AI & SOCIETY
https://doi.org/10.1007/s00146-022-01493-0

OPEN FORUM

Artificial intelligence and democratic legitimacy. The problem of publicity in public authority

Ludvig Beckman[1] · **Jonas Hultin Rosenberg**[2] · **Karim Jebari**[3]

Received: 21 January 2022 / Accepted: 12 May 2022

Abstract
Machine learning algorithms (ML) are increasingly used to support decision-making in the exercise of public authority. Here, we argue that an important consideration has been overlooked in previous discussions: whether the use of ML undermines the democratic legitimacy of public institutions. From the perspective of democratic legitimacy, it is not enough that ML contributes to efficiency and accuracy in the exercise of public authority, which has so far been the focus in the scholarly literature engaging with these developments. According to one influential theory, exercises of administrative and judicial authority are democratically legitimate if and only if administrative and judicial decisions serve the ends of the democratic law maker, are based on reasons that align with these ends and are accessible to the public. These requirements are not satisfied by decisions determined through ML since such decisions are determined by statistical operations that are opaque in several respects. However, not all ML-based decision support systems pose the same risk, and we argue that a considered judgment on the democratic legitimacy of ML in exercises of public authority need take the complexity of the issue into account. This paper outlines considerations that help guide the assessment of whether a ML undermines democratic legitimacy when used to support public decisions. We argue that two main considerations are pertinent to such normative assessment. The first is the extent to which ML is practiced as intended and the extent to which it replaces decisions that were previously accessible and based on reasons. The second is that uses of ML in exercises of public authority should be embedded in an institutional infrastructure that secures reason giving and accessibility.

Keywords Machine learning · Public authority · Democratic legitimacy · Publicity · Opaque

1 Introduction

An ever-larger share of human decisions are made by or with the support of sophisticated machine learning algorithms (ML). Since the breakthrough of ML techniques, this development has accelerated. More specifically, ML is increasingly used by public authorities in the application of law and the pursuit of public policy. Algorithmic advice based on this technology is now found in courtrooms, employment service offices, and other public institutions (Djeffal 2020).

When assessing the use of ML, we should distinguish between the consequences of artificial intelligence (AI) in public decision-making and the consequences of AI as a tool for private actors, such as companies and consumers. Though many problems and challenges—threats to privacy, for example—apply equally to both, public authorities are subject to additional demands. Public authorities not only have a duty to do no wrong, but they must also conform to principles of legitimate decision-making.

The literature attentive to the specific issues raised by the democratic legitimacy of involving AI in public decision-making is scant despite the fact that the consequences of AI for democratic politics generally are widely debated (e.g., Feldstein 2019; Zuboff 2019). Exceptions include

✉ Ludvig Beckman
Ludvig.beckman@iffs.se

Jonas Hultin Rosenberg
jonas.hultin_rosenberg@statsvet.uu.se

Karim Jebari
Karim.jebari@iffs.se

[1] Institute for Futures Studies Stockholm and Department of Political Science, Stockholm University, Stockholm, Sweden

[2] Department of Government, Uppsala University, Uppsala, Sweden

[3] Institute for Futures Studies, Stockholm, Sweden

Paul Nemetz (2018) and Simon Chesterman (2019), arguing that AI decisions that replace decisions made by humans in public authority "must thus be checked against higher law and the basic tenets of constitutional democracy"; Karl de Fine Licht and Jenny de Fine Licht (2020), who discuss the importance of transparency for perceived legitimacy; and Christoffer Starke and Marco Lünich (2020), who consider perceived legitimacy and the use of algorithmic decision-making in the EU. "Democracy" and "accountability" figure into the ethical guidelines and declarations on AI adopted by the Council of Europe (2019); by the European Commission (2018a), the Ethical Charter adopted by the Commission (2018b), and the subsequent EU ethics guidelines for AI (2019); and by the European Parliament (2020).

The main concern in these documents is with securing government overview and responsibility for applications of AI in society generally. Only the Montréal Declaration of Responsible AI (2018) mentions democratic legitimacy in connection to AI in public decisions. According to the declaration, algorithms used by public authorities should be transparent and "justifiable in a language that is understood by the people who use them or who are subjected to the consequences of their use" (Article 5). What this implies is left unanswered.

This paper suggests that the democratic legitimacy of public decision-making made or supported by ML raises other, more fundamental, concerns. According to some of the most prominent political theories (see, e.g., Christiano 2008, citizens have a fundamental interest in not only being treated fairly, but also in "seeing justice done." This idea is not alien to the principles embedded in constitutional democracies. The principle of publicity figures into national and European legal and constitutional frameworks.[1] The implication is that public decision-making should conform to the *principle of publicity,* which consists of the following two requirements:

1. Reason giving. This connects a decision with some features of the world (e.g., laws, relevant facts) and the specifics of the case.
2. Accessibility. This means that the reasons are available to the directly affected party and to other relevant stakeholders.

