Human Control and Discretion in AI-driven Decision-making in Government

Mitrou, Lilian; Janssen, Marijn; Loukis, Euripidis

DOI
10.1145/3494193.3494195

Publication date
2021

Document Version
Final published version

Published in
Proceedings of the 14th International Conference on Theory and Practice of Electronic Governance, ICEGOV 2021

Human Control and Discretion in AI-driven Decision-making in Government

Lilian Mitrou
University of the Aegean, Samos, Greece
L.Mitrou@aegean.gr

Marijn Janssen
Delft University of Technology, Delft, The Netherlands
M.F.W.H.A.Janssen@tudelft.nl

Euripidis Loukis
University of the Aegean, Samos, Greece
eloukis@aegean.gr

ABSTRACT

Traditionally public decision-makers have been given discretion in many of the decisions they have to make in how to comply with legislation and policies. In this way, the context and specific circumstances can be taken into account when making decisions. This enables more acceptable solutions, but at the same time, discretion might result in treating individuals differently. With the advance of AI-based decisions, the role of the decision-makers is changing. The automation might result in fully automated decisions, humans-in-the-loop or AI might only be used as recommender systems in which humans have the discretion to deviate from the suggested decision. The predictability of and the accountability of the decisions might vary in these circumstances, although humans always remain accountable. Hence, there is a need for human-control and the decision-makers should be given sufficient authority to control the system and deal with undesired outcomes. In this direction this paper analyzes the degree of discretion and human control needed in AI-driven decision-making in government. Our analysis is based on the legal requirements set/posed to the administration, by the extensive legal frameworks that have been created for its operation, concerning the rule of law, the fairness – non-discrimination, the justifiability and accountability, and the certainty/ predictability.

KEYWORDS

AI, discretion, decision-making, accountability

ACM Reference Format:
Lilian Mitrou, Marijn Janssen, and Euripidis Loukis. 2021. Human Control and Discretion in AI-driven Decision-making in Government. In *14th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2021), October 06–08, 2021, Athens, Greece.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3494193.3494195

1 INTRODUCTION

Decision-making is at the core of administration. Decision-making is based on the available data, the regulations and rules, and public servants' discretion to understand the situation at hand and interpret the legislation to make the most appropriate decisions. Decision-making is changing by the vast amount of data and the ability of Artificial Intelligence (AI) to process these data (Bullock, 2019; Bullock et al., 2020). Even if the use of technologies for decision-making in government is far from new, recently public bodies invest increasingly on AI-based computational algorithms, including machine learning, to automate human decision-making (Cobbe, 2019). The use of algorithms changes the materiality of governance in which people, data, algorithms and systems are involved (Janssen & Kuk, 2016). The more use of data influences not only the use of algorithms, but also the systems and the discretion of public servants in making decisions.

The European Commission in the Proposal for an Artificial Intelligence Act (presented on 21/04/2021) states that the definition, to be adopted for regulatory purposes, should be as technology-neutral and future proof as possible, taking into account the fast technological and market developments related to AI. So, the Commission suggests as AI system to be defined "'software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" (Article 3 (a)). In the normative view of European Commission, AI systems can be designed to operate with varying levels of autonomy and be used on a stand-alone basis or as a component of a product, irrespective of whether the system is physically integrated into the product (embedded) or serve the functionality of the product without being integrated therein (non-embedded).

While the first generation of AI was based on logic and rules explicitly pre-defined by humans ('Symbolic AI'), the second generation of it was based on logic and rules extracted automatically by computers through advanced processing of past historic data ('Statistical AI'), from which models or sets of rules are constructed, that enable on one hand deeper insights (e.g. concerning associations among important variables) and on the other hand making predictions of important variables (Duan et al., 2019; OECD, 2019). In this second generation of AI the most representative and widely used techniques are definitely the Machine Learning (ML) ones. They enable exploiting historic past data we possess for a number of units (e.g. individuals, firms, etc.) concerning the value of an important dependent variable (usually an outcome one) and also a set of independent variables (that might be possible causes of this outcome or factors affecting it), by processing them through various advanced algorithms, and finally extracting knowledge from them, usually having the form of a model or a set of rules, concerning the relationships between the independent variables and the dependent one (this is usually referred to as 'training'). This knowledge can be used then for gaining deeper understanding and insight about these relationships, as well as for predicting the value of the dependent variable for new units (for which we have the values of the

ICEGOV 2021, October 06–08, 2021, Athens, Greece

ACM ISBN 978-1-4503-9011-8/21/10…$15.00
https://doi.org/10.1145/3494193.3494195

independent variables), which can be quite useful for supporting relevant decisions or optimizing actions.

