Eur J Secur Res (2018) 3:181–203
https://doi.org/10.1007/s41125-018-0031-2

ORIGINAL ARTICLE

On Chances and Risks of Security Related Algorithmic Decision Making Systems

Katharina A. Zweig[1] · **Georg Wenzelburger**[2] · **Tobias D. Krafft**[1]

Received: 2 November 2017 / Accepted: 12 April 2018 / Published online: 20 April 2018

Abstract Recently, various decisions in security-related processes are assisted by so-called algorithmic decision making (ADM) systems, e.g., for predicting recidivism rates of criminals, for assessing the risk of a person being a terrorist, or the prediction of future criminal acts (predictive policing). However, the quality of such risk assessment is dependent on many modeling decisions. Based on requirements of proper democratic processes, especially security related ADM systems might thus require societal oversight. We argue that based on democracy-based processes it also needs to be discussed and decided, how aspects of its quality should be assessed: e.g., neither the proper measure for racial bias nor the one for its overall accuracy of prediction is decided on today. Finally, even if the ADM system would be as objective and perfect as it can be, its embedding in an important societal process might have severe side effects and needs to be controlled. In this article, we analyze the situation based on a political science view. We then point to some crucial decisions that need to be made in the planning stage, questions that need to be asked when purchasing a system, and measures that need to be implemented to measure the overall quality of the societal process in which the system is embedded in.

✉ Katharina A. Zweig
zweig@cs.uni-kl.de

Georg Wenzelburger
georg.wenzelburger@sowi.uni-kl.de

Tobias D. Krafft
krafft@cs.uni-kl.de

[1] Department of Computer Science, Algorithm Accountability Lab, TU Kaiserslautern, Gottlieb-Daimler-Straße 48, 67663 Kaiserslautern, Germany

[2] Department of Political Science, Policy Analysis and Political Economy, TU Kaiserslautern, Pirmasenser-Straße 65, 67663 Kaiserslautern, Germany

Keywords Algorithmic accountability · Algorithmic decision making · Democracy · Justice system · Recidivism

1 Introduction

The integration of algorithms to handle or support security related decisions is steadily increasing for at least two reasons: (1) human expertise is costly, time consuming, and often rare; (2) algorithmic decision making (ADM) systems are assumed to render decisions more objective and transparent.[1] Common examples of ADM systems are predictive policing systems, which predict where which kind of crime is to be expected (see also Egbert 2018). In countries such as the USA, different security related ADM systems are developed or in use. For example, a system called SKYNET was presented to the NSA in a power point slide deck; with the help of decision trees, SKYNET tried to identify terroristic couriers based on smartphone data (The Intercept 2015). Similarly, various tools involving ADM systems are used to assess risk for recidivism of offenders (Desmarais and Singh 2013), of which Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is a popular example (Northpointe 2012). According to NGOs such as the Electronic Privacy Information Centre (EPIC), all US states use some ADM system at some point during the criminal justice process and the type of application as well as the evaluation of these systems vary greatly between states (EPIC 2017).

However, while ADM systems are increasingly used in the criminal justice system, the rules of the game, e.g., how ADM systems actually work, who decides upon their application, and to what extent their results are used in actual criminal justice decision-making are often nebulous or, at least, not very transparent. In part, this is due to the field of their application because the algorithmic systems support institutions in the criminal justice system which is, by nature, very secretive. In the literature, those challenges related to the use of ADM systems in the criminal justice system have been mostly discussed from the perspective of law (Steinbock 2005; Pasquale 2016; Kroll et al. 2017) and criminology (e.g., Wormith 2017) as well as from a Science and Technology Studies angle (e.g., Ananny and Crawford 2017). What is largely missing, however, is a political science perspective which emphasizes how ADM systems challenge basic principles of democracies. In fact, the increasing use of ADM systems is, from a political science point of view, problematic, because it is both very difficult to hold these systems accountable for their decisions, especially for the society at large (Diakopoulos 2015), and almost impossible to collect best practices on how to design and deploy these systems in an accountable way. Accountability is, however, a key feature of decision-making processes in

[1] The assumption that an ADM system is in and of itself objective and transparent ist not correct. In any case it is possible to construct them such that their decision process is transparent, sometimes only by the cost of simplifying their decision structure. They are objective only in the sense that people with the same properties will be judged the same, independent of the time of day or other typically human biases.

developed democracies (Przeworski et al. 1999). Accountability is especially crucial if decisions are of great importance for the life chances of individuals (Lischka and Klingel 2017). This is why the use of ADM systems in the criminal justice system, where individual civil rights like personal freedom are concerned, is of particular interest.[2] With this paper, we therefore aim at contributing a particular and new perspective to the discussion about the ins and outs of the use of ADM systems in the public decision-making process on the example of criminal justice.

In doing so, we explicitly take a realistic perspective on the question of how ADM systems can be assessed from a perspective of democratic theory. We are aware of the fact that this view is not entirely compatible with the social constructivist strand of science and technology studies. The latter has its clear merits given that technology cannot be seen as detached from the social context in which it is used, evaluated and interpreted (see the debate on a technological deterministic vs. a relational social-constructivist perspective on the interplay between technology and society (e.g., Parchoma 2014; Hutchby 2001). However, given that we use a theoretical lens which is rooted in a criteria-based model of evaluating democracies empirically, the Democracy Barometer, it seems appropriate to also discuss the ins and outs of ADM systems in a way that acknowledges the fact that there may be undeniable characteristics of such systems that challenge certain criteria that can be used to evaluate the "quality of democracy".

On the first glance, the criminal justice system might seem to be a perfect field of application for ADM systems, because many decisions require a categorization of people into two or more groups or a ranking of persons by some possibly problematic behavior or anticipated behavior. ADM systems promise to solve these problems by classification and scoring. An example for a classification problem is the above-mentioned categorization of criminals in classes with different risks of recidivism, an example for a scoring problem is a ranking of persons that need to be observed based on the probability with which they might carry out a terrorist attack. Any wrong solution to either of these problems can inflict two kinds of costs: if a potentially dangerous person is not detected, it might incur large costs on society (s. Fig. 1). If, however, an innocent person is accused and scrutinized based on a false alarm, this will incur great costs on that individual. Thus, especially for all security related ADM systems, decisions need to be made with a delicate balance between sensitivity and specificity, i.e., detecting most of the people with a potentially dangerous behavior and inflicting little cost on those people with no dangerous behavior. However, in most cases, there are no 100% rules by which persons can be separated into potentially dangerous and innocent persons or by which people can be ranked unambiguously by the potential of damage they might inflict on society. Thus, there will always be mistakes in either direction. Under the assumption that

[2] Admittedly, the criminal justice system is not the only field in which ADM systems are increasingly used by public actors. Other fields are, for instance, decision-making processes in social policies (Niklas et al. 2015) or selection-processes in higher education (Frouillou 2016). Nevertheless, security-related decisions seem to the most researched area in this connection, which is not surprising given the importance of these decisions for a human's life chances.

