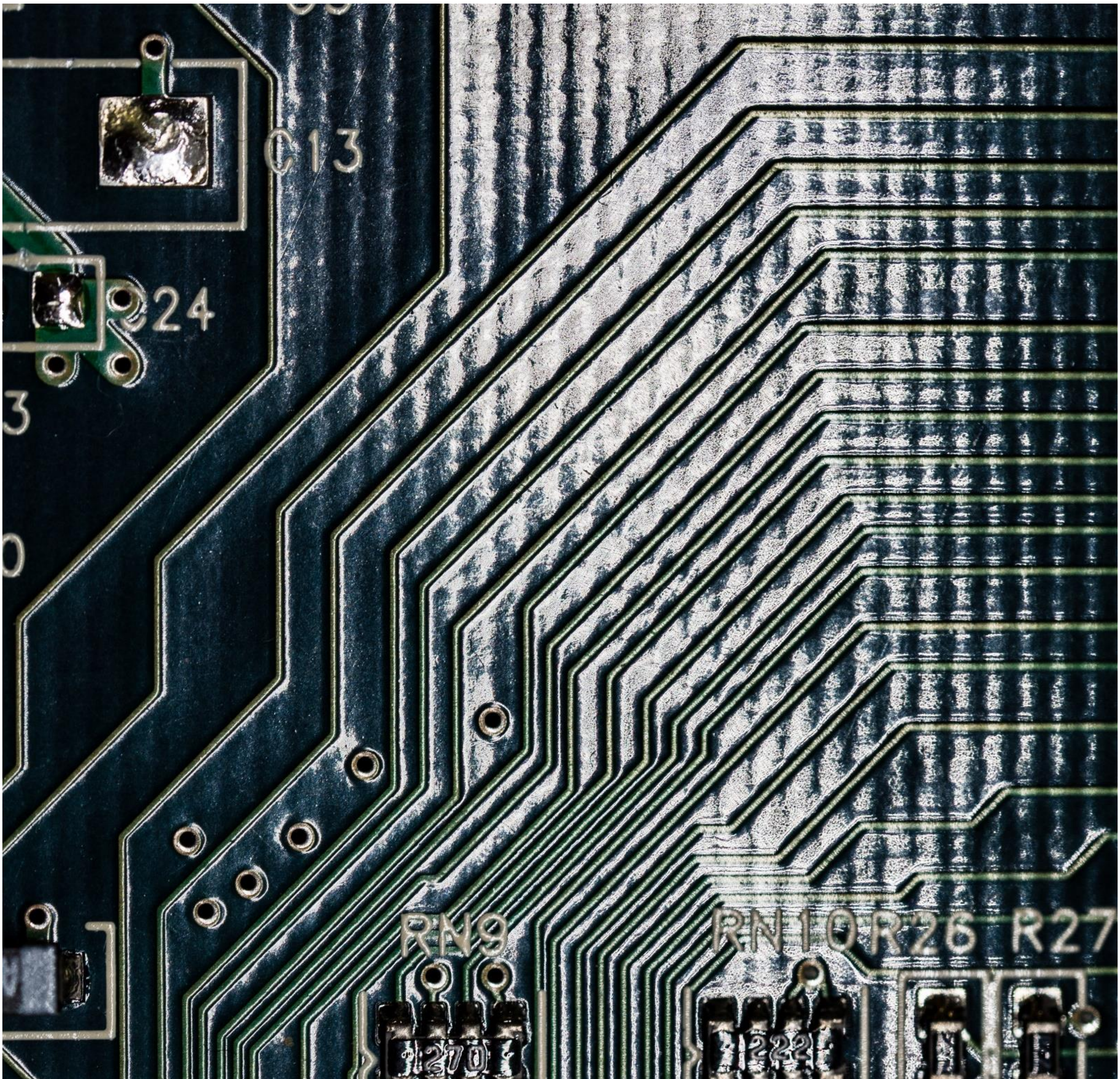
Discussion paper Ethics of Algorithms #2



Calculated participation

How algorithmic processes impact opportunities to be part of society

BertelsmannStiftung

Legal Notice

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh
www.bertelsmann-stiftung.de

Responsible for content
Konrad Lischka
Ralph Müller-Eiselt

Authors
Kilian Vieth
Dr. Ben Wagner

DOI 10.11586/2017027 https://doi.org/10.11586/2017027

Calculated participation

How algorithmic processes impact opportunities to be part of society

Kilian Vieth
Dr. Ben Wagner
on behalf of the Bertelsmann Stiftung

Contents

1 Preface

What is the difference between the following two programs: a software application that calculates the cost of car insurance based on the policyholder's driving behavior, and an application that predicts the likelihood of burglaries taking place on certain streets, thereby helping the police plan their patrols?

One intuitive answer might be that it seems more menacing to monitor the driving behavior of each policyholder and reward it with a favorable rate than to predict the crimes that will take place on certain streets. Some people, conversely, might reject both algorithm-driven scenarios, preferring decisions made by humans instead. No matter what the immediate reaction, machine-based systems, known as algorithmic decision-making (ADM) processes, are already observing and evaluating people and their behavior in many areas of life. ADM processes now help determine which verdicts are issued in court, who is awarded a loan, who is admitted to university and how much attention customers receive when they contact a call center. Such machine-based decision-making systems are already widespread in some countries, and the wide range of applications demonstrates the importance of having an informed and, above all, well-structured debate on this topic. This working paper is intended to contribute to that debate.

One key feature used to differentiate between ADM processes is their potential social impact and, more specifically, their effect on opportunities for participation. What influence do ADM processes have on whether individuals and organizations can participate equally in political decision-making and will-building processes? And how do they impact the chance everyone has to be part of social, cultural and economic developments?

The authors of this paper, Kilian Vieth and Ben Wagner, introduce a tool which uses seven characteristics of ADM processes to determine and tentatively quantify their potential impact on participation. Users answer a short questionnaire which, unlike intuitive assessments, can be applied consistently across a range of situations. The tool does not evaluate the quality of the processes per se, but offers a relative assessment of their potential impact, whether positive or negative. The higher this impact is, the higher the standards an ADM process must meet. More accountability is required, for example, when processes manage police activities than when they insure cars ().

The aforementioned tool represents a first attempt to systematically examine ADM processes, and we are documenting it in this working paper, which is published under a free license (CC BY-SA 3.0 DE). Our aim is to contribute to a rapidly-evolving field, with the expectation that others might want to continue the conversation we have begun. We therefore welcome adaptations and further development of the proposed methodology, as well as applications to other situations involving ADM processes. We also welcome suggestions and criticisms of all kinds.

The analysis by Vieth and Wagner forms part of an exploration of "Participation in the Age of Algorithms and Big Data," in which the Bertelsmann Stiftung is focusing on how phenomena taking place in the digital sphere affect social participation. In addition to the analysis of international case studies involving ADM processes (), other working papers will examine related aspects, such as error sources in ADM processes and interactions between algorithmic systems and psychological mechanisms.

Ralph Müller-Eiselt
Senior Expert
Digitalization Taskforce
Bertelsmann Stiftung

Konrad Lischka
Project Manager
Digitalization Taskforce
Bertelsmann Stiftung

2 Introduction

What do credit assessments, border controls and search engines have in common? They all use algorithmic methods to prepare and/or make decisions. The algorithmic judgments which determine who gets a loan, who is allowed to enter a country and which hits a search query displays have become a central feature of digitized society.

Every aspect of our day-to-day lives is shaped by and takes place using digital technology. Algorithms that determine how we communicate, work and move are a core component of all the computers that surround us. Digitalization generates increasingly large amounts of data and gives rise to new business models and platforms, all of which are controlled by algorithms in important ways. It is impossible to imagine a digital society without algorithms: we need them to make sense of our data-rich environment. “If we do not want to limit ‘the good life’ to a select few, then we need a better energy supply, better transport concepts and better resource management” (Stalder 2017). Even if the development of numerous “smarter” products sometimes turns out to be an empty promise, one thing is clear: without algorithms as an integral part of social processes, we would be unable to resolve pressing problems and enable fair participation in society. This applies to all processes in which large volumes of data are processed using software in order to make decisions or to lay the foundation for making those decisions.

Algorithms are increasingly making decisions with, for and about us – thereby raising new questions about participation. This paper advances a number of initial suggestions for structuring and classifying the participation-related issues stemming from algorithmic decision-making (ADM). Moreover, we must free ourselves from the “technological determinism” that is widespread today and begin focusing on possibilities for shaping digital technology. Technology is adaptable, so the first step must be to clarify the targets to be pursued through the use of algorithmic methods. The framework conditions for using technology must be developed not as a result of supposed technical constraints, but following a discussion of the desired outcomes. In today’s digitized knowledge society, shaping technology is becoming a fundamental form of power. And if knowledge is power, then algorithms are becoming today’s instruments of power. To what degree is it acceptable and desirable for algorithms to have an impact on the lives of individuals and on society as a whole? And which aspects of ADM must we consider more closely if we want to benefit from the potentials and minimize the risks to the greatest degree possible?

This paper offers a general overview of ADM processes by explaining key concepts (chapter 3), delimiting the relevant area of analysis of ADM processes (chapter 4) and presenting typical scenarios and functions relating to the use of these processes (chapter 5). With this as a basis, a system for classifying ADM processes is then proposed (chapter 6). The objective is to make it easier to evaluate and compare the potential impact of ADM processes on participation by using fewer criteria. This, in turn, will facilitate a prioritization and preparation of more in-depth research on the subject..

3 Conceptual basis

3.1 Participation

Within the framework of this study, the term “participation” denotes the equal inclusion of individuals and organizations in political decision-making and will-building processes, and the fair inclusion of all in social, cultural and economic developments. This means, first, participation in democratic processes – i.e. political equality – and, second, participation in society’s achievements: everything from “suitable living and housing conditions, social and health protection, adequate and universally accessible educational opportunities, and inclusion in the labor market to various opportunities for spending one’s leisure time and determining how one lives one’s life” (Beirat Integration 2013).

One prerequisite of participation in this sense is that the material resources available to all are above the minimum level required for ensuring everyone can be part of society. The guarantee of social and political participation thus presupposes a “basic equality of social necessities” (Meyer 2016). Elements of this basic equality are described, for example, in the Universal Declaration of Human Rights and in the International Covenant on Economic, Social and Cultural Rights (Bundesgesetzblatt 1966). Targeted investments in the development of individual skills are necessary to enable equal participation in this sense (Bertelsmann Stiftung 2011: 31). It is the responsibility of the state and of the community to continually empower each individual to take advantage of the opportunities available to him or her.

3.2 Algorithms

Algorithms are now the subject of various debates, but “algorithm” has different meanings depending on the context and reference group. The concept is used differently in computer science and mathematics than it usually is in the public sphere or in political discourse. To be able to classify debates on algorithms better, it is always important to take into account the speaker and the audience being addressed. Particularly when it comes to analyzing the social effects of algorithms, a distinction must be made between the formal definition of the term and its general use. Many debates are not concerned with a narrowly-defined algorithm, but rather with the role of technology in society (Bucher 2016).

The term “algorithm” can be traced back to the Persian-Arabic mathematician Muhammad ibn Musa al-Khwarizmi. Algorithms are not only software, but are generally regarded as programmed processes which convert a given input (a value or group of values) into a desired output (for a classic textbook definition, see Cormen et al. 2001). These processes use set calculations to solve a specific task in precisely defined steps (Gillespie 2014: 167). Algorithms therefore consist of a precisely-defined sequence of steps used to achieve a certain result (Diakopoulos 2015: 400). A simple task might be: “Sort the numbers in this sequence in ascending order.” A number of sorting algorithms can be used to solve this problem. One solution is to compare neighboring pairs of numbers in sequence and to swap them if they are in the wrong order. This process is then repeated until no further number swaps are necessary. This usually requires several iterations.