Through the exploitation of digitalization of information and advances in ICT and infrastructure across the various e-Government phases, public authorities and bodies become more data-driven and use digital transformation to extract value from large datasets in order to increase their capacity for problem-solving (Barcevičius et al., 2019). Current public sector AI applications in Europe include tax or welfare fraud detection (Systeem Risico Indicatie/ SyRi in the Netherlands), use of algorithms in the area of social welfare systems, such as house benefits or automated unemployment systems (AuroraAI in Finland, automated public services for social assistance in Trelleborg/Sweden, AMAS in Austria, Robot Tengai in Sweden, Powiatowe Urzędy Pracy-Public Employment Services in Poland), policing (VeriPol in Spain) or the use of pre-emptive tools for child protection (Kind en Gezin in Belgium, Cladsaxe model in Denmark), as well as use in healthcare for supporting diseases' diagnosis and treatment planning (Misuraca, G., and van Noordt, 2020, Kuziemski and. Misuraca 2020, Veale and Bras, 2019; DeSousa et al., 2019; Sun and Medaglia, 2019; Loukis et al., 2020; European Union Agency for Fundamental Rights, 2020, Digital Future Society, 2021). Some – mostly local – authorities are introducing predictive analytics and decision support systems for supporting intervention in geographic or individual cases (Vogl et al., 2019). What they have in common is that black-boxed algorithms are used for making decisions on a large scale. This changes the discretionary role of public servants.

In this direction this paper analyzes the degree of discretion and human control needed in AI-driven decision-making in government. Our analysis is based on the legal requirements set/posed to the administration, by the extensive legal frameworks that have been created for its operation, concerning the rule of law, the fairness – non-discrimination, the justifiability and accountability, and the certainty/predictability. So an important question is do AI applications empower public bodies to provide better services or be more efficient without challenging the normative structure underlying our understanding of law or its implementation (Zalnieriute et al. 2019), e.g. concerning the abovementioned requirements, and how can human control, intervention and discretion contribute to addressing possible problems.

A main issue is if/how could/should AI guide or challenge the way decisions are taken and respectively their output/ their outcome. Public administration is bound by and subject to the law and fundamental rights. It is not free to make choices, as they have to be enshrined in the mandatory regulatory framework and the legal responsibilities it is bound by/ carries. However, administration practices discretion both with regard to specifying policy goals and implementation and for individual decision-making processes. Treating each case as unique by taking into consideration the specific situation and facts belongs to the core features of administrative law. Furthermore, the lawful, fair and effective use of discretion relates quite directly to the quality of administration (Bullock, 2019).

Can this requirement, i.e., to decide on each case separately while preserving equity and transparency, be reconciled with machine learning applications and statistical regularities? Is AI likely to weaken discretion thus resulting into less equitable, biased or discriminating decisions or does it foster/ fostering a human-bias free and neutralized decision making? Can the appropriate level of human control, intervention and discretion in AI-driven decision-making help for counter-balancing such possible negative 'side-effects' of AI usage in government?

In this paper we analyze the role of discretion of public servants in AI-based decision systems from the above legal requirements' perspective. This paper is structured as follows. In the next section 2 we discuss the levels of discretion and the role of AI, while in section 3 we analyze the association between the rule of law and the use of AI for decision-making in government. Then in section 4 we discuss the transparency, justifiability and accountability challenges, and in section 5 we focus on discretion and human control in AI-driven decision-making in government. Finally, in section 6 conclusions are summarized.

2 THE LEVELS OF DISCRETION AND AI

The extent and level of discretion, defined as 'the latitude afforded individuals with delegated responsibilities to use their judgement when making a decision' (Bullock et al., 2020), its use and its outcome depend on the national legal order. Normally, administrative authorities are given discretionary powers and are controlled by the judiciary, if the limits of discretion are exceeded. In some European legal systems, discretion has to be explicitly granted to an authority, while in other countries the courts recognize discretion also where the application of norms requires the assessment of complex sets of facts or the prognosis of future developments (Nolte, 1994). Often provisions for discretion in decision making by public authorities is provided by law to respond to the need of decisional flexibility, "especially where it is difficult to foresee every scenario that could arise, conditions are changeable and/or a comprehensive set of statutory prescriptions would produce unfairness" (Varuhas, 2020), while part of European courts recognize discretion even in the application of indefinite (legal) concepts by an authority. In case of discretionary administrative acts, the reasons must also include the aspects on the basis of which the authority has exercised its discretion.

Any administrative decision is subject to the legality as well as the proportionality principles. The *principle of legality* states the requirement that public authorities decide or enact measures in conformity with the legal system's hierarchy of norms and the principle of judicial protection, while according to the *principle of proportionality*, an administrative decision has to satisfy the criteria of adequacy (for achieving the goal), necessity (less intrusive/restrictive measure) and proportionality stricto sensu (disproportionate or excessive with regard to the goal to be achieved).

