Institut Barcelona d'Estudis Internacionals
Central European University
Academic Year 2019 – 2020

Co-funded by the Erasmus+ Programme of the European Union

In Machines We Trust:

Political Legitimacy and the Case of Automated Decision-Making at Europe's Borders

Dissertation submitted by
Daniel Cohen

in partial fulfillment of the requirements for the degree of
ERASMUS MUNDUS MASTER in PUBLIC POLICY

Supervisors: Prof. Simon Rippon (CEU) and Prof. Aina Gallego Dobón (IBEI)

I hereby certify that this dissertation contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I hereby grant to IBEI and the Mundus MAPP Consortium the non-exclusive license to archive and make accessible my dissertation in whole or in part in all forms of media, now or hereafter known. I retain all ownership rights to the copyright of the dissertation. I also retain the right to use in future works (such as articles or books) all or part of this dissertation.

Name: Daniel Cohen

Signature: Daniel Cohen

Location and Date: Barcelona, Spain - July 31, 2020

Number of Words with Bibliography: 13380

Abstract

In this essay, I question the relationship between political legitimacy and the use of Automated Decision Making (ADM) systems for public service provision. Although algorithmic governance (AGOV) mechanisms are still in their infancy, they indicate a trend of increased cognitive outsourcing to efficient and cost-effective AI Assistants. Increased automation throughout the public service changes the dynamic between the citizen and the political authority. This paper considers the ways by which political legitimacy can be measured to develop policies that build trust in algorithmic governance. With the assistance of expert interviews from AI policy practitioners, I build on a framework of political legitimacy to assess the input, output and throughput legitimacy of ADM systems. The framework is applied to the case of the Horizon 2020 pilot project iBorderCtrl, which uses Machine Learning (ML) to assist border control guards with a risk assessment of incoming travelers. I find that a non-binary approach legitimacy, which includes the black-box of decision making, offers new avenues to study the ways by which throughput legitimacy can offer solutions to both the input of citizens' voices into the political decision-making process as well as fair, equal and socially beneficial outcomes. The paper concludes by offering mitigation strategies at different levels of the algorithmic development cycle to address potential legitimacy issues.

Keywords: political legitimacy, technocracy, AI assistants, automated decision-making, AI ethics, trust in AI

Acknowledgements

Since the early days in Simon's office in Budapest, this research process tested my ability to think critically and objectively about my impassioned ideas. Simon, your guidance and patience taught me a new way of constructing my thoughts which I hope will stay with me for a long time. Aina, though this year was turbulent, I appreciate the insights, ideas and time you gave me in the middle of a pandemic.

To my many friends that endured endless rants about outsourcing our decisions to AI assistants, I want to thank Antonia, Bogdan, Nic, Sacha, Setu, Stefano and Valentino. A special grazie to Laurella for the times you gave me the ears I needed to sort out my thoughts and for the delicious foods that kept the engine running. Thank you, Lewis, for making sure my ideas stayed within the bounds of reality. Awaiting this project in Toronto is my family, Alain, Lotte and Emma, who have supported me lovingly along the way.

This thesis would not have been possible without the invaluable professional guidance from friends in the AI community in Israel, Vanessa, Ofer and Shahar as well as those around the world, Charlotte, Elizabeth, Jared, Seth and Yolanda. Thank you, Amit and my Unbirthday family for setting off the chain of introductions.

This project has allowed me to develop a passion for the objectively unclear and subjectively terrifying AI systems expected to enter society. I will take this as motivation with me on my journey to align AI with human values and avoid the technological dystopia.

Academia, it's been great!

Morality rests on human shoulders, and if machines changed the ease with which things were done, they did not change the responsibilities for doing them. People have always been the only 'moral agents.' Similarly, people are largely the objects of responsibility. There is a developing debate over our responsibilities to other living creatures, or species of them.... We have never, however, considered ourselves to have 'moral' duties to our machines, or them to us.

J. Storrs Hall

Table of Contents

List of Figures

1. Introduction

Big data, machine learning (ML) and artificial intelligence (AI) are effective tools for individuals to make complicated processes simpler. Experts predict that there is a "50% chance of AI outperforming all human tasks in 45 years and automating all human jobs in 120 years" (Grace et. al 2018). For now, AI systems are commonplace in the private sector, yet governments are slowly catching on as they see the vast potential for the use of big data and Automated Decision Making (ADM) in government, or as it is being termed, algorithmic governance (AGOV). ML tools and AI assistants that rely on big data to support public servants in making more efficient and accurate decisions are rapidly emerging through public private-partnerships (Mikhaylov et al, 2018; Gomes et al, 2019). Decisions previously made by civil servants are outsourced to AI assistants in police departments, judicial courts, labor departments and smart cities among several others.

The expansion and integration of AI systems into the public sphere has raised discussions around Building Trust in AI to ensure its acceptability and fairness in society. More and more, decisions are expected to be measured by powerful machine processes. Public perception survey studies collected by the Euobarometer demonstrates increasing skepticism to the transparency, accountability and fairness of AI (Eurobarometer, 2019). Many governments and civil society actors have focused on topics related to the future of work, in which millions of jobs will be automated in the near future, while others look at opportunities for AI integration in education, healthcare, urban design, judicial systems, policing systems. This area of research is considered short-term AI. Others take futuristic approaches to the field, conceptualizing Artificial General Intelligence or super intelligent machines as a likely outcome in the next century (Bostrom et al, 2018).

Rational Choice Theory holds that policy makers base their decisions and actions on a systemic analysis of costs, benefits and risks of alternative courses of action with time constraints, imperfect information and cognitive biases (Majchrzak and Markus, 2013). ADM, a key function of algorithmic governance, is precisely aimed at making the most rational choice given the available data. A human mind is incapable of processing an equal amount of information and is likely to come to a less objective or rational result.

The importance of ethical AI in the private sector has been well documented, however less literature has spoken to the relationship between AI use in the public sector and its acceptability in the eyes of citizens. In large part, this is due to the visible impacts of automation in labor intensive jobs, medical services, social media, supply chains, to name a few. The slower moving public sector is increasingly making use of big data analysis for integration into decision making processes. As democracies are increasingly challenged to make better policy decisions, ADM appear as a cost-effective solution for efficient decisions. How does the turn to more technocratic governance by machines affect citizens' relationship with political authority? Political legitimacy is a widely discussed concept that is fundamental to democracy. There are various understandings of what makes a government or government process legitimate in a traditional democratic sense. However, this paper will try to understand how the introduction of AGOV may be reshaping the relationship between the electorate and the public sector. My research question is as follows: What is the effect of the increased use automated decision making by public bodies on the political legitimacy of public services?

This thesis will introduce the topic of algorithmic governance (AGOV) and provide some clarity as to the origins and potential of automated decision-making (ADM) systems for ameliorating governance mechanisms. This section will be accompanied by raising some of the ethical concerns regarding the development and implementation of the algorithms such as Fairness, Accountability, Transparency, Autonomy, Bias, Explicability and Trust. Second, I will explore the literature on political and democratic legitimacy in order to situate its value and importance in democratic society. Next, I will connect AGOV to a previously developed normative political legitimacy frameworks in order to explore how this early paradigm shift may affect the political legitimacy of ADM in public service provision. With the help of 3 semi-structured realist expert interviews from a short-term AI Policy expert (Future Society), a long-term AI Policy expert (Future of Life Institute) and a Senior Public Sector Policy Practitioner (anonymous), I will seek to refine my theoretical framework in order to strengthen it as a tool by which I can analyze my case study This framework is then applied to the case of the EU Horizon 2020 pilot project iBorderCtrl, an ADM system that assists border control officers to assess the risk of incoming travelers. I argue that ethical concerns related to ADM systems are likely to affect political legitimacy and trust in democratic institutions and that a conceptual framework can facilitate the identification of necessary policy responses.

2. Literature Review

2.1What do we mean when we say Algorithmic Governance?

2.1.1 Artificial Intelligence

A general understanding of AI is necessary in order to delve deeper into the benefits it can provide as a tool to public sector decision makers. Russell and Norvig have laid out the most basic description of AI in which it is broken down into eight definitions and split into four categories: think like a human, act like a human, think rationally and act rationally (Russell and Norvig, 2010). More recently, academics have honed in on the latter part of the definition pertaining to rationality which states that AI is a system "that acts so as to achieve the best outcome, or when there is uncertainty, the best expected outcome" (Russell and Norvig, 2010; Danaher, 2016). The emphasis is made on the 'expected' outcome because of the assessment of a variety of potential options against each other.

Over the last decade developments of new technologies have led to an explosion of AI systems throughout our society. Big data refers to the "volume, velocity, variety and complexity" (Desouza and Jacob, 2017) of data which is collected throughout our society. Our actions throughout our digital surroundings are recorded at unprecedented levels and the limits of what data is collected and retained appears unbounded, in that its limits are unclear (Ibid.). With Big Data, datasets are capable of storing much larger quantities of data and algorithmic processes are capable of deciphering connections between complex, varied and massive amounts of data at unprecedented rates. Particularly, this is one area where civil society has been very vocal calling for increased privacy and data rights as massive amounts of data is collected on citizens and their behavior (Manheim and Kaplan, 2019).

Zarsky (2011) explains the process of data mining as "the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data" (Zarsky, 2011). Beyond the collection of these large data pools, which is impressive in and of itself, mathematical algorithms are used to extract the information and draw connections between specific data points that would otherwise go unnoticed by the human eye.

When we refer to algorithms in the context of AI and Big Data, we are not referring simply to the mathematical construct of an algorithm, but rather "the implementation and interaction of one or more algorithms in a particular program, software or information system" (Mittelstadt et al, 2016). These algorithms are applied to Big Data datasets in which data mining can occur with both descriptive and predictive functions. This is to say that a descriptive function would allow for an algorithm to search through a given data set and explain what has already happened, such as going through financial records in order to detect fraud (Danaher, 2016). Data mining and Machine Learning (ML) takes a greater leap when it begins to use Deep Neural Network (DNN) algorithms to make predictions about future outcomes based on historical data, this is called predictive analytics. Examples of this could be recent use cases of predictive policing, credit, insurance and employment screening (Zarsky, 2011; Mittelstadt, 2016) or as I will later discuss, border control.

Basic tree model algorithms are explainable and understandable by the human minds which program them and they are technically replicable by a human. However, ML is one area in which algorithmic governance creates legitimacy questions for public servants' decision making processes (Yampolskiy, 2019). ML is the technology which allows for the machine to go beyond the original tasks set out in the codes that instruct it how to act. Van Otterlo (2013) describes the process of ML as "any methodology and set of techniques that can employ data to come up with novel patterns and knowledge and generate models that can be used for effective predictions about data". ML goes beyond the programmed functions of the algorithm to gather more information, learn and recognize new and unexplainable patterns using DNN (Yampolskiy, 2019). These are also referred to as Artificial Neural Networks (ANN), however DNN involves a process called Deep Learning (DL). This allows for more accurate predictions about uncertain future outcomes.

2.1.2 AI Assistants and Automated Decision Making

AI assistants have been generously welcomed into our lives in the form of digital tools which help us "search, plan, message, schedule and so on" (Danaher, 2018). AI assistants are commonplace in our daily lives as we interact with our cell phones and web services that can help us discover music, purchase items, and tailor our digital existence to our personal preferences (Gal, 2017). Danaher sets out a proper definition of personal AI assistants that I will refer to throughout this paper, in which they are defined as "any computer-coded software/program that can act in a goal directed manner … that can set some target output and can select among a range of options that optimizes (according to specific metrics) for that output" (Danaher, 2018).