Baking recipes illustrate the concept of algorithms, as they also consist of a series of steps which must be executed in order to achieve a specific goal. Data (the ingredients) are fed in and then processed step by step to produce a predefined result (e.g. a cake). However, unlike in baking, all sub-steps must be expressed in a way that can be directly interpreted by a computer. A computer cannot add “a pinch of salt,” only an exact amount, e.g. “2 grams of salt” (cf. Zweig 2016).

3.3 Algorithmic decision-making processes

Algorithmic decision-making describes the entire process from collecting and analyzing data to interpreting the results and identifying a decision or a decision-making recommendation on the basis of these results. The decision-making process is predetermined and can treat any number of data points. The algorithm acts as a (semi-)automated tool or as a sole decision maker. ADM thus comprises several steps and sub-processes, from the development of the algorithms themselves to the modeling and interpretation of the output.

3.4 Data mining

Data mining or data analysis involves examining (large) data sets for patterns. The type of pattern varies: for example, it is possible to search for events that often occur together. Data can also be automatically sorted into predefined classes (e.g. groups of people) on the basis of specific characteristics.

Pattern recognition is generally based purely on correlation between data points and says nothing about causal relationships. When using pattern recognition, therefore, it is always important to avoid confusing the identified correlation with a cause-effect relationship. The human brain, however, has a strong tendency to interpret correlations as causal statements and meaningful narratives (Kahneman 2012).

A famous example of successful data mining in large data sets comes from the US retailer Target, where it was found that pregnant women often buy certain products (Hill 2012). The purchases made by pregnant women are so unusual when taken together that the company is able to identify them with a certain degree of accuracy. The pattern identified here is therefore specific consumer behavior, which correlates with the pregnancy of the purchaser.

3.5 Learning algorithms

“Machine-based learning” or “learning algorithms” are methods in which the algorithm itself derives rules from consideration of many data points or examples. Calculating the result involves two phases: first, the algorithm tries to find patterns in the data. To do this, the algorithm is trained using known data sets to recognize a desired result. For example, an algorithm can “learn” to recognize faces in images by observing large volumes of images. Whether an image contains a face is indicated in the training data. Through the repeated comparison of images with and without faces, the algorithm “learns” rules that it can use to detect a face. It is therefore looking for patterns in the structure of the image (e.g. edges and shading) that indicate the presence of a face.

In the second phase, these rules are applied to new, unknown images. Using the rules which it has previously learned, the algorithm can now determine with a high degree of accuracy which images contain faces and which do not. What sets learning algorithms apart, therefore, is that the rules which they use for decision-making are not programmed directly, but instead derived from the data.

3.6 Modeling

When an algorithm is used to answer a question, it does not process a direct image of reality, but a model of reality specifically designed for the procedure at hand. There are always possibilities for modeling when an algorithm is being developed or implemented, depending on the sub-steps into which the problem is broken down. Often concepts must be made measurable first, for example friendship, or the relevance of a message (Zweig 2017: 15). It is important to decide which characteristics are relevant to the solution, depending on the objective, and which can be ignored. A task can be mapped in an algorithm in several different ways. If, for example, a navigation system is to calculate a route from A to B, it must first be decided which means of transportation will be used and whether the shortest route (in kilometers), the fastest route (in minutes) or the cheapest route (e.g. without toll roads) should be selected. The task “Show me how to get from A to B” can thus be interpreted and mapped out in a number of different ways. A model always involves condensing the original, and thus represents a restricted or abstracted image of reality. Thus, the most important factor influencing participation is not usually the mathematical construction of an algorithm per se; more important in this regard are generally the operationalization of the concepts and of the questions to be answered, the implementation of a suitable algorithm, and the collection of suitable data.

3.7 Social embeddedness of algorithms

Often the decisions and predictions generated by algorithms are considered neutral or independent. However, ADM processes are influenced by people at many points: people set the targets, and design and implement the processes; people interpret the results and determine the consequences, or – more generally – the range of possible consequences. The fact that an ADM system is used does not mean that the overall process is any less vulnerable to errors or bias.

Processes of algorithmic decision-making are often used to prepare and support human decision-making, or to replace it altogether. As a result, key aspects of human decision-making processes no longer take place, or they are carried out by others or become more limited in scope. To give an example, it is not possible to use algorithms to automate the discretionary decisions made by humans in special cases (Spiekermann 2015).

Without interaction with humans, many of the algorithms described here would be useless. Technology must not be seen as a neutral object because its design can have far-reaching policy consequences (Denardis 2012). To avoid being led astray by the assumption of objectivity, algorithms and their technological embeddedness must be understood as social constructs (Winner 1980). Algorithms are conceived and designed by humans and, in turn, influence human behavior in their application. To understand the effects of algorithms on society, it is necessary to consider their social construction in the interaction between humans and technology (Brey 2005). Algorithms, therefore, are “technical objects [which] are achieving an ever higher level and greater degree of agency” (Rammert 2006: 28) and which are thus also becoming a more important part of the social structure.

ADM is always based on specific values and norms. As a result, the algorithm should never be examined in isolation; the social context in which it is embedded must also be considered. The effect of algorithms on participation is often visible only when the algorithm is perceived as a social construct. Even if we wish to evaluate whether ADM processes function “better” or “more efficiently” in certain cases than human decision-making processes, what remains important is the interaction between people and technology. A thorough understanding of social decision-making processes is therefore needed given the growing significance of algorithms.

3.8 Explaining algorithmic decision-making processes

Human and ADM processes are fundamentally different and result in different types of errors. This makes it difficult to evaluate the quality and value of each type of process and to compare them with one another. Explaining algorithmic decision-making is also complicated by three phenomena:

- Human decisions determine the design of the algorithmic systems in many areas (e.g. when making inputs measurable) (cf. chapter 3.7).
- ADM processes often evaluate data that originate from human behavior (e.g. use of social networks as an indicator of the relevance of postings).
- The calculations taking place in many common algorithmic systems have now become so complex that they can no longer be reproduced by humans. The patterns and logic behind the decision-making processes remain hidden, especially in the case of learning algorithms (see chapter 3.5). With such complex systems, changes to the algorithm and to other variables can have a huge impact on the behavior of the overall system. Even if a human makes the final decision – whether to invite a candidate to a job interview, for example, after an algorithm has pre-sorted the applications – the system leaves only limited room for maneuver. It is rather unlikely that a person will revise or accept only part of the preliminary decisions made by an algorithm (Hannah-Moffat, Maurutto and Turnbull 2009).

It is often not possible for users of ADM systems to review individual cases, because of a lack of transparency, time or expertise, for example. In many cases, it is impossible to clearly distinguish between a semi-automated and a fully-automated process. If a person is no longer able to base his or her decision on reasoning, and must instead rely on the output of a computer, the line between human and algorithmic decision-making becomes blurred.

4 Field of analysis: computer systems that evaluate people

ADM processes have a variety of applications. It is almost always possible to construct a participation-relevant scenario in individual cases. But when is participation fundamentally impacted by an ADM process? This paper proposes provisional, pragmatic criteria for assessing a process’s impact. The classification scheme does not offer an absolute measure of how participation is affected, but measures it on a scale instead. This relative scale serves as an initial tool for comparing how different algorithm-based processes and applications affect participation. The field of analysis must be limited to phenomena with existing or foreseeable impacts on participation. This subset can then be subjected to a more in-depth, qualitative analysis. In this paper, ADM processes are considered fundamentally relevant if a digital decision-making system (section 4.1) is used and if people or the characteristics attributed to them (section 4.2) are evaluated.

4.1 Digital decision-making systems

The first limitation of the field of analysis reflects a technical aspect: this paper only considers digital decision-making systems; algorithms also exist that are applied manually. Since algorithms are only clear descriptions of a finite series of actions, they do not necessarily need to be computer-controlled. A very complex questionnaire, a hygiene guide for employees or an analog script for administrative staff can also be understood as algorithms. In many formalized, bureaucratic procedures, an input is converted into a desired output on the basis of a defined sequence of steps. Analog algorithms are not new – on the contrary, they have been shaping how humans coexist for a long time. The first filter for the field of analysis is therefore: is the decision-making system digital?

A decision-making system is considered a digital when the underlying algorithm is executed by a computer. The restriction to digital ADM systems does not mean that the social embeddedness of these systems is ignored, however. Rather, there is a pragmatic focus on those cases that generate new issues with regard to participation.

4.2 Evaluating people

Those decision-making processes are especially relevant which evaluate people or the characteristics and products clearly associated with them. This includes systems that control the distribution and production of content, for example on social media. These can have a direct or indirect impact on participation. For example, discrimination on the basis of personal characteristics often only becomes visible by looking at societal phenomena. Content selected for a specific person – an individual’s Facebook newsfeed for example – exhibits virtually no recognizable discriminatory structures, as such content is meant to be personalized. It is only when we look at the overall picture and consider a large number of personalized decisions that discrimination becomes visible. The same is true if software pre-sorts job applications and regularly suggests that a disproportionate number of applicants from certain neighborhoods not be invited to an interview. In such cases it is impossible to discern on an individual basis whether the participation of a certain group of people may be adversely affected, for example due to the neighborhood in question being characterized by a low income level or a high level of immigration.

Algorithms that are used in automated production processes, such as in industry or agriculture, are not included here due to the “evaluating people” filter. To give another example, elevator software, which uses algorithms to automatically determine which elevator should move first and to which floor, will not be examined for its impact on participation in this paper. And although cases like the Volkswagen emissions scandal can, in isolated instances, be highly relevant for participation, such instances will not be considered in order to limit the field of analysis to a manageable size.

Our proposed classification thus analyzes computer systems that evaluate people. These pre-selection criteria are addressed in further detail in chapter 6. It could be useful to further limit the field of investigation for other purposes or contexts, something that can easily be done by narrowing how participation is defined.

5 Application scenarios for ADM processes

A number of application scenarios from various social spheres are given below to provide an empirical basis for the classification of ADM processes. The list of examples has been chosen to highlight how participation can be impacted and does not claim to be complete. The focus here is on using examples to map the range of applications and affected areas of life that play a role in terms of social participation (for a detailed analysis of case studies, see Lischka and Klingel 2017).

5.1 Employment

5.1.1 Platform work: when the boss is an algorithm

In the platform economy, algorithms are increasingly being used to distribute, evaluate and improve work (Agrawal et al. 2015; German Federal Ministry of Labour and Social Affairs 2015; Choudary, Alstyne and Parker 2016). Here, algorithms serve primarily to improve automation and efficiency. The concept of platform work encompasses a wide range of new employment models and management processes. An essential component of platform work involves connecting decentralized contractors with equally decentralized and distributed clients. New technical infrastructures allow for new communication channels, through which social resources can be shared and a variety of interests can be brought together (Hensel et al. 2016; Kocher and Hensel 2016). The sectors and areas of work range from simple to highly-qualified industrial work (e.g. Clickworker, TaskRabbit, Amazon Mechanical Turk), but also include household-related and personal services (e.g. Uber, Lieferheld, careship.de); for a detailed overview see Leimeister, Durward and Zogaj (2016) and Schmidt (2016).

