Information Polity -1 (2020) 1–14 1
DOI 10.3233/IP-200224
IOS Press

Soul of a new machine: Self-learning algorithms in public administration

Lasse Gerrits
Institute for Housing and Urban Development Studies, Erasmus University Rotterdam, P.O. Box 1935, 3000 BX Rotterdam, The Netherlands
Tel.: + 31 10 408 9825; E-mail: gerrits@ihs.nl

Abstract. Big data sets in conjunction with self-learning algorithms are becoming increasingly important in public administration. A growing body of literature demonstrates that the use of such technologies poses fundamental questions about the way in which predictions are generated, and the extent to which such predictions may be used in policy making. Complementing other recent works, the goal of this article is to open the machine's black box to understand and critically examine how self-learning algorithms gain agency by transforming raw data into policy recommendations that are then used by policy makers. I identify five major concerns and discuss the implications for policy making.

Keywords: Enmeshment, machine learning, big data, policy networks, policy analysis, agency

Key points for practitioners:

- Highlights fundamental weaknesses in self-learning algorithms that defeat traceability and political and administrative accountability.
- States reasons why the usefulness of self-learning algorithms in public administration is much more limited than is often claimed.
- Combines different strands of literature from various disciplines to provide a more comprehensive picture of how self-learning algorithms combine with public administration.

1. Introduction

The arrival of digital technologies in general, and that of self-learning algorithms in particular, has given rise to much scholarly debate about its implications for the public domain, e.g. with regard to policy analysis, policy making and governance (e.g. Danaher et al., 2017; Gil-Garcia et al., 2018; González-Bailón, 2013; Hilbert, 2016; Janssen et al., 2015; Just & Latzer, 2017; Katsonis & Botros, 2015; Kosorukov, 2017; Landsbergen & Wolken, 2001; Todorut & Tselentis, 2018; Yeung, 2018, and others). Digitization, as an overarching term for various techniques and methods, is often seen as a useful tool that could improve policy analysis and decision making. An oft-repeated argument is that digitization can unlock insights from big data sets at speeds unattainable for human operators. As such, it could simultaneously offer an unprecedented access to information to guide governmental action, and makes sense of a daunting abundance of data about the state of society. It could be used to model and simulate decisions to explore possible outcomes of certain policies (Janssen et al., 2015), or to predict certain societal dynamics (e.g. crime in public transportation, see Kouziokas, 2017) so that authorities can deploy their resources more efficiently. No wonder then, that great things are expected from such techniques.

However, there is a growing body of literature that demonstrates that there are severe and persistent issues with the use of big data and self-learning algorithms in public administration. As will be elaborated later in this paper: such algorithms are partial and subject to self-confirmation, which can have real and sometimes dire consequences (e.g. Bellovin, 2019). For example, software used in Florida to assess the likelihood of sentenced criminals to become repeat offenders only predicted 20% of such occurrences correctly, and it was particularly likely to falsely predict black defendants as future criminals (at a twice higher rate than white defendants), while severely underestimating the likelihood of white defendants to become repeat offenders (Angwin et al., 2016). Naturally, a self-learning machine may be corrected using additional search and classification rules to counter bias (e.g. Caliskan et al., 2017) but the development and successful implementation of such a technique is strongly dependent on many factors. Besides, bias is only one of the major challenges in getting self-learning algorithms to work in public administration. There are also many practical, operational, legal and ethical questions (e.g. Mittelstadt et al., 2016) surrounding those techniques when used in policy analysis and policy making.

There appears quite some optimism, and also some hype, about the promises of digitization and the use of self-learning algorithms in public administration (e.g. Agarwal, 2018; Cervalán, 2018; Keast et al., 2019; Maciejewski, 2017). The many serious concerns that are starting to surface in literature provide ample reasons for a more critical approach (see e.g. Wirtz et al., 2019, 2020). For example, one could reconsider the ways in which administrations are organized, and the ways in which administrators use digital technology (e.g. Giest, 2017; Lindgren et al., 2019; Veale & Brass, 2019), as well as consider new practices for transparent data management and review (e.g. Janssen et al., 2020; König & Wenzelburger, 2020). The logic behind such reconsiderations is that the machine is becoming increasingly enmeshed in the practices of policy analysis and policy making. Consequently, it has a real, tangible effect on the formulation and execution of policies (e.g. Kolkman, 2020; Valle-Cruz et al., 2020). Indeed, any enmeshed technology has transformative capacities (Dolata, 2013), which implies that such a technology may gain agency in that enmeshment with policy making. This point is not novel in itself but is often discussed at a conceptual level. How it works within the machine itself is less articulated.

While there is a growing body of literature on what digitization could mean for policy making – as summarized above – and a vast literature on the technicalities of self-learning algorithms (e.g. Conway & White, 2012; Flach, 2012; Mitchell, 2013), there is scant literature that bridges the two realms (see e.g. Sun & Medaglia, 2019; Valle-Cruz et al., 2020). The goal of this article is to open the machine's black box in order to understand and critically examine how self-learning algorithms gain agency by transforming raw data into policy recommendations that are then used by administrators, and how this impacts policy making. Such a critical reflection is necessary because scholars in public administration should not only get a better understanding the implications of those techniques, they should also gain an understanding of the machine's main operations in order to understand how those techniques work. To this end, the paper follows the process of transforming large data volumes into policy recommendations. In doing this, I draw from various bodies of knowledge, most notably science and technology studies, information theory, and the sociology of technology. Mackenzie's ethnographic studies of digitization and machine learning (most notably Mackenzie, 2015, 2017) are central to this argument. I attempt to link these various literatures to the study of public administration and public policy, in the hope to enhance those literatures with some of the principles of machine learning. In addition, the references may also be used as pointers for public administration scholars who wish to do more in-depth research in this direction.