The advanced algorithms for AI Assistants are based on coded decision trees that will assign various weights to different decision-making inputs which set the parameters for optimized recommendations (Gal, 2017). As the AI assistant 'learns', the algorithm "self-adjusts based on its own analyses of data previously encountered, freeing the algorithm from predefined preferences" (Ibid). The overall aim of the AI assistant is to personalize and facilitate a decision-making process for the user, based on previous data which demonstrates what the most probable, rational and best option would be (Waldman, 2019).

AlgorithmWatch defines ADM systems as "procedures in which decisions are initially partially or completely—delegated to another person or corporate entity, who then in turn use automatically executed decision-making models to perform an action" (Speilkamp, 2019) . It is important to note in this definition that the idea of partial or complete delegation to an automated system is categorized in the literature as in-the-loop, on-the-loop and out-of-the-loop (Rahwan, 2017). This differentiation was first talked about in the military context for human oversight of automated drones. ADM systems and AI Assistants will be referred to interchangeably throughout the paper as they stem from separate literature but imply the same meaning.

The contribution which I seek to make in the literature stems from the expansive shift of personal AI assistants from the private sector, that is for commercial uses, to the implementation of AI assistants for services provided in the public domain. Mittelstadt et al (2016) and Danaher et al. (2017) have also set out along this path of highlighting the usage of AI assistants in the public sphere and attempted to discuss some of the potential issues associated with this level of cognitive outsourcing.

2.1.3 Algorithmic Governance

Applications of Big Data and AI are widely used throughout the private sector and the slow-moving nature of the public sector has made the emergence of AI products for evidence-based decision making a more recent development (Algorithm Watch Report 2019). Data governance is not new per se, the integration of AI and Big Data extend its abilities to make more accurate predictions as a result of data mining and ML's predictive capacities (Mehr, 2017). Giest (2017) describes data culture in government as "the capacity of both individual civil servants as well as the organizations as a whole to collect, merge and utilize big data and the institutional structure supporting this through training civil servants and open data initiatives". The vast data which is collected throughout government domains such as tax systems, social programs and health records can now be digitized and used by decision makers when formulating policies in education, economics, health and social welfare (Giest, 2017). The combination of ML with big data allows governments to make more accurate decisions (Gal, 2016).

Digital era governance emerges as a technological opportunity for governments to transition towards a more service oriented and accountable actor to its citizenry (Giest, 2016; Mehr, 2017; Poel et al. 2018; Stockmann, 2018). Some of the initial ways in which these machines will be able to support government have been outlined by Engin and Treleavan (2019) among others. Advancements include: Government Data Facilities, Internet of Things (IoT), Big Data, Behavioural/Predictive Analytics and Blockchain Technologies. From these large pools of data and metadata, behavioural and predictive analytics are capable of synthesizing this information and uncovering hidden patterns, unknown correlations and personal preferences, something which would be impossible for a human (Engin and Treleavan, 2019).

2.1.4 Building Trust in AI

AI has garnered lots of excitement into its ability to solve many of the worlds problems, yet many are fearful of its obscure methods of producing results. There is an underlying assumption that the world's problems are quantifiable into data. The idea that, if provided with the right amount of data, machines can find solutions to complex problems, is sometimes contestable. Governments, NGOs, International Organizations, Supranational organizations and private entities have released reports on how the pervasiveness of AI throughout society will be so extreme that ethical values must be instilled to combine the benefits of innovation with the technology's trustworthiness (Spielkamp, 2019; European Commission 2019; European Commission, 2020; European Parliament, 2019; Fjeld et al, 2020; Whittaker et al., 2019; OECD, 2019; White House, 2020).[1] The field of AI Ethics, or AI Safety, has emerged to understand how the integration of AI into society should be conducted. Throughout the AI Principles reports mentioned above, as well as in the academic literature, key areas of concern stand out. I will briefly highlight some of the important concepts in AI ethics which I will later discuss in more detail throughout the case study. *Algorithmic bias* refers to the ways in which an algorithm can be biased, either through the training data or the developer's lines of code (Pasquale, 2015; Bostrom et al, 2018; Barocas and Selbst, 2016; Perry, 2013, O'Neil, 2016). *Transparency* and *accountability* are two central concerns when it comes to AI. These refer to the *explicability* of how AI produces obscure decisions inside a ML

[1] For a more complete discussion of AI Ethics, the Berkman Klein Center for Internet and Society at HarvardUniversity report by Fjeld et al. 2020, analyzes 36 prominent AI Principles documents from around the world to uncover sectoral norms.

‘*black box*’ (Citron and Pasquale, 2014; Engin et al. 2019; Datta et al, 2016; Mittelstadt, 2015; Diakopolous, 2016, Annany and Crawford, 2016). Some argue that *explicable ML* (xML) is not possible because it would never be comprehensible to a human mind (Yampolskiy, 2020; de Fine Licht, 2011). *Human oversight* is proposed so that a human is involved in the final decision-making process, in case of an *algorithmic malfunction* (Danaher, 2015, Brundage et al. 2020). Others such as Gal and Elkin-Koren (2017) argue that we increasingly *cognitively outsource* our autonomy to make decisions, by trusting AI assistants and limiting our choice field of what decisions are available. These concerns have transformed into a larger movement aimed at *Building Trust in AI* (Fjeld et al, 2020) in order to ensure the safe integration of AI in society.

I will now discuss the literature on political legitimacy which has also documented the importance of trust in government (Feldman, 1983; Blind, 2006, Mény, 2002) as a means by which citizens determine the acceptability of decisions and their adherence to them.

2.2 Political Legitimacy

The concept of political legitimacy has been discussed in various contexts, some with a descriptive approach (Weber, 1964) and others with a normative focus that attempts to determine its value in society. For the state, political legitimacy is associated with the way in which citizens legitimize and justify political authority (Locke, 1980, Rawls, 2007). Both Rawls and Locke emphasize the need for consensual political authority, whereby citizens give a political body the right to enforce coercive measures over an individual.

2.2.1 Descriptive Legitimacy

Descriptive legitimacy finds its roots in Weber (1964) who argued that legitimacy is an important explanatory variable for social science because it relates to the faith which citizens have in a particular social order (Peter, 2017). The descriptive form of legitimacy gives credence to the historical reasons which justify authority, not simply an ethical or value-based perspective on legitimacy. On the other hand, normative approaches (Rawls, 1993; Ripstein 2004, Raz 1986) see legitimacy as the "justification for political coercion". In this sense, political legitimacy is a "benchmark of acceptability" (Peter, 2017) in which citizens determine whether coercive decisions are permissible to them. Simmons (2001) distinguishes between the justifications of states and the moral argument for legitimacy. This is an important distinction because I do not wish to focus on the authority of the state to use Automated Decision Making in government.

2.2.2 Normative Legitimacy

Democratic legitimacy has been primarily talked about from two sides. Proceduralism, understands moral authority and legitimacy to be achieved through democratic processes regardless of their outcomes (Manin, 1987; Gaus, 2010; Buchanan 2002; Christiano, 2004; Peter, 2008; Estlund, 2008ab; Kolodny, 2014). Others have referred to this same concept as input legitimacy (Scharpf, 1970), whereby democratic legitimacy is achieved through direct democracy, elections, deliberative democracy (Manin, 1987; Bohman and Rehg, 1997, 1997; Pettit and Rabinowicz, 2001) or public reason (Rawls, 1986; Gaus, 2008, 2010). On the other hand, some have focused on good policy outputs or social fairness as the means by which we judge legitimate authority (Arneson, 2003; Wall, 2007; Landemore, 2012; Weatherford, 1992). Pitkin (1967) discusses distributional authority insofar as it looks to the overall successes and failures of political decision making in its capacity to achieve desired outcomes. In some circles this has also been referred to as output legitimacy in which society shall continuously judge political authority on the basis of its ability to provide fair decisions for the public good (Scharpf, 1970). Few scholars genuinely argue that instrumental authority or output legitimacy can be achieved in a non-democratic context. This would be called pure instrumentalism (Peter, 2017). Many make the argument for epistemic or instrumental democracy, where democracy has the highest possibility of achieving the desired outputs based on the assumption that a citizen is most likely to vote for a correct choice (Wall, 2007; Landemore 2009; List and Godin, 2001). The third form of legitimacy which has emerged more recently is that of throughput legitimacy. Schmidt (2013, 2019) makes the argument that there is a 'black box' element of political decision making in which governance bodies must be accountable and engage with civil society for decisions to be legitimate in the eyes of the public. Throughput legitimacy, as she terms it, is considered to be a procedural middle ground between input and output. It is an context wherein political bodies should be accountable to the decisions they make and transparent about bureaucratic processes of those decisions once in power.

Attempts to study political legitimacy of different institutions or governance mechanisms have been used in a variety of contexts. A large part of this literature has focused on the legitimacy of the EU as a multi-level governance structure. Drawing on AI policy researchers' encouragement to investigate legitimacy questions of AI (Danaher, 2015) I connect previous concepts used to study political legitimacy with AGOV. I explore if a political legitimacy framework can uncover new strategies for regulating the integration of automated decision making into the public sector and build trust in AI.

My question is then, what is the effect of a transition to increasingly automated decision-making systems on the political legitimacy of public service provision?

2.2.3 Technocracy, Legitimacy and AGOV

AGOV is likely less democratic (Katzenback and Ulbricht, 2019; Saetra, 2020) because there is less opportunity for citizens to decide on the public good themselves when relying on technologically coded authority (Aneesh, 2002). The vast capabilities of AGOV have reignited discussions of technocratic governance, by which a technical expert can make more accurate, rational and beneficial decisions than a democratic process (Feenberg, 1994, 2005; Aneesh, 2002; Just and Latzer, 2017 Saetra, 2020). A more futuristic approach, which I will not specifically be discussing, has been termed as algocracy, or the rule by algorithms (Aneesh, 2002; Danaher, 2016; Saetra, 2020). However, the concept of technocratic governance in many senses challenges the concept of political and democratic legitimacy[2]. Bertsou and Pastorella (2017) speak to the inverse relationship between technocracy and democracy. As governments begin to use advanced algorithms and ADM processes, how does this affect the political legitimacy of increasingly technocratic decisions?

AGOV systems act as technocratic authority due to their capacity to exceed human ability in many settings that will make political and coercive decisions (Rahwan, 2017). Understanding the political legitimacy of increasingly technocratic AI systems in the public sector is important for ensuring trustworthy policy-making and improving the relationship between citizens and the state

[2] For a deeper discussion on the growth of technocratic government and its impact on democracy and be see Meynaud (1968) *Technocracy*.

(Danaher, 2015; Helbing et al.2019). If I can demonstrate the association between AGOV and political legitimacy, this could serve as a toolkit for ensuring the responsible use of ADM systems in government. In the long run, civil servants' cognitive outsourcing of decisions to AI assistants could impact the acceptability of coercive rules and trustworthiness in government if the decisions are not perceived to be legitimate.