Crowdwork companies use software to automate and design workflows for employees. Algorithms issue instructions, grant access, choose participants and even regulate controlling mechanisms. Assessment and scoring systems are also used whose evaluations partially determine how future contracts are awarded. Lee et al. (2015) describe the coordination of work using such intermediaries as “algorithmic management.” This includes management and coordination between platform providers, clients, employees and end users, generally via a centralized platform. The platform algorithms therefore assume traditional management functions, such as the coordination of work, the optimization of workflows, the evaluation of services and the planning of shifts.

When algorithms coordinate workflows, it has an impact on several fundamental rights, including the right to privacy in the workplace and the right to employee representation, i.e. co-determination and the collective advancement of workers’ interests. Decentralizing the division of labor and making people dependent on the communication structures used by the platforms has the potential to undermine the need for protection workers have vis-à-vis platform operators and clients (Arthurs 2011; Freedland and Kountouris 2011).

Platform work does not fit into the classical dichotomy between the principles of “hierarchy” (paid employment) and “market” (self-employment). Rather, it represents a hybrid principle which is essentially characterized by algorithms (Aneesh 2009). Work platforms take into account a variety of features when evaluating employees.

There are also specific mechanisms for providing feedback about work that include clients in the review process, and there are rating systems that sometimes replace traditional instruction (De Stefano 2015; Lingemann and Otte 2015). The fact that working conditions are set by the platforms also raises questions about employees' self-determination. Many employees appreciate the high degree of flexibility that allows them to work when and where they wish. Yet as a small part of a large "crowd," people who work on platforms are often treated no better than "humans as a service" (Irani 2015). The virtually seamless interchangeability of professionals is the result of global competition among crowd workers and the low level of effort required for becoming familiar with and carrying out many tasks, among other factors. Without any requirement to cater to individual or collective interests, algorithms define the limits of employees' autonomy.

5.1.2 Candidate selection: who should be invited to the job interview?

Surveys indicate that in the UK and in the United States 60 to 70 percent of applicants are subjected to automated selection procedures and tests (Lischka and Klingel 2017: 21). The effectiveness of the personality tests used in these countries to predict work performance is controversial, and methods are not transparent. There is evidence that some procedures disadvantage certain groups (people from poor neighborhoods or with medical conditions). The primary function of the algorithms in such cases is to increase efficiency.

Automated procedures are used to pre-select applications. A portion of the latter are immediately rejected on the basis of online tests, even before a human decision maker has seen the applications. The employer can determine what percentage of the applications received are immediately refused. One test provider puts the percentage of automated rejections at around 30 percent (Weber and Dwoskin 2014).

In principle, the use of such systems offers the opportunity to reduce discrimination if the assumptions underlying the selection process are made transparent and reviewed. For certain groups of people, however, such automated procedures can make access to the labor market more difficult, or even impossible. An existing form of discrimination – such as rejecting people with "foreign" sounding names (cf. Schneider, Yemane and Weinmann 2014) – thus risks being replaced by another. Since the technology is used mainly in the low-wage sector, companies are unlikely to invest in testing and improving the systems, since "the systems do not have to find the best of the best, they just have to be more efficient than the previous selection system. Investments in the calibration of systems and in the continuous testing and updating of the decision-making logic and data stock will never be as important in this sector as in those with a low supply of labor, high demand and correspondingly high salaries" (Lischka and Klingel 2017: 13).

5.2 Public safety

5.2.1 Police: which areas should be patrolled?

Police departments employ algorithms to predict areas with particularly high crime rates. Essentially, the aim is to better anticipate crime hotspots through the use of algorithms. These calculate the location of hotspots for certain offenses, such as burglaries. The results are used to set priorities when deploying patrols. Such systems are regularly used in Europe, for example in Zurich (Baumgartner, 2015), Kent (Mohler et al. 2015) and Milan (Mastrobuoni 2015). In Germany, this form of "predictive policing" is currently in use or under development in 14 pilot projects and tests (Pilpul 2016). Predpol in the US and Precobs in Germany are among the most widely-known analysis programs which use this approach.

Authorities and criminologists have been working with geographical patterns of criminal activity since the nineteenth century, when this practice was developed in London (Gluba 2014: 5). Long before software was deployed, human analysts were using crime statistics to identify high-crime areas. To provide some background: "The prediction of hotspots for specific offenses is based on 'near repeat' theory. This criminological approach assumes that following criminal offenses such as car theft, and domestic and car burglaries, the likelihood of further offenses in the immediate area will increase" (Lischka and Klingel 2017: 13). Empirical studies from the

UK, US, Netherlands, New Zealand and Australia show statistically significant near-repeat patterns for domestic burglaries (Ferguson 2012: 19).

Precobs also analyzes a small number of parameters, such as the time of the offense, the type of stolen goods (e.g. cash, till), the type of building (e.g. business premises or residential building) and the break-in method (e.g. lever, pressure, kicking) in order to recognize the patterns of serial offenders (Brühl 2014). The prediction of crime hotspots is therefore just one among many applications of predictive-policing technologies. In Chicago, prediction technologies are also used to identify potential individual offenders (Davey 2016).

5.2.2 Courts: who should be released on parole?

In many US states, software is used to predict the likelihood of offender recidivism (Barry-Jester, Casselman and Goldstein 2015). More than 60 predictive tools are available on the market, many of which are supplied by private companies, including the widely-used COMPAS system from Northpointe. These software products are used to anticipate behavior and to classify people.

The COMPAS system assesses offenders on the basis of 21 categories, including the risk of recidivism for violent acts. Values for all categories are given on a scale of 1 to 10. According to the user manual, the probability of recidivism is “high” when the overall score is between 8 and 10 (Northpointe 2015: 11). The COMPAS system determines scores based on answers to 137 questions. “The weighting of the answers in the calculation has not been made public” (Lischka and Klingel 2017: 8).

Preliminary investigations of the prediction procedures indicate racial discrimination: the proportion of people of color with a high score predicting recidivism but who do not re-offend within two years is twice as high as for white people (Angwin et al. 2016).

5.3 Public opinion

5.3.1 Newsfeeds: which content is shown?

Intermediaries – such as social networks and search engines – use ADM processes to create personalized content for each user. They are based on assumptions about user preferences and make predictions about user interests. The algorithms used in such instances therefore serve the core function of personalization. Providers collect and interpret a wide range of user data in order to generate personalized newsfeeds. The feed of images, videos and news displays only a selection of all available content, meaning that social media platforms, especially those with a wide reach, have an editorial function (Helberger and Trilling 2016). Newsfeed algorithms like Facebook’s EdgeRank select and sort the content that is visible to users. Such is the power of social media: quietly deciding for users which content to prioritize and which to ignore (York 2010).

The algorithmic selection of content directly affects freedom of expression; it can also have an indirect impact on fundamental rights and basic democratic principles. In a 2010 experiment, Facebook showed that it has the ability to influence election behavior using its newsfeeds (Bond et al. 2012). In the experiment, some Facebook users were shown how many of their friends had already voted on November 2, 2010, the day of the US midterm congressional elections. Users were also given the option to click a button to indicate that they had voted. The intervention increased the likelihood of clicking on the “I voted” button for some 2 percent of the users who saw their friends had voted. Not all of those who clicked actually went to vote, however. A comparison of Facebook data and actual voter turnout indicated an increase of just 0.4 percent. The control group showed no increase in voter turnout. In the case of a close result, such an effect could theoretically be large enough to decide the outcome of an election, particularly in first-past-the-post systems (Tufekci 2014; Zittrain 2014). A targeted manipulation would be difficult to prove, as this would only be discernible by looking at a high number of profiles. Another experiment was also able to demonstrate the effect of search results on the voting decision of undecided voters (Epstein and Robertson 2015). The highest ranking results in the searches had the greatest impact on the choices made by test subjects.

5.3.2 Content control: which content is removed?

The impact of intermediaries is not limited to the content they display and how this is prioritized. The content that is not displayed or that is purposefully removed is equally relevant to participation. Many intermediaries, e.g. Facebook, Google and Microsoft, use automated or semi-automated processes to deal with reported content and requests for removal (Wagner 2016; Zhang, Stalla-Bourdillon and Gilbert 2016). Many online services rely on automated solutions for filtering and blocking content, whether to protect against spam or illegal content such as abusive language or hate speech.

When algorithms are used to review and remove content, this has a direct impact on freedom of expression. In many cases, content control is not based on legal parameters, but rather on the general business conditions of the respective platform operator. This creates tension between what is permitted by law and what intermediaries permit as content. Large service providers are increasingly under pressure – particularly from legal authorities – to remove content on the basis of vague or ambiguous criteria such as “terrorist propaganda” (Brown and Cowls 2015; Europol 2015). Law enforcement agencies shift the responsibility for the control of undesirable content onto private operators, which often results in the loss of freedom of expression.

5.4 Public services

5.4.1 Job placement: which employment measure is the right one?

It is evident that public administrations are increasingly using automated procedures (Eubanks 2013; Tufekci et al. 2015; van Haastert 2016). In Poland, ADM systems are used to find jobs for the unemployed. Software groups the unemployed into three categories, which correspond to three different government-run employment programs (Niklas 2017). The data used to evaluate the unemployed is collected by means of an online survey, which serves as the basis for calculating a score. The agency responsible for this initiative has been compiling profiles since 2014. Data relating to age, gender, physical or mental disability and family status are included in the decision-making system. Researchers from the Panoptykon Foundation have investigated the procedure (Niklas, Sztandar-Sztanderska and Szymielewicz 2015). They highlight the fact that the use of algorithms in the provision of public services entails major challenges in terms of transparency and accountability. The use of such methods must be closely monitored and continuously evaluated, as the selection of training programs, for example, has a marked impact on opportunities for participation.

5.4.2 Health protection: which houses should be renovated first?

Lead poisoning in children is still a major problem in Chicago, as significant portions of the city’s housing stock date from a time when lead-based paint was still in use. In 2013, 10 percent of children under six years of age had blood lead levels higher than the limit set by the Centers for Disease Control and Prevention (Hawthorne 2015). The city, together with the University of Chicago, has therefore developed software to predict which children in which buildings are at the highest risk of lead poisoning. The goal is to carry out early, targeted and, thus, cost-effective interventions. Ranking buildings according to the risk of lead poisoning helps to prioritize measures for protecting the public (Potash et al. 2015), allowing high-risk buildings to be checked more quickly by inspectors and renovated when limits are exceeded.

The following data are available to researchers: 2.5 million blood test results, giving the date and the test subject’s identity, age and place of residence, as well as the results of 120,000 building inspections from the same period, giving the date and location. Previous research has shown that information relating to age, inspection results and building condition is particularly important for improving the quality of predictions at address level (ibid.: 2044). This process serves as an example of how public administration can benefit from algorithmic decision-making. Of course, it would be even more effective to use additional funds to renovate all of the housing stock. Since resources are scarce, however, risk-ranking undoubtedly represents a major improvement in public-healthcare provision.

5.5 Marketing

5.5.1 Recommendation systems: what will you be interested in next?

Many online services use recommendation systems to suggest specific content or products to individual users, such as books or videos. Highly-detailed user profiles can be created using cookies and technologies such as browser fingerprinting, and data from search queries, social networks and smartphone apps (Tene and Polonetsky 2011). Even logged-out users are increasingly shown personalized website content, making it difficult to “opt out.”

Algorithms are used to analyze the collected information in order to make personalized recommendations. Personalization was previously a luxury often only accessible to a select few due to its high cost. Algorithmic recommendation systems can provide cost-effective personalization for all. The question arises, however, of whether these recommendations actually reflect the preferences of users, and what effect they have on the diversity of media and opinions, for example.