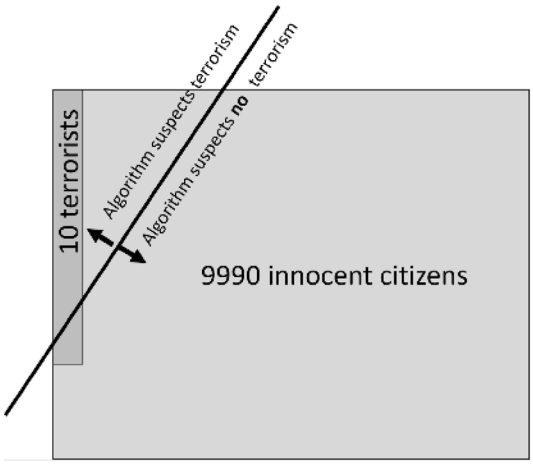

Fig. 1 If there are two types of citizens, terrorists and innocent people, an Algorithmic Decision Making System for terrorist identification searches for pattern in data where it is known which person belongs to which category. It deduces rules regarding the most important properties associated with terrorists. Given new data on people, the algorithm will decide for some that they are suspicious and for others that they are not. The percentage of found terrorists of all terrorists is called the sensitivity of the system, the percentage of correctly announced non-terrorists of all non-terrorists is called the specificity of the algorithm. Figure by Algorithm Accountability Lab [Prof. Dr. K. A. Zweig]/CC BY

an ADM system is an effective and acceptable way to make a decision, evaluating a system's success requires a quality measure by which society can measure whether the ADM system ranks people meaningfully or categorizes them correctly. For example, the above named COMPAS is mainly evaluated by the ROC AUC, a well-known measure in machine learning (Brennan et al. 2009), while the main quality aspect of SKYNET was measured by its low false alarm rate (The Intercept 2015).[3]

This introduction shows that there are multiple decisions that need to be made, and assumptions that need to be tested in order to design and deploy a normatively acceptable and effective ADM system. In the following, we will demonstrate some of the problems associated with such choices. As indicated above, democratic political systems follow certain well-established rules that are, to name two basic ingredients, based on accountability and the rule of law (see below). It is therefore crucial to evaluate the functioning of ADM-systems against a set of such standards in order to decide, whether features of ADM-systems are compatible with the basic rules of high-quality democracies.[4] In this commentary, we organize general problems in designing and deploying ADM systems according to a structure we call the long chain of responsibility (Zweig 2016a, b),

[3] It needs to be noted that in computer science, most systems are evaluated by a single measure rather than by an array of different and possibly conflicting measures. This is an intrinsic feature of all processes that use computers to find an optimal ADM system.

[4] See below for a more thorough discussion of the "quality of democracy".

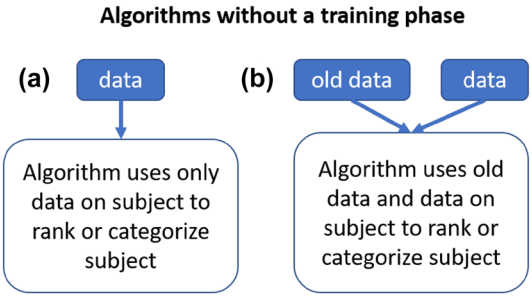

Fig. 2 Two algorithms without a training phase, a) uses only the new data, while b) uses old data for comparison. Such algorithms are said to be unsupervised. Figure by Algorithm Accountability Lab [Prof. Dr. K. A. Zweig]/CC BY

defined in Sect. 2. Section 3 then deals in more detail with the design of security relevant ADM systems. Section 4 summarizes the main sources of errors within the phases of designing and deploying an ADM system. This raises the question of how to deal with the implementation and embedding of security-related ADM systems in a democracy—this is discussed in Sect. 5. Finally, Sect. 6 gives a short discussion and hands out some recommendations which relate to the implementation process and transparency aspects.

2 Definitions

For this article, we will define an ADM system as any software that contains algorithms and produces a single output based on a set of input variables. The algorithmic component can contain a static algorithm, e.g., based on the rules of decision of experts in the field. An example for this are most German car insurance algorithms which categorize the owner of a car into different "Schadensfreiheits"-classes: it relies on a formula where the number of years without a car accident is the most important input, but, e.g., gender is not an input variable. In other cases, however, an algorithm is contained in the ADM, which uses a data set with known outcome, i.e., a data set in which the correct classification or the correct ranking is known. For this article, we differentiate two ways in which algorithms make use of the data: either directly or by training a so-called model on them (unsupervised and supervised machine learning, respectively).

For the first case (s. Fig. 2a), a typical example is the k-nearest neighbor algorithm. For any given data point, it searches for the k nearest data points in the data set with known classifications or rankings. From this, it directly computes the result on the subject at hand (s. Fig. 3).

In most cases, however, modern algorithmic decision making systems contain an algorithm from the class of supervised machine learning. They are assumed to be the most important class of algorithms as they are able to find intricate patterns of correlations within large sets of data. These are especially powerful, as they can deduct different kinds of rules and correlations from a large data basis to classify the corresponding entities. Again, their power relies on data with a known classification

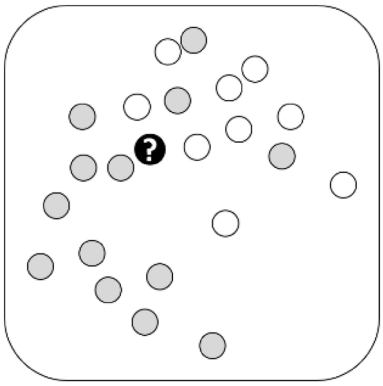

Fig. 3 The principle of *k*-means-clustering—an unsupervised learning algorithm—is based on data with a known categorization, symbolized by two different colors. Any new data point (black data point) is evaluated by its *k* closest neighbors. In this case, the nearest neighbor is grey, so for $k=1$, the black point would be assigned to the category represented by the grey color. However, for $k=3$, the majority of the neighbors would be white, so the black data point would be assigned to the "white" category. Figure by Algorithm Accountability Lab [Prof. Dr. K. A. Zweig]/CC BY

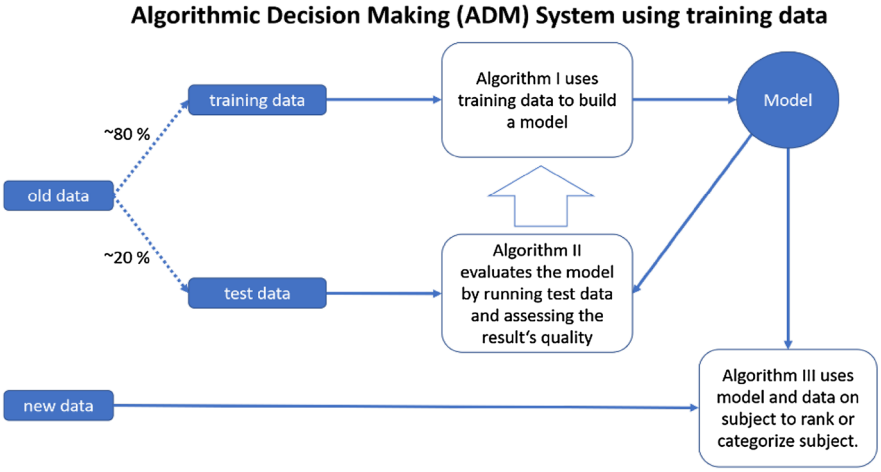

Fig. 4 ADM System where a model is first trained by using the feedback of a quality assessing algorithm and then used to actually compute the categorization or ranking. Figure by Algorithm Accountability Lab [Prof. Dr. K. A. Zweig]/CC BY

or ranking—the so-called training data set. With the deducted rules, new data points can then be classified. Thus, in this second class of algorithms, there are basically three subalgorithms involved in building and running the ADM system (s. Fig. 4):

1. The first one searches for structure in a set of data and deduces rules which categorize people or which rank them. This might be any classic algorithm from artificial intelligence or machine learning, e.g., a *k*-means clustering algorithm,

a random forest builder, or a neural network. Most of these algorithms behavior is determined by a set of tuning parameters which can be changed. Let now $P(h)$ denote the set of properties of some human h of a set of humans H. Then, a classifier is an algorithm which builds a structure that can be used to return one of a set of classes. A ranker is an algorithm which assigns a score to all elements—the score is used to sort and rank the elements. The result of these algorithms is a trained model, i.e., a way to determine the output given any possible input.