Does the interpretation of the legal rules and the subsumption of facts to these rules presuppose the participation of humans or can be entrusted to AI to apply the law to the specific case and come to a decision that is compliant with the law? The answer depends on the degree of discretionary power conferred to the administrative agency and the use of algorithm-based decision process as a complement or substitute to human-made decisions.

With regard to the first element, it is crucial to consider the context of discretion and decision-making. According to quite recently adopted German administrative procedure law, the adoption of a fully automated administrative decision is subject to

two conditions: (1) the existence of a legal ground and (2) the absence of discretion and room for maneuver/ discretion neither with the interpretation of the statutory conditions nor as to the decisions to be taken if the statutory conditions are met (VwVfG—Verwaltungsverfahrensgesetz (Etscheid, 2019).

AI-supported systems are proposed to be used to automate decision-making processes (or parts/components thereof) that rely "on clear, fixed and finite criteria", a detailed legal regime that demands no executive discretion. Zalnieriute et al. (2019) cite as an example the expert system that may "picks up" the cases/individuals that meet the criteria set explicitly and exclusively by law for receiving a benefit. The type of task performed is also of high importance: tasks that are routinized and simple are more likely to be completed by a machine (Bullock, 2019). However, some scholars (Veale and Bras, 2019) argue that such decision support systems may hide discretionary activities and power, revealing possible lacks with regard to oversight of the administrative action.

However, it is necessary to take into account that there are various types of AI algorithms. Some are self-learning from previous historic data, and the decision can change over time, as more data are accumulated and from them more learning/knowledge is extracted, whereas others have pre-defined rules, so they result in the same outcomes when the input data is the same. The application of the latter may produce consistent outputs, but at the same time leaves the particularities of each case out of consideration, and also does not exploit valuable previous knowledge and experience that historic data contain (and can be extracted with appropriate algorithms).

Automated decisions become more difficult if individual cases' details or still- unstructured information (that cannot be part of the historic data from which rules are extracted) has to be taken into account. Although AI is increasing in capacity, its use seems to be mostly rejectable/questionable when it comes to accomplishing discretionary tasks, or where there is a need for structuration of information or assessment (Etscheid, 2019). Even if AI-supported systems may operate adequately with explicit law-authored rules and do more of the tasks that are in the domain of human actors, the situation becomes less clear when machine learning establishes processes by which a system will learn patterns and correlations to improve performance to achieve specific goal (Zalnieriute et al., 2019). These systems are self-learning and automate the construction of criteria and rules (from historic data) to reach a decision. The determination of applicable norms is also disputable, as it necessitates a full understanding of the facts and complexity of the case at stake.

As Hagendorff and Wezell (2020) note, such systems can only operate on the basis of the given information, using existing circumstances to learn, and extrapolate the appropriated patterns along predefined terms and lines. However, even law and moreover its interpretation by the Courts is not stable and can change over time, reflecting changes in societal values and needs. As reported by the European Fundamental Rights Agency, an AI tool used on a pilot basis by a public body to process applications and subsequently support its staff in making decisions on housing benefits, has failed and the project has been terminated, as it wasn't possible to use AI in practice and estimate income in advance because of the frequent changes in the legislation (European Union Agency for Fundamental Rights, 2020).

Additional complexity may occur where an automated system has to take not only a data-driven but a values-driven decision. Rule-based systems may be embedded into an AI decision-making application but this is hardly the case when the public administration has to strike a balance between competing rights and interests, not to mention the difficulty to ensure that a principles-driven system, "programmed with slave-morality" (Wirtz et. all, 2019) does not result in moral or legal rigidity with regard to individual circumstances. Applying uniform rules and criteria across all decisions, without variation or particular consideration of the specific situation, may constitute a misuse of discretionary power that leads to an unlawful decision (Cobbe, 2019).

Therefore, in cases of complex government decisions, in which i) there are a lot of input data items to be taken into account, with some of them being unstructured, or even not known in advance; ii) in case of rules extracted from historic data, if these do not include all these numerous data items (structured and unstructured); iii) it is not clear which law has to be taken into account; iv) a balance between competing values, rights and interests, then increasing human control and discretion is required (so AI should be used as a decision-support, which provides recommendations to human actors, or at least humans should have a role 'in the loop').