Profiling entails the risk that data will be taken out of their original context. Recommendation systems construct a “digital double,” which is not representative of the person it is meant to represent (Bauman et al. 2014). Such systems can also result in filter bubbles (Pariser 2011), wherein people are only exposed to content which echoes their own point of view. The actual impact of filter bubbles resulting from recommendation systems is disputed, but it is nonetheless important to look closely at whether they exist and, if so, how they come about (O’Callaghan et al. 2016). The customization of prioritized content according to personal characteristics, such as gender or income, can have discriminatory effects. One study found that Google plays fewer ads for high-paying jobs on computers linked to Google advertising profiles declared as female than on computers linked to male profiles. The researchers were unable to determine how this difference occurs (e.g. through targeted bookings) (Datta, Tschantz and Datta 2015).

5.5.2 Insurance: who pays which premiums?

The Admiral Group, the third largest car insurance company in the UK, announced in November 2016 that it wanted to analyze its customers’ Facebook profiles. The data were to be used to determine the pricing of insurance for female novice drivers (Stalder 2017). As no data on driving behavior or accident history are currently available for this group, the insurance company wanted to use Facebook data to identify personality traits that are associated with safe driving. The algorithm used for this purpose looks for correlations between data from social media and existing data relating to claims (Ruddick 2016a). Participation in the program was to be voluntary and discounts were offered to participants. At the last moment, however, Facebook refused the use of its data; as an alternative, the insurance company now wants to collect data on novice drivers via a questionnaire (Ruddick 2016b). On the one hand, the individualization of insurance rates can reduce costs; however, it can also conflict with the principles of solidarity and lead to discrimination against certain groups.

Many young drivers might feel compelled to participate in the program if they cannot afford to do without the offered discount. Other insurance rates take into account data relating to driving behavior. For example, BonusDrive, a telematics plan from Allianz,[1] is aimed at young drivers who do not yet have any accident history. BonusDrive uses an app to transmit data about driving behavior to the insurance company, and people who drive safely – by keeping their distance and sticking to the speed limit, for example – receive a discounted rate. Specific details of how driving is assessed, however, have not been disclosed, nor has their influence on pricing.

[1] https://www.allianz.de/auto/kfz-versicherung/telematik-versicherung/ (downloaded March 1, 2017)

5.6 Preliminary conclusion: what do algorithms actually do?

These application scenarios illustrate the broad spectrum of participation-relevant algorithms. The following offers a brief summary of the tasks executed by the ADM processes outlined in the examples. The methodology used expands on the work of Gillespie (2014) and Diakopoulos (2016) and summarizes the most important activities carried out by ADM systems.

5.6.1 Automation

The central purpose of ADM systems is automation, i.e. acting without human intervention (Winner 1978). Their ability to act and react in an automated manner allows computer systems to replace people in an ever-increasing number of situations. Automated systems are used for both simple and complex tasks. Decisions that are automatically made by algorithms are not always predictable – or even explainable – to humans. Input data must be prepared and adapted if they are to be processed automatically by an algorithm. Automation should not, however, be equated with autonomy. Decision-making systems are never fully autonomous: they always operate within a prescribed framework and social environment. Limits are always set during development and implementation which specify the algorithm’s function and concrete task (Stalder 2017).

5.6.2 Increasing efficiency

Automation is closely linked to the goal of making processes more efficient. Efficiency gains result from the acceleration of processes (technical efficiency) and from the improved scalability of algorithmic processes (economic efficiency). Better scalability generally reduces marginal costs to a very low level and allows a large number of cases to be handled with relatively few resources. Cost reductions also result, however, because human labor is being replaced, thereby lowering personnel costs. It also means that decisions are made by one centralized system instead of at numerous widely distributed locations.

5.6.3 Prioritization

Many decisions influenced by algorithms result in prioritization or hierarchization. These judgments are, in principle, equivalent to “subjective decisions,” in which an assessment is made for which there is no obvious correct or incorrect answer, no clear “yes” or “no” (Pasquale 2015: 8). These include questions such as “Should this employee take a break?” or “Which content should be displayed at the top of a newsfeed on election day?” Thus, algorithms help define what is relevant. In terms of filtering information, prioritization is part of everyday life, since without it we could not process the flood of information we encounter on a daily basis. Algorithms allow us to use our time, attention, money and other scarce resources in a more purposeful manner (Diakopoulos 2016: 57).

5.6.4 Classification

Many ADM processes group objects, people, concepts and other elements into specific categories. Classification refers to sorting information into predefined categories based on typical characteristics. The sociologist David Lyon refers to the sorting of people into groups as “social sorting” (Lyon 2003). Classification into a particular group, for example into the category “is likely to re-offend,” can result in the need for multiple subsequent decisions. Classifications almost inevitably entail “false positive” or “false negative” groupings. If a person who is classified as “will not re-offend” does in fact commit another crime, this is considered a false negative.

5.6.5 Individualization

When systems are tailored to individual people or individual accounts, this is referred to as individualization. Customization on the basis of user data is usually carried out incrementally, by adding further data about the user over time. Algorithms enable the individualization of services and other goods in new industries and sectors. However, this does not always result in more autonomy for users. New forms of discrimination can arise if algorithms adapt a platform’s terms of use or pricing on an individual basis.

5.6.6 Anticipation

Algorithms are also used to make statements about the future. These predictions are based on statistical calculations. The goal is to predict certain patterns or events with increased accuracy. Anticipation assumes that a specific phenomenon, such as the correlation between two or more variables, will happen in the future too. The accuracy of predictions only extends to abstractly defined groups, however; precise predictions for individual cases cannot be made.

6 Calculating the potential impact of algorithmic decision-making on participation

The greater the potential impact on participation, the more closely a procedure must be examined. This also applies in the positive sense: the greater the potential impact on participation, the higher the potential for social improvement. Here too it is worth taking a closer look. In any case, the greater an algorithmic process's impact on participation, the higher the standards it must meet and the greater the need to reflect in detail on its possible consequences.

Assessing the impact on participation – an undertaking for which this paper proposes a structure – provides a basis for further steps, be they more detailed analyses or regulatory measures. If a considerable impact on participation is anticipated, efforts must be consistently made to ensure that the positive effects are fully exploited and that any negative effects, such as discrimination, are prevented. Thus, when algorithms greatly affect participation, it is often necessary to make certain that diversity, fair competition and the rule of law are not diminished. The higher the standards that participation-relevant ADM processes must meet, the more comprehensive the measures that must be considered.

6.1 Methodological approach

There can be no binary decision-making criteria when classifying the potential impact on participation, as this would ignore the many socially relevant logics embedded in ADM processes.

A variety of criteria can influence the degree to which participation is impacted. Furthermore, the selection and weighting of these criteria depend on the definition of participation that is used. It is not a matter of assessing whether ADM processes are good or bad; what is needed, rather, is a relative evaluation of their potential impact on participation, regardless of the form it takes.

The criteria are divided into three groups: *actors, social embeddedness and consequences.* Thus, the groups do not reflect technological factors, but are an attempt to document political and social considerations.

The actors-related criteria examine the actual economic and political power of those supplying and/or operating the decision-making processes. The criteria used for social embeddedness reflect mutually reinforcing social interdependencies, both intended and unintended. Finally, the potential consequences – along with those that have already become apparent – are analyzed in terms of basic political and social rights. Existing legal norms are used as the guidelines for evaluating the impact on participation. The starting point is the German Antidiscrimination Act (§1 and 2 AGG) and its area of application. Discrimination is thus prohibited in the following areas in particular: recruitment, hiring, working conditions, membership in trade unions, and occupational training. Fair access to public goods such as education, social security, health care and housing must also be ensured.

As already described in chapter 4, a (theoretical) case study could, in principle, be designed for almost every ADM process in which a potential impact on participation is discernible. We therefore suggest assuming there is an impact on participation only after a certain threshold has been reached. The relative consequences for participation can be summarized and compared using the criteria sketched out below. This makes it possible to rank algorithms more simply and manageably in order to prioritize further steps, where needed.

Each of the test criteria listed below can increase the *potential impact on participation* (PIP) (PIP = +1). If a criterion is not applicable or cannot be evaluated as part of the quick-check procedure, no point is given for the potential impact ($PIP_{neutral}$ = 0). Half points (PIP = +0.5) can also be assigned, but these are limited to no more than two per criterion (PIP_{max} = +2). The values thus range from zero to two points per criterion. The higher the total value, the greater the process's potential impact on participation.

Test questions are used to evaluate the criteria, and answering them is not always easy. The various applications are highly diverse and, on occasion, obscure, which is precisely why the system of abstract classifications is beneficial: it reduces the complexity of a highly complex network of issues. What might limit or promote participation? To answer that question, a person or organization uses the evaluative structure to analyze one or more relevant processes. After applying the criteria to specific cases and considering the results, they can assess which aspects of the decision-making process are most prominent; they can also compare different processes with each other.

6.2 Actors

A1 Competition: Who is supplying/operating the algorithm? How much political and economic power do they have? Is there real competition between multiple suppliers/operators (oligopolies, monopolies, cartels, dominant market position, sovereign power)?

Example – Border controls: Even if private companies produce or supply software for granting visas, the software's use is an expression of sovereign power. No one can escape the state's power to grant visas, and to that extent a private supplier working on behalf of the state must also be seen as a monopoly. The actual organizational type or legal status of the supplier/operator is therefore less relevant for the impact on participation.

- ➢ If competition between several supplier/operators is weak, the impact on participation is increased (PIP = +1).
- ➢ If the supplier/operator is part of an oligopoly or is a monopoly, any potential impact on participation is likely to be quite high (PIP = +2).

A2 Dependency: Is it a problem when the process or product is no longer available or access to it is denied? Can the product be readily replaced? Do substitute processes/products exist? What would it cost to switch to another process? Which organizations or groups are structurally dependent on the process? Do the system's construction and functioning result in an informal dependency? Which technical or social lock-in effects exist?

Example – Crowdwork: The market for crowdworking platforms is highly competitive. Several platform providers compete for both clients and crowdworkers. It is nevertheless possible that employee dependency could arise if job history and reputation are linked to the platform's specific algorithmic evaluation procedures. A similar lock-in effect can also occur in other assessment systems or social media.

- ➢ If a system creates dependencies, the potential impact on participation is increased (PIP = +1).
- ➢ If dependency is very high because changeover costs are prohibitively high, then any potential impact on participation will be especially strong (PIP = +2).

6.3 Social embeddedness

SE1 Self-determination: To what extent does the process serve (only) as a preliminary step for making a decision? Does the user retain any right to influence the process? Does the system decide (de facto or officially) on its own? How much freedom to change or influence an algorithmic decision does the user have? Do time pressure and the process's practical implementation affect the user's autonomy?

Example – Content control: The way software is designed for monitoring potentially illegal content can have a major impact on user autonomy. Time pressure (e.g. fast pace) and technical limitations can, de facto, greatly reduce the possibility of human input, even if a human will officially be the one to make the final decision.

- If the user remains autonomous (SE1), participation will not be affected (PIP = 0).
- If people only have a perfunctory influence on decisions, which are actually (largely) automated, the potential impact on participation is increased (PIP = +1 or +2).

SE2 Adaptation: How do people adapt to the algorithmic process? What impact does the decision-making system have on a practical level? Which interactions between humans and computers can potentially change the outcome? What is the process's actual power in social terms? Which (social) significance is ascribed to the process? Does it define and/or lead the discussion? How inherently influential are the process outcomes?