I hope to contribute to emerging literature by connecting the public management research on political and democratic legitimacy with the salience of AGOV. The analytical framework of political legitimacy will more clearly identify policies for building citizens' trust in the responsible use of AI in government. Evidently, the attempt to establish a functional framework to assess political legitimacy is not original (Bowman et al 2005). As such, I will rely on previous works and more recent literature in order to apply it to AGOV. Rather than looking at political legitimacy from the perspective of the authority of the state (Greene, 2016), I make the assumption that the state as a political entity holds authority. In the context of AGOV, I want to analyze more specifically what legitimizes democratic decisions and enables citizens to adhere to automated political decisions.

3. Theory and Expectations

3.1 Measuring Political Legitimacy

Weatherford argues that research on legitimacy and its relation with theories often makes the mistake of beginning with "measures and tries to fit theoretical inferences to them rather than the reverse, and it promotes the question of policies versus incumbents by construing legitimacy in terms of public approval for governmental outputs, rather than the more central issues of how citizens evaluate the system's procedural efficiency or distributive fairness" (Weatherford, 1992: 152). Following in Weatherford's cautionary words, I build off of public management literature that has used a normative framework of political legitimacy to analyze the European Union (EU)'s multi-level governance system (Scharpf, 1999; Schmidt, 2013; Geeart, 2014). The relevant concepts of input, output and throughput legitimacy have been used as a standard form of constructively analyzing legitimacy (Schmidt, 2013). Although this method of analysis has been implemented mostly in EU studies (Olsen et al., 2000, Della Salla, 2010), the explicit connection to AGOV has not yet been made.

3.2 Toolkit for Analyzing Input, Throughput and Output Legitimacy

3.2.1 Input Legitimacy

The concept of input legitimacy stems from Easton (1965) who identified the way in which citizen's interests and demands enter the political system, whether through voting or the citizens' perspective on the legitimacy of the system. For Scharpf (1999), input legitimacy refers to "the participatory quality of the process leading to laws and rules". As was covered in parts of the literature review on political and democratic legitimacy, input legitimacy is concerned with the procedural elements of achieving the citizen's democratic voice to guide decision making. The view of input legitimacy that I refer to in the paper, is consistent with Scharpf (1999) and Schmidt's (2013, 2019) description as a political criterion focused on citizen's political participation and government's responsiveness. I will limit the scope, to the point at which government processes are in line with the preferences of the public.

The relationship with algorithmic governance is made by understanding at which point the citizens' voice is considered in the algorithmic development lifecycle. I contend that the areas of interest here are the data collection process and the development of algorithms by private entities as either simple automation processes or ML. The private entity refers to those which the government outsources to for the design of the Application Programming Interface (API).

The data which is collected represents the interests of citizens as it defines which persons or values are in consideration. Large scale big-data goes beyond traditional public sector data collection from citizens, such as a census, because it requires monitoring and tracking of much more specific and detailed personal data. Since data mining and programming of algorithms is the basis of developing an AI assistant, I offer these two variables as my starting point of assessment. First, data collection from citizens shapes what the algorithm can output. Secondly, the developers who create the algorithms which make political decisions are in a sense coding the law (Lessig, 1999). The rules which they give to a coded algorithm will be responsible for how the law is applied in the political setting. In many instances in which a public service is assisted by an ADM system, governments outsource the project to a private entity to develop the technology (Spielkamp, 2019). In order to produce these technologies, private firms will need access to citizens' personal data, for training and later in the roll out stage of the service. In turn, the private entity becomes the owner of the citizens' personal data.

3.2.2 Throughput Legitimacy

The more recent addition to the analysis of political legitimacy has been that of throughput legitimacy. It is interesting for this analysis because it discusses how decision-making processes are legitimate in and of themselves. The concept is separate from input legitimacy, wherein the citizen's political voice is achieved whether through direct democracy, elections or deliberative democracy. It also distinguishes itself from the outputs produced by a political body which are judged on the basis of their fairness and efficiency. Throughput legitimacy speaks to the middle ground between these two legitimizing forces. Schmidt and Wood (2019) argue that throughput does not replace, but rather serves as a distinct procedural criterion which can strengthen both the inputs and outputs. This form of legitimacy is achieved through transparency and accountability, inclusiveness and openness with civil society (Schmidt 2013, Schmidt and Wood, 2019). It tackles what Schmidt considers the black-box of decision making in governance. In Geearart and Leuven's (2014) in depth analysis on throughput legitimacy in the EU, they suggests that throughput legitimacy can increase both input legitimacy and output legitimacy. It has the ability to increase input legitimacy because certain "processes or deliberative interactions" (Geerart and Leuven, 2014) improve input participation. While on the other hand, certain governance processes can increase the output performance of government.

The technologically coded algorithms which are used to determine the outputs of public services are thus the point by which I will assess the throughput legitimacy of ADM. Specifically, the opacity or 'black-box' of ANNs' decisions creates the linkage to throughput legitimacy. The AI assistant effectively makes a political decision on its own rationale which is path dependent after its deployment. It contributes to the political process of making a decision for the provision of said public service. The accountability of the AI assistant in this case would refer to its ability to explain its actions and have check and balances in place in case of unreasonable outputs. In a similar tone to how we measure the throughput legitimacy of a traditional bureaucratic decisions, it is important to look at the transparency of the algorithm which provides the end user, or in this case the public servant, with a suggested action. The public would need access to exploring what data is collected and how decisions are produced by the ADM system. The increase or decrease of political legitimacy will thus be measured on the basis of its accountability for the actions taken, the transparency of information used to build the algorithmic system and the public accessibility to engage with the development and usage of the technology.

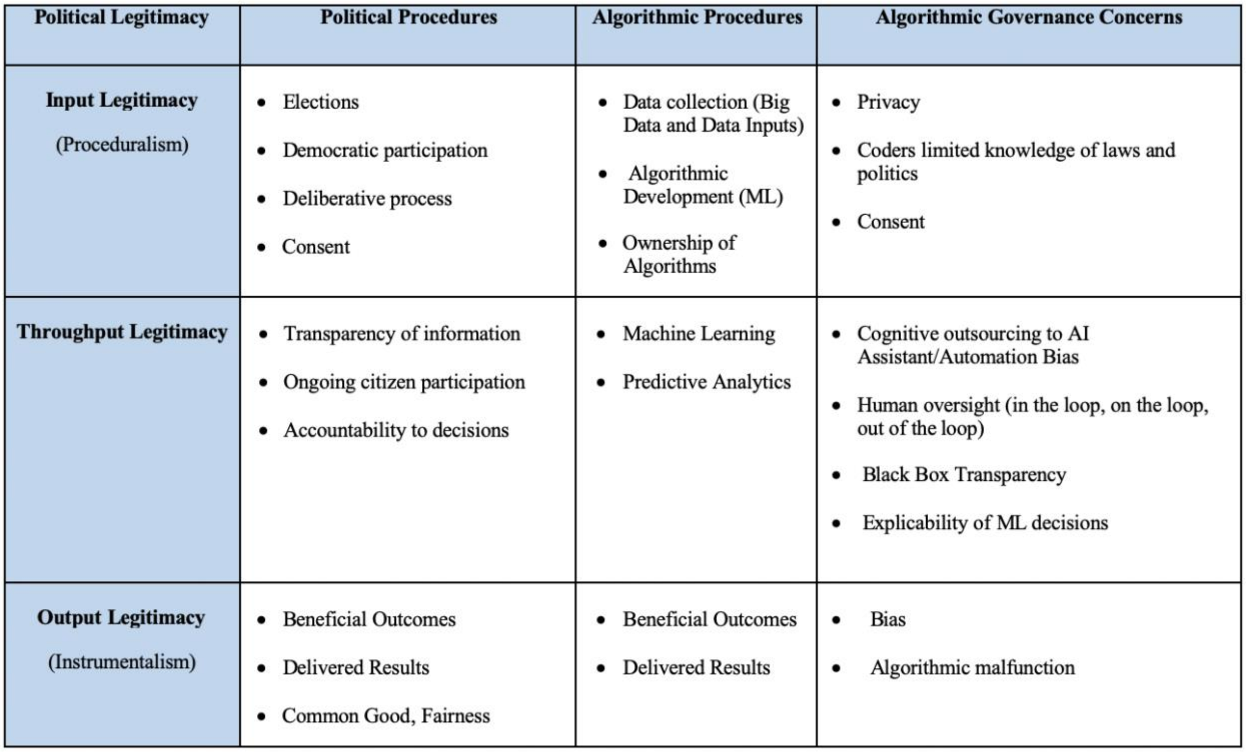
3.2.3 Output Legitimacy

The last form of legitimacy which I will assess, is the instrumental aspect of output legitimacy. The instrumental value of algorithmic governance is by and large the basis of the excitement as it reduces costs, increases capacity and increases the objectivity of decision-making. Output legitimacy refers to the legitimacy that is achieved as a result of efficient policy-making. It is what Scharpf (1999) refers to as *government for the people*, whereby "the policies adopted will generally represent effective solutions to common problems of the governed". Scharpf argues that the EU lacks in input legitimacy due to the complexity of the balance of power between institutions, rather it retains legitimacy through its output to keep peace and prosperity across Europe (Schmidt, 2013).

In the context of AGOV and ADM, output legitimacy is an extension of the same concept. The decisions produced by AI assistants are assumed to be more accurate because of the programmed rationality which they employ. In order for the ADM system to increase its output legitimacy, its decisions must be fair and serve the public good.

This model will connect the traditional concepts of normative political legitimacy as introduced by Scharpf (1999) and Schmidt (2013). In order to incorporate the complexity and automated processes that characterize ADM, the model cannot be limited to inputs and output legitimacy but should also look at Schmidt's (2013, 2019) throughput 'black box' legitimacy. With recent concerns of AI being focused on their fairness (Binns, 2018), accountability and transparency, there is a clear connection with the addition of throughput legitimacy to the traditional input-output analytical framework of political legitimacy. Through a review of the literature on the performance and integration of short-term AI applications into society as well as their algorithmic development lifecycle, I categorized the different elements of an ADM tool for public sector use into their respective legitimacy requirements. See Figure 1.

Figure 1 Political Legitimacy Framework



Political Legitimacy	Political Procedures	Algorithmic Procedures	Algorithmic Governance Concerns
Input Legitimacy (Proceduralism)	• Elections • Democratic participation • Deliberative process • Consent	• Data collection (Big Data and Data Inputs) • Algorithmic Development (ML) • Ownership of Algorithms	• Privacy • Coders limited knowledge of laws and politics • Consent
Throughput Legitimacy	• Transparency of information • Ongoing citizen participation • Accountability to decisions	• Machine Learning • Predictive Analytics	• Cognitive outsourcing to AI Assistant/Automation Bias • Human oversight (in the loop, on the loop, out of the loop) • Black Box Transparency • Explicability of ML decisions
Output Legitimacy (Instrumentalism)	• Beneficial Outcomes • Delivered Results • Common Good, Fairness	• Beneficial Outcomes • Delivered Results	• Bias • Algorithmic malfunction

3.3 Expectations

Overall this model of categorizing algorithmic governance in terms of its political legitimacy allows us to reconceptualise the AI application's development lifecycle in terms of how it relates to traditional bureaucratic decision-making. In analyzing the different steps of the algorithmic development cycle, the model is more sensitive to the different impacts the technology may have on society. Through the application of this model to the iBorderCtrl case, I expect to find that throughput legitimacy is negatively impacted the most as a result of the inexplicability and unaccountability of ML. Output legitimacy is positively increased due to its cost-effectiveness and more rational decision making in public services. Further, this allows for more nuanced understanding of which areas may require the attention of policy makers to build trust in automated public service provision in the eyes of society.