Yet, the particular properties of ML seem to violate both these requirements in some cases. ML provide reasons that are "statistical" and therefore not sufficiently determinate for the individual case, and ML are moreover "opaque" in the sense that their operations cannot be fully explained (Burell 2016).

The principle of publicity should be distinguished from the requirement that government and law making should be transparent. The value of transparency in government is widely accepted (Florini 2007; Hood and Heald 2006). However, transparency is primarily concerned with access to government information to empower citizens as participants in the democratic process. Publicity is, alternatively, concerned with the citizens as subjects to public authority; it addresses their interests as law *takers* rather than as law *makers*. Whereas transparency requires access to information about legislative and government decision processes, the principle of publicity requires access to reasons in the exercises of public authority.

The principle of publicity raises the following two issues: (i) the nature of opacity in ML—to what extent is it different from human decision-making, where motivations and implicit assumptions may also be hidden for external observers?, (ii) the normative significance of the principle of publicity—specifically, the importance of access to reasons for public decisions should be assessed from the perspective of the potential benefits of algorithmic exercises of public authority. The aim of this paper is to advance our knowledge of the democratic legitimacy of ML in public decisions by clarifying the nature of the problem and to identify the normative conflicts that need to be addressed. This requires an investigation into the opacity in ML and the relation between opacity and publicity. Moreover, we will argue, in line with Dourish (2016), that no systematic evaluation of the normative implications of ML can be adequate without taking into consideration how these systems are implemented in a particular context. For example, the question of opacity is not merely a property inherent in ML algorithms, it is also a result of social dynamics and human psychology.

The paper is structured as follows: After introducing the theoretical framework by explicating the importance of publicity for the legitimacy of administrative and judicial exercises of public authority, we engage with the challenge posed by ML for reason giving and accessibility. We argue that the use of ML in the exercise of administrative and judicial authority risk violating the reason giving component of the principle of publicity by basing decisions or recommendations solely on statistical similarities. Furthermore, we argue that decisions and recommendations made by ML

[1] Fundamental treatises of the European Union; "Legal acts shall state the reasons on which they are based" (Article 296). Swedish Law on Public Administration; "A decision that can be expected to affect a person in non-marginal ways should include a clear justification, unless clearly unnecessary" (§ 32 Förvaltningslag 2017:900). These examples are from Europe but the scope of the argument in this paper extends to all states that aspire to be democratic. Reason giving and accessibility are requirements for democratic legitimacy even if these requirements are not embedded in the constitutional framework of every existing democracy.

often violates the accessibility component due to the multiple reinforcing layers of opacity in ML. We conclude the paper with a discussion about two main considerations that are pertinent to the normative assessment of the democratic legitimacy of ML in public authority. The first is practical and concerns the extent to which ML is practiced as intended and the extent to which it replaces decisions that were previously accessible and based on reasons. The second is to insist that ML in public authority is embedded in an institutional infrastructure that secures reason giving and accessibility.

2 Democratic legitimacy and the importance of publicity

The scholarly literature on democratic legitimacy usually focuses on the conditions for the legitimate authority of legislative bodies, that is, legitimate law making. The focus in this paper is different, however, as we address the legitimacy of ML in administrative and judicial decision-making, not in law making. Though the basic principles of democratic legitimacy inform all exercises of public authority, it is important to keep the distinction between legislative and administrative decision-making in mind in what follows. The claim that public authority is legitimate either means that it permissibly wields coercive power or that the subjects of public authority are morally bound to comply with it (Peter 2017). In the present context, not much depends on the distinction, and for reasons of simplicity, we presume, therefore, that legitimate public authority implies that subjects are morally bound to comply (see, for example Estlund 2008: 2; Viehoff 2014: 340; Wall 2007. This means that those who are subjected to legitimate decisions must abide by them even if they believe that these decisions are wrong.

The relation between legitimate authority and democracy has often been neglected in democratic theory (Wall 2007). One difficulty relates to the extent to which democracy is sufficient for legitimate public authority. Even if democracy is a necessary precondition for legitimate law making, it arguably does not follow that subjects have a moral obligation to comply with the law regardless of its substance (see Christiano 2008: Ch. 7). For example, there may be no moral obligation to comply with democratic laws that violate basic rights to democratic participation, personal liberty, or the enjoyment of a social minimum. Moreover, no existing democratic state is likely to be fully legitimate, though it is arguably part of the nature of legal systems that they *claim* legitimate authority (Raz 2009).

An additional difficulty is to explain how democracy contributes to the legitimacy of public decisions. Democracy is sometimes understood as the right of the majority to impose its will. But the fact that a law or a rule expresses the will of the majority does not imply that the minority is morally bound to comply. As early as the 14th century, the notion emerged that "will" was insufficient for legitimate authority and that public decisions must be based on *reasons* (Pennington 1993).

On the account developed in the liberal tradition, reasons are essential for legitimate public authority as only reasons can secure public justifiability (Friedrich 1968; Waldron 1987). Legitimate exercises of public power must be based on reasons that are *acceptable to subjects*.