Example – Trending topics: The "trending topics" on Twitter are not infrequently seen as a social "reality." Although it is not generally known how the ranking is created, it still affects the public discourse and different actors adjust their behavior to reflect the logic used by the "trending topics" algorithm.

- If the decision-making process has social authority, the potential impact on participation is increased (PIP = +1).
- If users adapt their behavior to the decision-making system to a significant degree, then the potential impact on participation is particularly pronounced (PIP = +2).

6.4 Consequences

C1 Reach: How many people are affected by the decision-making process (e.g. user population)? Is the process's reach known and/or can it be limited? How great is it?

Example – Candidate selection: If only one company uses a software application for choosing job candidates, then the application's reach is relatively small. If, however, it becomes a core decision-making tool for many companies, its impact on participation increases accordingly, for example if the logic used by one system predominates when employees are hired in a particular region or industry.

- The greater the reach of decision-making processes, the stronger the potential impact on participation (PIP = +1 or +2).

C2 System change: Does the decision-making process undermine principles of solidarity? Would individualization change the system? Could the process result in an unintended transformation of the impacted (social) environment?

Example – Health insurance: If the costs for public health insurance are partially decided by algorithms and are subsequently individualized using a new logic, then a system which has been based on social solidarity will no longer be to some degree. Instead of having everyone pay only according to their ability to do so, at least part of each person's contributions will be determined by individual factors.

- ➢ If the process changes the basic principles of its (social) environment, it can potentially increase the impact on participation (PIP = +1 or +2).

C3 Discrimination: To what degree could people be disadvantaged by the decision-making process? Do the results exhibit a pattern of discrimination or might such a pattern be expected, e.g. based on race, ethnic or cultural background, gender, sexual identity, physical disability, age, religion or world view?

Example – Automated job listings: An analysis of Google's employment advertisements revealed that users classified as men were shown more highly paid positions than women were. In this case, the process has an impact on job market participation.

- ➢ If the process structurally disadvantages people or groups or could do so, the potential impact on participation is increased (PIP = +1 or +2).

Table 1: Overview of criteria for assessing the potential impact on participation

Dimension	**Criteria**	**Explanation**
Actors	**A1** Competition	Who is the operator of the algorithm? How much political and economic power do they have? Is there effective competition between several operators (oligopolies, monopolies, cartels, market dominance, sovereign power)?
	A2 Dependency	Is it a problem if the process or the resulting product is lost or access to it is denied? Can the product be replaced immediately? Do alternative procedures/offers exist? What is the cost of switching to an alternative procedure?
Social embeddedness	**SE1** Self-determination	Does the procedure serve (only) to prepare for a decision? Does the user retain autonomy? Does the system decide (de facto or officially) on its own? How much freedom to change or influence an algorithmic decision does the user have? Do time pressure and the process's practical implementation affect the user's autonomy?
	SE2 Adaptation	How do people adapt to the process? How does the decision-making system work in practice? Which interactions between humans and computers might change the result under certain conditions?
Consequences:	**C1** Reach	How many people are affected by the decision-making process (e.g. user population)? Is the process's reach known and/or can it be limited? How great is it?
	C2 System change	Does the decision-making process undermine principles of solidarity? Would individualization change the system?
	C3 Discrimination	To what degree could people be disadvantaged by the decision-making process? Do the results exhibit a pattern of discrimination, or might such a pattern be expected?

6.5 Summary: determining the potential impact on participation

The overall potential impact on participation (PIP) of any new algorithmic decision-making process can therefore be represented as follows:

$$PIP = VA1 * VA2 * (A1 + A2 + SE1 + SE2 + C1 + C2 + C3)$$

VA1 and VA2 represent the pre-selection criteria explained in chapter 4. The filter criteria “Digital decision-making systems” (VA1) and “Evaluating people” (VA2) can be given either a value of 0 (not true) or a value of 1 (true). The value for the potential impact on participation (PIP) is the product of the two pre-selection criteria and the sum of the participation-impact criteria. If one of the factors has a value of 0, this also directly affects the overall result (PIP = 0). Otherwise, the specific PIP value corresponds to the sum of the test criteria (A1 + A2 + SE1 + SE2 + C1 + C2 + C3).

The overall result (PIP) indicates the degree to which an ADM process can affect participation. The scale ranges from 0 to 14. The results can be used to compare the impact potential of different processes, and the relative outcomes can be used to prioritize processes for further investigation. For example, a case study with a PIP of 11 is a more relevant candidate for further examination than one with a PIP of 3. At the same time, the tool was not developed to make absolute judgments. It does not allow absolute statements to be made, e.g.: “Processes with a PIP of 6 or above are particularly relevant for participation.”

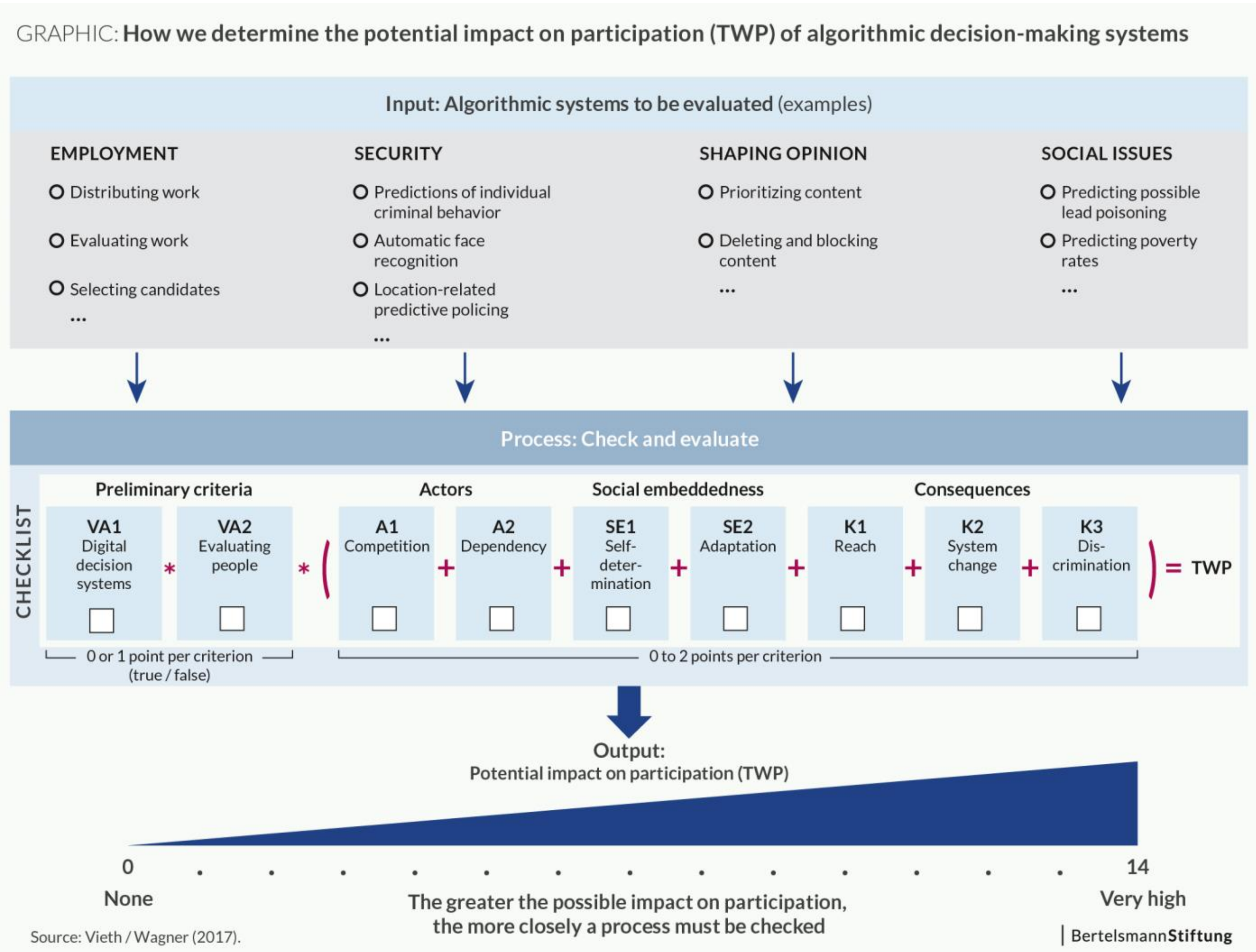


Figure 1: Calculating the impact potential of algorithmic processes on participation

The above criteria can be used to develop a fast and relatively simple testing procedure, allowing companies and evaluators to check their ADM processes using specific criteria in just two or three pages.[2] Individuals and organizations can employ such criteria to evaluate ADM processes for their relevance to participation. To that extent, the process is a simple and quick introduction to understanding ADM processes. At the same time, it is not a substitute for a more comprehensive evaluation. The actors themselves could also conceivably define threshold values for the impact on participation when analyzing processes. In such cases, the underlying understanding of participation must be defined and operationalized in a transparent way.

6.6 Example calculations

6.6.1 Spell checking

The checking of spelling and grammar in word processing software is based on a digital decision-making system (VA1 = 1). The correct spelling and sentence structure are selected automatically. This process does not, however, lead to the evaluation of human characteristics (VA2 = 0). This automatically results in a participation impact potential of 0, and no further testing is needed.

PIP = 1 * 0 * (x)

PIP = 0

6.6.2 Predictive policing

What is the PIP of predictive policing (see chapter 4.2.1)? Predictive policing involves a digital decision-making system that evaluates human characteristics (VA1 = 1; VA2 = 1). In terms of the actors involved, we can say that, given their sole right to the use of force, the police can have a significant impact on participation (A1 = +2). Moreover, everyone in society is very dependent on the “commodity” of public safety (A2 = +2). The social embeddedness of predictive policing software is difficult to evaluate using the quick-check procedure. As far as one can tell, police officers are able to act on their own judgment, at least to some extent (SE1 = 0.5). Yet the outcome has social authority since people generally adjust their actions to reflect it (SE2 = +1). In terms of reach, the consequences – which can be ascertained based only on a small number of pilot projects carried out in cities – are moderate (C1 = +1). The use of predictive-policing applications has the potential to alter police work in structural terms (C2 = +2). The potential to reduce or exacerbate discrimination is high (C3 = +2). This yields the following result:

PIP = 1 * 1 * (2 + 2 + 0.5 + 1 + 1 + 2 + 2)

PIP = 10.5

With a score of 10.5 out of a maximum of 14 points, it can be expected that predictive policing will have a very high impact on participation. The attendant opportunities and risks must therefore be examined in a particularly attentive and detailed manner.

6.6.3 Platform work

In assessing how platform work impacts participation (see chapter 4.1.1) it is difficult to classify actors’ characteristics. It should be noted, first of all, that the pre-selection criteria have both been met (VA1 = 1; VA2 = 1). The operators to date have been very diverse and there is robust competition among them (A1 = 0). Nevertheless, dependencies can result from reputations and work histories linked to platforms (A2 = +1). The sovereignty of the people using the platform software is significantly affected (SE1 = +1.5), and considerable

[2] See for example the “Quick Check” from the Danish Institute for Human Rights: .

adjustments to the algorithmic working environment are to be expected (SE2 = +2). The reach of platforms used in multiple sectors cannot be definitively assessed using the quick-check procedure (C1 = 0). However, system-changing effects are likely. For example, platform work could have an influence on areas relevant to participation such as employee representation (C2 = +2). The first discriminatory consequences have already been identified, and an increased impact potential must therefore be assumed (C3 = 1).

$$PIP = 1 * 1 * (0 + 1 + 1.5 + 2 + 0 + 2 + 1)$$

$$PIP = 7.5$$

With 7.5 out of 14 points, the impact of platform work on participation is, in abstract terms, roughly in the middle of the scale. A detailed, industry-specific investigation would be appropriate.