I start with a brief explanation of what is meant by the enmeshment of the machine in policy making. Since 'digitization' is often used as an umbrella term, it tends to obfuscate various differences between related techniques. It is therefore necessary to discuss three main aspect of digitization, namely big data (as volume as well as diversity), algorithms (to sort, structure and synthesize data) and machine learning (to develop policy recommendations with a large degree of automated autonomy). The final part of the paper is dedicated to an in-depth and critical examination of how these three aspects impact policy making. I argue that these aspects lead to five major concerns that hamper the seamless enmeshment of machine and administration as advocated by certain authors. First, self-learning algorithms require dichotomization of data at various levels to produce output, which means that the machine transforms data before it gives a recommendation. Second, machine output is likely to be biased because of the way the algorithms were trained. Third, machine learning may lead to normalization, as such confirming that the machine was correct even if it has generated biased output. Fourth, machines may learn from data but unlearning it is much more difficult, even though such unlearning is important in generating more accurate production. Fifth, the machine can't print intelligible output by itself. Taken together, these five points imply that humans are poor monitors when using machines in policy making. Machine learning may possibly have something to add to public administration but an uncritical embrace of the technology is not justified.

2. Enmeshment

Public policies are developed within networks of actors (e.g. Wachhaus, 2009). Actors in such networks can be individuals but more often concern collectives as actors such as a Ministry, a government agency or a stakeholder group (Klijn & Koppenjan, 2015). However, digital technologies can also be considered as actors in such governance networks. The technology derives its actor quality not only from its own capabilities but above all from the way it interacts with other actors in the network – in the same way that e.g. a group of stakeholders can gain agency from its interactions with public officials. It is in networks that agency is created.

The idea that agency is a network attribute is articulated most prominently in Actor-Network Theory or ANT (Latour, 1991, 2005; Law, 1992; Venturini et al., 2017). Naturally, ANT resonates strongly in the digital age where social life has become increasingly dependent on all sorts of digital technologies (e.g. Bächle, 2016; Bellanova, 2017; Haque & Mantode, 2013; Schmidgen, 2011; Stanforth, 2007). Thinking in terms of actor-networks that include actors of any type (not just humans) is somewhat underrepresented in policy and governance theories (Ludmilla et al., 2014; O'Brien, 2015). In ANT, all technologies can have actor qualities, even simple technologies such as the hotel room key in Latour's 1991 example. While above I wrote about digitization in broad terms, for the present purpose I focus on self-learning algorithms that are used to generate predictions. Since it encompasses a variety of computational techniques that are combined in order to achieve learning and prediction, I refer to this actor as the 'machine' in the network. Importantly, this goes beyond the use of computers for e.g. registering and keeping track of data (although data management can be part of the machine) or programs made to streamline public service delivery.

The example of predicting whether convicted people may become repeat offenders as given above serves to demonstrate how much agency a machine may achieve. In that particular case, it recommended which convicts could be considered eligible for parole, a recommendation that was usually followed-up. Thus, the machine developed agency. Another example concerns the predictions of livestock disease outbreaks (Kroschewski et al., 2006). Here, the machine calculates the likelihood of such diseases spreading from farm to farm. Using input such as contagiousness of the disease, density of the area etc., it recommends different scenarios. For example, it may recommend quarantining a farm or an entire region, or even the destruction of all animals in the infected area. As with the example of repeated crime prediction, the machine gains considerable agency if its recommendations are acted upon. This is a recurring theme when it comes to the role of the machine in policy making, even though that role is still poorly understood (Janssen et al., 2020; Lindgren et al., 2019; Valle-Cruz et al., 2020).

The machine is an assemblage of hardware and software, of external input and self-generated learning mechanisms, of predefined schemes for structuring data and autonomous generation of recommendations. It is a set of different technologies that combine with human input and subsequent action to generate a certain outcome, such as a decision to turn down a request for probation or the decision to clear a farm of live-stock. This agency is characterized by *invisibility* (human operators at least partially unaware of how data is processed and recommendations are made) and *impact* (the machine's output has an actual outcome on the real world). Although I won't deploy the entire apparatus of ANT in this paper, I use the main idea as a search light to discuss how the machine operates and interacts with human operators to bring about policies. I focus on self-learning algorithms as arguably the most far-reaching role a machine can obtain in policy making. In literature, various terms such as machine learning, artificial intelligence and automation are often used interchangeably (see e.g. Etscheid, 2019) so there is a need to first clarify the main techniques and concepts, and to map how they relate (see also boyd & Crawford, 2012; Manovich, 2012). These are: big data, algorithms and machine learning. Of the three, big data may be the most loosely defined.

3. Big data

Big data concerns data that is not only characterized by its volume but above all by its unsorted type diversity as well as granular diversity. While conventional policy analysis works with theoretical frames and sets of assumptions about relationships between variables to collect and categorize key data such as census data or data for certain socio-economic key variables, big data sets are principally unsorted and lacking in predefined structuring (Manovich, 2012). In such large, diverse and unstructured data sets, each utterance is considered (potentially) valuable data and each piece of information forms a variable, regardless of its form (Mackenzie, 2015). The key to working with this daunting abundance of information is categorization, i.e. the sorting and labelling of every piece of data such that those pieces can become related through statistical operations. Since every piece of data becomes a variable, the entire data set forms a very-high dimensional space where countless pieces of data are related to other countless pieces of data, i.e. many vectors are formed within this space (Mackenzie, 2017). The question which variables are to be related to others is determined in the statistics used (Hastie et al., 2009). Compare this with more conventional approaches to policy analysis, in which the data space is predefined by a limited set of variables and their relationships, usually in the shape of correlation. In big data, the properties of the vector space may, and usually does, change because of the sorting and labelling that takes place Big data spaces contain all contextual, indexical, symbolic or lived differences in data (Mackenzie, 2015). By implication, the data set is dynamic. That is: new data can enter or leave the space continuously. The new data is not merged with a given pre-defined causal structure. Instead, it may change the causal structure if the new data or the discarding of old data provide reasons for doing that. As such, new data becomes part of the vector space and the causal structure may change continuously.

It is important to note that the diversity of the data in big data sets doesn't only stem from the nature of the input data and the continuous flow of new data (and discarding of data considered no longer relevant in the light of what has been 'learnt') but also from the juxtaposition of entire but seemingly incompatible data sets. The 'remixing' (Mackenzie, 2017) of different types of data and of various collections or sets of data is one of the key aspects of big data, as is the transformation of entire data sets when combined with other sets in various ways (Mackenzie, 2012).