4. Research Design

The thesis seeks to highlight areas in which AGOV can be assessed in accordance with traditional conceptions of political legitimacy in order to improve AI's responsible integration into government. In order to achieve this, I used qualitative analytical methods: conceptual framework construction, semi-structured expert interviews and a case study. The conceptual framework is developed as an interlinkage of political legitimacy and algorithmic governance concepts in order to "provide a comprehensive understanding of a phenomenon" (Jabareen, 2009). I demonstrate how a theoretical framework of political legitimacy that analyzes the input, output and throughput legitimacy of ADM systems in the public sector can provide policy-oriented insights into how best to ensure citizens' trust and compliance.

4.1 Expert Interviews

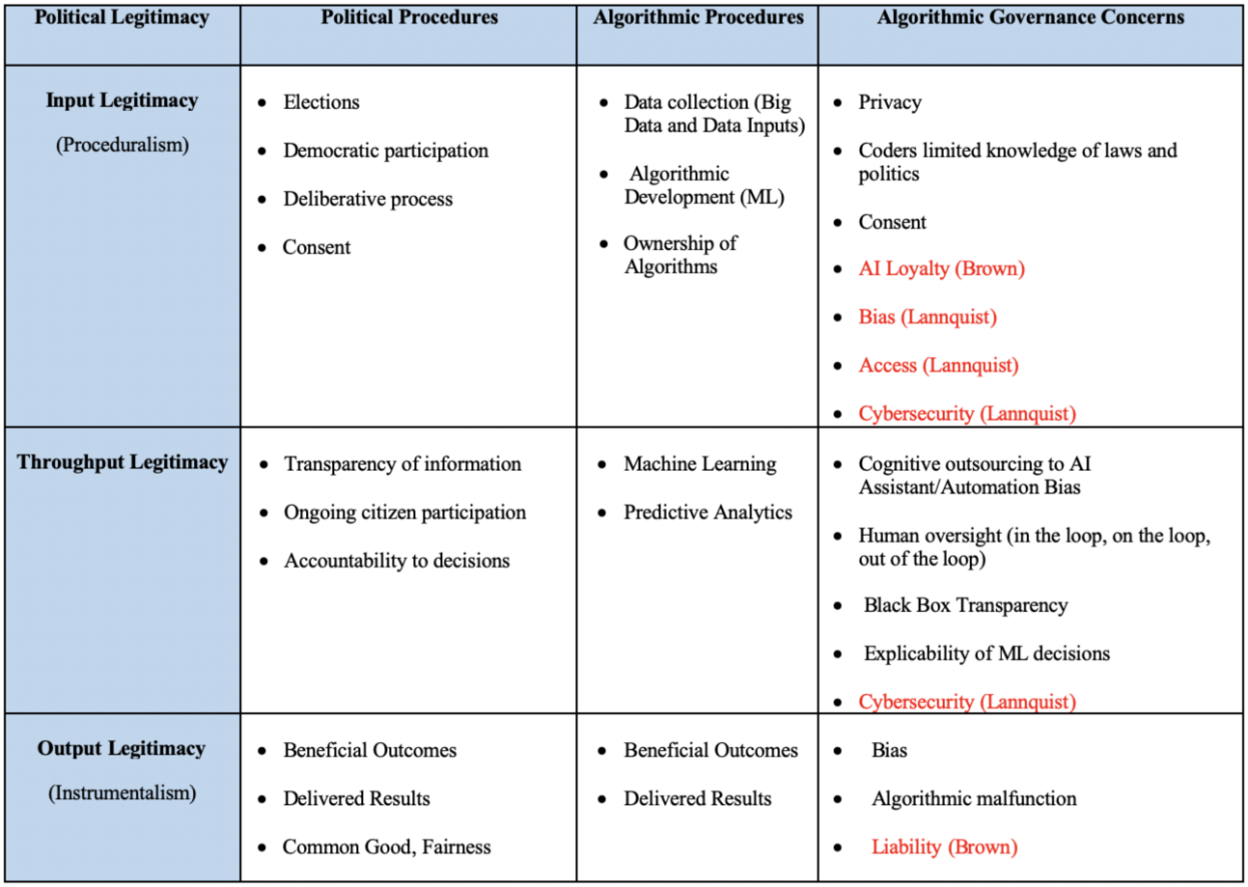
In order to validate and refine my theoretical framework I conducted 3 semi-structured expert interviews. In selecting the interviewees, I ensured that each would speak from a different perspective on algorithmic governance so as to give a more holistic understanding to the subject matter. The different angles I wanted expert knowledge from included: the short-term AI community, the long-term AI community and the public sector itself. The short-term AI community refers to those individuals working on the current applications of AI as they relate to its integration to society today. This relates to more tangible and contemporary applications of AI such as AI strategies or white papers for labor markets, public health or military. The long-term AI community refers to those conducting research on future predictions of the massive capabilities of AI, such as Superintelligence (Bostrom et al, 2018) and AI Existential Risk (Baum, 2020).

Lastly, I thought it would be important to provide elite insight from the public sector itself to gain insights into how they may perceive their role in delivering public services effectively and responsibly. I provided each interviewee with the same preliminary conceptual framework and asked them to fill in the sections as they saw fit. The interviewees provided responses, that served as theory refinement (Manzano, 2016) by speaking to their ongoing projects, emerging debates in the field, and opinions about where the industry is heading. Theory refinement interviews are helpful to provide the evaluator with hard to observe information and make adjustments to the interviewer's framework (Manzano, 2016). Since the field of AI and the technology itself is rapidly evolving, expert insights were a useful tool to ensure that the conceptual analytical framework included all the relevant components.

4.1.1 Refinement with Expert Interviews

The interviews were conducted in a semi structured fashion. The first part of the discussion was set for reviewing the framework itself, while the second half discussed novel mitigation strategies. Throughout the case study analysis, I will refer to some of the mitigation strategies across the algorithmic development lifecycle. Now, I will briefly highlight the main findings of the expert interviews.

Figure 2 Political Legitimacy Framework Revised by Policy Experts



Political Legitimacy	Political Procedures	Algorithmic Procedures	Algorithmic Governance Concerns
Input Legitimacy (Proceduralism)	• Elections • Democratic participation • Deliberative process • Consent	• Data collection (Big Data and Data Inputs) • Algorithmic Development (ML) • Ownership of Algorithms	• Privacy • Coders limited knowledge of laws and politics • Consent • AI Loyalty (Brown) • Bias (Lannquist) • Access (Lannquist) • Cybersecurity (Lannquist)
Throughput Legitimacy	• Transparency of information • Ongoing citizen participation • Accountability to decisions	• Machine Learning • Predictive Analytics	• Cognitive outsourcing to AI Assistant/Automation Bias • Human oversight (in the loop, on the loop, out of the loop) • Black Box Transparency • Explicability of ML decisions • Cybersecurity (Lannquist)
Output Legitimacy (Instrumentalism)	• Beneficial Outcomes • Delivered Results • Common Good, Fairness	• Beneficial Outcomes • Delivered Results	• Bias • Algorithmic malfunction • Liability (Brown)

For input legitimacy, it was stressed that many of the concerns related to ADM systems should be addressed in their early development stage. The short-term AI policy expert discussed how algorithmic bias could not only be included in the output stage but should be considered as the algorithms are being developed. She emphasized the importance of considering the ownership of the technologies. “(The) Private sector has misaligned interests to the public sometimes” (Lannquist and Cohen, 2020). The long-term AI policy expert discussed his concerns of AI Loyalty. He explained “Fiduciary responsibility – it has to behave in your interest – or in societies interest” (Brown and Cohen, 2020). The conversation looked more in detail at examples in which it is important to know who’s interests are pre-programmed into an AI’s codes. Furthermore, Lannquist notes that the accessibility of the technology is important because the training data needs to be representative and diverse (Ibid). The cybersecurity concern was added because massive data sets are collected on personal information and behaviour.

In terms of the throughput legitimacy, there was near unanimous agreement that the concerns had mostly been covered. Lannquist, added cybersecurity again as she explained that there are cyber vulnerabilities at all stages of the algorithmic development cycle. If a private entity’s ADM system is hacked by a hostile intruder and adjusts the ML code to tailor results in their own interests the dangers are severe.

Lastly, the only addition that was made to the output legitimacy were liability concerns. Brown explained the importance of liability to determine who is responsible when ADM systems malfunction. This debate is often discussed in autonomous vehicles, whether the driver, developers or the private company are responsible for mistakes. There is an evolving policy debate on who bears the burden of responsibility and the importance and risks of a strict liability regime. He states:

> “if there's the strict liability regime … then you're going to see increased sensitivity to safety and not make any mistakes. Because they'll get sued and they'll have to pay for it. … on the negative side of that coin … pressing the analogy a bit, if you have too strict of a liability regime then you squash innovation, which is sort of true, but the idea being that, nobody's going to take a chance to throw out this new tech if it has a possibility of going wrong because then they'll get sued and go under” (Brown and Cohen, 2020)

With a more refined framework, the next section will now conduct the case study analysis on iBorderCtrl.

4.2 Case Study Selection

Lastly, to demonstrate the applicability and usefulness of the model I selected an ongoing case study (King et al., 1994) of an ADM system tested for implementation in the EU. The representative case was selected “to represent a broader population of cases in some relevant respect, which may be descriptive or causal” (Gerring and Cojocaru, 2015). The case under study is the iBorderCtrl – Intelligent Portable Border Control System, which is an EU funded project currently being tested in Hungary, Greece and Latvia (iBorderCtrl, 2020). The particular case was selected for two primary reasons. Firstly, because it uses ML tools that exemplify ADM processes in other public services. In order to test the political legitimacy framework on AGOV, it was important that the case would be translatable into other contexts. Secondly, due to the fact that implementation of ML algorithms for public service provision is still in its very early stages, this specific case was unique in that it had publicly available primary source documents with sufficient information to conduct my research. AlgorithmWatch.org, a non-profit NGO which monitors AI's integration into society, published a report documenting all of the current known ADM systems currently in use across the EU (Spielkamp, 2019). iBorderCtrl stood out due to the availability of documents related to the organization and the use of its ML technologies. Riccardo Coluccini from the Hermes Center for Transparency and Digital Human Rights made two Request for Information applications to the EU for the disclosure of confidential published documents(Kampis, 2018). Many documents remain confidential due to their sensitive nature; however, 5 were released, albeit some included blacked out content. These documents in combination with 2 scientific papers on the technologies used in the Horizon 2020 project made the research possible.

The technology used in iBorderCtrl applies a broad range of biometrics tools and behavioral analytics to make risk estimations on travellers entering the EU (iBorderCtrl, 2018b). The multi-layered technology serves as an AI assistant for border patrol officers. The ADM system can make more efficient decisions, limit subjective biases of the officers and save time for travelers at the border (Ibid). The technologies used, particularly behavioral analytics and facial scanning can be found in other public services like predictive policing (Brantingham, 2018; Richardson et al, 2019) and automated risk assessments (Dressel and Farid, 2018) or in supporting healthcare workers throughout their interactions with patients. Although the results of this case will not speak for all instances of ADM in public service provision, it serves to demonstrate how the analytical framework can be a tool to identify areas where regulation and other policy instruments may be better able to increase public trust in ADM. If effective, more research should be done on other uses of ADM in the public sector. The next section will explain changes made to the constructed framework as a result of the expert interviews.