6.6.4 Telematics insurance

What is the potential impact of telematics car insurance on participation (see chapter 4.5.2)? Here, too, the pre-selection criteria – a digital decision-making system that evaluates people – are clearly met (VA1 = 1; VA2 = 1). There is competition among operators/providers of the applications and no significant dependencies are expected at this point (A1 = 0). Lock-in effects could potentially occur at certain insurance companies, however, for example in cases where it is not possible to transfer a policyholder's digital driving history to another company (A2 = +0.5). Customer autonomy does not seem to be affected (SE1 = 0). Adaptation effects, such as a more cautious driving style, are desirable and may potentially arise (SE2 = +1). The reach remains very limited, as the product is only offered to novice drivers and participation is voluntary (C1 = 0). Telematics car insurance has the potential to alter its operating environment, since individualized data analysis could be carried out in the event of an accident, for example (C2 = +0.5). Finally, the focus on novice drivers results in potential discrimination, i.e. based on age (C3 = +0.5).

$$PIP=1 * 1 * (0 + 0.5 + 0 + 1 + 0 + 0.5 + 0.5)$$

$$PIP = 2.5$$

With only 2.5 points, it would not be unreasonable to exclude this application from a more in-depth investigation.

7 Conclusion and outlook

Linking two complex topics such as participation and ADM is challenging, but opens up exciting possibilities, particularly through the combination of qualitative and quantitative approaches. In the classification put forth in this paper, the criteria focus on the actors, the relationship of the ADM process with its social context, and the potential consequences for fundamental rights. Other criteria are naturally also conceivable. Limiting them, however, offers some advantages, as we have done here for the purposes of offering a simple introduction to what is a complex topic. Many of the test questions can be answered with relatively little research. The classification system can be quickly applied and the results are easily comparable. The criteria are designed to reduce complexity and increase flexibility. The test method's basic structure is relatively easy to adjust. If key components are felt to be missing, they can be added without undue effort. In addition, the simplified test criteria are well suited to visualizing inputs and results, for example in the form of bar graphs or tables.

The results could also be revealing when different people calculate the potential impact of an ADM process on participation: How do their answers to the test questions differ? What expectations do they have? How does the score change when people of different backgrounds assess the criteria? The scoring system presented here is thus also suitable for use in workshops and as a basis for discussion during strategic consultation sessions.

The debate about algorithms and their influence on participation has just begun. Providing a coherent overview of the subject is not easy given the wide range of applications and tasks handled by ADM systems. When employed as decision-making aids, algorithms are not good or bad per se. Whether their potential impact on participation is used to provide more equitable access to opportunities, to strengthen basic rights and to bring about *better* decision-making processes will depend on the algorithms' objectives and how they are implemented.

In many cases, ADM systems can indeed be the better solution, since they make it possible to avoid the inconsistencies and biases that occur when humans make decisions (Kahneman 2012: 233; Meehl 1954). Yet technology itself cannot achieve the goal of promoting participation; it must be negotiated and shaped through social debate. It is therefore imperative that participation-related algorithms be critically examined – a step that cannot be automated.

What is crucial here is that the pursued objectives be explicitly identified. Anyone who fails to do so will be making implicit assumptions and running the risk of limiting participation for others. Assumptions and expectations must be clarified if the impact on participation is to be positive. The classification criteria presented in this paper can help move the debate forward on social goals and policy responses. This is particularly true when different organizations or individuals evaluate specific criteria differently. After all, it is disagreement that reveals the full spectrum of options available to us for taking action.

8 Bibliography

Agrawal, Ajay, John Horton, Nicola Lacetera and Elizabeth Lyons (2015). "Digitization and the contract Labor Market: A Research Agenda". *Economic Analysis of Digital Economy*. Pub.: Avi Goldfarb, Shane M. Greenstein and Catherine E. Tucker. Chicago IL: University of Chicago Press. 219–250. http://www.nber.org/chapters/c12988.pdf (downloaded May 8, 2017).

Aneesh, Aneesh (2009). "Global Labor: Algocratic Modes of Organization". *Sociological Theory* (27) 4. 347–370. https://doi.org/10.1111/j.1467-9558.2009.01352.x(downloaded May 8, 2017).

Angwin, Julia, Lauren Kirchner, Jeff Larson and Surya Mattu (2016). "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks". May 23, 2015. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (downloaded December 11, 2016).

Arthurs, Harry (2011). "Labour Law After Labour". *The Idea of Labour Law*. Pub.: Guy Davidov and Brian Langille. Oxford, New York: Oxford University Press. 13–29.

Barry-Jester, Anna M., Ben Casselman and Donna Goldstein (2015). "The New Science of Sentencing". *The Marshall Project.* 4.4.2015. https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing#.xXEp6R5rD (downloaded April 24, 2017).

Bauman, Zygmunt, Didier Bigo, Paulo Esteves, Elspeth Guild, Vivienne Jabri, David Lyon and R. B. J. Walker (2014). "After Snowden: Rethinking the Impact of Surveillance". *International Political Sociology* (8) 2. 121–144. https://doi.org/10.1111/ips.12048 (downloaded May 8, 2017).

Baumgartner, Fabian (2015). “Deutliche Zunahme im Jahr 2015: Wieder mehr Einbrüche in der Stadt Zürich” (Sharp increase in 2015: More burglaries in Zurich again). *Neue Zürcher Zeitung* October 27, 2015. http://www.nzz.ch/zuerich/stadt-zuerich/wieder-mehr-einbrueche-in-der-stadt-zuerich-1.18636278 (downloaded April 24, 2017).

Beirat Integration (German Advisory Council on Integration) (2013). “‚Soziale Teilhabe‘ Handlungsempfehlungen des Beirats der Integrationsbeauftragten”. (‘Social Participation’ Recommended action from the Advisory Council on Integration) Die Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration (Federal Government Commissioner for Migration, Refugees and Integration). http://www.bagiv.de/pdf/soziale-teilhabe-empfehlungen-beirat.pdf (downloaded May 8, 2017).

Bertelsmann Stiftung (2011). *Soziale Gerechtigkeit in der OECD – Wo steht Deutschland? (Social justice in the OECD – Where does Germany stand?) Sustainable Governance Indicators 2011*. Gütersloh. http://news.sgi-network.org/uploads/tx_amsgistudies/SGI11_Social_Justice_DE.pdf (downloaded May 8, 2017).

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Carlow Marlow, Jaime E. Settle and James H. Fowler (2012). “A 61-million-person experiment in social influence and political mobilization”. *Nature* (489) September 13, 2012. 295–298. https://doi.org/10.1038/nature11421(downloaded May 8, 2017).

Brey, Philip (2005). “Artifacts as social agents”. *Inside the Politics of Technology: Agency and Normativity in the Co-Production of Technology and Society*. Pub.: Hans Harbers. Amsterdam: Amsterdam University Press. 61–84.

Brown, Ian, and Josh Cowls (2015). *Check the Web – Assessing the Ethics and Politics of Policing the Internet.* Oxford: Oxford Internet Institute and VOX-Pol Network of Excellence. http://voxpol.eu/wp-content/uploads/2015/11/VOX-Pol_Ethics_Politics_PUBLISHED.pdf (downloaded May 8, 2017).

Brühl, Jannis (2014). “Ermitteln mit ‚Predictive Policing‘-Algorithmen: Polizei-Software soll Verbrechen voraussagen“. (Investigating with predictive policing algorithms: police software to predict crime) *Süddeutsche Zeitung* September 12, 2014. http://www.sueddeutsche.de/digital/ermitteln-mit-predictive-policing-algorithmen-polizei-software-soll-die-zukunft-voraussagen-1.2121942 (downloaded April 24, 2017).

Bucher, Taina (2016). “The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms”. *Information, Communication & Society* (20) 1. 1–15. https://doi.org/10.1080/1369118X.2016.1154086 (downloaded May 8, 2017).

Bundesgesetzblatt (German Federal Legal Gazette) (1966). Internationaler Pakt über wirtschaftliche, soziale und kulturelle Rechte (International Covenant on Economic, Social and Cultural Rights). Bundesgesetzblatt (BGBl) 1976 II, 428. December 19, 1966. http://www.institut-fuer-menschenrechte.de/fileadmin/user_upload/PDF-Dateien/Pakte_Konventionen/ICESCR/icescr_de.pdf (downloaded May 8, 2017).

Choudary, Sangeel Paul, Marshall W. van Alstyne and Geoffrey G. Parker (2016). *Platform Revolution: How Networked Markets Are Transforming the Economy – And How to Make Them Work for You* (1st edition). New York NY: W. W. Norton & Company.

Cormen, Thomas H., Clifford Stein, Charles E. Leiserson and Robert L. Rivest (et al) (2001). *Introduction to Algorithms*. Cambridge MA: MIT Press.

Datta, Amit, Michael Carl Tschantz and Anupam Datta (2015). “Automated Experiments on Ad Privacy Settings”. *Proceedings on Privacy Enhancing Technologies* 1. 92–112. https://doi.org/10.1515/popets-2015-0007 (downloaded May 8, 2017).

Davey, Monica (2016). “Chicago Police Try to Predict Who May Shoot or Be Shot”. *The New York Times* May 23, 2016. http://www.nytimes.com/2016/05/24/us/armed-with-data-chicago-police-try-to-predict-who-may-shoot-or-be-shot.html (downloaded April 24, 2017).

De Stefano, Valerio (2015). “The Rise of the ‘Just-in-Time Workforce’: On-Demand Work, Crowd Work and Labour Protection in the ‘Gig-Economy’”. *Comparative Labor Law & Policy Journal* (37) 3. 471–504.

Denardis, Laura (2012). “Hidden Levers of Internet Control”. *Information, Communication & Society* (September). 37–41.

Diakopoulos, Nicholas (2015). “Algorithmic Accountability”. *Digital Journalism* (3) 3. 398–415. https://doi.org/10.1080/21670811.2014.976411 (downloaded May 8, 2017).

Diakopoulos, Nicholas (2016). “Accountability in Algorithmic Decision Making”. *Communication of the ACM* (59) 2. 56–62. https://doi.org/10.1145/2844110 (downloaded May 8, 2017).

Epstein, Robert, and Ronald E. Robertson (2015). “The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections”. *Proceedings of the National Academy of Sciences* (112) 33. E4512–E4521. https://doi.org/10.1073/pnas.1419828112 (downloaded May 8, 2017).

Eubanks, Virginia (2013). “Caseworkers vs. Computers”. *Poptech* 11.12.2013. https://virginiaeubanks.wordpress.com/2013/12/11/caseworkers-vs-computers/ (downloaded May 8, 2017).

Europol (2015). “Europol’s Internet Referral Unit to Combat Terrorist and Violent Extremist Propaganda” (Press Release). July 1, 2015. https://www.europol.europa.eu/newsroom/news/europol%E2%80%99s-internet-referral-unit-to-combat-terrorist-and-violent-extremist-propaganda (downloaded May 8, 2017).

Ferguson, Andrew Guthrie (2012). “Predictive Policing and Reasonable Suspicion”. *Emory Law Journal* 259. https://papers.ssrn.com/abstract=2050001 (downloaded April 24, 2017).