The data itself is also transformed upon entering the vector space. The machine operates by the grace of digitization. This requires data to become encoded into bits. The encoding essential for the operation of the machine but it also causes dichotomization of the data at the micro level (Mackenzie, 2017). Data that may appear as gradients to the naked eye needs to be cut up into discrete values before it can be put to work. Naturally, dichotomization of data itself is nothing new. Academic and policy researchers do it all the time whenever they decide that a certain observation falls into one category or another one. The same goes for policy makers when they try narrowing the complexity of real-world issues into categories that can be processed in bureaucracies, even if it is understood that such simplification violates the actual complexity of those issues (Boisot, 2004, 2006; Boisot & Child, 1988; Gerrits, 2012). The difference between such instances on the one hand, and dichotomization of big data on the other, is that the dichotomization of *all* data is a necessary step before it can be processed to form a vector space in the latter.

The dichotomization of data takes place at various level. At the micro level, it is encoded in bits. Above that level, each piece of data is classified into (emerging) categories (Mackenzie, 2017). The exact consequences of this dichotomization for encoding and vectorization are hard or perhaps even impossible to assess on a case by case basis. This is not much of an issue in instances of discrete data but becomes more pressing when the data is ambiguous and open to multiple interpretations. Data that is not easily classified is still forced into a category, or it may become a new category on its own. Either way, the ambiguities that are real to social data are hard to deal with in big data sets.

Exactly how data as variables in the vector space relate will emerge once sufficient data have been collected and labelled – which is why such data sets tend to be enormous. Naturally, it is considered impossible to sort those data manually and to discern patterns that matter. This where algorithms and machine learning come into play.

4. Algorithms

If data forms into vector spaces, algorithms can be said to inhabit those spaces (Mackenzie, 2017). In their very basic form, algorithms are nothing but if-then rules applied to the data in order to form the vector space by classifying, structuring and relating said data (Cormen et al., 2007), e.g. the rule that if a piece of new data appears similar (in properties) to a piece of data that is already classified, the new data will be classified in the same way. Machines may feature many of such algorithms. They can be complicated, and they can be combined at will. Many decisions in conventional policy analysis can be considered algorithmic, too, for example when all instances having a certain set of attributes are considered to fall under the scope of a particular rule. However, there are some important differences, in particular when it comes to the number and diversity of algorithms that can be combined, and the speed with which the data can be processed.

When it comes to the algorithms deployed in big data sets, a principal distinction can be made between reactive systems, i.e. algorithms that trigger an automated response; and pre-emptive systems, i.e. algorithms that utilize historic data to infer predictions about future behaviour (Yeung, 2018). An example of the first would be a speed camera monitoring car drivers on a road. Once someone drives faster than the pre-set limit, it will register that driver as an offender. An example of the algorithms that generate predictions – which is what I'm after in this article – would be an algorithm that sorts through (seemingly) unrelated data to establish vectors in order to predict an outcome. An example of this is China's 'Situation-Aware Public Security Evaluation' (SAPE) platform that is developed for the prediction of terrorist attacks (Wu et al., 2016). This machine combines different data from different sources, including (but not limited to) money transfers that appear irregular in size and in sender-receiver patterns, and overseas calls by citizens with no relatives outside of China (Gallagher, 2016). This data is collected on a daily basis and the results are compared to similar patterns in such data preceding terrorist attacks, as registered in the Global Terrorism Database.[1] The outcomes are tailored with the help of data from over 10.000 'public security events' as registered by Chinese provinces, in order to account for regional differences. In fitting the curve to the data, authorities may be able to predict that people with certain characteristics are be more likely to engage in acts of terrorism.

The initial sorting of the data can be done manually in order to provide the machine with an anchoring point about what constitutes a fit. This is called supervised learning. A sample from an existing data set may be assessed, sorted and categorized by human operators, giving the machine some basis for the accurate processing of the rest of the data. The remaining sorting, categorization and relating of data is done by algorithms that become more able as more data is processed and checked against what has happened in the real world. A simple algorithm can be told to label all instances of a particular word in communications as a possible indicator for social security fraud, and another one to check if those words correspond with actual fraud as detected in the real world. The data can be matched to pre-defined data and the outcomes can be checked and adjusted in the light of known outcomes. Over time, the algorithms can be made to learn that certain instances in the data, and the way they relate, also co-occur with given outcomes. As such, there is not necessarily a need for continuous human oversight. If algorithms are capable of going through this entire process from sorting to predicting all by themselves, this is called machine learning.

5. Machine learning

Machine learning enables the machine to develop categories and labels for data all by itself, as such actively sorting and relating data without much prior instruction as how this should be done exactly. In other words, the machine will try out in what ways the best fit with real world outcomes can be created. This is called unsupervised learning. The basic principle of (unsupervised) machine learning constitutes a positive feedback loop. The data are labelled and related, and the outcomes are then tested to see if the sorting and structuring have indeed generated the correct prediction. If not, the data will be related repeatedly until its output starts to approach known reality, i.e. fit has been reached. Once the resulting predictions are confirmed, the machine will be better able to sort new incoming data and entire data sets. In other words: the more a machine knows, the more it can know, i.e. generalization through mobilization (Mackenzie, 2015). The inclusion of additional data may improve the capacity of the machine to learn and to get better at sorting data and predicting outcomes. An evolutionary approach sees the machine pitching alternative, competing algorithms that label and sort the data in different ways and check their predictions against outcomes. The algorithm or combinations of algorithms approximating known reality the best will be kept and the other ones discarded (Salcedo-Sanz et al., 2014). Not only will this enable the machine to make better predictions, it will also become increasingly more efficient at making such predictions. It actively selects and shapes the algorithms that work the best, i.e. it is capable of enhancing its own learning capacities.

[1] https://www.start.umd.edu/gtd/.

While this certainly looks impressive, there is no black magic involved in this process. Machine learning runs on a collection of known statistical techniques to do the labelling and sorting (Hastie et al., 2009). The apparent magic derives from the speed with which these enormous amounts of data are labelled, sorted, tested, and resorted and relabelled until they produce output that starts appears to get closer to reality. The important point is that it is impossible for human operators to track and trace how the machine traversed the highly-dimensional vector space in order to come up with a given output (Latour et al., 2012; Mackenzie, 2015; Mittelstadt et al., 2016). That is, the self-selection of algorithms on the basis of the machine's learning curve is invisible to the human operator. In that sense, the machine is indeed a black box, the capacities of which are to be assessed by its output, i.e. its capacity to predict, but not necessarily by the way it achieves its predictive capacity.