3 AI AND THE RULE OF LAW

The use of AI tools in administrative decision-making shapes what we understand as rule of law. Theory and Courts, especially the European Court of Human Rights (ECHR), have developed various substantive concepts, guarantees and requirements that may be inferred from the notion of the rule of law, one important aspect thereof consisting in foreseeability and consistency of legislation and governmental/ administrative action. The rule of law implies a system of certain and due process, including that all individuals are subject to the same rules of justice. *"The notion of the Rule of Law requires a system of certain and foreseeable law, where everyone has the right to be treated by all decision-makers with dignity, equality and rationality and to have the opportunity to challenge decisions before independent and impartial courts through fair procedures"* (European Commission for democracy through law (the "Venice Commission"), 2016). Respect of rule of law becomes much more critical and imperative when the public authority enjoys discretionary powers, a "room for manoeuvre" when applying the law in the specific case.

The rule of law, conceived as predictability, seems not to be affected by automated decision making, if the latter follows a series of "pre-programmed" process, as decisions may be grounded on predefined and known factors (Zalnieriute et al., 2019). On the contrary, proponents of AI use argue that decision process with baseline programming responds to the rule of law, as due to the lack of human discretion and emotions it may lead to more objective and/or rational decisions (Wirtz et. all, 2019, Vogl et al., 2019; Araujo et al., 2020), promulgate "algorithmic bureaucracy", that "could be more calculable than ever before, generating predictable results that are sensitive to contextual factors" (Etscheid, 2019) and recognizes the possibilities of (weak) AI, especially if it was developed for a specific task. Yet, practice is more stubborn than those

idealistic views, as AI systematically introduce inadvertent bias, reinforce historical discrimination, favor a political orientation or reinforce pre-existing undesired practices (Janssen & Kuk, 2016), if the AI rules are extracted on historic data so the rules incorporate experience from the past, but also biases, stereotypes and possible bad practices from the past.

However, this is not the case if the decision process is designed to learn continuously from new data fed into the system, especially if AI tools are applied to administrative tasks with high complexity and uncertainty (Wirtz et. all, 2019; Bullock, 2019). These systems rely heavily on the use of data which are dynamic and can change over time resulting in different and even undesired outcomes (Janssen et al., 2020). AI systems that may develop their own decision-making process, based on their own values-assessment and interpretation schemes may be proved to be at the odds of the rule of law and harmful for humans.

Harm, meant as material or non-material damage, may derive from/as breach of equity/equality principle, according to which identically situated individuals should not be treated differently either by humans or by the "machines". The obligation to respect the principle of non-discrimination is enshrined in Articles 2 and 10 of the Treaty for European Union, Articles 20 and 21 of the EU Charter for Fundamental Rights and Freedoms as well as Article 14 of ECHR.

AI advocates rely on its use in administrative decision-making systems to rationalize them and face the various cognitive and motivational biases of public servants (Hermstrüwer, 2020), thus reducing their margin of appreciation (Veale and Bras, 2019) and preserving a fair and non-arbitrary decision. According to EU Agency for Fundamental Rights, algorithmic data analysis may produce results that could - under certain circumstances – contribute to the reduction of biases and stereotyping and dispel prejudicial attitudes, by reducing reliance on subjective human judgements (European Union Agency for Fundamental Rights, 2020). In Sweden, the municipality of Upplands-Bro has introduced (since June 2019) on an experimental basis the robot Tengai to deal with recruitment processes with the aim to make the recruitment process less biased than traditional interview practices would do.

On the other side, "it is a mistake to assume [these big data analysis techniques] are objective simply because they are data-driven" (White House Report on Big Data, 2016). Data inherits the bias from the past (Janssen & Kuk, 2016). Serious concerns are expressed with regard to the "algorithmic neutrality" as the use of data as input to an algorithm, the very purpose (especially) of machine learning algorithms (categorise, classify, separate) and the inner working of the algorithm itself that may result into discriminatory decisions. We should not ignore that humans are not deprived from personal beliefs and biases that may affect their assessment (Vogl et al., 2019). However, people are most concerned about bias in the application of algorithms and direct or indirect discrimination is considered as one of the most crucial challenges in the use of AI-driven tools for decision-making areas.

Various aspects of discrimination, including gender or race discrimination, can occur for several reasons. The selection and the quality of data fed into the system (lack of representativeness and accuracy), the data samples used to train and test algorithmic systems, choices of features, metrics and analytic structures that reproduce the designers' perceptions and biases may predefine the outcome to be produced (Leslie, 2019). Richardson et al. (2019) suggest that in fact human prejudices are reinforced and consolidated into the AI systems that, moreover, influence the effectiveness, as for example in the context of predictive policing police officers are more likely to stop or arrest people because of expectations raised by the system's analysis and prediction, rather than the actual circumstances on the ground (Babuta and Oswald 2019). Some variables used in AI modelling can be proxies for race, ethnicity, gender and other protected categories. The complexity and the obscurity of the algorithms produces additional difficulties to identify and remove such biases.