Freedland, Mark R., and Nicola Kountouris (2011). “The Legal Characterization of Personal Work Relations and the Idea of Labour Law”. *The Idea of Labour Law.* Pub.: Guy Davidov and Brian Langille. Oxford, New York: Oxford University Press. 190–208.

German Federal Ministry of Labour and Social Affairs (2015). *Work 4.0 (Green Paper)*. Berlin. http://www.bmas.de/EN/Services/Publications/arbeiten-4-0-greenpaper-work-4-0.html;jsessionid=88D6D2043941997AE96A4D881CA66A01 (downloaded May 8, 2017).

Gillespie, Tarleton (2014). “The Relevance of Algorithms”. *Essays on Communication, Materiality, and Society.* Pub.: Tarleton Gillespie, Pablo Boczkowski and Kirsten Foot. Cambridge MA: MIT Press. 167–194.

Gluba, Alexander (2014). *Predictive Policing – eine Bestandsaufnahme*. (*Predictive Policing – Taking stock)* LKA Niedersachsen: Hannover. https://netzpolitik.org/wp-upload/LKA_NRW_Predictive_Policing.pdf (downloaded April 24, 2017).

Hannah-Moffat, Kelly, Paula Maurutto and Sarah Turnbull (2009). “Negotiated Risk: Actuarial Illusions and Discretion in Probation”. *Canadian Journal of Law and Society* (24) 3. 391–409. https://doi.org/10.1017/S0829320100010097 (downloaded April 24, 2017).

Hawthorne, Michael (2015). “Could Chicago prevent childhood lead poisoning before it happens?”. *Chicago Tribune* July 16, 2015. http://www.chicagotribune.com/news/ct-lead-poisoning-solutions-20150707-story.html (downloaded January 3, 2017).

Helberger, Natali, and Damian Trilling (2016). “Facebook is a news editor: the real issues to be concerned about”. 26.5.2016. http://blogs.lse.ac.uk/mediapolicyproject/2016/05/26/facebook-is-a-news-editor-the-real-issues-to-be-concerned-about/ (downloaded May 8, 2017).

Hensel, Isabell, Jochen Koch, Eva Kocher and Anna Schwarz (2016). “Crowdworking als Phänomen der Koordination digitaler Erwerbsarbeit – Eine inter-disziplinäre Perspektive”. (Crowdworking as a phenomenon of coordinating digital employment – an interdisciplinary perspective) *Industrielle Beziehungen* (23) 2. 162–186.

Hill, Kashmir (2012). “How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did”. February 16, 2016. http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/ (downloaded May 8, 2017).

Irani, Lilly (2015). “Difference and Dependence among Digital Workers: The Case of Amazon Mechanical Turk”. *South Atlantic Quarterly* (114) 1. 225–234. https://doi.org/10.1215/00382876-2831665 (downloaded May 8, 2017).

Kahneman, Daniel (2012). *Thinking, Fast and Slow*. London, inter alia: Penguin.

Kocher, Eva, and Isabell Hensel (2016). “Herausforderungen des Arbeitsrechts durch digitale Plattformen – ein neuer Koordinationsmodus von Erwerbsarbeit”. (Challenges of labor law via digital platforms – a new coordination mode of employment) *Neue Zeitschrift für Arbeitsrecht NZA* (33) 16. 984–990.

Lee, Min Kyung, Dabiel Kusbit, Evan Metsky and Laura Dabbish (2015). “Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers”. ACM Press. 1603–1612). https://doi.org/10.1145/2702123.2702548 (downloaded May 8, 2017).

Leimeister, Jan Marco, David Durward and Shkodran Zogaj (2016). *Crowd Worker in Deutschland eine empirische Studie zum Arbeitsumfeld auf externen Crowdsourcing-Plattformen*. (Crowdworking in Germany, an empirical study on the working environment of external crowdsourcing platforms) Düsseldorf: Hans-Böckler-Stiftung. http://hdl.handle.net/10419/146407 (downloaded May 8, 2017).

Lingemann, Stefan, and Jörn Christoph Otte (2015). “Arbeitsrechtliche Fragen der ‚economy on demand‘”. (Labour law issues of the economy on demand) *Neue Zeitschrift für Arbeitsrecht NZA*. 1042–1047.

Lischka, Konrad, and Anita Klingel (2017). *Wenn Maschinen Menschen bewerten (*When Machines Judge People). Gütersloh: Bertelsmann Stiftung. https://doi.org/10.11586/2017025 (downloaded May 29, 2017)

Lyon, David (2003). *Surveillance as social sorting: privacy, risk, and digital discrimination*. New York NY: Routledge. http://public.eblib.com/choice/publicfullrecord.aspx?p=240591 (downloaded May 8, 2017).

Mastrobuoni, Giovanni (2015). “Crime is terribly revealing: Information technology and police productivity”. Unpublished Paper. http://cep.lse.ac.uk/conference_papers/01_10_2015/mastrobuoni.pdf (downloaded April 24, 2017).

Meehl, Paul E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis MN: University of Minnesota Press. https://doi.org/10.1037/11281-000 (downloaded May 8, 2017).

Meyer, Thomas (2016). “Gleichheit – warum, von was und wie viel?” (Equality – why, what and how much?) *Neue Gesellschaft/Frankfurter Hefte* 11. 42–46.

Mohler, George O., Martin B. Short, Sean Malinowski, Mark Johnson, George E. Tita, Andrea L. Bertozzi and P. Jeff Brantingham (2015). "Randomized controlled field trials of predictive policing". *Journal of the American Statistical Association* (110) 512. 1399–1411. https://doi.org/http://dx.doi.org/10.1080/01621459.2015.1077710 (downloaded April 24, 2017).

Niklas, Jedrzej (2017). "Can an algorithm hurt? Polish experiences with profiling of the unemployed". March 2017. https://cihr.eu/can-an-algorithm-hurt/ (downloaded May 8, 2017).

Niklas, Jedrzaj, Karolina Sztandar-Sztanderska and Katarzyna Szymielewicz (2015). *Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making*. Warsaw: Panoptykon Foundation. http://panoptykon.org/biblio/profiling-unemployed-poland-social-and-political-implications-algorithmic-decison-making (downloaded May 8, 2017).

Northpointe (2015). "Practitioners Guide to COMPAS Core". http://images.google.de/imgres (downloaded April 24, 2017).

O'Callaghan, Derek, Derek Green, Maura Conway, Joe Carthy and Pádraig Cunningham (2016). "Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems". http://doras.dcu.ie/20280/1/Down_the_Rabbit_Hole_DORAS_Version.pdf (downloaded May 8, 2017).

Pariser, Eli (2011). *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. London. Penguin Publishing Group.

Pasquale, Frank A. (2015). *Privacy, Autonomy, and Internet Platforms* (SSRN Scholarly Paper No. ID 2737413). Rochester NY: Social Science Research Network. http://papers.ssrn.com/abstract=2737413 (downloaded May 8, 2017).

Pilpul (2016). "Wo Predictive Policing eingesetzt wird". (Where predictive policing is used) December 2016. https://blog.pilpul.me/wo-predictive-policing-eingesetzt-wird/ (downloaded April 24, 2017).

Potash, Eric, Joe Brew, Alexander Loewi, Subhanrata Majumdar, Andrew Reece, Joe Walsh, Eric Rozier, Emile Jorgenson, Read Mansour and Rayid Ghani (2015). "Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning". *KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2039–2047. https://doi.org/10.1145/2783258.2788629 (downloaded April 24, 2017).

Rammert, Werner (2006). "Technik, Handeln und Sozialstruktur: Eine Einführung in die Soziologie der Technik". (Technology, Action and Social Structure: introduction to sociology of technology) *Technical University Technology Studies, Working Papers*, *TUTS-WP-3-2006*. Berlin.

Ruddick, Graham (2016a). “Admiral to price car insurance based on Facebook posts”. November 2, 2016. https://www.theguardian.com/technology/2016/nov/02/admiral-to-price-car-insurance-based-on-facebook-posts (downloaded May 8, 2017).

Ruddick, Graham (2016b). “Facebook forces Admiral to pull plan to price car insurance based on posts”. *The Guardian* November 2, 2016. https://www.theguardian.com/money/2016/nov/02/facebook-admiral-car-insurance-privacy-data (downloaded May 8, 2017).

Schmidt, Florian A. (2016). *Arbeitsmärkte in der Plattform-ökonomie – Zur Funktionsweise und den Herausforderungen von Crowdwork und Gigwork*. (Labor markets in the platform economy – functions and challenges of crowdwork and gigwork) Berlin: Friedrich-Ebert-Stiftung. http://library.fes.de/pdf-files/wiso/12826.pdf (downloaded May 8, 2017).

Schneider, Jan, Ruta Yemane and Martin Weinmann (2014). *Diskriminierung am Ausbildungsmarkt Ausmaß, Ursachen und Handlungsperspektiven*. (Discrimination on the training market – extent, causes and prospects for action) Berlin: Sachverständigenrat deutscher Stiftungen für Integration und Migration. https://www.svr-migration.de/wp-content/uploads/2014/11/SVR-FB_Diskriminierung-am-Ausbildungsmarkt.pdf (downloaded May 8, 2017).

Spiekermann, Sarah (2015). *Ethical IT Innovation: A Value-Based System Design Approach*. Toronto: Apple Academic Press. https://www.crcpress.com/Ethical-IT-Innovation-A-Value-Based-System-Design-Approach/Spiekermann/p/book/9781482226355 (downloaded October 3, 2016).

Stalder, Felix (2017). “Algorithmen, die wir brauchen”. (Algorithms we need) January 6, 2017. https://netzpolitik.org/2017/algorithmen-die-wir-brauchen/ (downloaded May 8, 2017).

Tene, Omer, and Jules Polonetsky (2011). *To Track or “Do Not Track”: Advancing Transparency and Individual Control in Online Behavioral Advertising* (SSRN Scholarly Paper No. ID 1920505). Rochester NY: Social Science Research Network. https://papers.ssrn.com/abstract=1920505 (downloaded May 8, 2017).

Tufekci, Zeynep (2014). “Engineering the public: Big data, surveillance and computational politics”. *First Monday* (19) 7. http://firstmonday.org/ojs/index.php/fm/article/view/4901 (downloaded May 8, 2017).

Tufekci, Zeynep, Jillian C. York, Ben Wagner and Frederike Kaltheuner (2015). “The Ethics of Algorithms: from radical content to self-driving cars”. Berlin: European University Viadrina. 1–18. https://cihr.eu/publication-the-ethics-of-algorithms/ (downloaded May 8, 2017).

van Haastert, Hugo (2016). “Government as a Platform: Public Values in the Age of Big Data” (OII IPP 2016 Conference). Oxford Internet Institute.

Wagner, Ben (2016). “Algorithmic regulation and the global default: Shifting norms in Internet technology”. *Nordic Journal of Applied Ethics* March 22, 2016. https://www.ntnu.no/ojs/index.php/etikk_i_praksis/article/view/1961 (downloaded May 8, 2017).

Weber, Lauren, and Elizabeth Dwoskin (2014). “Are Workplace Personality Tests Fair?”. *Wall Street Journal* September 30, 2014. http://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257 (downloaded April 24, 2017).

Winner, Langdon (1978). *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought* (1st US-1st Printing edition). Cambridge MA inter alia: MIT Press.

Winner, Langdon (1980). “Do Artifacts Have Politics?”. *Daedalus* (109) 1. 121–136.