The best way of telling that the machine has learned is by looking at its ability to generalize (Burrell, 2016). There are two issues with this generalization. First, the resulting model may adapt itself too closely to the current data set and subsequently fail to generalize (excessive fit), or may not be complex enough, subsequently representing too little and performing poorly in generalization (underfit). Again, this is not dissimilar from what administrators also do when they try to match real world issues to predefined bureaucratic categories (Boisot, 1998) but the speed at which it happens is unmatched, and the impromptu flexibility imposed on existing categories is virtually non-existent in bureaucracies. Second, the learning works well as long as the object it is learning about remains more or less static. A static object allows the machine to fine-tune its model and to become increasingly good at making predictions. However, every change in the object of interest requires a new iteration and possibly a change of the predictive model. By implication, machine learning has a hard time keeping up with the complexity of social reality (Mackenzie, 2017). Naturally, this also goes for humans (Ang, 2011). The difference here is that machines can iterate at a much higher rate than humans can do. Regardless, Mackenzie and others are correct in saying we are still far off self-learning algorithms that respond adequately to the fluidity of human complexity.

6. Transformations

Following the process from the processing of raw data to policy recommendations, it appears that the data is transformed in many and profound ways before it reaches the policy maker's desk. These transformations are non-trivial in the sense that they alter the lived experience into analytical and bifurcating units, which is not a difference in degree but a difference in kind (Savage, 2009). Data is dichotomized at the micro-level and vectorized in as many ways as is necessary in order to produce an output that appears sensible. The inclusion of different types of data in one data set (e.g. quantitative gradients vs. qualitative gradients, categorical vs. ordinal, etc.) requires transformations before these data can vectorized, can be made to relate. Aggregated and transformed data are restructured in all different types of arrays or schemata such as dendrograms, trees, scatter plots and NK-models. Every traversing of the (dichotomized) data, every production of curves approximating fit involves a transformation. As such, there is a considerable difference between the real world and the representation thereof as generated by the machine – even if its models and predictions appear to resemble social reality. It is this altered reality that is used for guidance in their policy making. Naturally, the machine may learn from its own mistakes by developing and then selecting competing algorithms for their best performance. But even that can be considered a process related only indirectly to human operations. There is no exaggeration in saying that there is considerable autonomy in, and subsequent agency of, the machine (Stampfl, 2013). At the same time, however, machine learning would be a dead artifact were it not for the ways in which it is deployed by human operators in general (e.g. Markham, 2013; Matzner, 2019; following ANT), and by administrators in particular (e.g. Bellanova, 2017).

Admittedly, the discussion above only scratches the surface of how the machine operates in generating predictions on the basis of self-learning algorithms because I can only cover a tiny fraction of the techniques used. However, the main point is that transformations are real. The remainder of the article continues to follow the process and examines how that transformation links to practices in public administration, at which point the machine attains full agency.

7. The machine enmeshed in public administration

The use of computers in policy making goes a long way back. Early versions saw computers computing the input given by human operators, for instance to assess the possible effects of a certain policy measure given known facts as collected and structured by those operators (e.g. Kaufmann, 1968). Although becoming increasingly advanced over time, these models can be considered conventional in that the input is pre-structured in sets of variables and the relationships between them on the basis of the operators' prior knowledge of the subject matter. In those instances, the computations are essentially passive. That is: the models produce outputs in exactly the same was they were told to produce. Algorithms are in place – otherwise there would be nothing to compute with – but they are not self-learning algorithms so it is not about machine learning and big data sets. The juxtaposition of those two make the difference between a machine that produces a complicated, but essentially traceable output, and a machine that produces outputs no longer (directly) traceable for human operators. In fact, machine learning also means that the type of output generated may change over time as new data enters the vector space.

Machines of the latter kind are becoming increasingly popular in administration and policy making, and the number of applications seem to grow year by year (see Yeung, 2018; for an overview). One could argue that any policy recommendation that is enacted by policy makers allows the machine to gain agency because it has an impact on the real world. Establishing this fact is a first step. In the following, I will examine that enmeshment more critically. There are five main critical concerns that come with this enmeshment.

First, the data transformations as discussed above mean that, contrary what is often believed, the machine is not generating a true representation. The dichotomization of data when it is classified and clustered can be a clumsy affair. For example, Ku and Leroy (2014) demonstrated that a human expert could be more accurate than a machine that was trained to generate automated classifications of anonymous crime reports. The machine struggled to see differences in between two types of crimes if the reports were highly similar in other aspects, which is that kind of ambiguity that a human expert (a crime analyst in this case) has no trouble in dealing with (Ku & Leroy, 2014). While the machine was faster, the expert was more precise.

Second, the machine is prone to bias (Kolkman, 2020). This can happen in both supervised learning, if the trainers confirm the bias knowingly or unknowingly, and in unsupervised learning. Prominent and pressing examples can be found in predictive policing. For example, Ferguson (2017) showed how an operator can distort the output of the machine if it is told to correlate poor neighbourhoods with crime rates. While certain neighbourhoods may co-occur with high crime rates, the actual dynamics that produce the crime rates remain invisible. All that the machine achieves is to make it seem as if the people living in that neighbourhood are more likely to commit crimes. This may also happen in unsupervised learning. The nature of self-learning algorithms is such that they need historical data to develop an explanatory model of the subject matter, thus confirming existing biases more than discovering new causal relationships. If the machine recommends policy to patrol a certain area more heavily – therefore increasing the likelihood that a larger portion of all people arrested are from that area – the machine will train itself that it has made the correct prediction even if it is a heavily biased prediction. After all, self-learning algorithms are seeking increased fit and not a particular policy outcome, such as a just policing system.