An AI-based decision may be discriminatory as far as the analysis is focused or even restricted on pattern recognition methods instead of recognition and assessment of causalities and causal relationships (Cobbe, 2019). In this respect the example cited by FRA demonstrates the risk of group discrimination: When scrutinising their algorithms, a public administration body found a higher degree of errors in tax declarations among recently issued national identification numbers, which have almost always been attributed to immigrants, and as FRA notes "this is also an example of proxy information, where parts of a number could indicate immigrant status". Moreover (and especially) by supervised machine learning applications the analysis may "perpetuate the past" and possibly discriminatory outputs: "computer outputs typically reflect what is already given, and not what could or should be, what is new, surprising, innovative or deviant. . . ..ML applications calculate a future which is like the past (Hagendorff and Wezel, 2020). Extracting patterns from existing data and use thereof for predicting the future results to a technology-driven "affirmation of the given" (Horkheimer, 2007) and cements existing bias and injustices. This assumes that all decisions in the past were fully accurate and correct and data is perfect, which of course is not always the case. Whether an automated decision-making process system is discriminatory is a question to be also answered by reference to the decisions produced by the system in much the same way as for human decision (Cobbe, 2019) (often humans as well tend to make decisions that are similar with previous decisions of other humans concerning similar cases).

The response to discrimination challenges also relates to the way AI has to be viewed and governed, the main choice is between a precautionary and a reactive approach. The European Commission seems to suggest a distinct path in the Proposal for an 'Artificial Intelligence Act' (European Commission, 2021): according to the Explanatory Memorandum, the proposal "complements existing Union law on non-discrimination with specific requirements that aim to minimise the risk of algorithmic discrimination, in particular in relation to the design and the quality of data sets used for the development of AI systems complemented with obligations for testing, risk management, documentation and human oversight throughout the AI systems".

The Draft- Regulation follows a risk-based approach, differentiating between uses of AI that create (i) an unacceptable risk, (ii) a high risk, and (iii) low or minimal risk. Among the list of prohibited practices in Title II are to find AI systems whose use is considered unacceptable as contravening Union values, for instance by violating fundamental rights. The proposal prohibits AI-based social

scoring for general purposes done by public authorities, as they may lead to discriminatory outcomes and the exclusion of certain groups. The Commission in Recital 17 points out that such systems "may violate the right to dignity and non-discrimination and the values of equality and justice. Such AI systems evaluate or classify the trustworthiness of natural persons based on their social behaviour in multiple contexts or known or predicted personal or personality characteristics. The social score obtained from such AI systems may lead to the detrimental or unfavourable treatment of natural persons or whole groups thereof in social contexts, which are unrelated to the context in which the data was originally generated or collected or to a detrimental treatment that is disproportionate or unjustified to the gravity of their social behaviour".

The above analysis of the use of AI-driven decision-making in government from a 'rule of law' perspectives reveals further needs for human control, intervention and discretion in order to avoid the use of data-sets that are non-representative, inaccurate, or reflect undesired past practices (or past practices that are not appropriate for the current context), or even reflect historic discriminations; this will help avoiding 'perpetuation' of negative past situations, and cementing pre-existing biases and injustices. Furthermore, human control, intervention and discretion might be necessary if there might be problems of lack of consistency and predictability of the decisions provided by AI, especially in cases where new data are used for continuous learning of the AI algorithms (which might result in AI providing different decisions for similar cases in different time points, because the algorithm has been slightly modified based on new data).

4 THE TRANSPARENCY, JUSTIFIABILITY AND ACCOUNTABILITY CHALLENGE

At the end the main goals consist in enhancing efficiency of administrative action while ensuring that the law is properly forced in/by the AI systems. Public bodies and decision-making procedures and processes must be visible and explainable in a way to be understandable so to preserve their auditability and accountability. Preliminarily, it has to be clarified these requirements have to be addressed by the public authorities, "as humans remain responsible for the consequences and risks associated with technology and the law is concerned with the activities of natural or legal persons and doesn't directly address the actions of machines" (Cobbe, 2019).

A crucial issue is how to implement the principle of accountability and the -interrelated- principle of transparency and explainability of AI-based administrative action. Responding to this need is of utmost importance, as individuals must be able to hold directly accountable the public bodies for classifications, predictions or /and decisions produced by AI that produce legal effects concerning them. In this case individuals must be able to confirm whether a decision that affects them has been taken (un)lawfully and to bring judicial review if they want (Cobbe, 2019). It is clear that public bodies are subject to more intensive transparency and accountability obligations, if they are acting with discretion (Leslie, 2019).