2. The second algorithm determines how well the trained model is able to categorize the data. Its result is fed back to the designer of the ADM system and might lead to further refinement of the values of the tuning parameters. This algorithm contains the choice of a set of quality indices and this choice is quite independent of the first algorithm, but dependent on the type of algorithm system (i.e., a ranker vs. classifier).
3. The third algorithm uses the final, trained model to classify or rank new data.

In this article, we will concentrate on those algorithmic decision making systems of the second type, like recidivism risk assessment systems or terrorist identification systems.

3 The Design of General and Security-Relevant ADM Systems

The nature of ADM systems in security relevant contexts is by definition secretive. Thus, not much is known about who orders them based on which decision process, how requirements on the software system are posed, what the design process looks like, and how it is evaluated whether the resulting system is good enough to be implemented in these societally important processes. In the following, some of these steps are discussed on the example of a terroristic courier identification system. After that, we introduce an abstract process model that depicts the general sequence of design phases and typical mistakes in designing ADM systems—note that this latter is not an ideal model but rather a description of how we perceive current design processes.

3.1 Terrorist Courier Identification by an ADM System

It is best to illustrate the challenges in designing and deploying ADM systems with an example: there is a set of Power Point slides from the Snowden leak, which describes an ADM system to identify terroristic couriers (The Intercept 2015). In these slides, the results of a classification algorithm that takes data points representing behavioral properties on the basis of smartphone-usage related data of 55 million inhabitants of a country is reported.[5] Finally, the system is asked to put each

[5] Such a classification can be based on a scoring algorithm where each person is assigned a score or probability to be a terrorist. In most cases, institutions will then define a threshold which defines the two classes: people with a score higher than the threshold and those with a score of at most the threshold.

data point (i.e., human) into one of two classes: suspect or non-suspect. The ADM system has the telling name "Skynet" after the fictional neural net-based conscious group system that features centrally in the Terminator franchise. From the slides, it can be deduced that there is a very small set of known "terroristic couriers" among the 55 million data sets—the slides call it 'a handful'. On this small set of known terroristic couriers, the algorithm designers trained a so-called random forest, a machine learning algorithm which results in a number between 0 and 1 for all data points. This means that the algorithm is a ranker. Thus, it is not enough to just train the random forest, it also needs to be decided above which value people are suspected of terrorism. This turns the ranker into a classifier. Based on the report, the value was chosen such that 50% of the known terrorist would be classified as "terrorists". I.e., given all 55 million data points sorted by the number assigned to them by the random forest and the known terroristic couriers, the threshold is chosen such that 50% of the known terrorists have a number at least as high as this threshold. Thus, the sensitivity of the algorithm is by design set to 50%. However, when this first version was evaluated according to step 2 above, it turned out that the results were not yet good enough as measured by the false positive rate, also called the false alarm rate: That is the fraction of the data points (i.e., humans) whose number is also greater than the threshold, but which are not known terrorists. The false-alarm rate is 0.18% for this first algorithm. Since the algorithm is applied to 55 million data sets of which almost all represent humans which are no terrorists, this means that about 99,000 people are falsely assumed to be terrorists. Thus, the first algorithm is neither very sensitive (detects only 50% of the terrorists) nor specific (many innocents are falsely detected). It seems that this first data base, containing only a handful of terrorists to let the algorithm learn from, was too unbalanced to generate a better distinction. In a second approach, the designers have increased the data base by also including people only suspected of terrorism but not yet known to be terrorists. The slides do not give a rationale for this decision other than the unsatisfying quality of the first trial's result. The size of this increased data set is unknown but when a random forest is trained on this new data set, the false alarm rate drops to 0.008%. This still means that out of 55 million inhabitants, 4,400 might be falsely suspected to detect only 50% of the known and suspected terrorists. However, as it is unclear how large the number of known terroristic couriers is, it is difficult to assess whether this number is reasonable. And, from a democratic perspective, it is very doubtable whether decisions based on such alarm rates would be normatively acceptable if the persons suspected have no further indication (1) why they are suspected and how they could appeal and (2) if the broader democratic public has no idea about the functioning of such systems (see below).

While the amount of information on the leaked slide desk is limited, it allows an insight into the quality evaluation and tinkering of the algorithm's parameters until a better quality is reached. For other ADM systems, as the COMPAS algorithm, not much is known about the process of its design and its evolution. However, from other, non-sensitive applications the general design process can be abstracted as discussed in the next section.

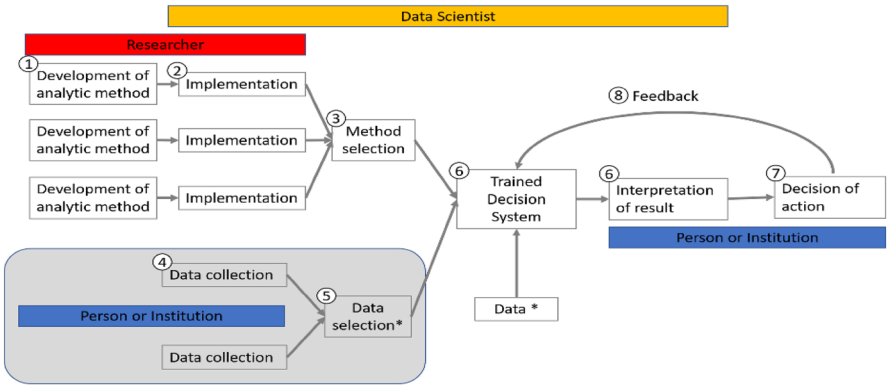

Fig. 5 The design of an algorithmic decision making (ADM) system requires the interaction of various persons and institutions. In a long chain of responsibilities various decisions are made that all have an influence on the final quality of the resulting system. Figure by Algorithm Accountability Lab [Prof. Dr. K. A. Zweig]/CC BY

3.2 Phases of Design and Deployment of ADM Systems

In the following, we describe an abstract phase model of how algorithmic decision making systems are constructed today. It has been abstracted from personal discussions with ADM designers, and written documentation of parts of this process.

In general, there are multiple persons and institutions involved in the design of ADM systems (s. Fig. 5). First, there is the institution demanding the development of an ADM system. In the above example, it might have been directly ordered and/or developed by a secret service, while predictive policing solutions are often bought from companies and applied by cities, while recidivism risk assessment is ordered by state governments. The institution in need for an ADM system often possesses the basic data on which the ADM system is later trained, together with the knowledge of which person in the data set showed which kind of behavior. In most cases, persons from the relatively new and largely undefined field of data science, so-called data scientists, are hired to actually build the ADM system. The data scientist normally then relies on software packages of some dozen classic algorithms, first described by researchers and implemented by computer scientists. In most cases, the algorithmic decision making system does not directly make a decision but supports the decisions of people in the institution that ordered the ADM system. For example, in Wisconsin in the case State vs. Loomis the judges are encouraged by the Wisconsin Supreme Court to use the recidivism risk assessment tool COMPAS before deciding upon a sentence (Freeman 2016, p. 89). The results of a predictive policing algorithm are used in police stations to organize the patrols but do not directly dictate them. In the terrorist identification, it is also most likely that the results are perused by a human decision maker to decide on some action against the suspects (like human surveillance or tapping in on communication). However, it is at least possible that such

an algorithm is working in an automatic drone which would trigger a gun when a person is "identified" with high enough probability.