Third, and following from the previous point, the machine's predictions may lead to normalization of a situation because humans act upon the recommendations (Coglianese & Lehr, 2017). As such, there may be a convergence between human behaviour and machine-generated predictions (Mackenzie, 2015). While the machine itself runs on a feedback loop between the computation of predictions and the matching of those predictions to reality (establishing fit), there is a second feedback loop that runs between the generated recommendations and the conforming behaviour of humans. The predictive policing example given above illustrates this. Ultimately, all machine learning is geared towards ordering, transforming, and shaping unstructured data in such a way that it can detect patterns that would neither be visible to the naked eye nor accessible through conventional statistical methods used in isolation with more limited data sets (Mackenzie, 2015). Some of the obvious errors can be corrected (e.g. prohibiting the machine to use the label 'ethnicity' when traversing crime statistics), provided that the human operator can be vigilant enough. The keyword, then, is traceability (alternatively: followability; being intelligible). One can, and should, ask how machines arrive at their recommendation (Coglianese & Lehr, 2017) but this may be extremely complicated and in many cases impossible. The weak spot may not rest with the machine itself – it just does what it can – but in how humans interact with machines (Gross, 2013, 2015; Mcsherry, 2005; Pu & Chen, 2007). Even if the machine could share the reasons for its recommendation, there is no guarantee that human operators would be able to understand the reasons given.

Fourth, while much attention is given to how the machine can learn, 'unlearning' is considerably less developed (Bourtoule et al., 2019). The machine's algorithms are trained on existing data sets. As mentioned before, these data bases may be dynamical with new data added continuously. But while older data can be discarded, the machine cannot stop knowing what it has gathered from those older data. That is, the older data may remain present in the shape of how the algorithms are trained even when the original data on which the machine was trained is no longer present. Among others, it implies that a request to pull data (e.g. under the General Data Protection Regulation (EU) 2016/697; or GDPR) does not mean that the machine has forgotten what the original data meant. This can be dealt with in various ways, ranging from discarding the machine's algorithms and retraining them from scratch using new data, to marking data such that one can determine the ways in which algorithms were affected by that particular piece of data. All of those options are inconvenient and requiring considerable work. For example, marking data requires the operator to understand the importance of each data point in constructing the final model, which is a tall order in big data sets (Bourtoule et al., 2019).

Fifth, policy makers essentially deal with machines that do not know how to print an intelligible, followable output suitable for the human operator requiring that information (Norman, 1989). This is already an issue when the machine works with a crisp database (Beierle et al., 2003; Clancey, 1983; Puppe et al., 2013) but becomes even more complicated when the database is ambiguous and the information needs not clearly defined a priori (Mast et al., 2016), and the dichotomization is applied autonomously – as is the case with the machines described above. On top of that, the ex-post explanation is still an aggregate of various algorithms are human operators are unlikely to observe the machine working through each bit of data. There is ample evidence that humans perform poorly in the role of monitor. Getting the machine enmeshed has the advantage of analysing heaps of unstructured data that cannot be processed by humans alone. The disadvantage is that it induces passivity because humans will no longer actively be involved in structuring data and creating outputs. Such passivity impacts awareness to such an extent that humans may not be comprehend the output even if was produced in a comprehensible way (Dixon & Wickens, 2006; Endsley, 1995, 1996). Moreover, information is irretrievably lost if no initial attention is paid (Peterson, 1985) and humans have struggle to process complex information, regardless of how it is produced and presented (Gerrits, 2012).

The enmeshment of machines and human operators stem from the interaction between the both, where machines build on prior human knowledge, which then leads to real-world consequence, and, subsequently, more 'learning' on behalf of the machine. The five main concerns highlight how that enmeshment can also create warped realities in public administration. This means that an uncritical embrace of the machine in public administration is not warranted.

8. Conclusions: The machine and its administration

While the machine has already gained agency in policy making because of its autonomy in developing recommendations from unsorted data, we are still a long way off building a seamless mesh of humans and machines (Pantic et al., 2006). Some authors (e.g. Coglianese & Lehr, 2017) have argued that legal authority and accountability still rests with humans as they are the ones that make the actual decisions following recommendations by the machine. As such, the decisions are to be subjected to the usual requirements for sound decision making, including due diligence. From that perspective, the machine doesn't change existing questions about transparency and accountability, it only adds a novel technical layer (König & Wenzelburger, 2020). But that is only half of the story. I argue that such legal aspects are difficult to uphold if the administrator can't retrace the operations that lead to the machine's recommendation. Certainly, an administrator can choose to follow or ignore a recommendation but won't be able to state the reasons of ignoring or following it, outside of one's own consideration regarding justice, fairness, representativeness etc. because the machine's workings remain a black box. In other words, the fact that there are laws that apply to the administrator in order to maintain accountability doesn't solve the problematic deus ex machina. As such, I fully agree with Valle-Cruz (2020) that administrations should not embrace machine learning uncritically.

While data can certainly make sense of itself, as per Anderson's famous opinion piece (2008), this not necessarily produce sensible and traceable outputs that could be used for policy making. A machine to structure heaps of data into patterns by means of statistics but such patterns may not reveal actual truths, let alone present something that policy makers can follow blindly. Data volume does not equate objectivity, complete datasets are as difficult to deal with as are incomplete ones, and vectors in the data are not necessarily insights (Mackenzie, 2017). Besides, contexts remain as important as ever (boyd & Crawford, 2012; Margetts, 1991). Last but not least, all machine learning needs to negotiate the same gap between intension (the attributes an object must feature in order to fit a concept) and extension (the class of objects referred to) that any policy research encounters. That is: more abstract concepts generate generic statements that apply to many cases but without specific detail, and more precise concepts generate more precise outcomes that cover fewer instances (Boisot, 1998; Toshkov, 2016). Consequently, its reliance on sheer volume may render the machine less competent in the face of complex reality.

Contemporary big data and machine learning repertoires can be useful in digging up patterns as long as the transformations that take place inside of the machine are understood. Ultimately, those transformations are simplifications of complex, real-world problems. The most pressing problem lies in the machine's opacity – that derives from the self-learning and self-selecting of unsorted and highly diverse data – when rendering a recommendation that leads to a decision that leads to a material change. It can't be expected from administrators to first develop deep knowledge about machine learning in order to work with the recommendations that the machine has generated. Likewise, it can't be expected that the machine will function as an accountable (and perhaps even better) partner in administrative decision making (Hofstetter, 2014). Of course, the machine will become rather more enmeshed than less, which implies that the challenges identified here are likely to become more prominent in the future. The positive feedback loop central to machine learning and the normalization of situations when machine learning is enacted in actual policy analysis and policy making means that scholars are looking at a reality partially generated by the machine itself. With the machine in the loop, reality becomes recursive. Administrations better be prepared for this.