AI-based decisions are heavily influenced by the selection of data(sets), the training data, the design, construction and training of (statistical) models. The so-called "accountability gap" seems to be inherent in AI systems, as the outcomes they produce are not self-justifiable and - consequently - "accountable" and "responsible" in the same lawfully and morally sense as human actors. Additional difficulties arise due to the lack of transparency or – prima facie- explainability and comprehensibility of such processes, as it is nearly impossible – for an outstander/outsider to review the decision process and the basis of calculation and output. Transparency and – consequently- accountability seem difficult to be achieved, as far as it concerns AI systems that "learn" while they are in operation and act in that way autonomously, so that their developers or operators may not be capable of predicting, controlling or explaining their subsequent behaviour (Johnson, 2015). The shortcomings with regard to transparency and scrutiny undermine not only the right to challenge administrative decisions but also the acceptability of AI systems and the outcomes they produce (Yeomans et al. 2019), which may lead to "algorithmic aversion" (Dietvorst et al., 2015).

The need to provide transparency and explainability and -at the end- auditability and accountability has to be faced when it comes to decisions assisted or produced by an AI system. In order to enable auditability and accountability what is required is designing hardware, software and process in a way that enables end-to-end oversight, review and scrutiny. Very useful for this purpose can be the research conducted in the area of explainable/interpretable AI (Du et al. 2020), which has already developed two types of techniques for this purpose: intrinsic interpretability/explainability and post-hoc interpretability/ explainability ones. Intrinsic interpretability/explainability is achieved by constructing self-explanatory models which incorporate interpretability directly to their structures (such as decision tree, rule-based model, linear model and attention model algorithms). On the contrary, the post-hoc interpretability/explainability is based on building a second model to provide explanations for the decisions provided by the primary model. Processes and metrics for transparent and accountable AI - systems do not always translate easily to legal frameworks and administrative procedure. (Cobbe, 2019). The lack of understanding between technical approach and administrative law does not facilitate the achievement of a balanced response.

Therefore, the requirements set by law for transparency, explainability and auditability of administrative decisions, and therefore accountability for them, necessitate also some degree of human control and discretion in AI-driven decision-making in government.

5 DISCRETION AND MEANINGFUL HUMAN CONTROL

However, even if AI systems may not be subject to direct human control, humans and in our context authorities remain accountable for the behaviour of AI systems despite their lack of control and influence. The principle of accountability demands to establish a continuous chain of human responsibility across the entire AI – system (Leslie, 2019). Does it mean that public authorities should abstain from deploying AI systems as it seems difficult to bear responsibility for processes and outcomes?

The answer depends on the task, the level of discretion and the impact of decisions on their addressees. As Araujo et al. (2020) state studies have also shown differences in the perceptions for automated decision-making for objective or subjective decisions (Logg, 2017; Logg et al., 2019), or for management decisions requiring

human or mechanical skills (Lee, 2018). The type of task and the type of (human) decision-maker is important for cultivating trust and social acceptability: machines were more trusted than human non-experts, but less trusted than human experts.

Fully AI-based decision making seems to be acceptable only in cases that there is low level of discretion or absence of discretion, e.g. when the public body are strictly bound by the law in respect of the provisions and the procedures to be followed and no negative effects for individuals and /or groups are reasonably presumed. In case that the level of discretion is high, due to specific data and circumstances that have to be taken into consideration, AI - systems may serve as support for classification and prediction of needs for improving public service delivery. AI systems can also support the human decision-making in such procedures, for example by automatically collecting information, graphically processing information or by giving suggestions for evaluation (Etscheid, 2019). For instance, in social policy, AI is being used to support the prediction of high-risk youth for targeting interventions (Sun and Medaglia, 2019) or to enable more accurate predictions to detect day-care services to children and families that may require further inspection (System Kind en Gezin developed by the Flemish Agency for Child and Family).

However, AI systems might result in less discretion, as less and less humans are involved in decision-making and decisions made by AI might be perceived as objective. Furthermore, an important reason for introducing these systems might be the need for reducing cost and reducing staff. This might have negative consequences for administrative decision making. Criado et al. (2020) analyze the implications in decision-making processes supported by the use of an AI-based system and the effects in the discretionary power of public employees involved in its implementation. They found a positive impact of algorithms on the work and discretionary power of civil servants, as AI was viewed as a decision support system instead of a decision-making system. Hence, the role of humans and discretion needs to be strengthened instead of dismissed.

Humans remain accountable for the systems. Given the deficiency of AI systems, there is a huge need for *meaningful human control of AI systems*. In such situations, humans control the input data, information processing and output results and have the discretion to deviate from the suggested decisions by AI. This is different from classical discretion in public administration, and knowledge of algorithms, legislation and the situation at hand is needed. This makes this even more challenging.