5. Case Study Analysis

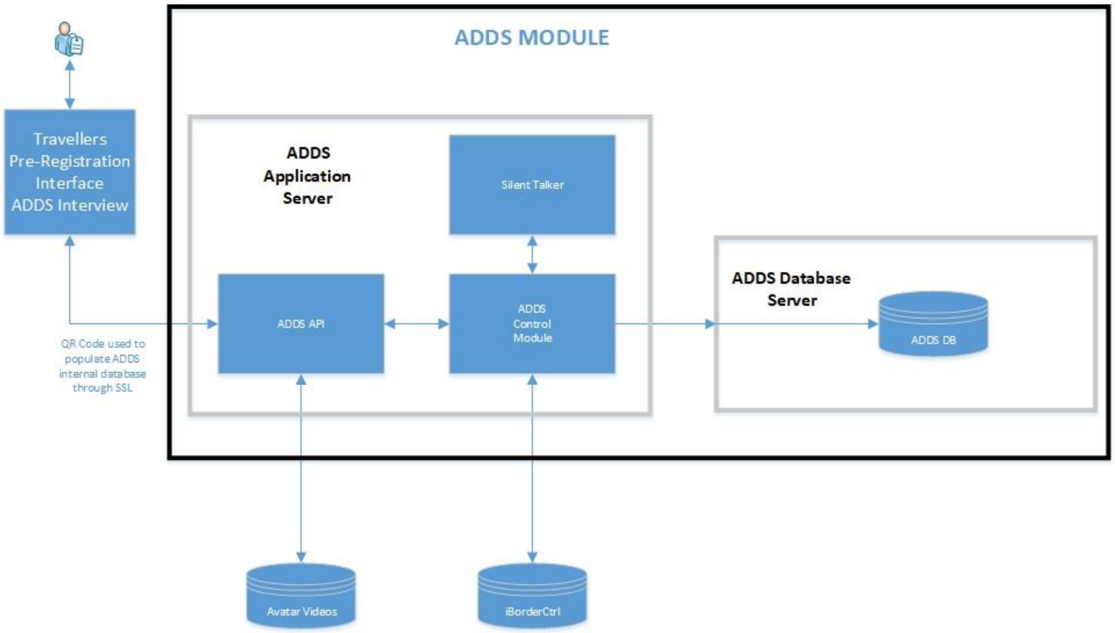
5.1 What is iBorderCtrl?

With over 700 million people entering the EU each year, pressure is building on its borders (European Commission, 2018). The capacity of border control guards to efficiently check the personal documents and biometrics of each passenger is increasingly tested. As part of the Digital Single Market Horizon 2020 Secure Societies project, the EU has heavily invested in supporting the development and integration of new technologies to secure its' physical and digital borders. The European Commission has proposed €34.9 billion euros for border control and migration management between 2021-2027 (Sanchez-Monedero and Dencik, 2020). These investments are aimed to support innovative technology solutions to "address security gaps and lead to a reduction in the risk from security threats" (European Commission, 2018). Under the pressure of increased movement throughout Europe, the EU operates with a mindset that new technologies can reduce the costs associated with securing its borders, while also increasing the efficiency by which it does so. As part of the H2020 project, iBorderCtrl was developed by researchers at Manchester Metropolitan University as an ADM system to support border control guards in making accurate judgements of third-country nationals wishing to enter Europe's borders. iBorderCtrl completed its testing stage between 2016-2019, in which the technology was developed and tested in 3 countries: Greece, Hungary and Latvia (iBorderCtrl, 2020). The pilot project awaits approval and implementation at Europe's borders.

iBorderCtrl – the Intelligent Portable Border Control System, is an ADM system that pre-screens individuals wishing to enter the EU, using biometrics, facial recognition and automated risk assessments. iBorderCtrl is an example of an AI Assistant for a border control officer. The system is broken down into a variety of different technological tools: Automatic Deception Detection System (ADDS), Biometrics Module, Face Matching Tool, Document Authenticity Analytics Tool (DAAT), External Legacy and Social interfaces system (ELSI), Risk Based Assessment Tool (RBAT), Integrated Border Control Analytics Tool (BCAT), and a Hidden Human Detection Tool (HDD) (iBorderCtrl Consortium, 2017). The iBorderCtrl system uses the combination of these tools to collect information, conduct a virtual interview and provide a verdict risk assessment of the individual to the border control officer's handheld device. The final decision is made by the border guard and is considered a human-in-the-loop ADM system.

The ADDS is a configuration of ANNs used to detect non-verbal behavior in the form of micro gestures over short periods of time (O'Shea et al, 2018). The avatar which conducts the virtual interview is "personalized to communicate with the traveler including utilizing subtle non-verbal communication cues to stimulate richer responses from them" (Ibid). The avatar will also adapt its own behavior as it reads the non-verbal responses of the traveler. It may become more inquisitive if it senses that an individual may be lying. The developers go on to suggest that the border control officers are often fatigued and prone to subjective opinions, for which an automated solution is desirable. The information collected from the personal documents and the avatar interviews feed into the ADDS API. The Silent Talker ANN is used to feed a response into the ADDS Control Module, which then outputs the risk assessment to the iBorderCtrl system that is used as a handheld device by the border control officer. The data is later stored into the ADDS Database Server which is retained as the property of iBorderCtrl. They state the "automated system, which utilizes a few minutes of traveler time at the pre-crossing stage without increasing the amount of time they spend with a border control agent, could thus potentially increase efficacy while reducing cost" (Ibid).[3] Figure 3 demonstrates the process of the ADDS Module.

Figure 3 ADDS Module (iBorderCtrl Consortium 2018b)



5.2 Input Legitimacy

5.2.1 Procedures: Data Collection/Data Inputs; Algorithmic Development; Ownership of Algorithms

In the pre-screening phase, the DAAT uses document scanners "to achieve unparalleled scrutiny of travel documents for signs of falsification or counterfeiting" (iBorderCtrl Consortium, 2018b). The Biometrics Module validates the traveler's identity by analyzing fingerprints stored in their

[3] A sample of the iCtrlBorder border guard avatar posing a question can be found here: http://stremble.com/iBorderCtrl/1/1/1/1.mp4) (O'Shea, 2018)

travel document and comparing with uploaded fingerprints. “If needed, this module can also use fingerprints retrieved from different national or European databases” (Ibid.). The interview process also uses the FMT which includes video and photos of the interviewee’s face to create a biometric signature and provide a matching score. Furthermore, the ELSI will crosscheck the traveler’s information from social media platforms. These elements of data collection, though more invasive than current methods of identity verification at border crossings, still appear to be within the reasonable realm of data collection by state authorities. As I will now discuss, the ADDS is the more controversial ADM system that is used by iBorderCtrl.

The original scientific paper released about Silent Talker explains the contention in the literature related to the psychology and physiology of micro-gestures as a means of determining emotional reactions, particularly deceptive behavior (Rothwell et al, 2006). Silent Talker “assumes that certain mental states associated with deceptive behavior will drive an interviewee’s NVB” (O’Shea, 2018).

5.2.2 Concerns

The implications for input legitimacy stem primarily from the novel development of the ADDT that uses the Silent Talker technology. As it is patented, there is limited access to what can be said about how it is created. The ADDT will store the personal data of all citizens entering through border crossings in their databases. It is crucial that the iBorderCtrl database is *securely encrypted* to ensure that no malicious actors can penetrate their systems that contain sensitive data.

In the interviews, the concept of AI Loyalty (Cohen, 2020) was discussed, whereby we need to ensure that an AI system is built in the interest of society. Although the developers assure that economic interests are not at play, the Commission report concludes “the partner organizations of IBORDERCTRL are likely to benefit from this growing European security market—a sector predicted to be worth USD 146 billion (EUR 128 bn) in Europe by 2020” (European Commission, 2020). When firms are competing to release their new ADM systems, checks and balances should be in place so that ethical concerns related to untrustworthy technologies can be considered before deployment.

From the scientific papers it can be seen the many assumptions are made with regard to the physiological science that supports the basis of the datafication of deceptive behavior. Polygraph tests have notoriously debated whether physiological responses can produce accurate results (Rothwell et al, 2006). The developers of Silent Talker claim that, as a machine learning system, “it takes a set of candidate features as input and determines itself which interactions between them, over time, indicated lying” (O’Shea, 2018).

The training data is built on a group of 22 men and 10 women, of which 22 are white European and 10 are classified as Asian/Arab. The original SilentTalker scientific paper concedes that “the system has a truthful bias and is less able at detecting patterns that indicate deception when used with person types it has not been trained upon” (Rothwell et al., 2006). Although this is briefly mentioned in iBorderCtrl’s scientific paper (O’Shea et al, 2018), the argument is made that with more training data, its predictive capacity will increase.

Data mining conducted by the ML algorithm establishes a path dependency. The algorithm will be limited to producing decisions based on which data has been collected. For this reason, emphasis is often placed on the *inclusiveness* of data (Lannquist and Cohen, 2020). If that data were incomplete or biased in any way due to underlying collection methods, sometimes referred to as 'dirty data' (Richardson et. al, 2019), it may not include the interests of all citizens equally. Marginalized citizens may not be included in the training data which, in turn, affects the accuracy of the risk assessment. The decision to push forward demonstrates the *lack of political sensitivity* that the algorithmic developers may have to the contentious nature of the migrant crisis in Europe. In contrast, throughout the interviews it was mentioned that a key concern is the *public servant's limited understanding* of the ADM system (Lannquist and Cohen, 2020). In this case, the border control officer receives training on how to use the handheld device. However, no mechanisms are in place to engage the end-user to understand the technological details regarding how the ADM produces its risk assessment (iBorderControl Consortium, 2017).

Travelers are notified of the collection of their data, yet there is no option to opt-out if you are intent on entering the EU. This issue was also highlighted in the interview with the anonymous public sector policy analyst who stated "we are in a social contract with the government and the way that our privacy laws, our public privacy law, federally is set up with the (government) is that they don't have to rely on our *consent*. So long as the activity in pursuit is authorized by legislation." (Anonymous and Cohen, 2020). The General Data Protection Regulation (GDPR) is not relevant in the context of protecting your data from the public body when it comes to national security. Personal data is collected upon entry when you pass through any airport. We quickly transitioned from removing our shoes before boarding a plane to having a biometric scan of your movement and behavior to better train an algorithm and secure borders.

5.3 Throughput Legitimacy

5.3.1 Procedures: Machine Learning and Predictive Analytics

Throughput legitimacy refers to the 'black box' of decision making in both the democratic procedural sense as well as in the context of ADM. In the traditional bureaucratic setting, this is achieved through transparency of information, accountability for decisions and ongoing citizen participation/engaging civil society. In iBorderCtrl the black box decisions exist in its ML ADDT that uses predictive analytics to produce a risk score.

The ADDT uses the following process to collect data on non-verbal behavior (O'Shea, 2018):

1. Object Locators: Each object locator finds the position of a particular object in the current video frame

2. *Pattern Detectors*: Pattern detectors detect particular states of objects located by the object-locators. Left eye closed given score of 0/1 left eye half closed given score of 0/1

3. Channel Coder: The variations in the state of an object determined by specific pattern detector are referred to as a channel. Channel coding is the process of collecting these variations over a specific time period.