Acknowledgments

I'd like to thank Sebastian Hemesath (University of Oldenburg) for the initial literature search, as well as the students at the University of Bamberg who joined my course 'Governance in the Digital Age'. The discussions we had there have informed the argument in this paper. I would also like to thank Prof. Dr. Tom Gross and Prof. Dr. Diedrich Wolter of the Department of Applied Informatics (University of Bamberg) for their inspiring conversations.

Conflict of interest

No financial interest or benefit has arisen from the direct applications of this research.

Funding

None.

References

Agarwal, P.K. (2018). Public administration challenges in the world of AI and bots. *Public Administration Review*, *78*(6), 917-921. doi: 10.1111/puar.12979.

Anderson, C. (2008, June 23). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. https://www.wired.com/2008/06/pb-theory/.

Ang, I. (2011). Navigating complexity: from cultural critique to cultural intelligence. *Continuum*, *25*(6), 779-794. doi: 10.1080/10304312.2011.617873.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine Bias* [Text/html]. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Bächle, T.C. (2016). *Digitales Wissen, Daten und Überwachung zur Einführung*. Junius Verlag.

Beierle, C., Kern-Isberner, G., Bibel, W., & Kruse, R. (2003). *Methoden wissensbasierter Systeme: Grundlagen, Algorithmen, Anwendungen* (2., überarb. u. erw. Aufl. 2003). Vieweg+Teubner Verlag.

Bellanova, R. (2017). Digital, politics, and algorithms: governing digital data through the lens of data protection. *European Journal of Social Theory*, *20*(3), 329-347. doi: 10.1177/1368431016679167.

Bellovin, S. (2019, January 24). *Yes, "algorithms" can be biased. Here's why*. https://arstechnica.com/tech-policy/2019/01/yes-algorithms-can-be-biased-heres-why/?comments=1&post=36727351.

Boisot, M. (1998). *Knowledge Assets*. Oxford University Press. https://www.beck-shop.de/boisot-knowledge-assets/product/694829?campaign=webcode&utm_source=offline&utm_medium=webcode&utm_campaign=print-stuff&pac=weco_sem.

Boisot, M. (2004). *Exploring the information space* (No. WP04-003; 1-26). Sol Snider Center for Entrepreneurial Research (The Wharton School, University of Pennsylvania).

Boisot, M. (2006). Moving to the edge of chaos: bureaucracy, IT and the challenge of complexity. *Journal of Information Technology*, *21*(4), 239-248. doi: 10.1057/palgrave.jit.2000079.

Boisot, M., & Child, J. (1988). The iron law of fiefs: bureaucratic failure and the problem of governance in the Chinese economic reforms. *Administrative Science Quarterly*, *33*(4), 507. doi: 10.2307/2392641.

Bourtoule, L., Chandrasekaran, V., Choquette-Choo, C., Jia, H., Travers, A., Zhang, B., Lie, D., & Papernot, N. (2019). Machine Unlearning. *ArXiv:1912.03817 [Cs]*. http://arxiv.org/abs/1912.03817.

boyd, danah, & Crawford, K. (2012). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662-679. doi: 10.1080/1369118X.2012.678878.

Burrell, J. (2016). How the machine 'thinks': understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 205395171562251. doi: 10.1177/2053951715622512.

Caliskan, A., Bryson, J.J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183-186. doi: 10.1126/science.aal4230.

Clancey, W.J. (1983). The epistemology of a rule-based expert system – a framework for explanation. *Artificial Intelligence*, *20*(3), 215-251. doi: 10.1016/0004-3702(83)90008-5.

Coglianese, C., & Lehr, D. (2017). Regulating by robot: administrative decision making in the machine-learning era. *The Georgetown Law Journal*, *105*, 78.

Conway, D., & White, J.M. (2012). *Machine Learning for Hackers*. O'Reilly & Associates Inc.

Cormen, T.H., Leiserson, C.E., Rivest, R., Stein, C., & Molitor, P. (2007). *Algorithmen – Eine Einführung* (2., korr. A.). Oldenbourg Wissenschaftsverlag.

Corvalán, J.G. (2018). Digital and intelligent public administration: transformations in the era of artificial intelligence. *A&C – Revista de Direito Administrativo & Constitucional*, *18*(71), 55-87. doi: 10.21056/aec.v18i71.857.

Danaher, J., Hogan, M.J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S.-M., Morison, J., Murphy, M.H., O'Brolchain, N., Schafer, B., & Shankar, K. (2017). Algorithmic governance: developing a research agenda through the power of collective intelligence. *Big Data & Society*, *4*(2), 205395171772655. doi: 10.1177/2053951717726554.

Dixon, S.R., & Wickens, C.D. (2006). Automation reliability in unmanned aerial vehicle control: a reliance-compliance model of automation dependence in high workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(3), 474-486. doi: 10.1518/001872006778606822.

Dolata, U. (2013). *The Transformative Capacity of New Technologies. A theory of sociotechnical change*. Routledge.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32-64. doi: 10.1518/001872095779049543.

Endsley, M.R. (1996). Automation and situation awareness. In *Automation and Human Performance: Theory and Applications*, pp. 163-181. Lawrence Erlbaum Associates, Inc.

Etscheid, J. (2019). Artificial Intelligence in Public Administration. In Lindgren, I., Janssen, M., Lee, H., Polini, A., Rodríguez Bolígar, M.P., Scholl, H.J., & Tambouris, E., (Eds.), *Electronic Government*, pp. 248-261. Springer International Publishing. doi: 10.1007/978-3-030-27325-5_19.

Ferguson, A. (2017, December 29). Is "Big Data" racist? Why policing by data isn't necessarily objective. *Ars Technica*. https://arstechnica.com/tech-policy/2017/12/is-big-data-racist-why-policing-by-data-isnt-necessarily-objective/.