Meaningful human control is particularly important when there might be possible failures. Humans can play different roles and can deviate from decisions suggested by AI. This would require a sound argumentation and can be related to the input data, the algorithms used, the understanding of part of the context not captured by data, the interpretation of regulations, or other aspects not captured by AI. Indeed, meaningful human control has its limitations as a recently published experiment showed (Janssen et al., 2021). In particular, it was shown that even with rule-based algorithms suggesting to humans decisions, these humans were not able to detect all mistakes. Its findings suggest that explainable AI combined with experience helps human decision makers detect incorrect suggestions made by algorithms, however even experienced persons were not able to identify all mistakes. So, additional measures might be needed, such as four-eyes principles, education is crucial; nevertheless, it might be impossible to avoid all mistakes.

6 CONCLUSIONS

AI-based systems are more and more used for decision-making in government. Yet humans remain accountable and therefore, discretion and meaningful human of the AI systems are needed. There are various levels of discretion, and the role of discretion will likely change. From our analysis from a legal perspective presented in the previous section it can be concluded that discretion will be particularly important:

- for administrative decisions in which there are a lot of input data items to be taken into account, with some of them being unstructured, or even not known in advance; this is going to be even more important if decision rules are extracted from historic data, which do not include all these numerous data items (structured and unstructured);

- for administrative decisions for which it is not clear which law has to be taken into account, or a balance has to found between competing values, rights and interests.

Also, some degree of human control, intervention and discretion might be necessary:

- in order to avoid the use of data-sets that are non-representative, inaccurate, or reflect undesired past practices (or past practices that are not appropriate for the current context), or even reflect historic discriminations;

- if there might be problems of lack of consistency and predictability of the decisions provided by AI, especially in cases where new data are used for continuous learning of the AI algorithms;

- for increasing transparency, explainability and auditability of administrative decisions, and therefore accountability for them.

Further empirical research is required for the evaluation of cases in which the public authorities may exercise discretion, as well as of cases in which this is not necessary. Also, further research is needed in order to measure and assess the distribution of decisions that could/would be different depending on the method/decisive process, and the use of AI. As proposed by Hermstrüwer (2020) the obvious – and simplest? - way were to replace machine learning with human judgement randomly for a sufficiently large sample in order to measure the impact of the variables that drive and explain potential differences between human judgement and the AI-driven outcome. Finally, it would be interesting to analyze further the provisions included in the recent proposal of Artificial Intelligence Act (AIA) of the European Union, and follow the respective scientific and policy discussion during the drafting period. This can be done in comparison with the analysis of national legal frameworks introduced for the purposes of automated administrative decision-making diverges from the procedural principles applicable to classic (non-automated, human-based) administrative decision-making.

REFERENCES

[1] Araujo T., Helberger N., Kruikemeier S., de Vreese C. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI & Society, 35, 611–623.

[2] Babuta A. and Oswald M. (2019). Data Analyticsand Algorithmic Bias in Policing, Briefing paper,Royal United Services Institute for Defence and-Security Studies. UK government's Centre forata Ethics and Innovation.