This "acceptability requirement" can be understood in different ways. According to the consent theory of authority, actual agreement is a necessary requirement for authority (Beran 1977). On this account, the exercise of political and legislative power has legitimate authority only if those who are subjected have consented to this exercise of power. A less demanding version of the acceptability requirement necessitates acceptance only from a subset of those who are subjected, namely from those who are subjected and reasonable (for versions of this qualified acceptability requirement, see Estlund 2008; Rawls 1993; cf. Enoch 2015). In any version of the acceptability requirement, democracy is necessary for legitimate authority since only democracy provides reasons for public authority that subjects can or could accept.

2.1 The legitimate authority of democracy

Reasons are thus fundamental to the legitimacy of public authority. The question then is why democracy provides reasons that make authority legitimate. There are three main answers in the literature. The first is that democratic procedures are fair and that subjects therefore have reason to accept the legitimacy of decisions made by a democratic public authority (see for an overview Viehoff 2014). The second answer is that democratic decisions tend to be substantively good decisions and that subjects therefore have reason to accept decisions made by democratic authorities as legitimate (Estlund 2008; Wall 2007). The first is commonly referred to as proceduralism and the second as instrumentalism. A third alternative is to reject the monistic claim of both of them that legitimate authority derives from a single source (Christiano 2004; see also Wall 2007). The third view, most consistently pursued by Christiano (2004, 2008), is that the reasons for the acceptability of public authority need to attend both to its tendency to generate good decisions and the procedural qualities of how decisions are made.

Which account is the correct one is among the perennial normative issues in democratic theory that cannot be resolved here. Given our focus on the democratic legitimacy of ML in administrative and judicial authority, there are nevertheless methodological reasons to adopt

the third account. ML technology is typically introduced to improve the quality of public decisions both in terms of procedures and outcomes. Assuming that the legitimacy of public decisions depends exclusively on one of these dimensions would therefore limit our capacity to evaluate the impact of ML in public decision-making. Thus, we believe that democratic legitimacy should provisionally be conceived of as grounded in both procedural and instrumental reasons. To this end, Christiano's dualistic account of democratic legitimacy is a particularly promising candidate for present purposes.

The moral foundation of Christiano's view is the *principle of equal advancement of interests,* according to which a just society (i) advances the interest of all its members and (ii) advances the interest of all its members equally (Christiano 2008: Ch. 1). In the domain of social justice (e.g., justice of institutions and interactions among persons), the principle of equality of advancement of interests requires that equality be publicly realized: "it must not only be the case that people are treated as equals, they must be able to see that they are treated as equals'' (Christiano 2008: 46). In the domain of social justice, the principle of equality requires *public* equality. According to Christiano, publicity is important given an assumption about a circumstance of disagreement, diversity, fallibility, and cognitive bias (Christiano 2008: 46).

People have a fundamental interest in being treated as equals. Under idealized circumstances, it is conceivable that everyone's interests can advance without publicity. But under circumstances of disagreement about what justice requires, fallibility in moral judgment, and cognitive bias that distorts the interpretation of the interests of others, people also have a fundamental interest in seeing that they are being treated as equals (Christiano 2008: 56–59).

On this account, publicity is not the same as governmental transparency. Public equality does not imply that political and legal institutions are transparent in all respects and in all parts. Full transparency is not always desirable from the perspective of the principle of equality. Secrecy may in some cases be important for the equal advancement of interests (for example, when children are defendants in a court case). A public principle is a principle that a person, with normal cognitive faculties, who makes a reasonable effort, *can see* to be realized. The principle of equality is publicly realized in public institutions when citizens are able to confirm that they are treated as equals (Christiano 2008: Ch. 2). Hence, publicity is not in itself an independent good. Unjust decisions are not more so just because they are public. Publicity is a dimension of social justice in the sense that a just decision is defective if it is not public. Democratic decisions have legitimate authority because democracy uniquely satisfies public equality. According to Christiano, democracy is necessary to "the public realization of equal advancement of interests" (Christiano 2008: 249). To this end, decisions need to be public in a practical sense too, which implies that not only rules and procedures allow the public to access information, but also that the public can understand and ultimately evaluate the reasons that determine exercies of public authority.

2.2 Democratic legitimacy in administrative and legislative authority

Christiano's account of legitimate democratic authority applies primarily to legislative institutions (Christiano 2008: 245). However, the present assessment is not about ML technology in law making but about ML technology in the administrative and judicial exercises of public authority. Christiano indicates, in his discussion of the complexity of authority, that the authority of administrative agencies and courts is primarily instrumental: "they are meant primarily to serve the aims of the democratic assembly and to protect the realization of public equality" (Christiano 2008: 258). The authority exercised by administrative and judicial agencies should thus be assessed based on its tendency or ability to serve the aims of the democratic assembly and on its tendency to protect the realization of public equality. If we assume the principle of equality of advancement of interests and the circumstance of disagreement, diversity, fallibility, and cognitive bias, democracy is required for legislative authority. Administrative and judiciary authority is related to democracy in the sense that this authority depends on the tendency to serve what has been democratically decided and to protect the realization of the value that democracy uniquely realizes. The instrumental merit, the legitimate authority, of the administrative and judiciary system thus depends on the connection between the decisions taken by these agencies and the legislative decisions made by the democratic assembly. The realization of what is democratically decided and of public equality depends on the efficiency and accuracy of the administrative and judicial system.