4. Group Channel Coders: Process of amalgamating and statistically summarizing the information from the individual channel coders to form a summary vector, which can be input to the final deception classifier.

5. Deception Classifiers: the deception classifier is a single ANN trained to classify input vectors from the group channel coders as either truthful or deceptive.

The ANN uses the vectors made up of the data collected on micro gestures and begins to learn which series of movements constitute deceptive behavior (Ibid).

5.3.2 Concerns: Liability, Cognitive Outsourcing, Transparency and Accountability

There are several ways by which the accountability of the ADM system must be analyzed. First, for final decisions to be held accountable in a legal sense. In the interview, Brown mentioned on several occasions the importance of *AI Liability* laws. This speaks to the complications that arise when an AI has made a decision, and no human can understand why such an action was taken. Secondly, human oversight has been often cited as a good measure of holding AI accountable, as a human actor can identify malfunctions. Another form of accountability comes in the form of the system's *explicability*, that is, the ADM system's ability to provide justifications for its outputs (Gilpin et al, 2019). In the case of algorithmic governance, accountability and transparency are interlinked. The general consensus is that more transparency and human oversight leads to more accountability.

In terms of human oversight, iBorderCtrl evades liability in cases of false classifications, by suggesting that there is a "*human-in-the-loop principle*" (Spielkamp, 219) which ensures that the risk score provided by the AI Assistant is subjected to review and a final human decision. Gal (2017) has focused on ways which users of an AI assistant cognitively outsource their decision-making capacity as they become increasingly trustworthy of its efficient capacity. Throughout the varying levels of human oversight of an ADM system, Gal argues that even when humans make the final decision their choice field has been limited. This is important because if a border guard makes a false decision with limited knowledge of the algorithmic development of the system, they as the end user, have become so accustomed to trusting the rational AI assistant instead of contesting its output. xAI is important for a citizen to understand why their unemployment insurance was deemed fraudulent, why their bail plea was rejected, or why they were denied entry at a border. Depending on the human oversight and level of automation, whether it be in the loop, on the loop or out of the loop (Rahwan, 2018) plays a significant role in how we hold those decisions accountable. As the border guard increasingly outsources their own decision-making ability to an AI assistant, there may be liability questions related to the responsibility for unjustifiable decisions.

Furthermore, iBorderCtrl concludes in their assessment report of ADDT "since this Deliverable indicates the successful completion of WP3 as a technical development WP, there would be no other opportunity to show the algorithms and the analysis that was carried out and corresponds to the "background" of what will be visible to the border guards and the travelers as final end users." (iBorderCtrl Consortium, 2018b)

Annany and Crawford (2016) suggest that even if civil society was given access to all the data, including the training data, the 'black box' of an ANN makes it impossible to understand how a decision was made. Though the accuracy of predictions or outputs may be more rational or effective than that of a human, the machine's decision-making process is opaque to human understanding (Yampolskiy, 2019). The machine is given input variables with different weighted values and through ANN and deep learning (DL) it determines the most logical output based on the training data which teaches it how to think. This learning process is done between the inputs and outputs, which obscure the interpretability of how a machine achieved a given decision.

As both the academic literature and the developers of Silent Talker argue, if the data inputs and weighted values are publicly available, "a subject could learn how to manipulate (or hide) a particular channel that is known to be important" (Rothwell et al, 2006). This level of transparency is not desirable. The interviewees agreed that a stronger approach is a third-party audit of the development and gathering of data from the beginning of the system's development lifecycle (Lannquist and Cohen, 2020).

The iBorderCtrl communication strategy (iBorderCtrl Consortium, 2019) emphasizes the importance of engaging civil society actors and attending academic conferences. It should be noted that a majority of conferences attended are focused on border security rather than technology and ethics (iBorderCtrl Consortium, 2017, 2018a).

One major complication when it comes to the transparency of information of ADM, are the intellectual property rights of the algorithmic developer. In this case, the Hermes Center for Transparency and Digital Human Rights made a Request for Documents to the Commission for all 24 confidential publications, that include: internal reports relating to the hardware and software technologies used, ethics assessments, annual reports, progress reports (Kampis, 2018). The Commission granted the release of 5 of the requested documents, many of which were largely blacked out. The blacked out sections which seemed most relevant followed the headings that detail the ADDT. The rejection of the other documents was justified “based on the exceptions relating to the protection of the privacy and integrity of the individual and/or commercial interests of a natural or legal person, laid down respectively in Articles 4(1)(b) and 4(2), first indent, of Regulation (EC) No 1049/2001” (Tachelet, 2019). The letter describes that there is no overriding public interest in disclosure, “such an interest must, firstly, be public and, secondly, outweigh the harm caused by disclosure” (Ibid). The Commission argues that public disclosure gives competitors an advantage to develop competing services. Furthermore, they explain that a disclosure of the intricacies of the technology would allow end-users to trick the system. Similarly, Diakopolous (2016) discusses how full transparency is undesirable because people could “game the system”. Copyright law is important to safeguard trade secrets and foster innovation. Establishing a balance between business’ intellectual property and society’s interest could be found in the form of third-party auditors.

5.4 Output Legitimacy

5.4.1 Procedures: Beneficial Outcomes and Fairness

Output legitimacy is judged on the basis of the ability to produce desirable results which are fair and serve the common good of society. iBorderCtrl completed its live testing phase at border crossings in Greece, Hungary and Latvia. Information on the results of those tests remain confidential. Secondly, it is reported that the tests in the three countries were not conducted on incoming third-country nationals, but rather on the border guards themselves (iBorderCtrl, 2020).

The legal framework is not yet in place to allow for real-life testing. As such, I make inferences on the results achieved in the scientific papers by both Silent Talker and iBorderCtrl.

The developers of Silent Talker celebrate that "the system is non-invasive, operates in real-time, does not rely upon a small number of channels/cues and, as an ANN-based system, provides the long-sought objectivity required in profiling" (Rothwell, 2006). The success rates are based on early experiments. When Silent Talker was completed it boasted classification rates between 74% and 87% ($p<0.001$), however the average test accuracy was measured to 73.66 and 75.55% for truthful tests (Rothwell, 2006; O'Shea, 2018). These tests are measured against the likelihood that the machine assessed the deceptiveness of the individual by chance as opposed to the successful ML algorithm. Both papers speak to the success of the ANN at classifying above the statistical likelihood of a chance-based correct classification.

5.4.2 Concerns: Bias, Algorithmic Malfunction and Liability

In Trial C (Rothwell, 2006) there was a specific attempt to investigate how gender and ethnic/cultural differences could play a role in the outputs made by the ANN. The skewed training sets which consisted of mostly men and mostly white Europeans were tested by *only* feeding the ANN training data on the male and white European participants. Following this, they attempted to classify deceptive and truthful behavior of European women, non-European men and non-European women. The trial, which was repeated 64 times, demonstrated that classification accuracy fell drastically (Rothwell, 2006). The developers supported these results by providing literature on the differences between the behavior of men and women, and the behavior of people of different cultures. They state "the system has a truthful bias and is less able at detecting patterns that indicate deception when used with person types it has not been trained upon" (Rothwell, 2006).

The group of developers at iBorderCtrl acknowledge this shortcoming stating, "the unbalanced dataset in terms of ethnicity and gender might influence the deception classification network performance. For instance the deceptive dataset consists of 4 Asian/Arabic participants compared to 13 of White EU" (O'Shea, 2018). As border guards increasingly trust the claimed objectivity of an ADM, the data should not reconfirm pre-existing social biases or stereotypes. Particularly in the iBorderCtrl case, the majority of training is done on white European males, and the technology is deployed on the borders of Europe. Oftentimes there will be non-white migrants or travelers entering through a system which is less capable of detecting their micro gestures. The result is likely to demonstrate that non-white males and females are subject to a less accurate technology that is prone to producing false positives and incorrect classifications.

Though the subject of biased output decisions by ADM is contentious, there are those who argue that a machine will not be more biased than a human. In this case, the argument would be such that a border agent is more likely to hold implicit biases on the basis of their political, social and cultural upbringing. If we know that humans make poor decisions, why should we expect more from an ADM (Cave et al, 2018). Miller (2018) describes a study on a job-screening algorithm at a software company in which the algorithm "favored non-traditional candidates" in particular those without top ranking universities on their CV. A study at the National Bureau of Economic Research analyzes the way in which a behavioral/predictive algorithm could replace the job of a judge that makes bail decisions. The study found that the algorithm could achieve significantly more equitable decisions with “jailing rate reductions of up to 41% with no increase in crime rates” (Kleinberg et al, 2017). The report goes on to suggest that the jailing rate reductions can be achieved “while simultaneously reducing racial disparities” (Ibid.). The debate around algorithmic bias implies that increased objectivity and lesser bias is possible but must be worked on.

6. Discussion and Conclusions

The thesis has investigated the relationship between political legitimacy and the algorithmic lifecycle of ADM tools for civil servants to make more efficient, objective and cost-effective decisions. The categorization of AGOV into a political legitimacy framework contributed how the algorithmic development cycle is subject to the interplay of input, output and throughput legitimacy concerns. Input legitimacy is particularly affected by the ways in which data is collected on the individual as well as the ownership rights to the algorithms and personal data. Outsource public services to apolitical developers in turn code the law inside of algorithms. Throughput legitimacy is most drastically impacted due to the opacity of decision-making inside of DNN. This form of layered learning creates concerns for the way by which society can hold decision makers accountable. To judge the viability of output decisions, early detection of algorithmic concerns are needed at the input and throughput levels respectively.

