceptable stereotypes end up in an algorithm (consciously or unconsciously) via the programmer(s), or via others involved in the development of the algorithm. When self-learning algorithms are used, there is also a danger that the data that is used to train, validate, or test the algorithm are insufficiently representative of the group of people about whom a decision is taken. Another risk of discrimination arises when the data used are a reflection of social stigmatization, stereotyping or bias. When the algorithm is applied and new data are collected that are subsequently used to further train the algorithm, a feedback loop can occur that reinforces the discriminating effect.

The **right to legal protection** can come under pressure because of difficulties in explaining the functioning of algorithms. This can hinder transparency of the decision-making process and as a result also hinder possibilities to object to a decision. This difficulty in explaining how an algorithm works can be the result of several interconnected (rule-based) algorithmic systems, but it can also result from the use of self-learning systems that are inherently difficult to explain. It is also important to realize that the number of questions with regard to responsibility and accountability will increase as our society digitalizes further and interactions between systems increase. Moreover, such questions will then become more complex and therefore harder to answer. As a result of this lack of clarity concerning accountability and responsibility in algorithmic decision-making, or the decision-making chain, risks can likewise emerge in relation to the realization of effective legal protection.

4. Resilience of legal frameworks

Since opportunities and risks from using algorithms are very much related to the domain in which they are applied, the resilience of the legal frameworks should also be assessed while taking account of the specific needs of these domains. In this respect it should be noted that a specific application of an algorithm in a decision-making process is always partly regulated by domain-specific legal frameworks and partly by general legal frameworks.

It appears that the specific legal frameworks examined in the case studies do not directly stand in the way of realizing opportunities related to the public values and interests involved in this study. It is possible, however, that the legal frameworks do not provide an adequate basis to be able to benefit from the advantages of using algorithms.

In addition, it appears from the case studies that the specific legal frameworks examined in many cases are broad enough to avoid or mitigate the identified risks of algorithmic decision-making. However, it is often necessary that interpretations of the existing broadly defined standards are tailored to the specific use of the algorithm and the applicable values and interests. Nevertheless, the study also shows that specific legal frameworks, or parts of these frameworks, in some cases fail to prevent or mitigate risks.

As regards general legal frameworks such as the General Data Protection Regulation (GDPR), the General Administrative Law Act (in Dutch, Awb) and the general norms of tort law, from this study it can be concluded that they can accommodate technological developments related to algorithmic decision-making. Consequently, these general legal frameworks offer the room to be shaped further in a gradual and flexible manner in line with such developments. This requires, however, that courts and regulatory authorities provide clarifications or interpretations of the general frameworks that are tailored to the use of algorithms. The first tentative and important steps in this direction have already been taken.

Although there are no clear structural problem areas, various obstacles have been identified in relation to specific rules in the general frameworks. For example, the strict rules in relation to the use of certain personal data, such as data on ethnic origin and gender, can actually impede the detection of discriminating effects in algorithms. In addition, in relation to Article 22 GDPR, it is not clear to what extent a lack of human involvement is necessary in order to have protection from fully automated individual decision-making and to be able to invoke human rights.

5. Conclusion

The possible regulation of the use of algorithms calls for an integral (policy) assessment. Such an assessment should be aimed at establishing both the relevant opportunities and risks for public values and interests, and to determine whether – in light of requirements, standards, values, interests and context of a specific domain – it is possible to mitigate or avoid risks, while realizing the opportunities. The weighing of opportunities and risks and striking a balance between them is ultimately a political and policy-related process. The legal framework is the result of this process and can regulate and control whether and how algorithms are applied and thus influence the extent to which opportunities and risks are realized or avoided.

It is therefore important that legislation is formulated in such a way that responsible innovation in the field of algorithmic decision-making is possible. In particular, it is necessary to ensure that legislation does not focus too heavily on existing technologies in such a way that no room is left for future developments. Such legislation, after all, can only provide legal certainty and protection as long as the specific technology is actually regulated by the established rules. As technological developments occur, it is possible that the rules would no longer apply to the current situation and therefore no longer provide guidance in the new technological context. Furthermore, adopting technology-specific rules could obscure the underlying guiding principles which could ultimately have a detrimental effect on the protection of individuals. The public values that the legislature intended to protect, would then be vulnerable due to the fast pace of technological developments.

To conclude, the findings from this study suggest that it is not worthwhile setting up legal frameworks to focus on algorithmic decision-making in a general sense. The current general frameworks appear to have no significant shortcomings or structural problems when it comes to protecting the studied public values and interests. On the contrary, the study established that such frameworks appear to have considerable capacity for absorbing new developments. When new technologies are developed and applied, these general frameworks provide a considerable degree of direction. Furthermore, the case studies demonstrate that the opportunities and risks for all public values and interests examined, depend to a large extent on the domain and the organizational context in which an algorithm is applied. Thus, additional general legislation aimed at algorithms to combat the risks to public values and interests would have little added value. The opportunities and

risks for the realization of public values and interests must be primarily analyzed and weighed in light of the dynamics of a specific area and the character of the relationships between the parties involved. Therefore, where problems are experienced, these should be addressed as far as possible in relation to the domain in which they occur. Only then will it be possible to take proper account of the specific opportunities and risks related to the particular public values and interests that are at stake.