Hence, to know whether administrative and judicial decisions accord with the requirements for democratic legitimacy, we need to be able to determine the extent to which these decisions serve the aim of the democratic assembly and protect the realization of public equality. For this, it is required that these decisions be based on *reasons* that connect the administrative and judicial decision with the democratically decided laws, rules, or regulations and the specific case and that these reasons be *accessible* to those subjected to these decisions as well as to relevant third parties.

Moreover, it is also possible to challenge Christiano's view on the strict separability between the legislative domain and the execution of the legislative assembly's

will. A law is not something that comes into existence, fully formed, by an act of an assembly. Rather, a law takes form and develops as it is implemented by jurists, government agencies, and grass-roots bureaucrats. For example, precedents can extend or interpret a law. When judges, civil servants, and other public officeholders act, they are not only implementing the will of the legislative assembly but they are also adding their own will to the outcome. The boundary between law making and law enacting is in practice vague and constantly changing. Thus, the principle of publicity proposed by Christiano could be extended to decisions made by the judiciary, by government agencies, and so on. In this view, publicity for public decision-making is important not only because it allows us to make sure that the decisions serve the aim of the legislative assembly but additionally because these decisions are also part of the legislative process, broadly speaking. These decisions can only be just if they are publicly just for the very same reasons that legislative decisions can only be just if they are publicly just.

Based on this understanding of the instrumental and legislative authority of administrative and judicial agencies, the implications of the implementation of ML-based decision support systems are ambiguous. On one hand, ML have the potential of contributing to the realization of the aims of the democratic assembly and to the protection of the realization of public equality by contributing to the efficacy and accuracy of the administrative and judicial system. On the other hand, the opacity of ML makes it difficult to determine the extent to which these decisions serve the aim of the democratic assembly and protect the realization of public equality. The use of ML in administrative and judicial public decision-making could thus be incompatible with the requirements for democratic legitimacy either by failing in the realization of what is democratically decided; by not connecting the decision in the individual case to the relevant facts of the case and the relevant democratically decided rules and laws; or by failing to make the connection between the decision, the relevant facts, and the applicable rules and laws available to the relevant parties.

As understood here, the principle of publicity has two components—a reason-giving component and an accessibility component—and both need to be satisfied in order for public decisions to be fully democratically legitimate. The use of ML in administrative and judicial public decision-making should be assessed based on its accordance with both of these components. However, as noted above, the purpose of the principle of publicity is to enable citizens to assess the extent to which administrative institutions pursue the aims decided by the democratic law maker. In sum, the present account stipulates that ML technologies in the exercise of public authority are legitimate by democratic standards if and only if (i) they serve the ends of the democratic law maker, (ii) they are based on *reasons* that align public decisions with the aims that are democratically decided, (iii) and the reasons are *accessible* to the subjects of public authority. Clearly, tensions and conflicts can emerge between the three distinct normative criteria of democratic legitimacy. Our ambition here is not to provide a definitive resolution of these conflicts but to take a first step in identifying what the relevant questions ought to be.

3 The challenge posed by ML

ML is a type of AI that has recently made huge advances (Marcus and Davis 2019). ML is now widespread in recommendation systems, that is, software that recommends one or more options from a larger set according to certain criteria. For example, Google Search and YouTube rely on ML-based recommendation algorithms to suggest search results and videos to watch (Covington et al. 2016). ML are different from earlier forms of AI, often referred to as "good old-fashioned AI" or "expert systems" (Russell and Norvig 2020). These early AI systems consist of two main parts: a set of logical rules, such as "if condition A is not fulfilled, then the person is not eligible for intervention B," and a knowledge base with structured information. The advantage with this older generation of AI is its relative transparency. The rules are coded by programmers and could be thought of as codified laws and practices. By contrast, ML create a mathematical model based on training data (Bishop 2006).

The most prevalent form of ML technique is called "supervised learning." Here, the algorithm is presented with labeled training data, such as a number of facts about a person and the information that the person defaulted on a home loan (Agrawal et al. 2018). For each item, a neural net learns to associate a certain pattern of features with a certain category, such as "defaulter" and "non-defaulter." After being shown a large number of items, the algorithm is able to classify objects that are not in the training set (Russell and Norvig 2020). This means that ML categorizes an object depending on whether it shares features with other objects that are already categorized. For example, a person could be categorized as having a "high risk for recidivism" if they share some features with other people that did reoffend. Whereas expert systems only categorize individuals according to a number of clear rules, ML makes a categorization based on statistical similarity (O'Neil 2016). While ML have a significant potential to assist in public decision-making, the intrinsic features of such systems pose a significant practical and normative challenge for decisions that ought to be constrained by the principle of publicity.