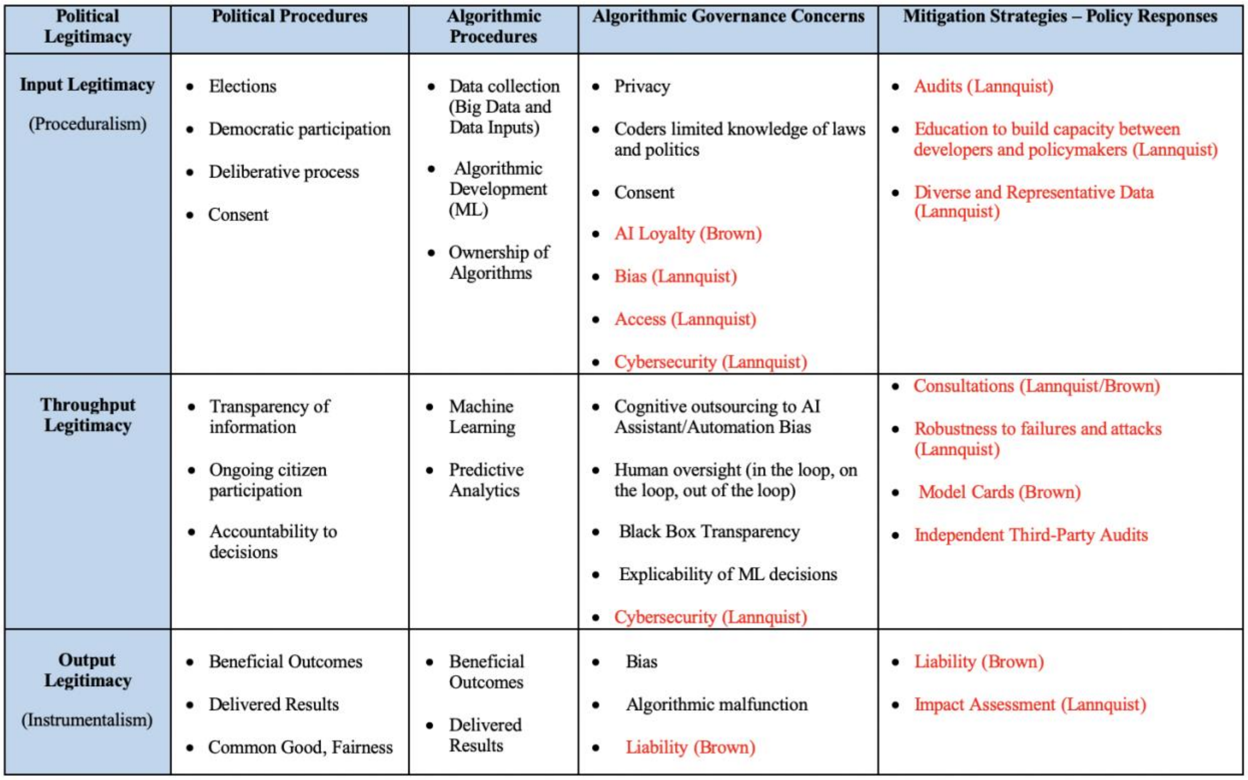
Contributions

The non-binary approach to this legitimacy framework, which includes the black-box of decision making, offers new avenues to study the ways by which throughput legitimacy can support the input of citizens' voices into the political decision-making process as well as developing fair, equal and socially beneficial outcomes. Only regulating outputs in the hopes that this will avoid innovation deterrence is not a viable strategy given the interconnectivity of legitimacy concerns. This framework has demonstrated that regulation is required at the various stages of the algorithmic development cycle.

The case specific contribution demonstrates the pressing reality of the integration of AI into government. iBorderCtrl demonstrates the importance of building trust in ADM and possible avenues for doing so. Scrutiny of AI technologies uncover higher stakes when used for delivering public services. Citizens' ability to consent to an engagement with a private entity is different than the social contract and checks and balances are needed to ensure ADM serves societies interest.

Mitigation Strategies

Figure 4 Suggestions Algorithmic Legitimacy Framework with



Political Legitimacy	Political Procedures	Algorithmic Procedures	Algorithmic Governance Concerns	Mitigation Strategies – Policy Responses
Input Legitimacy (Proceduralism)	• Elections • Democratic participation • Deliberative process • Consent	• Data collection (Big Data and Data Inputs) • Algorithmic Development (ML) • Ownership of Algorithms	• Privacy • Coders limited knowledge of laws and politics • Consent • AI Loyalty (Brown) • Bias (Lannquist) • Access (Lannquist) • Cybersecurity (Lannquist)	• Audits (Lannquist) • Education to build capacity between developers and policymakers (Lannquist) • Diverse and Representative Data (Lannquist)
Throughput Legitimacy	• Transparency of information • Ongoing citizen participation • Accountability to decisions	• Machine Learning • Predictive Analytics	• Cognitive outsourcing to AI Assistant/Automation Bias • Human oversight (in the loop, on the loop, out of the loop) • Black Box Transparency • Explicability of ML decisions • Cybersecurity (Lannquist)	• Consultations (Lannquist/Brown) • Robustness to failures and attacks (Lannquist) • Model Cards (Brown) • Independent Third-Party Audits
Output Legitimacy (Instrumentalism)	• Beneficial Outcomes • Delivered Results • Common Good, Fairness	• Beneficial Outcomes • Delivered Results	• Bias • Algorithmic malfunction • Liability (Brown)	• Liability (Brown) • Impact Assessment (Lannquist)

Based on the iBorderCtrl case, preliminary mitigation strategies speak to the interconnectedness of the different levels of input, throughput and output legitimacy. In order to address output concerns of fairness and equality, the collection of data and training data as well as the development of algorithms should be subject to third party audits. The scientific papers demonstrated that a technology, based on questionable scientific findings, is set to be implemented into society on the premise of objectivity and efficiency. Audits could promote AI loyalty towards citizens instead of self-interested private entities. The accountability issues related to cognitive outsourcing and liability, transparency and explicability of the ADDS are difficult to manage. It is not clear if explicability as a goal can ever be achieved, which reinforces the need for more transparent and ethical data mining activities. Model cards have been proposed in the AI community to publish how algorithmic models are built so that a scientific method could allow others to replicate the same experiment. Implementing ADM systems which reinforce pre-existing social biases from technocratic authority should be subject to public scrutiny. Closing the information gap between policy-makers, civil servant end-users and ML developers will be essential to human-in-the-loop accountability. At the output level, impact assessments could prove to be useful in order to have checks and balances on the ways which ADM systems affect citizens across the board. Policies should strive to secure AI as a technology that promotes the safe and fair development of society, not one that reinforces a feedback loop of historical prejudices.

Future Research

Although the case study is not representative of all ADM systems set to enter society for public service provision, the politicization of the algorithmic development cycle helps policymakers understand the complexity of needed policy responses. Areas for future studies could go in two especially helpful directions. First, having a larger sample size of experts could lead to a more refined framework. This could include end user civil servants to better understand how they see their role with regard to their AI assistant. A particular limitation of this paper included the omission of algorithmic developers working in partnership with governments. This could have provided deeper perspectives into the development of the legitimacy framework as well as the role and responsibility of developers. Second, as more examples of AI assistants in the public service emerge, applying this model to more cases would help to uncover more varied solutions. Further, interesting research could empirically calculate the benefits from AI assistants in contrast with human decision makers.

Final Thoughts

In a broader sense, as ADM becomes further integrated into society, it will be important to understand the role of the state in relation to citizens subjected to this technocratic authority. With limited access to data training sets and the algorithms themselves, the increased reliance on technocratic AI systems could very well lead society down a path towards rule by algorithms, algocracy. Human-centric and ethical AI in the private sector has been well documented, however this paper has attempted to provoke thought on how AI will fundamentally change the unique relationship between citizen and state. The social contract is much different than a private contract between an individual and a business. Today, consent to the provision of your data for the betterment of society is taken as a given. We must question technological band-aids to complex political issues. A paradigm shift towards is underway as citizens interact with their government as if it were another service provider.

Abbreviations

ADDS:	Automatic Deception Detection System
AGOV:	Algorithmic Governance
API:	Application Programming Interface
AI:	Artificial Intelligence
ANN:	Artificial Neural Networks
ADM:	Automated Decision Making
DL:	Deep Learning
DNN:	Deep Neural Networks
DAAT:	Document Authenticity Analytics Tool
EU:	European Union
ELSI:	External Legacy and Social interfaces system
FMT:	Face Matching Tool
GDPR:	General Data Protection Regulation
HDD:	Hidden Human Detection Tool
BCAT:	Integrated Border Control Analytics Tool
IO:	International Organization
ML:	Machine Learning
NGO:	Non-Governmental Organization
RBAT:	Risk Based Assessment Tool
xAI:	Explicable AI

Bibliography

"A Comprehensive European Industrial Policy on Artificial Intelligence and Robotics." 2019. European Parliament.

Ananny, Mike, and Kate Crawford. 2016. "Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability." *New Media and Society*, 1–17.

Aneesh, Aneesh. 2002. "Technologically Coded Authority." In . Austria: Stanford University.

Anonymous, and Daniel Cohen. 2020. Public Sector Anonymous Interview: Transcript. Google Meet.

Barocas, Solon, and Andrew D. Selbst. 2016. "Big Data's Disparate Impact." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2477899.

Baum, Seth D. 2020. "Forthcoming: Artificial Interdisciplinarity: Artificial Intelligence for Research on Complex Societal Problems." *Philosophy & Technology*, July. https://doi.org/10.1007/s13347-020-00416-5.

Bertsou, Eri, and Giulia Pastorella. 2017. "Technocratic Attitudes: A Citizens' Perspective of Expert Decision-Making." *West European Politics* 40 (2): 430–58. https://doi.org/10.1080/01402382.2016.1242046.

Binns, Reuben. 2018. "Fairness in Machine Learning: Lessons from Political Philosophy." In *Machine Learning Research*, 11.

Blind, Peride. 2006. "Building Trust in Government in the Twenty-First Century - Review of Literature." In *Building Trust in Government*, 31. Vienna, Austria.

Bohman, James, and William Rehg, eds. 1997. *Deliberative Democracy: Essays on Reason and Politics*. Cambridge, Mass: MIT Press.

Bostrom, Nick, Allan Dafoe, and Carrick Flynn. 2018. "Policy Desiderata for Superintelligent AI: A Vector Field Approach," 29.

Bowman, Kirk, Fabrice Lehoucq, and James Mahoney. 2005. "Measuring Political Democracy: Case Expertise, Data Adequacy, and Central America." *Comparative Political Studies* 38 (8): 939–70. https://doi.org/10.1177/0010414005277083.

Brantingham, P. Jeffrey, Matthew Valasik, and George O. Mohler. 2018. "Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial." *Statistics and Public Policy* 5 (1): 1–6. https://doi.org/10.1080/2330443X.2018.1438940.

Brown, Jared, and Daniel Cohen. 2020. Senior Advisor for Government Affairs Jared Brown (FLI) Interview: Transcript. Skype.

Brundage, Miles, Shahar Avin, Jasmine Wang, Haydn Belfield, Gretchen Krueger, Gillian Hadfield, Heidy Khlaaf, et al. 2020. "Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims." *ArXiv:2004.07213 [Cs]*, April. http://arxiv.org/abs/2004.07213.

Buchanan, Allen. 2002. "Political Legitimacy and Democracy." *Ethics* 112 (4): 689–719. https://doi.org/10.1086/340313.

Cathy O'Neil. 2016. *Weapons of Math Destruction*. Crown Books.

Christiano, Thomas. 2004. "The Authority of Democracy." *Journal of Political Philosophy* 12 (3): 266–90. https://doi.org/10.1111/j.1467-9760.2004.00200.x.

Citron, Danielle Keats, and Frank Pasquale. 2014. "The Scored Society: Due Process for Automated Predictions." *Washington Law Review* 89: 33.

"Commission White Paper: Artificial Intelligence - A European Approach to Excellence and Trust." 2020. European Commission.

Danaher, John. 2016. "The Threat of Algocracy: Reality, Resistance and Accommodation." *Philosophy & Technology* 29 (3): 245–68. https://doi.org/10.1007/s13347-015-0211-1.

Danaher, John. 2018. "Toward an Ethics of AI Assistants: An Initial Framework." *Philosophy & Technology* 31 (4): 629–53. https://doi.org/10.1007/s13347-018-0317-3.

Danaher, John, Michael J Hogan, Chris Noone, Rónán Kennedy, Anthony Behan, Aisling De Paor, Heike Felzmann, et al. 2017. "Algorithmic Governance: Developing a Research Agenda through the Power of Collective Intelligence." *Big Data & Society* 4 (2): 205395171772655. https://doi.org/10.1177/2053951717726554.

Datta, Anupam, Shayak Sen, and Yair Zick. 2016. "Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems." In *IEEE*, 20.

Della Sala, Vincent. 2010. "Political Myth, Mythology and the European Union*: POLITICAL MYTH, MYTHOLOGY AND THE EUROPEAN UNION." *JCMS: Journal of Common Market Studies* 48 (1): 1–19. https://doi.org/10.1111/j.1468-5965.2009.02039.x.