Flach, P. (2012). *Machine Learning: The Art and Science of Algorithms that make Sense of Data*. Cambridge University Press.

Gallagher, S. (2016). *China is building a big data platform for "precrime"*. https://arstechnica.com/information-technology/2016/03/china-is-building-a-big-data-plaform-for-precrime/.

Gerrits, L. (2012). *Punching clouds: An introduction to the complexity of public decision-making*. Emergent Publications.

Giest, S. (2017). Big data for policymaking: fad or fasttrack? *Policy Sciences*, *50*(3), 367-382. doi: 10.1007/s11077-017-9293-1.

Gil-Garcia, J.R., Dawes, S.S., & Pardo, T.A. (2018). Digital government and public management research: finding the crossroads. *Public Management Review*, *20*(5), 633-646. doi: 10.1080/14719037.2017.1327181.

González-Bailón, S. (2013). Social science in the era of big data: social science in the era of big data. *Policy & Internet*, *5*(2), 147-160. doi: 10.1002/1944-2866.POI328.

Greenwald, A.G. (2017). An AI stereotype catcher. *Science*, *356*(6334), 133-134. doi: 10.1126/science.aan0649.

Gross, T. (2013). Supporting effortless coordination: 25 years of awareness research. *Computer Supported Cooperative Work (CSCW)*, *22*(4–6), 425-474. doi: 10.1007/s10606-013-9190-x.

Gross, T. (2015). Supporting informed negotiation processes in group recommender systems. *Icom*, *14*(1). doi: 10.1515/icom-2015-0008.

Haque, A., & Mantode, K.L. (2013). Governance in the Technology Era: Implications of Actor Network Theory for Social Empowerment in South Asia. In Dwivedi, Y.K., Henriksen, H.Z., Wastell, D., & De' R., (Eds.), *Grand Successes and Failures in IT. Public and Private Sectors*, Vol. 402, pp. 375-390. Springer Berlin Heidelberg. doi: 10.1007/978-3-642-38862-0_23.

Hastie, T., Tibrshirani, R., & Friedman, J.H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hilbert, M. (2016). Big data for development: a review of promises and challenges. *Development Policy Review*, *34*(1), 135-174. doi: 10.1111/dpr.12142.

Hofstetter, Y. (2014). *Sie wissen alles: Wie intelligente Maschinen in unser Leben eindringen und warum wir für unsere Freiheit kämpfen müssen* (Originalausgabe). C. Bertelsmann Verlag.

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. doi: 10.1016/j.giq.2020.101493.

Janssen, M., Wimmer, M.A., & Deljoo, A. (2015). *Policy Practice and Digital Science: Integrating Complex Systems, Social Simulation and Public Administration in Policy Research*. Springer.

Just, N., & Latzer, M. (2017). Governance by algorithms: reality construction by algorithmic selection on the internet. *Media, Culture & Society*, *39*(2), 238-258. doi: 10.1177/0163443716643157.

Katsonis, M., & Botros, A. (2015). Digital government: a primer and professional perspectives. *Australian Journal of Public Administration*, *74*(1), 42-52. doi: 10.1111/1467-8500.12144.

Kaufmann, A. (1968). *The Science of Decision-Making*. World University Library.

Keast, R., Koliba, C.J., & Voets, J. (2019). *Expanded Research Pathways, Emerging Methodological Opportunities and Responsibilities*. International Research Society for Public Management, Wellington.

Klijn, E.H., & Koppenjan, J. (2015). *Governance Networks in the Public Sector*. Routledge.

Kolkman, D. (2020). The usefulness of algorithmic models in policy making. *Government Information Quarterly*, *37*(3), 101488. doi: 10.1016/j.giq.2020.101488.

König, P.D., & Wenzelburger, G. (2020). Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics. *Government Information Quarterly*, *37*(3), 101489. doi: 10.1016/j.giq.2020.101489.

Kosorukov, A.A. (2017). *Digital Government Model: Theory and practice of modern public administration*. *20*(3), 10.

Kouziokas, G. (2017). The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment. *Transportation Research Procedia*, *24*, 467-473. doi: 10.1016/j.trpro.2017.05.083.

Kroschewski, K., Kramer, M., Micklich, A., Staubach, C., Carmanns, R., & Conraths, F.J. (2006). Animal disease outbreak control: the use of crisis management tools: -EN- utilización de herramientas de gestión de crisis para luchar contra brotes zoosanitarios -FR- Le contrôle des foyers de maladies animales: utilisation des outils de gestion de crise -ES-. *Revue Scientifique et Technique de l'OIE*, *25*(1), 211-221. doi: 10.20506/rst.25.1.1657.

Ku, C.-H., & Leroy, G. (2014). A decision support system: automated crime report analysis and classification for e-government. *Government Information Quarterly*, *31*(4), 534-544. doi: 10.1016/j.giq.2014.08.003.

Landsbergen, D., & Wolken, G. (2001). Realizing the promise: government information systems and the fourth generation of information technology. *Public Administration Review*, *61*(2), 206-220. doi: 10.1111/0033-3352.00023.

Latour, B. (1991). Technology is society made durable. In Law, J. (Ed.), *A Sociology of Monsters: Essays on Power, Technology, and Domination*, p. 273. Routledge.

Latour, B. (2005). *Reassembling the social: An introduction to actornetwork-theory*. Oxford University Press.

Latour, Bruno, Jensen, P., Venturini, T., Grauwin, S., & Boullier, D. (2012). 'The whole is always smaller than its parts' – a digital test of Gabriel Tardes' monads. *The British Journal of Sociology*, *63*(4), 590-615. doi: 10.1111/j.1468-4446.2012.01428.x.

Law, J. (1992). Notes on the theory of the actor-network: ordering, strategy, and heterogeneity. *Systems Practice*, *5*(4), 379-393. doi: 10.1007/BF01059830.

Lindgren, I., Madsen, C.Ø.l., Hofmann, S., & Melin, U. (2019). Close encounters of the digital kind: a research agenda for the digitalization of public services. *Government Information Quarterly*, *36*(3), 427-436. doi: 10.1016/j.giq.2019.03.002.