3.1 Reason giving

The principle of publicity requires not only that decisions be fair but that they are arrived at in the right way. For example, if a judge were to declare someone guilty as charged on the basis of a coin toss, this would not be acceptable even if the charged person did actually commit the crime. Part of what it means for a decision to be reached in the right way is that it be motivated by the right kind of reasons. From the perspective of democratic legitimacy, this consists in making decisions based on the relevant facts of the individual case and the will of the democratic assembly as expressed in democratically decided rules and laws. Decision support systems based on ML can come in conflict with this requirement.

First, to have a decision be motivated by the right kind of reasons involves that the reasons concern the particular case in relation to laws, regulations, and procedural norms. For example, a motorist can be fined for speeding if the traffic police can show that the measured speed of the car exceeded the speed limit in a particular case. It would violate the reason-giving component of the principle of publicity to punish a motorist simply because they drove a car model that was statistically associated with speeding, such as a sports car. Only an appeal to reasons that concern an individual case are consistent with the principle of publicity.

However, since ML typically arrive at a categorization based on statistical similarity between an object and objects in a particular category, their assessment, if applied uncritically by a public official, could in some circumstances violate the reason-giving component of the principle of publicity. For example, if a court were to use a ML that delivers an assessment of the risk of recidivism for a defendant, this assessment would not be based on the individual's risk for reoffending but rather on the individual's statistical similarity with people that did reoffend and were caught doing so. Thus, if there is a statistical discrepancy between reoffenders that avoid the justice system and reoffenders that fail to do so, then this discrepancy will be carried over to those deemed to be of "high risk" for recidivism.

Second, the decision can only be motivated by reasons deemed to be relevant in a particular case. The fact that a person is male or has low-income neighbors may be statistically correlated with the risk of that person being criminal. But to base a public decision, such as denial of probation, on such reasons would be a form of discrimination. This is a general problem for the democratic legitimacy of the use of ML in the exercise in administrative and judicial public authority. It is also something that risk further disadvantage already disadvantaged groups (O'Malley and Smith 2020). The risk of biases and discrimination in AI-systems in general has been frequently acknowledged in the scholarly literature. These biases will become increasingly hard to identify and control as these systems become more complex and advanced (see Mann and Matzner 2019).

Of course, the risk that public decisions are not grounded in reasons that are relevant to the particular case is not unique to decisions supported by ML. Public officials can make decisions based on routine without taking into consideration the specifics of the individual case. More seriously, they may deliberately ignore the specifics of the case because of some statistical theory. This is manifested in the problem of *profiling,* which is a recurrent theme in debates about the legitimate means of law enforcement by security personnel, the police, and anti-terrorism agencies (Becker 2004; Rudovsky 2001). Profiling is premised on the existence of a correlation between the observable features of individuals and their propensity for criminal behavior, derived either from stereotypes or statistical models. As such, profiling further exemplifies deviations from the requirement that public decisions should be based on reasons relevant to the individual case.

Yet, the introduction of ML in public decision-making does not merely replicate these risks but also inflates them. Whereas the main issue in profiling is that decisions are made on the basis of assumptions about the "socially salient" features of individuals (e.g., race, gender, sexual orientation), ML enables statistical reasons based on any feature (Lippert-Rasumussen 2014). Given access to data about the social, economic, or physical properties of individuals, ML has unlimited potential to make decisions based on statistical similarities applicable to individual cases.

3.2 Accessibility

The principle of publicity also requires that the reasons that motivate public decisions be accessible. This is not feasible if the ML system that is making a decision or making an assessment that is an important part of a decision is opaque. Yet, not all forms of opacity are necessarily problematic. For example, human decision makers have brains, and these are also opaque to some extent. While human decision makers can explain the reasons on which they based a particular decision, it is not always feasible to know the real reasons that motivated a human decision maker. Moreover, sometimes the reasons for a decision are not possible to explain. For example, when a police officer identifies a person as a suspect from a crime scene, the police officer cannot explain how he/she was able to make the identification other than "I recognize him from the crime scene." Though human decision-making can be opaque, ML-based systems are opaque in at least five ways that are of concern when used as decision support tools in public decision-making. When multiple dimensions of opacity are present, they tend to compound the uncertainty and severely limit the scope for

explainability. Consider, for example, the following case in which human public decision-making is highly opaque:

- The defendant does not know what crime they are being prosecuted for.
- The laws are secret.
- The evidence is secret.
- The judge communicates in highly legalistic jargon.
- The identity of the decision makers is unknown to the defendant.
- The sentence is classified.

Whereas any of these circumstances would be unacceptable on their own, together they offer little recourse for the defendant to understand and challenge the reasons for the decision. Each circumstance adds to the opacity of the process, and the opacity is compounded. With regard to ML systems, there is a similar concern. The multiple dimensions of opacity in these systems make the decisions much murkier than decisions made by ordinary human decision makers would be when reasonable democratic procedures are employed. Unfortunately, the ML used in public decision-making often have multiple reinforcing layers of opacity.