These actors are now involved in different phases of the ADM design, implementation, and deployment. Over the last years, we have observed that the way ADM systems are designed, proceeds along different phases as detailed in the following (s. Fig. 5):

1. *Algorithm development* The basic algorithms for finding rules and correlations in data are in most cases already well-known. Many of them were developed in the last 20 years, some are much older (e.g., linear regression as a very simple predictor).
2. *Algorithm implementation* After publication, the algorithm is programmed in some programming language and possibly released in some software package or a general software. This is most often not done by the person initially describing the algorithm but by various programmers from private researchers to doctoral students or professionals.
3. *Algorithm selection* The data scientist selects the most suitable algorithm for the application. In the SKYNE example, they chose random forests, but there are dozens of other algorithms available.
4. *Data collection* The data collection might be specifically designed for the task of training the algorithm or can be of a more varied nature where various data sources are combined to a new data base. In the example above, the data seems to have been collected for the general purpose of the NSA, but not specifically for the task at hand.
5. *Data selection* Optional step if the data collection was not specifically designed for the task of the ADM. In that case, not all available data might be meaningful for the task, so a selection has to be made.
6. *Design and training of the ADM system* In this step, the data scientist(s) decide which parameters to set and train the model, based on the chosen algorithm and input set. After each training, the model is evaluated by some quality measure.

As seen above, the steps 4–6 may repeatedly be applied as long as the quality measure is not yet high enough. It is important to note that in almost all cases, it is not possible to find rules that allow a perfect classification or a perfect ranking. This is inherent to the social problem at hand: neither is there a set of properties that make a human a criminal or a terrorist nor is there a clear-cut behavior of criminals that would allow to perfectly predict where the next crime will happen. Thus, the resulting ADM will almost always make mistakes and neither be 100% sensitive nor 100% specific. After the ADM system is fully trained, the following two steps are necessary:

7. *Embedding the ADM system in the societal process* In this step, the system is implemented in the institution that wants to use it. This step entails, e.g., training of the users in feeding data to the system and interpreting its results.

8. *Feedback* This is an optional step mainly for classifiers in which the real behavior of a person is compared to the predicted behavior and fed back to the ADM system. Thus, it can further learn from new evidence.

It is important to note that we call phase 8 optional because not all ADM systems seem to be automatically or even regularly re-evaluated based on feedback on their decision. For search engines which recommend 'relevant' websites based on some search, each user provides direct feedback on the recommendation by either clicking on the offered links or not. This feedback seems to be regularly incorporated by which the ADM system adapts to the different users. However, in the case of recidivism risk assessment, Oregon seems to use a static formula—presumably the result of a logistic regression—based on 16 variables.[6]

It can easily be seen that in this long process, there are multiple actors on an individual and institutional level that need to make diverse decisions, which might interact with each other and influence the quality of the final ADM system both individually and on a societal level. This is why we call it the long chain of responsibility. This large set of actors and the many steps of the process make it difficult to hold accountable individual decision makers and to guarantee the best ADM system.

4 Possible Problems in the Design and Deployment of an ADM System in Security

This section describes the potential problems of designing and deploying an ADM system in general, which leads to various recommendations for this process. The phases described above apply for ADM systems in general and there are multiple problems related with them (Zweig 2016a). Among them are erroneous data, wrong proxies for behavior which is difficult to measure, a wrong method selection, wrong quality measure selection, and missing evaluation of the embedding of the ADM system into the societal process of interest (Zweig 2016a). But there are also some specific problems associated with designing and running an ADM system for security related issues.

First of all, there are conceptual problems that run through the modeling process. The simplest one may only be an implementation error. However, this is very difficult to detect, especially with proprietary systems, where there is no transparency of the subroutines[7] used and no declaration of their correct behavior.

However, even if the subroutines like the *k*-means-clustering (s. Fig. 3) or regression approaches work correctly, the societal problem may have been mathematically wrongly modeled which is why the use of these well studied methods is already flawed. In other cases it might not even be meaningful to model it as a mathematical

[6] The formular can be viewed here: https://drive.google.com/file/d/0B8KbLffq9fg5cS0zbzF2VkY1dEpzZW4tZUttT3hVY29LUkhv/view (downloaded last on 28th of January, 2018).

[7] Modern Software is designed in a modular way, where a subroutine encapsulate small, well-defined functionalities.

problem to be solved. For example, it is difficult or even impossible to solve multi-causal problems with a simple logistic regression (Press and Wilson 1978). However, at least one of the recidivism risk assessment systems in use today is based on a logistic regression, namely the one mentioned above which is applied in Oregon (Diakopoulos 2016). In general in recidivism risk assessment, it may occur that two completely different profiles of offenders have to be recognized, e.g., drug addicts and non-drug addicts. Such a necessary distinction of cases is not possible in a classical logistic regression and each possible distinction needs to be incorporated by hand, involving many design decisions. Especially, when design teams are not truly interdisciplinary, the data scientist might not have the information that the process at hand is likely to be multi-causal, leading to a wrong method selection. A lack of understanding of data collection or problem definition can have such an impact on results that it is essential for the transparency and accountability of ADMs to work in multidisciplinary teams with a high level of communication.

There are also some problems with the data used for training an ADM system. As the above discussed example SKYNET has shown, it is often not possible to collect a sufficiently large and accurate ground truth[8] in security relevant areas (The Intercept 2015). Even if more than 55 million cell phone records were used, the number of real terrorist couriers in this dataset was so low that, additionally, suspects were called in. It also has to be noted that a ground truth in this area is usually neither complete nor accurate.

The accuracy of the input data is also limited by the problem of acquisition. Since an algorithm can only work on digitally available data, there are also major problems regarding the operationalization of important human behavior and other societal aspects (see also Pelzer 2018). For example, the designers of COMPAS opted for 137 questions in the 'CORE risk assessment' (Angwin et al. 2016) questionnaire. It shows a multitude of questions that try to quantify the defendants perspective on crime, their social environment, or their financial background. It is important to note here that COMPAS was first developed for post-trial decisions, e.g., the assignment of drug therapy places. Now, the same algorithms are also used for pre-trial criminals, based on questions such as:

- "How many of your friends/acquaintances have ever been arrested?"
- "How often do you have barely enough money to get by?"
- "How often do you get bored?"

While these questions might help to better predict the recidivism risk of a person, their usage in court needs to be made transparent and the decision-making process accountable to the democratic society, as discussed in Sect. 5. Once the design decisions have been made and the ADM system is trained, it is up to the society to select the quality criteria they want to evaluate the results with. In earlier work, we have shown that classical quality measures like the ROC AUC from machine learning

[8] A data set in which the class assignment which should be predicted (e.g., terroristic courier or not) is already indicated.

should not be blindly applied, as underlying model assumptions are not necessarily fulfilled (Krafft 2017). Similarly, there are different measures of fairness which contradict each other (Kleinberg et al. 2017)—again, it needs to be discussed which is the best one. There needs to be a discourse on the full range of possible evaluation criteria and what they really mean. For example, the quality of SKYNET's prediction were said by the slide authors to show a low "false alarm rate at 50% miss rate of 0.008%". However, while the percentage itself seems to be low, applied to 55 million inhabitants of a country of which almost all are innocent, it results in 4400 wrong suspects.