Ludmilla, M., Montenegro, & Bulgacov, S. (2014). Reflections on actor-network theory, governance networks, and strategic outcomes. *BAR – Brazilian Administration Review*, *11*. doi: 10.1590/S1807-76922014000100007.

Maciejewski, M. (2017). To do more, better, faster and more cheaply: using big data in public administration. *International Review of Administrative Sciences*, *83*(1_suppl), 120-135. doi: 10.1177/0020852316640058.

Mackenzie, A. (2012). More parts than elements: how databases multiply. *Environment and Planning D: Society and Space*, *30*(2), 335-350. doi: 10.1068/d6710.

Mackenzie, A. (2015). The production of prediction: what does machine learning want? *European Journal of Cultural Studies*, *18*(4–5), 429-445. doi: 10.1177/1367549415577384.

Mackenzie, A. (2017). *Machine Learners*. MIT Press. https://mitpress.mit.edu/books/machine-learners.

Manovich, L. (2012). Trending: The Promises and the Challenges of Big Social Data. In Gold, M.K., (Ed.), *Debates in the Digital Humanities*, pp. 460-475. University of Minnesota Press. doi: 10.5749/minnesota/9780816677948.003.0047.

Margetts, H. (1991). The computerization of social security: the way forward or a step backwards? *Public Administration*, *69*(3), 325-343.

Markham, A.N. (2013). The Algorithmic Self: Layered Accounts of Life and Identity in the 21st Century. *Selected Papers of Internet Research*, 5.

Mast, V., Falomir, Z., & Wolter, D. (2016). Probabilistic reference and grounding with PRAGR for dialogues with robots. *Journal of Experimental & Theoretical Artificial Intelligence*, *28*(5), 889-911. doi: 10.1080/0952813X.2016.1154611.

Matzner, T. (2019). The human is dead – long live the algorithm! Human-algorithmic ensembles and liberal subjectivity. *Theory, Culture & Society*, *36*(2), 123-144. doi: 10.1177/0263276418818877.

Mcsherry, D. (2005). Explanation in recommender systems. *Artificial Intelligence Review*, *24*(2), 179-197. doi: 10.1007/s10462-005-4612-x.

Mitchell, T.M. (2013). *Machine Learning*. McGraw-Hill.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: mapping the debate. *Big Data & Society*, *3*(2), 205395171667967. doi: 10.1177/2053951716679679.

Norman, A. (1989). *Inappropriate Feedback and Interaction, Not 'Overautomation'* (No. 8904; 14). Institute for Cognitive Science University of California.

O'Brien, M.G. (2015). *Epistemology and networked governance: An actor-network approach to network governance*. Florida Atlantic University.

Peterson, S.A. (1985). Neurophysiology, cognition, and political thinking. *Political Psychology*, *6*(3), 495-518. JSTOR. doi: 10.2307/3791084.

Pu, P., & Chen, L. (2007). Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, *20*(6), 542-556. doi: 10.1016/j.knosys.2007.04.004.

Puppe, F., Gappa, U., Poeck, K., & Bamberger, S. (2013). *Wissensbasierte Diagnose- und Informationssysteme: Mit Anwendungen des Expertensystem-Shell-Baukastens*. Springer-Verlag.

Salcedo-Sanz, S., Del Ser, J., Landa-Torres, I., Gil-López, S., & Portilla-Figueras, J.A. (2014). *The Coral Reefs Optimization Algorithm: A Novel Metaheuristic for Efficiently Solving Optimization Problems* [Research article]. The Scientific World Journal. doi: 10.1155/2014/739768.

Savage, M. (2009). Contemporary sociology and the challenge of descriptive assemblage. *European Journal of Social Theory*, *12*(1), 155-174. doi: 10.1177/1368431008099650.

Schmidgen, H. (2011). *Bruno Latour zur Einführung*. Junius Verlag.

Stampfl, N.S. (2013). *Die berechnete Welt: Leben unter dem Einfluss von Algorithmen* (1., Auflage). Heise Zeitschriften Verlag.

Stanforth, C. (2007). Using actor-network theory to analyze e-government implementation in developing countries. *Information Technologies and International Development*, *3*(3), 35-60. doi: 10.1162/itid.2007.3.3.35.

Sun, T.Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. *Government Information Quarterly*, *36*(2), 368-383. doi: 10.1016/j.giq.2018.09.008.

Todorut, A.V., & Tselentis, V. (2018). Digital technologies and the modernization of public administration. *Quality – Access to Success*, *19*, 73-78.

Toshkov, D. (2016). *Research Design in Political Science*. Macmillan International Higher Education.

Valle-Cruz, D., Criado, J.I., Sandoval-Almazán, R., & Ruvalcaba-Gomez, E.A. (2020). Assessing the public policy-cycle framework in the age of artificial intelligence: from agenda-setting to policy evaluation. *Government Information Quarterly*, *37*(4), 101509. doi: 10.1016/j.giq.2020.101509.

Veale, M., & Brass, I. (2019). Administration by Algorithm? Public Management meets Public Sector Machine Learning. In Yeung, K., & Lodge, M., (Eds.), *Algorithmic Regulation*, 31. Oxford University Press.

Venturini, T., Jacomy, M., Meunier, A., & Latour, B. (2017). An unexpected journey: a few lessons from sciences po médialab's experience. *Big Data & Society*, *4*(2), 205395171772094. doi: 10.1177/2053951717720949.

Wachhaus, A. (2009). Networks in contemporary public administration: a discourse analysis. *Administrative Theory & Praxis*, *31*(1), 59-77. doi: 10.2753/ATP1084-1806310104.

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. doi: 10.1080/01900692.2018.1498103.

Wirtz, B.W., Weyerer, J.C., & Sturm, B.J. (2020). The dark sides of artificial intelligence: an integrated ai governance framework for public administration. *International Journal of Public Administration*, *43*(9), 818-829. doi: 10.1080/01900692.2020.1749851.

Wu, S., Liu, Q., Bai, P., Wang, L., & Tan, T. (2016). SAPE: A System for Situation-Aware Public Security Evaluation. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*, p. 2.

Yeung, K. (2018). Algorithmic regulation: a critical interrogation: algorithmic regulation. *Regulation & Governance*, *12*(4), 505-523. doi: 10.1111/rego.12158.