[PDF] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/831750/RUSI_Report_-_Algorithms_and_Bias_in_Policing.pdf
[3] Barcevičius, E., Cibaitė, G., Codagnone, C., Gineikytė, V., Klimavičiūtė, L., Liva, G., Matulevič, L., Misuraca, G., Vanini, I., Editor: Misuraca, G., Exploring Digital Government transformation in the EU - Analysis of the state of the art and review of literature, EUR 29987 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-13299-8, doi:10.2760/17207, JRC118857
[4] Bullock, J. B. (2019). Artificial Intelligence, Discretion, and Bureaucracy. American Review of Public Administration, 49(7), 751–761.
[5] Bullock, J. B., Young, M. M. and Wang, Y.F. (2020). Artificial intelligence, bureaucratic form, and discretion in public service. Information Polity, 25, 491–506.
[6] Cobbe, J. (2019). Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making. Legal Studies , 39(4), 636 – 655.
[7] Criado, J. I., Valero, J., Villodre, J. (2020). Algorithmic transparency and bureaucratic discretion: The case of SALER early warning system. Information Polity, 25(4), 449-470.
[8] DeSousa, W. G., DeMelo, E. R. P., De Souza Bermejo, P. H., Sous Farias, R. A., Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. Government Information Quarterly, 36(4), 101392.
[9] Dietvorst B.J, Simmons, J. P., Massey C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology – General, 144, 114–126.
[10] Digital Future Society (2021) - Governing algorithms: perils and powers of AI in the public sector
[11] Duan, Y., Edwards, J. S., Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. International Journal of Information Management 48, 63–71.
[12] Du, M., Liu, N., Hu, X. (2020). Techniques for Interpretable Machine Learning. Communications of the ACM, 63(1), 68-77.
[13] Etscheid, J. (2019). Artificial Intelligence in Public Administration A Possible Framework for Partial and Full Automation. In Ida Lindgren, Marijn Janssen, Habin Lee, Andrea Polini, Manuel Pedro Rodriguez Bolivar, Hans Jochen Scholl, Efthimios Tambouris (Eds.), Electronic Government, 18th IFIP WG 8.5 International Conference, EGOV 2019 San Benedetto Del Tronto, Italy, September 2–4, 2019, pp. 248-261.
[14] European Commission for democracy through law (the "Venice Commission"). (2016). Rule of Law Checklist
[15] European Union Agency for Fundamental Rights (2020). Getting the Future Right - Artificial Intelligence and Fundamental Rights. Luxembourg, Publications Office of the European Union.
[16] European Commission (2021). Laying Down Harmonized Rules on Artificial Intelligence (ARrtificial Intelligence Act) and Amending Certain Union Legislative Acts. Brussels.
[17] Hagendorff T. and Wezel K. (2020). 15 challenges for AI: or what AI (currently) can't do. AI & Society (2020), 35, 355–365.
[18] Hermstrüwer, Y. (2020). Artificial Intelligence and Administrative Decisions Under Uncertainty. In T Wischmeyer, T Rademacher (Eds) Regulating Artificial Intelligence, 199-223.
[19] Horkheimer, M. (2007). Zur Kritik der instrumentellen Vernunft. Fischer Verlage, Frankfurt.
[20] Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. Government Information Quarterly, 33(3), 271-377.
[21] Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. Government Information Quarterly, 37(3), 101493
[22] Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Kuk, G. (2021). Will Algorithms Blind People? The Effect of Explainable AI and Decision-Makers' Experience on AI-supported Decision-Making in Government. Social Science Computer Review, 0894439320980118.
[23] Johnson, D. G. (2015). Technology with no human responsibility? Journal of Business Ethics, 127(4), 707–715.
[24] Kuziemski M. and Misuraca G. (2020). *AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings
[25] Lee M. K. (2018). Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management. Big Data Big Data & Society, January–June 2018, 1–16.
[26] Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529.
[27] Logg, J. M. (2017). Theory of Machine: When Do People Rely on Algorithms? Harvard Business School Working Paper. No. 17-086
[28] Logg, J. M., Minson, J.A., & Moore, D.A. (2019). Algorithm Appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90-103.
[29] Loukis, E., Maragoudakis, M., Kyriakou, N. (2020). Artificial Intelligence based Public Sector Data Analytics for Economic Crisis Policy Making', Transforming Government: People, Process and Policy, 14(4), 639-662.
[30] Misuraca, G., and van Noordt, C.., Overview of the use and impact of AI in public services in the EU, EUR 30255 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19540-5, doi:10.2760/039619, JRC120399
[31] Nolte, G. (1994). General Principles of German and European Administrative Law - A Comparison in Historical Perspective. The Modern Law Review, 57(2), 191-212.
[32] Organisation for Economic Co-operation and Development. (2019). Artificial Intelligence in Society. OECD Publishing, Paris.
[33] Richardson, R., Schultz, J. and Crawford, K.(2019). Dirty Data, Bad Predictions: How CivilRights Violations Impact Police Data, PredictivePolicing Systems, and Justice. 94 N.Y.U. L. REV.ONLINE 192. [online] Available at: https://ssrn.com/abstract=3333423
[34] Sun, T.Q. and Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. Government Information Quarterly, 36(2), 368-383.
[35] Varuhas, J. (2020). The principle of legality, Cambridge Law Journal,79(3), 578–614.
[36] Veale, M. and Bras, I. (2019). Administration by Algorithm? Public Management meets Public Sector Machine Learning. In: Karen Yeung and Martin Lodge (Eds), Algorithmic Regulation, Oxford University Press.
[37] Vogl, T. M. and Seidelin, C. and Ganesh, B. and Bright, J. (2019). Algorithmic Bureaucracy: Managing Competence, Complexity, and Problem Solving in the Age of Artificial Intelligence Available at SSRN: https://ssrn.com/abstract=3327804.
[38] Yeomans, M., Shah, A., Mullainathan, S., Kleinberg, J. (2019). Making sense of recommendations. Journal of Behavioural Decisio Making, 32(4), 403–414.
[39] White House Report on Big Data. (2016)
[40] Wirtz, B. W., Weyerer, J. C., Geyer C. (2019). Artificial Intelligence and the Public Sector-Applications and Challenges. International Journal of Public Administration, 42(7), 596-615.
[41] Zalnieriute M., Bennett Moses L. and Williams G. (2019). The Rule of Law and Automation of Government Decision-Making. The Modern Law Review, 82(3), UNSW Law Research Paper No. 19-14.