Finally, a phenomenon called "asymmetric feedback" needs to be mentioned, as discussed by O'Neil (2016): the problem is that many security based questions only provide unilateral feedback, e.g., a criminal offender who is not released on bail on the recommendation of an ADM system has no way to prove that he would not have recidivated. Despite this asymmetry in feedback, many approaches to machine learning try to exhibit a "lifelong learning", i.e., to evolve with current data (Michie et al. 1994). In order to avoid over-specialization in this direction, ADM systems in these areas are often no longer fed with current data and one of the core abilities of machine learning does not come into play. A farsighted approach is therefore required to consider these and other possible risks in the preparation and modeling of an ADM which also extends to the legal fields of application. The following section thus discusses the necessity of transparency and algorithm accountability from a political science perspective.

5 ADM Systems in the Justice System as a Challenge to Democracy?

In recent years, ADM systems have been criticized as a possible threat to democracy (O'Neil 2016; Pasquale 2016). Against the backdrop of the inadequacy of current legal frameworks to control algorithm-based decision-making (Kroll et al. 2017, p. 636), critics have demanded that the state needs to "provide an appropriate regulatory framework, which ensures that technologies are designed and used in ways that are compatible with democracy" (Helbing et al. 2017).

Admittedly, the question of how to "govern" (Danaher et al. 2017; Ziewitz 2016) or how to "regulate" (Yeung 2017) ADM systems can be analyzed from very different analytical perspectives. We have opted here for a political science perspective for two reasons: First, recent years have witnessed an increasing numbers of scholars criticizing ADM systems directly in relation to democracy, which is a key concept in political sciences; second, whereas several other fields of research, such as Science and Technology Studies (see for instance the 2016 special issue on "Governing algorithms" in Science, Technology and Human Values (Ziewitz 2016) or (Annany/Crawford 2016) but also media and communication science (Just and Latzer 2017), law (Hildebrandt 2016) or scholars from other fields of social sciences (Kitchin 2017), have already engaged in a scientific discussion on the implications of algorithms, political scientists have remained rather quiet in this respect (Richey and Taylor 2017). Hence, in our view, there is a need to relate questions of ADM systems to the core concepts of democratic theory in political science.

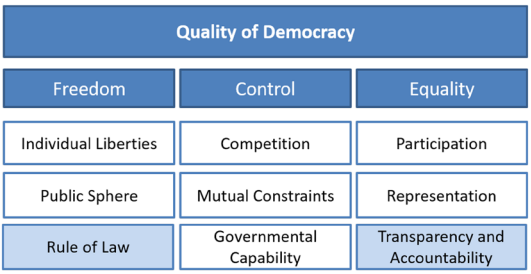

Fig. 6 The quality of democracy as measured by the "Democracy Barometer". Source: Democracy Barometer (Bühlmann et al. 2012) (with minor changes by the authors)

If one aims at evaluating the design and deployment of ADM systems in society from a political science perspective, it seems an appropriate starting point to discuss the relationship between ADM systems and society from the very foundations of how democratic political systems work.

Political scientists have a long history in studying the characteristics of Western democracies (Barber 2003; Dahl 1971; Schumpeter 1942). One of the main debates within this literature has been about the question of what defining characteristics a political system has to fulfil in order to merit the name democracy. On a very general level, one can differentiate between "thin" and "thick" definitions of democracy (Mair and Peters 2008). Schumpeter, in his groundbreaking work, defined democracy in a purely procedural (i.e., "thin" way) as "institutional arrangement for arriving at political decisions in which individuals acquire the power to decide by means of a competitive struggle for the peoples vote" (Schumpeter 1942, p. 269). In contrast, Dahl's concept of polyarchy is "thick" in that it comprises a list of minimal requirements political systems have to fulfil in order to be considered as a democracy (Dahl 1971). However, as the primary interest of these scholars was to differentiate between democratic and non-democratic (e.g., autocratic or totalitarian) systems, the different shades of grey within democratic systems can only be assessed to a limited extent with such concepts. This is why, following Beethams initial impetus (Beetham 1999), several researchers in the 2000s set out to look more closely at the quality of the democracy within well-established democratic countries, put forward different measures of the quality of democracy ("democratic audit") such as the Bertelsmann Sustainable Government Index (Empter and Novy 2009) or the "Democracy Barometer" (Bühlmann et al. 2012) and, indeed, found substantial differences. For our inquiry, such fine-grained concepts are the appropriate starting point, because they enable us to zoom in very closely onto the dimensions of democracies that are affected by the increasing use of ADM systems within society in general and the justice system in particular (such as the rule of law or accountability, see the shaded boxes in s. Fig. 6).

One of the most influential concepts in the last years has been Merkel's notion of an "embedded democracy" which disentangles different dimensions of democracy that can be more or less fulfilled in individual countries (Merkel 2010). According to this concept, five dimensions are key if one wants to assess the quality of a democratic system: (1) the electoral system; (2) the system of participation; (3) the civil rights system; (4) the system of horizontal accountability (e.g., institutional checks

and balances) and (5) the system of effective governance. In more recent work, the research group around Merkel has merged these five criteria into three core principles of freedom, control, and equality (Bühlmann et al. 2012). Moreover, each of the three principles comprises several sub-dimensions ("functions"), such as individual liberties, representation, rule of law or competition which are then, in the last step, measured by indicators and aggregated into one index, the "democracy barometer" (s. Fig. 6).[9]

On the basis of such a fine-grained concept of the quality of democracy, it is possible to identify individual dimensions that are affected by the rise of ADM systems in very different spheres of public life. Evidently, ADM systems challenge all three core principles of democracy (Helbing 2015; Helbing and Pournaras 2015). If, for instance, algorithms strongly intervene in public discourse and the formation of public opinion by filtering information for individuals and generating echo chambers and filter bubbles via social media (Flaxman et al. 2016), this clearly affects the foundations of democracy in the public sphere (third dimension of the principle of "freedom" as discourse about politics is a key ingredient to the functioning of modern democracy (Habermas 1992). Besides such general challenges that shall not be underestimated, ADM systems also have very direct effects if we zoom in on security-related issues in general and the criminal justice system in particular. Here, two components of democracy as conceptualized below move center stage: The dimension of "rule of law" as part of the principle of freedom as well as the component "transparency and accountability" as part of the principle of equality.[10] In the following, we will discuss how ADM systems in the justice system challenge these two dimensions and what possible ways forward might exist.

5.1 ADM Systems and the Rule of Law

The rule of law is one of the basic prerequisites of democracy (Habermas 1992). Citizens have to be certain that the same rules apply to everybody and that they are equal before the courts. A critical aspect in this respect is to guarantee the due process. However, it is at this point where ADM systems become problematic. As Steinbock has pointed out for the US (Steinbock 2005), the use of ADM systems in the criminal justice systems via data mining and data matching can pose serious problem to the due process principle within the criminal justice system. One of the

[9] For a discussion of the validity of this measure, see the debate between Merkel et al. (Merkel et al. 2013) on the one and Wagschal et al. (Jäckle et al. 2012, Jäckle et al. 2013) on the other hand.