First, the information in these systems is not organized and stored in symbolic representations but in sub-symbolic weights across an artificial neural network that are distributed in a seemingly haphazard pattern, making the code virtually impossible to access and read for a human. Whereas all the code of expert systems can (at least in theory) be inspected and evaluated, this cannot be done with the code of ML. ML can only be inspected and evaluated based on how well they perform on test data or on real-world applications. In other words, ML systems are *observationally opaque*. This is especially problematic when encountered with a situation where an AI performed well with the test dataset but failed when used in real-world applications, a very common problem (D'Amour et al. 2020).

Second, while ML has shown rapid progress and impressive results, very little is still understood about why and how these systems work so well. The epistemic state of the field was compared to "alchemy" by Ali Rahimi, a researcher at Google, in a talk at the Conference on Neural Information Processing, NIPS 2017 (one of the most prestigious conferences in the field) (Hutson 2018). For example, it is not generally known why certain net architectures are better than others for some problems. Moreover, it is generally not possible to know in advance how much training data are required for a certain task and a certain algorithm (Domingos 2012). Thus, ML systems are *theoretically opaque*.

Third, despite this ambiguity in recent progress, the public discourse remains as triumphant as it is lacking in a general understanding of what ML is capable of (Sumpter 2018). Politicians, respected news outlets, and other individuals with great public authority have repeatedly shown to have a very limited understanding of the basics of ML (Burell 2016; Dourish 2016). This lack of understanding makes our public institutions poorly equipped to understand, explain, or predict the behavior of ML. This lack of understanding makes ML systems *sociologically opaque*. The sociological opacity of ML systems risks making the public ill-disposed to assess decisions made by or with the help of algorithms and public officials ill-equipped to monitor and evaluate these decisions.

Fourth, the situation is further aggravated by the unintuitiveness of how ML work (Burell 2016; Dressel and Farid 2018). Human decision-making has the feature of *graceful failure*. That is, when human judgment deteriorates, the decrease in the quality of output is proportional to the severity of the failure, as compared to typical machine learning systems, in which even a small failure can cause total breakdown. This is of particular concern in life-critical systems. Moreover, failure in ML systems is sometimes a result of input that is not evident to a human observer. For example, adding a few pixels (invisible to the human eye) to an image can dramatically change the ML system's ability to identify the object. For this reason, ML systems can be said to be *psychologically opaque*.

Fifth, as Burell (2016) argues, many algorithms are proprietary and thus not available to relevant stakeholders to investigate. This means that even the parts of a decision that could be public, such as the input data, in some cases are deliberately kept secret. Thus, ML are in some cases *legally opaque*.

Thus, ML are opaque in multiple ways. *Observational opacity* makes it difficult to know whether an algorithm succeeded in a given case even if the algorithm has performed well on training data. *Theoretical opacity* makes it impossible to explain how an algorithm reached a decision in every step. *Sociological opacity* makes the public ill-disposed to assess algorithmic decision-making. *Psychological opacity* makes it difficult for humans to predict failures or to intervene before a total breakdown of the system. *Legal opacity* makes it difficult for the public to access information about a given algorithm, the data that were used, and how the algorithm took part in a decision process.

When the problem of opacity is combined with the problem of reason giving, it appears that ML poses a unique challenge to legitimate exercises of public authority. First, as shown in section (a) above, decisions taken by or with the help of a sophisticated algorithm do not usually provide sufficient reasons that apply to the individual case. Because the reasons provided by ML are statistical in nature, they are similar to profiling in the sense of identifying measures on the basis of general patterns rather than on individual facts. Second, because the operations of ML

are observationally, theoretically, sociologically, psychologically, and legally opaque, the statistical reasons upon which decisions are made are hard or even impossible to know and make publicly available. In contrast to profiling, citizens are, in the end, left in the dark about the model of statistical reasons that determine the public decisions to which they are subjected.

4 Discussion

The argument defended here is that ML technology-based decision support in public administration and law poses serious challenges to the democratic legitimacy of public authority. The democratic legitimacy of public authority depends on *serving the ends* of the democratic law maker, decisions *based on reasons* that align public decisions with democratic ends that are *accessible* to the general public. These requirements are not satisfied by decisions determined through ML as the reasons involved are statistical in nature and the operations opaque in several respects. In the final analysis, we argue that two main considerations are pertinent to the normative assessment of the democratic legitimacy of ML in public authority. The first is practical and concerns the extent to which ML is practiced as intended and the extent to which it replaces decisions that were previously accessible and based on reasons. The second is to insist that ML in public authority is embedded in an institutional infrastructure that secures reason giving and accessibility.