York, Jillian C. (2010). *Policing Content in the Quasi-Public Sphere* (Open Net Initiative Bulletin). Boston MA: Berkman Center, Harvard University. https://opennet.net/policing-content-quasi-public-sphere (downloaded May 8, 2017).

Zhang, Pei, Sophie Stalla-Bourdillon and Lester Gilbert (2016). “A content-linking-context model for ‘notice-and-takedown’ procedures”. Conference on WebSci ’16 Proceedings of the 8th ACM Conference on Web Science. http://eprints.soton.ac.uk/397678/ (downloaded May 8, 2017).

Zittrain, Jonathan (2014). “Engineering an Election”. *Harvard Law Review Forum* (127) 8. 335–341.

Zweig, Katharina Anna (2016). “2. Arbeitspapier: Überprüfbarkeit von Algorithmen”. (2nd Working Paper: Verifiability of algorithms) July 7, 2016. http://algorithmwatch.org/zweites-arbeitspapier-ueberpruefbarkeit-algorithmen/ (downloaded April 24, 2017).

Zweig, Katharina Anna (2017, in press). “Wo maschinelle Prognosen irren können” (Where machine forecasts can make mistakes). Bertelsmann Stiftung.

9 Executive Summary

Algorithms are increasingly making decisions with us, for us and about us – thereby giving rise to new questions about participation. This paper makes a number of initial suggestions for structuring and classifying the participation-related issues stemming from algorithmic decision-making (ADM). In today's digitized knowledge society, shaping technology is becoming a fundamental form of power. And if knowledge is power, then algorithms are becoming today's instruments of power. To what degree is it acceptable and desirable for algorithms to have an impact on the lives of individuals and on society as a whole? And which aspects of ADM must we consider more closely if we want to benefit from the potentials and minimize the risks to the greatest degree possible?

This paper offers a general overview of ADM processes by explaining key concepts, delimiting the relevant area of analysis of ADM processes and presenting typical scenarios and functions relating to the use of these processes. With this as a basis, a system for classifying ADM processes is then proposed. The objective is to make it easier to evaluate and compare the potential impact of ADM processes on participation by using fewer criteria. This, in turn, will facilitate a prioritization and preparation of more in-depth research on the subject.

An evaluation of the impact on participation – a structure for which we are proposing here – is the point of departure for further steps, be they detailed analyses or regulatory measures. In those cases where ADM is expected to affect participation significantly, then a thoroughgoing effort must be made to ensure that positive impacts are exploited as much as possible, while negative impacts, such as ADM-driven discrimination, are prevented. When algorithms greatly influence participation, it is often necessary to counteract that influence by ensuring diversity, fair competition and principles of due process. The more prevalent the participation-relevant ADM processes are, the more comprehensive the potential responses must be.

9.1 Methodological approach

The potential impact on participation cannot be classified using binary decision-making criteria, since that would overlook the many socially relevant logics embedded in ADM processes.

Numerous criteria can affect the relevance to participation. The selection and weighting of these criteria, moreover, depend on how participation is defined. The point here is not to judge ADM processes as good or bad, but to evaluate their potential impact on participation in relative terms, regardless of the direction that might take.

The criteria can be divided into three groups: *actors, social embeddedness* and *consequences.* In other words, the groups do not reflect technological factors, but are an attempt to document political and social considerations.

The *actors*-related criteria examine the actual economic and political power of those supplying and/or operating the decision-making processes. The criteria used for *social embeddedness* reflect mutually reinforcing social interdependencies, both intended and unintended. Finally, the potential *consequences* – along with those that have already become apparent – are analyzed in terms of basic political and social rights. Existing legal norms are used as the guidelines for evaluating the impact on participation. The starting point is the German Antidiscrimination Act (§1 and 2 AGG) and its area of application. Discrimination is thus prohibited in the following areas in particular: recruitment, hiring, working conditions, membership in trade unions, and occupational training. Fair access to public goods such as education, social security, health care and housing must also be ensured.

We suggest assuming there is an impact on participation only after a certain threshold has been reached. Using the criteria sketched out below, the relative consequences for participation can be summarized and compared. This makes it possible to rank algorithms more simply and manageably in order to prioritize further steps, where needed.

Each of the criteria listed below can increase the potential impact on participation (PIP, here PIP = +1). If a criterion does not apply or cannot be evaluated as part of the quick-check process, no point is given for the potential impact ($PIP_{neutral}$ = 0). Half points can also be given for each criterion (PIP = +0.5), and the maximum is two points ($PIP_{max} = +2$). The values thus range from 0 to 2 points per criterion. The higher the overall value, the greater the process's potential impact on participation.

Test questions are used to evaluate the criteria, and answering them is not always easy. The various applications are highly diverse and, on occasion, obscure, which is precisely why the system of abstract classifications is beneficial: It reduces the complexity of a highly complex network of issues. What might limit or promote participation? To answer that question, a person or organization uses the evaluative structure to analyze one or more relevant processes. After applying the criteria to a specific case and considering the results, they can assess which aspects of the decision-making process are most prominent; they can also compare different processes with each other.

9.2 Actors

A1 Competition: Who is supplying/operating the algorithm? How much political and economic power do they have? Is there real competition between multiple suppliers/operators (oligopolies, monopolies, cartels, dominant market position, sovereign power)?

<u>Example – Border controls:</u> Even if private companies produce or supply software for granting visas, the software's use is an expression of sovereign power. No one can escape the state's power to grant visas, and to that extent a private supplier working on behalf of the state must also be seen as a monopoly. The actual organizational type or legal status of the supplier/operator is therefore less relevant for the impact on participation.

A2 Dependency: Is it a problem when the process or product is no longer available or access to it is denied? Can the product be readily replaced? Do substitute processes/products exist? What would it cost to switch to another process? Which organizations or groups are structurally dependent on the process? Do the system's construction and functioning result in an informal dependency? Which technical or social lock-in effects exist?

<u>Example – Crowdwork:</u> Robust competition exists in the market for crowdworking platforms. Multiple platform operators compete both for contract suppliers and crowdworkers. Participants could nonetheless become dependent if job histories and reputations become linked to the platform's specific algorithmic assessment processes. A similar lock-in effect can also arise with other assessment systems or with social media.

9.3 Social embeddedness

SE1 Self-determination: To what extent does the process serve (only) as a preliminary step for making a decision? Does the user retain any right to influence the process? Does the system decide (de facto or officially) on its own? How much freedom to change or manage an algorithmic decision does the user have? Do time pressure and the process's practical implementation affect the user's autonomy?

<u>Example – Content control:</u> The way software is designed for monitoring potentially illegal content can have a major impact on user autonomy. Time pressure (e.g. fast pace) and technical limitations can, de facto, greatly reduce the possibility of human input, even if a human will officially be the one to make the final decision.

SE2 Adaptation: How do people adapt to the algorithmic process? What impact does the decision-making system have on a practical level? Which interactions between humans and computers can potentially change the outcome? What is the process's actual power in social terms? Which (social) significance is ascribed to the process? Does it define and/or lead the discussion? How inherently influential are the process outcomes?

Example – Trending topics: The “trending topics” on Twitter are not infrequently seen as a social “reality.” Although it is not generally known how the ranking is created, it still affects the public discourse and different actors adjust their behavior to reflect the logic used by the “trending topics” algorithm.

9.4 Consequences

C1 Reach: How many people are affected by the decision-making process (e.g. user population)? Is the process’s reach known and/or can it be limited? How great is it?

Example – Candidate selection: If only one company uses a software application for choosing job candidates, then the application’s reach is relatively small. If, however, it becomes a core decision-making tool for many companies, its impact on participation increases accordingly, for example if the logic used by one system predominates when employees are hired in a particular region or industry.

C2 System change: Does the decision-making process undermine principles of solidarity? Would individualization change the system? Could the process result in an unintended transformation of the impacted (social) environment?

Example – Health insurance: If the costs for public health insurance are partially decided by algorithms and are subsequently individualized using a new logic, then a system which has been based on social solidarity will no longer be to some degree. Instead of having everyone pay only according to their ability to do so, at least part of each person’s contributions will be determined by individual factors.

C3 Discrimination: To what degree could people be disadvantaged by the decision-making process? Do the results exhibit a pattern of discrimination or might such a pattern be expected, e.g. based on race, ethnic or cultural background, gender, sexual identity, physical disability, age, religion or world view?

Example – Automated job listings: An analysis of Google’s employment advertisements revealed that users classified as men were shown more highly paid positions than were women. In this case, the process has an impact on job market participation.

9.5 Summary: determining the potential impact on participation

The degree of the potential impact on participation (PIP) of a random new ADM process can thus be summarized as follows:

$$PIP = VA1 * VA2 * (A1 + A2 + SE1 + SE2 + C1 + C2 + C3)$$

For this paper, ADM processes are fundamentally relevant when a digital decision-making system is used and when people or characteristics attributed to people are being evaluated by that system. These two filtering criteria – “digital decision-making systems” (VA1) and “evaluation of people” (VA2) – can have either a value of 0 (not true) or 1 (true). The degree of the potential impact on participation (PIP) is the product of both preliminary criteria and the sum of the impact criteria. The PIP value is the sum of the test criteria (A1 + A2 + SE1 + SE2 + C1 + C2 +C3). Expressing the extent to which an ADM process can affect participation, the overall result (PIP) can range from 0 to 14. This testing structure makes it possible to filter out processes which do affect participation but whose potential impact is limited. This could be the case when PIP is less than 3, for example, although this threshold value can be adjusted to reflect different definitions of participation.

The above criteria can be used to develop a fast and relatively simple testing procedure, allowing companies and evaluators to check their ADM processes using specific criteria in just two or three pages.[3] Individuals and organizations can employ such criteria to evaluate ADM processes for their relevance to participation. To that extent, the process is a simple and quick introduction to understanding ADM processes. At the same time, it is not a substitute for a more comprehensive evaluation.

10 About the authors

Kilian Vieth

Kilian Vieth manages the "European Surveillance Reform Initiative" project at the Stiftung Neue Verantwortung in Berlin, where he is working on reform proposals for a more democratic and efficient security and surveillance policy in Europe. His research focuses on digital human rights, critical security studies, as well as political and social issues relating to algorithmic decision-making. Priviously, Kilian worked on different research projects at the Centre for Internet & Human Rights at the European University Viadrina in Frankfurt (Oder). He also worked for two years in the field of political campaigning for a communications consultancy firm in Berlin. Kilian Vieth studied Political Science and European Affairs at the Otto Suhr Institute at the Freie Universität Berlin and at Sciences Po Paris, as well as Political Sciences and Public Management the Zeppelin University in Friedrichshafen.

Dr. Ben Wagner

Dr. Ben Wagner is an Assistant Professor and Director of the Privacy & Sustainable Computing Lab at Vienna University of Economics and Business and a Senior Researcher of the Centre of Internet & Human Rights (CIHR). His research focuses on communications technology at the intersection between rights, ethics and governance. Ben holds a PhD in Political and Social Sciences from European University Institute in Florence. He was previously worked at the German Institute for International and Security Affairs, the University of Pennsylvania, Human Rights Watch and the European Council on Foreign Relations. His research has been published in Telecommunications Policy, JITP and the International Journal of Communications.

3 See the "Quick Check" developed by the Danish Institute for Human Rights:

Address | Contact

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh
Telephone +49 5241 81-81216

Konrad Lischka
Digitalization Taskforce
Telephone +49 5241 81-81216
konrad.lischka@bertelsmann-stiftung.de

www.bertelsmann-stiftung.de

| BertelsmannStiftung