[10] We do not discuss the case of "fairness" in algorithm-informed decision-making here, although it clearly is relevant in the context of judicial decisions. However, as we start from a broader framework of democratic theory, the issue of fairness will not be central here. Moreover, it has been widely discussed in the scientific debate around the use of algorithms (for a state of the art report, see (Berk et al. 2017)). It is important to note that there are also discussions on the question of fairness from the computer scientist perspective (Kleinberg et al. 2017; Angwin et al. 2016; Brennan et al. 2009). Both fields agree that the question for algorithmic or societal fairness is not yet fully solved.

most problematic points here clearly relates to the generation of falsely classified persons through ADM systems [for an overview of the problems and merits: (Duwe and Kim 2017)]. This is especially important, since as described above it is in most cases impossible to avoid mistakes. What seems to be necessary at least from a point of view of the rule of law is complete transparency[11] as to how the ADM systems work:

> "(I)ndividuals should have the opportunity to be notified at least of the process, the evidence it has produced against them, and the basis for the data match or profile before significant consequences are imposed as a result of computerized decision-making. No less is required by basic due process principles." (Steinbock 2005, p. 64–65).

More concretely speaking, one would expect a democratic system that follows the rule of law to give their citizens a chance to respond to results produced by an ADM system. If a person is, for instance, sorted out by an algorithm as a high-risk profile recidivist, she should have the chance to comment on the data that has been collected, on the indicators used as well as on the algorithm that seems to classify her as a probable recidivist. [For a suggestion of how to implement such measures, see Citron and Pasquale (2014)]. However, from a technical perspective, this will only be possible for the weakest and least powerful ADM systems. For example, a simple linear regression from the field of machine learning certainly produces explainable results, however, they will very likely also be too low in quality to be applied.

Another way forward that has been proposed is the creation of independent bodies that oversee the application of ADM systems in the criminal justice process (Steinbock 2005, p. 80; Pasquale 2016). However, while the advantages of such boards is evident if it is charged with evaluating the outcomes of ADM systems and has decision-making power about the future use of such programs, it also raises the question of democratic control. In fact, such boards only seem to make sense if they are sheltered from political and bureaucratic pressure, e.g., following the rules of nominations for constitutional courts (Kneip 2011). These boards will not always need access to the actual code of a system but in all cases they will need access to at least the same data that the algorithm designer used for training their ADM system. Furthermore, they need data on the final decisions been made to analyze for possible discriminations.

In sum, it seems to be clear from this brief discussion that the use of ADM systems in criminal justice systems necessitates to be accompanied by a set of rules guaranteeing that the basic principles of the rule of law are followed: judicially granted ways to appeal against decisions on the one hand and bodies or instruments of regular evaluation on the other hand seem to be important ingredients to making

[11] For a more critical view on this call for transparency, see the recent contribution by Ananny and Crawford (2017). In fact, the further discussion (see Sect. 5.0.2) touches on several of the limitations that the authors have put forward (e.g., the need for explanation in order to be held accountable and the need to have enforcement rules if a transparent process proves to be problematic).

the use of ADM systems in the judicial system compatible with high-quality democracies of the Western world.

5.2 ADM Systems, Transparency and Accountability

The second aspect of ADM systems that touches the foundations of democratic systems relates to transparency[12] and accountability. Democratic theory considers transparency as a key component of a high-quality democracy as it ensures that the ruling political elite can be held accountable for their actions see e.g., Hollyer et al. (2011). This is why the democracy barometer therefore treats transparency as one of three dimensions of the principle of "equality". However, the principle of accountability does not only relate to transparency. Following Schedler (1999, p. 14), one may distinguish three aspects of accountability: (1) subjecting power to the threat of sanctions ("enforcement"); (2) obliging power to be exercised in transparent ways ("monitoring"); and (3) forcing power to justify its acts ("justification"). The first aspect simply relates to the fact that accountable persons will be sanctioned and rewarded for their actions. There are many possible ways of enforcing accountability, the most widely accepted (and most hotly discussed) of which goes via elections. An enormous amount of studies has dealt with this fundamental question of democracy, namely, whether citizens actually hold governments accountable at the ballot box for their actions and the story is far from settled (Achen and Bartels 2016; Lee et al. 2017; Soroka and Wlezien 2010). The second and third aspects have to do with accountability in a more discursive way and deal with information and justification. However, both are important as prerequisites for well-reasoned sanctions and reward of the persons that are to be held accountable. On the one hand, the exercise of power has to be transparent and information has to be accessible for everyone. On the other, the exercise of power also has to be justified and explained, those holding power have to give reasons for their actions and these justifications have to withstand public debates. Building on these three criteria, one can think of accountability as a scale with full accountability on one side of it, if all three aspects are completely fulfilled, and a gradual shift toward less fulfillment if one or several criteria are not met (Schedler 1999, p. 18).

How can such an understanding of transparency as accountability be related to ADM systems in the justice system? First of all, the power relationship in the justice and security realm seems rather clear. On the one hand, actors from within the justice system be it police officers, prosecutors, or judges make decisions about others, for instance offenders or suspects. Hence, the actors of the justice system are the ones to be held accountable for their decisions. The complexity, however, starts if these decisions are partly based on ADM systems. An illustrative example is, as discussed above, the criminal justice decision of whether an offender will be released on parole. Such decisions are—in some countries such as the US—increasingly

[12] Clearly, transparency also matters for due process (see above). However, whereas the possibility to oppose decisions (e.g., via the creation of independent bodies), is the core of the argument on due process, accountability is not thinkable without transparency of rules and procedures.

dependent on a risk assessment based on algorithms which assess the risk of recidivism (Berk et al. 2017). Data collected on the offender will then be assembled and processed by an algorithm to help humans, e.g., a parole board, to render a decision. How will the long chain of responsibility work in such cases? And what if ADM systems are effectively used as sole basis of decision-making?

Following the concept of accountability as presented above, it is clear that access to the data used and transparency about the main decisions made in designing the ADM are key, as monitoring is only possible if the information on which a decision has been made is available. Related to this point is the aspect of justification, which means that there has to be an explanation and a possible debate about the outcome. From this perspective, it becomes clear that the release of information has to be accompanied by an explanation of how this information has been used and to what extent it has affected the decision-making process. In fact, both points quite nicely illustrate what Burrell (2016) has termed illiterate opacity and intrinsic opacity of machine learning algorithms: opacity of ADM systems because they are simply too complex to understand. Therefore, it is not enough to simply describe what happens, but to explain and to justify why certain outcomes are generated by an ADM system (see also: Ananny and Crawford (2017): 8–9). To make it more concrete and related to the above-mentioned case: Only if a debate on the decision of an ADM-generated risk assessment of recidivism of an offender can be started between non-technicians, only if the risk assessment has been justified to a broader public, can accountability in terms of "monitoring" and "justification" be reached. All these considerations mesh well with the idea of "procedural regularity" which has recently been proposed by Kroll et al. (2017). It means that ADM systems "prove to oversight authorities and the public that decisions were made under an announced set of rules consistently applied in each case." (Kroll et al. 2017, p. 637). Hence, this should also include a proper evaluation, documentation, and reporting of caveats and strengths, as well as indicating what explicit role the ADM system has played within the decision making context. Only with such a well-explained transparency and justification of choices guided by machines, justification seems possible to achieve.

Finally, enforcement is clearly the most challenging part of applying the concept of accountability as discussed above. Whereas several standard techniques of "enforcement" have to be debated in that context, the most drastic sanction clearly is: to stop using a certain ADM system within a certain decision-making context.[13] However, the foundation of such a sanction must be a criteria-based evaluation of the functioning of a system and its performance, a praxis which is not always used in reality in the justice system (according to a list of ADM-use in US states collected by EPIC 2017. Moreover, in the political and judicial system, decisions about stopping an ADM system are usually conflict-ridden. Bureaucracies have been adapted to the use of such systems and the commercial interests as well as budgetary costs will necessarily be weighed against the decision to stop the use of an ADM system.