The first consideration relates, in line with Dourish (2016), to the practical application of ML in public decisions. The extent to which ML is in fact substituted for human agency depends on the institutional and psychological setting of the decision-making. Observational and sociological opacity may obstruct our ability to monitor, understand, and explain instances where ML is used in public decision-making. Hence, tasks that are not designed to be executed by an automated process may inadvertently end up as an automated process. Humans may fail to intervene and supervise machines that are not sufficiently advanced for the task. For example, the car company Tesla has a self-driving system that has had a few notable accidents. These have often been caused by drivers behaving as if the self-driving system was fully autonomous, such as by watching Netflix while driving.[2] Overestimating algorithmic capacity is of particular concern when algorithms are introduced with the expectation that they will reduce costs of public administration. This ambition can lead to an increased workload for public officials, which may make it impossible for a human decision maker to exercise proper supervision and to make their own decisions and instead default to deferring to the judgement of the algorithm. As such, we need to consider the actual use of a certain software product, something that depends on both the level of technical sophistication, the workplace culture, and other institutional arrangements. Alternatively, the opposite eventuality is also possible. A user, such as a civil servant, may choose to intervene much more than intended for a certain system, as seems to be the case with a profiling tool used by the Swedish Public Employment Service (Bennmarker et al. 2021).

A second aspect of the first consideration is the extent to which ML replaces a step in public decision-making that used to be explained and/or justified. Often, public decisions involve multiple steps, and some of them are not typically explained. For example, for a police officer to write a speeding ticket, the officer needs to identify the driver from his/her driver's license. This step is a simple exercise in pattern recognition and is rarely explained further. To the extent that we find this acceptable, we should also accept when ML do the same thing. If a ML performed the same step with a roadside camera, there would be no further requirement to justify the decision to identify a person. In this case, replacing human decision-making with algorithmic decision-making is no worse in terms of reason giving.

Now, the final test case is whether ML in public authority is or can be democratically legitimate when intentionally introduced to replace decisions that used to be public and based on reasons. Of course, new technologies introduced in public administration and law can be legitimate only if they are effective in promoting the ends determined by the democratic law maker. But that requirement is barely sufficient for democratic legitimacy. Remember, reason giving and publicity are valuable features of public decision-making because they are necessary to estimate whether democratic ends are in fact served. Public authority must not only serve democratic ends but also be accessible in order for subjects to be able to verify that they are treated as democratic equals. From that vantage point, opaque and statistical models for decision-making appear inherently problematic.

Alternatively, democratic legitimacy may not necessarily require either that all decisions be based on reasons or that all reasons be accessible to the public. Instead, we argue that a necessary and sufficient condition for the democratic legitimacy of the use of ML in public authority is that these technologies be *embedded in an institutional infrastructure that is public and based on reasons*. First, the decision to introduce the use of ML at any particular level of public authority must itself be accessible and based on reasons. When that is the case, citizens are able to judge and evaluate the reasons for introducing technologies that are not themselves accessible and based on reasons.

[2] "Tesla driver killed while using autopilot was watching Harry Potter, witness says," *The Guardian,* July 1, 2016.

Second, the use of ML must not be introduced by institutions to which citizens turn for the purpose of appeal. The appeal process is a vital component for structures of public authority to achieve the ideals of rule of law. In order for the rule of law to be secured, the scrutiny of the primary exercises of public authority must be handled by procedures that are accessible to the public and grounded in reasons. This last point is important as it intersects with ongoing debates on the use of algorithms in judicial processes. The point is that in order for the appeal process to serve as a checker for automated decisions in public administration, it must not itself be executed by automated processes. Though this observation is pointing toward a broader set of issues that goes beyond the ambitions of the current paper (see Zalnieriute et al. 2019), it may be that the democratic legitimacy of ML in public authority significantly depends on the extent to which ML is employed by judicial institutions.

Author contributions All authors contributed equally to the conception, design and writing of the study. All authors read and approved the final manuscript. No data were used in preparing the manuscript and thus data availability requirements do not apply.

Funding Open access funding provided by Stockholm University. Jonas Hultin Rosenberg's work on this article was supported by the Marcus Wallenbergs Stiftelse [MMW 2019.0160]. Ludvig Beckman and Karim Jebari did not receive support from any organization for the submitted work. The authors have no relevant financial or non-financial interests to disclose.

References

Agrawal A, Gans J, Goldfarb A (2018) Prediction machines: the simple economics of artificial intelligence. Harvard Business Press, London

Becker S (2004) Assessing the use of profiling in searches by law enforcement personnel. J Crim Just 32(3):183–193

Bennmarker H, Lundin M, Mörtlund T, Sibbmark K, Söderström M, Vikström J (2021) "Krom – erfarenheter från en ny matchningstjänst med fristående leverantörer inom arbetsmarknadspolitiken." IFAU Rapport 2021:7

Beran H (1977) In defense of the consent theory of political obligation and authority. Ethics 87(3):260–271

Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin

Burrell J (2016) How the machine "thinks": understanding opacity in machine learning. Big Data Soc. https://doi.org/10.1177/2053951715622512

Chesterman S (2019) Artificial Intelligence and the Problem of Autonomy. Notre Dame Journal on Emerging Technologies. NUS Law Working Paper No. 2019/016

Christiano T (2004) The authority of democracy. J Polit Philos 12(3):266–290

Christiano T (2008) The constitution of equality: democratic authority and its limits. Oxford UP, Oxford

Council of Europe (2019) Ad Hoc committee on artificial intelligence–CAHAI. https://www.coe.int/en/web/artificial-intelligence/cahai. Accessed 1 Jan 2022

Covington P, Adams J, Sargin E (2016) Deep neural networks for YouTube recommendations. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp 191–198

D'Amour A et al (2020) Underspecification presents challenges for credibility in modern machine learning. ArXiv: 2011.03395

de Fine LK, de Fine LJ (2020) Artificial intelligence, transparency, and public decision-making. AI Soc 35(1–10):917–926. https://doi.org/10.1007/s00146-020-00960-w

Djeffal C (2020) Artificial intelligence and public governance. In: Wischmeyer T, Rademacher T (eds) Regulating artificial intelligence. Springer, Berlin

Domingos P (2012) A few useful things to know about machine learning. Commun ACM 55(10):78–87

Dourish P (2016) Algorithms and their others: algorithmic culture in context. Big Data Soc. https://doi.org/10.1177/2053951716665128

Dressel J, Farid H (2018) The accuracy, fairness, and limits of predicting recidivism. Sci Adv. https://doi.org/10.1126/sciadv.aao5580

Enoch D (2015) Against public reason. In: Sobel D, Vallentyne P, Wall S (eds) Oxford studies in political philosophy: volume 1. Oxford University Press, Oxford

Estlund DM (2008) Democratic authority: a philosophical framework. Princeton, Princeton UP

EU Ethics Guidelines for Trustworthy AI (2019). https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Accessed 1 Jan 2022.

European Commission (2018a) High-level expert group on artificial intelligence: draft ethics guidelines for trustworthy AI. Brussels

European Commission for the Efficiency of Justice (CEPEJ) European Ethical Charter on the Use of Artificial Intelligence (2018b) Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, December 3–4, 2018b)

Feldstein S (2019) The road to unfreedom: How artificial intelligence is reshaping repression. J Democr 30(1):40–52

Florini A (2007) The right to know: transparency for an open world. Columbia UP, New York

Friedrich CJ (1968) Authority, reasons and discretion. Harvard UP, Cambridge

Hood C, Heald D (eds) (2006) Transparency: the key to better governance? Oxford UP, Oxford

Hutson M (2018) Has artificial intelligence become alchemy? Science 360(6388):478

Lippert-Rasmussen K (2014) Born free and equal? A philosophical inquiry into the nature of discrimination. Oxford University Press, Oxford

Mann M, Matzner T (2019) Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination. Big Data Soc. https://doi.org/10.1177/2053951719895805

Marcus G, Davis E (2019) Rebooting AI. Ballantine Books Inc, New York
Nemitz P (2018) Constitutional democracy and technology in the age of artificial intelligence. Philos Trans R Soc A 376:20180089
O'Neil C (2016) Weapons of math destruction: how big data increases inequality and threatens democracy. Crown Publishers, New York
O'Malley P, Smith GJD (2020) "Smart" crime prevention? Digitization and racialized crime control in a Smart City. Theor Criminol. https://doi.org/10.1177/1362480620972703
Parliament E (2020) The ethics of artificial intelligence: Issues and initiatives. Eur Parliam Res Serv PE 634:452
Pennington K (1993) The prince and the law 1200–1600: sovereignty and rights in the Western legal tradition. California University Press, California
Peter F (2017) Political legitimacy. In: Zalta EN (eds) The Stanford encyclopedia of philosophy, summer 2017 edition.
Rawls J (1993) Political liberalism. Columbia University Press, New York
Raz J (2009) The authority of law. Oxford University Press, Oxford
Rudovsky D (2001) Law enforcement by stereotypes and serendipity: racial profiling and stops and searches without. University of Pennsylvania Journal of Constitutional Law Cause 3(1):296–366
Russel S, Norvig N (2020) Artificial intelligence : a modern approach, 4th edn. Pearson, London
Starke C, Lünich M (2020) Artificial intelligence for political decision-making in the European Union: effects on citizens' perceptions of input, throughput, and output legitimacy. Data Policy 2:E16. https://doi.org/10.1017/dap.2020.19
Sumpter D (2018) Outnumbered. From Facebook and Google to fake news and filter-bubbles—The algorithms that control our lives. Bloomsbury Sigma, London
The Montréal Declaration of Responsible AI (2018) Montréal declaration of responsible artificial intelligence 2018. https://recherche.umontreal.ca/english/strategic-initiatives/montreal-declaration-for-a-responsible-ai/. Accessed 20 Dec 2021
Viehoff D (2014) Democratic equality and political authority. Philos Public Aff 42(4):337–375. https://doi.org/10.1111/papa.12036
Waldron J (1987) Theoretical foundations of liberalism. Philos Q 37(147):127
Wall S (2007) Democracy and equality. Philos Q 57(228):416–438
Zalnieriute M, Moses LB, Williams G (2019) The rule of law and automation of government decision-making. Mod Law Rev 82:425–455
Zuboff S (2019) The age of surveillance capitalism: the fight for a human future at the new frontier of power. Profile Books, London