[13] Although it is probable that ADM systems will be used increasingly in the upcoming years, it has to be thinkable – from the perspective of democratic accountability– that a certain system in a specific decision-making context will be stopped.

Moreover, in cases like SKYNET, national security interests often trump democratic concerns—given that the reason of security can be used by political actors to move issues in the realm of "emergency politics" and to justify "special measures" (see: Buzan et al. 1998: 27). In terms of democracy, it seems most important to solely base decisions to implement or abandon ADM systems based on evaluations and validity something that can only be guaranteed by installing an independent oversight body which is shielded from political and economic pressure (see Ananny and Crawford 2017: 6).

6 Discussion

The use of ADM systems in the justice and security sector has been praised as a way to overcome the weaknesses of human decision making. Algorithms do not run for office [as do elected judges who may therefore render biased decisions (Huber and Gordon 2004)] and they do not suffer from the well-known psychological limits of human decision-making, such as status quo bias (Kahneman and Tversky 1979; Samuelson and Zeckhauser 1988). Insofar, ADM systems have a high potential to overcome some of the weaknesses that human decision-making is plagued with. However, ADM systems are not only solutions, but they also create new problems and for some critics even bigger ones (O'Neil 2016). In this open debate, we have put forward a political science perspective centered around the issue of democracy. We have found the rule of law as well as the criterion of transparency and accountability to be challenged by the implementation of ADM systems in the justice and security system. At the same time, it would be naive not to make use of the merits that algorithms have especially in terms of the synthesis of big data, where these merits are clearly observable and agreed upon. Therefore, we have laid out several ways forward, which might help to reconcile the democratic process with the use of ADM systems. Two aspects are key in our understanding:

The first recommendation is related to the implementation process of ADM systems in the security and justice sector. From the perspective of democratic theory, it seems reasonable to create independent bodies of ADM oversight that carefully assess the functioning of such systems in the decision-making process. In any case, it needs to be discussed which kind of ADM systems need to be assessed in which depth, depending on the potential damage for individual and society as a whole. The members of these bodies should be shielded from political and economic interests in a similar way as how constitutional courts work in some European countries, such as Germany.

The second recommendation deals with the concept of "procedural regularity" as proposed by Kroll et al. (2017). In order to fulfil the criteria of transparency and accountability, the decisions made in every step of the algorithmic process have not only to be made transparent, but also to be explained in a way understandable to non-specialists. Only if this is the case can a public debate about the advantages and disadvantages of such systems be set in motion which can, after reflection, yield to a sanction.

Admittedly, these two ways forward, which may help strengthening the principles of due process (dimension of the rule of law) as well as introducing more accountability and transparency to the system, are not a one-size-fits-all solution. And clearly, when progress is made on accountability and transparency issues, problems of due process may remain. Nevertheless, it seems at least that these propositions are first steps on a way toward a more careful use of ADM systems in the criminal justice system.

Nevertheless, these seemingly straightforward and not very radical recommendations are very difficult to implement in real-life politics. Calls for transparency and full disclosure of information about both the algorithm and the data it uses have been uttered repeatedly [e.g., Lepri et al. (2017, p. 9–10)]. However, these calls have only seldom been heard as there are also many reasons not to make such information publicly available. Economic reasons, because ADM systems are sold to bureaucracies by private companies that compete for the best product (i.e., the best algorithm); privacy reasons, because some data which are used in the criminal justice systems are confidential and not meant for public use; And secrecy reasons, because data used to trace offenders or terrorists serve best for policing if they are not available publicly. Hence, in the realm of justice and security, many questions related to transparency and accountability of ADM systems are intimately linked to broader questions of security and liberty (Waldron 2003, 2006). Western democracies have opted for rather different solutions regarding this goal (Epifanio 2011, 2014; Wenzelburger and Staff 2017) with Anglo-Saxon countries being considerably "tougher" in terms of law and order than Scandinavian countries (e.g., Lacey 2008; Wenzelburger 2018). This is why huge cross-country variance is also probable in terms of how ADM systems are regulated in different countries. The criteria of the democracy barometer can help sorting out to what extent countries follow a "high-quality" solution, based on how they regulate ADM systems with respect to the rule of law or accountability and transparency. At any rate, it seems hardly convincing to argue for the supremacy of data driven technocracy over democracy (Khanna 2017), given the challenges that come with the employment of ADM systems in democracy.

References

Achen CH, Bartels LM (2016) Democracy for realists. Princeton University Press, Princeton

Ananny M, Crawford K (2017) Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc 33:1–17

Angwin J, Larson J, Mattu S, Kirchner L (2016) Machine bias—there's software used across the country to predict future criminals. And its biased against blacks. ProPublica. Accessed 27 Oct 2017 **(Online)**

Barber B (2003) Strong democracy: participatory politics for a new age. University of California Press, Berkeley

Beetham D (1999) Democracy and human rights. Polity Press, Cambridge

Berk R, Heidari H, Jabbari S, Kearns M, Roth A (2017) Fairness in criminal justice risk assessments: the state of the art. arXiv:1703.09207

Brennan T, Dieterich W, Ehret B (2009) Evaluating the predictive validity of the COMPAS risk and needs assessment system. Crim Justice Behav 36(1):21–40

Bühlmann M, Merkel W, Mller L, Weels B (2012) The democracy barometer: a new instrument to measure the quality of democracy and its potential for comparative research. Eur Polit Sci 11(4):519–536

Burrell J (2016) How the machine thinks: understanding opacity in machine learning algorithms. Big Data Soc 3(1):2053951715622512
Buzan B, Wæver O, De Wilde J (1998) Security: a new framework for analysis. Lynne Rienner, Boulder
Citron DK, Pasquale FA (2014) The scored society: due process for automated predictions. Wash Law Rev 89:1–33
Dahl RA (1971) Polyarchy. Participation and opposition. Yale University Press, New Haven
Danaher et al (2017) Algorithmic governance: developing a research agenda through the power of collective intelligence. Big Data Soc. https://doi.org/10.1177/2053951717726554
Desmarais SL, Singh JP (2013) Risk assessment instruments validated and implemented in correctional settings in the United States—guide 028352. National Institute of Corrections, Washington
Diakopoulos N (2015) Algorithmic accountability: journalistic investigation of computational power structures. Digit Journal 3(3):398–415
Diakopoulos N (2016) We need to know the algorithms the government uses to make important decisions about us. The conversation. https://theconversation.com/we-need-to-know-the-algorithms-the-government-usesto-make-important-decisions-about-us-57869. Accessed 29 Dec 2017 (Online)
Duwe G, Kim K (2017) Out with the old and in with the new? an empirical comparison of supervised learning algorithms to predict recidivism. Crim Justice Policy Rev 28(6):570–600
Egbert S (2018) About discursive storylines and techno-fixes: the political framing of the implementation of predictive policing in Germany. Eur J Secur Res. https://doi.org/10.1007/s41125-017-0027-3
Empter S, Novy L (2009) Sustainable governance indicators 2009: policy performance and executive capacity in the OECD. Verlag Bertelsmann Stiftung, Gütersloh
EPIC (2017) Algorithms in the criminal justice system. https://epic.org/algorithmic-transparency/crim-justice/. Accessed 27 Oct 2017 **(Online)**
Epifanio M (2011) Legislative response to international terrorism. J Peace Res 48(3):399–411
Epifanio M (2014) The politics of targeted and untargeted counterterrorist regulations. Terror Polit Violence 28(4):1–22
Flaxman S, Goel S, Rao JM (2016) Filter bubbles, echo chambers, and online news consumption. Public Opin Q 80(S1):298–320
Freeman K (2016) Algorithmic injustice: how the Wisconsin supreme court failedo protect due process rights in State v. Loomis. N Carol J Law Technol 18:75–233
Frouillou L (2016) Post-bac admission: an algorithmically constrained "free choice". In: Justice spatialelspatial justice, no. 10. https://www.jssj.org/wp-content/uploads/2016/07/JSSJ10_3_VA.pdf. Accessed June 2016
Habermas J (1992) Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des Demokratischen Rechtsstaats. Suhrkamp, Frankfurt am Main
Helbing D (2015) The automation of society is next. Createspace, North Charleston
Helbing D, Pournaras E (2015) Build digital democracy. Nature 527(7576):33–34
Helbing D, Frey BS, Gigerenzer G, Hafen E, Hagner M, Hofstetter Y, van den Hoven J, Zicari RV, Zwitter A (2017). Will democracy survive big data and artificial intelligence? Sci Am. https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence. Accessed 25 Feb 2017
Hildebrandt M (2016) Law as information in the era of data-driven agency. Mod Law Rev 79(1):1–30
Hollyer JR, Rosendorff BP, Vreeland JR (2011) Democracy and transparency. J Polit 73(4):1191–1205
Huber GA, Gordon SC (2004) Accountability and coercion: is justice blind when it runs for office? Am J Polit Sci 48(2):247–263
Hutchby I (2001) Technology, texts and affordances. Sociology 35(2):441–456
Jäckle S, Wagschal U, Bauschke R (2012) Das Demokratiebarometer: basically theory driven? Z Vgl Polit 6(1):99–125
Jäckle S, Wagschal U, Bauschke R (2013) Allein die Masse machts nicht—Antwort auf die Replik von Merkel et al. zu unserer Kritik am Demokratiebarometer. Z Vgl Polit 7(2):143–153
Just N, Latzer M (2017) Governance by algorithms: reality construction by algorithmic selection on the internet. Media Cul Soc 39(2):238–258
Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47(2):263–292
Khanna P (2017) Technocracy in America: rise of the info-state. CreateSpace Independent Publishing Platform
Kitchin R (2017) Thinking critically about and researching algorithms. Inf Commun Soc 20(1):14–29

Kleinberg J, Mullainathan S, Raghavan M (2017). Inherent trade-offs in the fair determination of risk scores. https://arxiv.org/abs/1609.05807
Kneip S (2011) Constitutional courts as democratic actors and promoters of the rule of law: institutional prerequisites and normative foundations. Z Vgl Polit 5(1):131–155
Krafft TD (2017) Qualitätsmaße binärer Klassifikatoren im Bereich kriminalprognostischer Instrumente der vierten Generation. Master's thesis, TU Kaiserslautern. arxiv.org/abs/1804.01557v1
Kroll JA, Huey J, Barocas S, Felten EW, Reidenberg JR, Robinson DG, Yu H (2017) Accountable algorithms. Univ Pa Law Rev 165:633
Lacey N (2008) The prisoners' dilemma. CUP, Cambridge
Lee S, Jensen C, Arndt C, Wenzelburger G (2017) Risky business? Welfare state reforms and government support in Britain and Denmark. Br J Polit Sci. https://doi.org/10.1017/S0007123417000382
Lepri B, Oliver N, Letouzé E, Pentland A, Vinck P (2017) Fair, transparent, and accountable algorithmic decision-making processes. Philos Technol. https://doi.org/10.1007/s13347-017-0279-x
Lischka K, Klingel A (2017) Wenn Maschinen Menschen bewerten. Studie der Bertelsmann Stiftung, Gütersloh
Mair and Peter (2008) Democracies. Oxford University Press, Oxford
Merkel W (2010) Systemtransformation. Springer, Wiesbaden
Merkel W, Tanneberg D, Bühlmann M (2013) Den Daumen senken: Hochmut und Kritik. Z Vgl Polit 7(1):75–84
Michie D, Spiegelhalter DJ, Taylor CC (1994) Machine learning, neural and statistical classification. Ellis Horwood Ltd. ISBN-13: 978-0131063600
Niklas J, Sztandar-Sztanderska K, Szymielewicz K (2015) Profiling the unemployed in Poland: social and political implications of algorithmic decision making. Fundacja Panoptykon, Warsaw
Northpointe (2012). Practitioners guide to COMPAS core. http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf. Accessed 27 Oct 2017 **(Online)**
O'Neil C (2016) Weapons of math destruction. How big data increases inequality and threatens democracy, 1st edn. Crown, New York
Parchoma G (2014) The contested ontology of affordances: implications for researching technological affordances for collaborative knowledge production. Comput Hum Behav 37:360–368
Pasquale F (2016) The Black box society. The secret algorithms that control money and information. Harvard University Press, Cambridge **(first Harvard University Press paperback edition)**
Pelzer R (2018) Policing of terrorism using data from social media. Eur J Secur Res. https://doi.org/10.1007/s41125-018-0029-9
Press SJ, Wilson S (1978) Choosing between logistic regression and discriminant analysis. J Am Stat Assoc 73(364):699–705
Przeworski A, Stokes SC, Manin B (1999) Democracy, accountability, and representation, vol 2. Cambridge University Press, Cambridge
Richey S, Taylor JB (2017) Google and democracy: politics and the power of the internet. Routledge, Abingdon
Samuelson W, Zeckhauser R (1988) Status quo bias in decision making. J Risk Uncertain 1(1):7–59
Schedler A (1999) Conceptualizing accountability. Lynne Rienner, Boulder, pp 13–28
Schumpeter JA (1942) Capitalism, socialism and democracy. Harper & Brothers, New York
Soroka SN, Wlezien C (2010) Degrees of democracy: politics, public opinion, and policy. Cambridge University Press, Cambridge
Steinbock DJ (2005) Data matching, data mining, and due process. Ga Law Rev 40(1):1–84
The Intercept (2015) US government designated prominent Al Jazeera journalist as member of Al Qaeda. https://theintercept.com/document/2015/05/08/skynet-courier/. Accessed 27 Oct 2017 **(Online)**
Waldron J (2003) Security and liberty. J Polit Philos 11(2):191–210
Waldron J (2006) Safety and security. Neb Law Rev 85(2):454–507
Wenzelburger G (2018) Political economy or political systems? How welfare capitalism and political systems affect law and order policies in twenty western industrialised nations. Soc Policy Soc 17(2):209–226
Wenzelburger G, Staff H (2017) The third way and the politics of law and order: explaining differences in law and order policies between Blair's new labour and Schröder's SPD. Eur J Polit Res 56:553–577
Wormith JS (2017) Automated offender risk assessment: next generation or black hole. Criminol Public Policy 16(1):281–303
Yeung K (2017) Algorithmic regulation: a critical interrogation. Regul Gov. https://doi.org/10.1111/rego.12158

Ziewitz M (2016) Governing algorithms: myth, mess and methods. Sci Technol Human Val 41(1):3–16
Zweig KA (2016a) 2. Arbeitspapier: Überprüfbarkeit von Algorithmen. https://algorithmwatch.org/de/zweites-arbeitspapier-ueberpruefbarkeit-algorithmen/. Accessed 27 Oct 2016 **(Online)**
Zweig KA (2016b) Network analysis literacy: a practical approach to the analysis of networks. Springer, Berlin