Desouza, Kevin C., and Benoy Jacob. 2017. "Big Data in the Public Sector: Lessons for Practitioners and Scholars." *Administration & Society* 49 (7): 1043–64. https://doi.org/10.1177/0095399714555751.

Diakopoulos, Nicholas. 2016. "Accountability in Algorithmic Decision Making." *Communications of the ACM* 59 (2): 56–62. https://doi.org/10.1145/2844110.

"Directive on Automated Decision Making." 2018. Treasury Board of Canada.

Dressel, Julia, and Hany Farid. 2018. "The Accuracy, Fairness, and Limits of Predicting Recidivism." *Science Advances* 4 (1): eaao5580. https://doi.org/10.1126/sciadv.aao5580.

Easton, David. 1965. *A Systems Analysis of Political Life*. Prentice Hall.

Engin, Zeynep, and Philip Treleaven. 2019. “Algorithmic Government: Automating Public Services and Supporting Civil Servants in Using Data Science Technologies.” *The Computer Journal* 62 (3): 448–60. https://doi.org/10.1093/comjnl/bxy082.

Estlund, David M. 2008. “Democratic Authority.” In *Democratic Authority*, 21.

Estlund, David M. 2008. “Epistemic Proceduralism.” In *Democratic Authority*, 20.

Eurobarometer. 2019. “Standard Eurobarometer 92: Europeans and Artificial Intelligence.” European Commission.

European Commission. 2018. “Smart Lie-Detection System to Tighten EU’s Busy Borders.”

European Commission. 2020. “High-Level Expert Group on Artificial Intelligence.” Text. Shaping Europe’s Digital Future - European Commission. 2020. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence.

“Expert Group on Liability and New Technologies: Report Liability for AI and Other Digital Technologies.” 2019. EU Commission.

Feldman, Stanley. 1983. “The Measurement and Meaning of Trust in Government.” *Cambridge University Press* 9 (3): 15.

Fine Licht, Jenny de. 2011. “Do We Really Want to Know? The Potentially Negative Effect of Transparency in Decision Making on Perceived Legitimacy: Do We Really Want to Know? The Potentially Negative Effect of Transparency in Decision Making on Perceived Legitimacy.” *Scandinavian Political Studies* 34 (3): 183–201. https://doi.org/10.1111/j.1467-9477.2011.00268.x.

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. 2020. “Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI.” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3518482.

Gal, Michal S. 2017. “Algorithmic Challenges to Autonomous Choice.” *Michigan Telecommunications and Technology Law Review*, 40.

Gal, Michal S, and Niva Elkin-Koren. 2017. “Algorithmic Consumers.” *Harvard Journal of Law and Technology* 30: 45.

Gaus, Gerald. 2010. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511780844.

Gaus, Gerald F. 2008. “The (Severe) Limits of Deliberative Democracy as the Basis for Political Choice.” *Theoria* 55 (117). https://doi.org/10.3167/th.2008.5511703.

Geeraert, Arnout, and KU Leuven. 2014. “New EU Governance Modes in Professional Sport: Enhancing Throughput Legitimacy” 10 (3): 20.

Gerring, John, and Lee Cojocaru. 2015. “Case-Selection: A Diversity of Methods and Criteria,” 31.

Giest, Sarah. 2017. “Big Data for Policymaking: Fad or Fasttrack?” *Policy Sciences* 50 (3): 367–82. https://doi.org/10.1007/s11077-017-9293-1.

Gilpin, Leilani H., David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. 2019. “Explaining Explanations: An Overview of Interpretability of Machine Learning.” *ArXiv:1806.00069 [Cs, Stat]*, February. http://arxiv.org/abs/1806.00069.

Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. 2018. “When Will AI Exceed Human Performance? Evidence from AI Experts.” *ArXiv:1705.08807 [Cs]*, May. http://arxiv.org/abs/1705.08807.

Greene, Amanda. 2016. “Consent and Political Legitimacy.” In *Oxford Studies in Political Philosophy, Volume 2*, 1st edition. Oxford Studies in Political Philosophy. NewYork, NY: Oxford University Press.

Helbing, Dirk, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V. Zicari, and Andrej Zwitter. 2019. “Will Democracy Survive Big Data and Artificial Intelligence?” In *Towards Digital Enlightenment*, edited by Dirk Helbing, 73–98. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-90869-4_7.

iBorderCtrl Consortium. 2017. “H2020 – BES – 5 – 2015 Dissemination and Communication Plan 17-D7-3-Dissemination-and-Communication-Plan.” Ref. Ares(2017)4690395-26/09/2017.

iBorderCtrl Consortium. 2018a. “H2020 – BES – 5 – 2015 Dissemination and Communication Plan 21-D7-8-Dissemination-and-Communication-Plan-2.” Ref. Ares(2018)5019050-01/10/2018.

iBorderCtrl Consortium. 2018b. “H2020 – BES – 5 – 2015 Research Innovation Action - 9-D3-3 2nd Version of All Tech Tools and Subsystems for Integration.” Technical Description Ref. Ares(2018)562507309362.

iBorderCtrl Consortium. 2019. “H2020 – BES – 5 – 2015 D7 6 Yearly Communication Report Including Communication Material.”

Just, Natascha, and Michael Latzer. 2017. “Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet.” *Media, Culture & Society* 39 (2): 238–58. https://doi.org/10.1177/0163443716643157.

Kampis, Barbara. 2018. “European Commission: Application for Access to Documents on Project IBorderCtrl (700626) REA-Reply-Ares-2018-65003962,” December 17, 2018.

Katzenbach, Christian, and Lena Ulbricht. 2019. “Algorithmic Governance.” *Internet Policy Review* 8 (4). https://doi.org/10.14763/2019.4.1424.

King, Gary, Robert Keohane, and Sydney Verba. 1994. “The Science in Social Science.” In *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, New Jersey: Princeton University Press.

Kleinberg, John, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. “HUMAN DECISIONS AND MACHINE PREDICTIONS.Pdf.” *National Bureau of Economic Research*.

Kolodny, Niko. 2014. “Rule Over None II: Social Equality and the Justification of Democracy: Rule Over None II: Social Equality and the Justification of Democracy.” *Philosophy & Public Affairs* 42 (4): 287–336. https://doi.org/10.1111/papa.12037.

Krigel, Tina, RA Benjamin Schitze, and Jonathan Stoklas. 2018. “Legal, Ethical and Social Impact on the Use of Computational Intelligence Based Systems for Land Border Crossings.” In *2018 International Joint Conference on Neural Networks (IJCNN)*, 1–8. Rio de Janeiro: IEEE. https://doi.org/10.1109/IJCNN.2018.8489349.

Lafont, Cristina. 2015. “Deliberation, Participation, and Democratic Legitimacy: Should Deliberative Mini-Publics Shape Public Policy?: Deliberation, Participation & Democratic Legitimacy.” *Journal of Political Philosophy* 23 (1): 40–63. https://doi.org/10.1111/jopp.12031.

Landemore, Hélène. 2009. “Democratic Reason: Why the Many Are Smarter than the Few and Why It Matters.” In , 29.

Lannquist, Yolanda, and Daniel Cohen. 2020. Head of Research and Advisory Yolanda Lannquist (Future Society) Interview Transcript. In-person.

Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace*. Basic Books.

Levi, Margaret, Audrey Sacks, and Tom Tyler. 2009. “Conceptualizing Legitimacy, Measuring Legitimating Beliefs.” *American Behavioral Scientist* 53 (3): 354–75. https://doi.org/10.1177/0002764209338797.

List, Christian, and Robert E. Goodin. 2001. “Epistemic Democracy: Generalizing the Condorcet Jury Theorem.” *Journal of Political Philosophy* 9 (3): 277–306. https://doi.org/10.1111/1467-9760.00128.

Majchrzak, Ann, and M. Lynne Markus. 2013. “Make a Difference with Policy Research.” In *Methods for Policy Research*. 3 Applied Social Research Methods. Sage Publications Ltd.

Manheim, Karl, and Lyric Kaplan. 2019. “Artificial Intelligence: Risks to Privacy and Democracy.” *Yale Journal of Law and Technology* 21 (106): 83.

Manin, Bernard. 1987. "On Legitimacy and Political Deliberation." *Political Theory* 15 (3): 338–68. https://doi.org/10.1177/0090591787015003005.

Manzano, Ana. 2016. "The Craft of Interviewing in Realist Evaluation." *Evaluation* 22 (3): 342–60. https://doi.org/10.1177/1356389016638615.

Mehr, Hila. 2017. "Artificial Intelligence for Citizen Services and Government," 19.

Mény, Yves. 2002. "De La Démocratie En Europe Old Concepts and New Challenges.Pdf." In , 41:1–13. 1. Blackwell Publishing.

Mikhaylov, Slava Jankin, Marc Esteve, and Averill Campion. 2018. "Artificial Intelligence for the Public Sector: Opportunities and Challenges of Cross-Sector Collaboration." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376 (2128): 20170357. https://doi.org/10.1098/rsta.2017.0357.

Miller, Alex P. 2018. "ORLAGH MURPHY/GETTY IMAGES." *Harvard Business Review*, 5.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society* 3 (2): 205395171667967. https://doi.org/10.1177/2053951716679679.

Näsström, Sofia. 2007. "The Legitimacy of the People." *Political Theory* 35 (5): 624–58. https://doi.org/10.1177/0090591707304951.

"NSTC Select Committee on Artificial Intelligence November 2018 Update." 2018. White House.

Olsen, Johan P., Alberta Sbragia, and Fritz W. Scharpf. 2000. "Symposium: Governing in Europe: Effective and Democratic?" *Journal of European Public Policy* 7 (2): 310–24. https://doi.org/10.1080/135017600343214.

"Opinion of the European Economic and Social Committee on 'Artificial Intelligence — The Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society' (Own-Initiative Opinion)." n.d., 9.

Ortolano, Leonard, and Anne Shepherd. 1995. "ENVIRONMENTAL IMPACT ASSESSMENT: CHALLENGES AND OPPORTUNITIES." *Impact Assessment* 13 (1): 3–30. https://doi.org/10.1080/07349165.1995.9726076.

O'Shea, James, Keeley Crockett, Wasiq Khan, Philippos Kindynis, Athos Antoniades, and Georgios Boultadakis. 2018. "Intelligent Deception Detection through Machine Based Interviewing." In *2018 International Joint Conference on Neural Networks (IJCNN)*, 1–8. Rio de Janeiro: IEEE. https://doi.org/10.1109/IJCNN.2018.8489392.

Pagallo, U. 2012. "Cracking down on Autonomy: Three Challenges to Design in IT Law." *Ethics and Information Technology* 14 (4): 319–28. https://doi.org/10.1007/s10676-012-9295-9.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pateman, Carole. 1970. *Participation and Democratic Theory*. 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511720444.

Perry, Walt L. 2013. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Santa Monica, CA: RAND.

Peter, Fabienne. 2007. "Democratic Legitimacy and Proceduralist Social Epistemology." *Politics, Philosophy & Economics* 6 (3): 329–53. https://doi.org/10.1177/1470594X07081303.

Peter, 2017. "Political Legitimacy." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2017. Metaphysics Research Lab, Stanford University.

Pettit, Philip. 2012. "On the People's Terms A Republican Theory and Model of Democracy." *Cambridge University Press*, 3.

Pettit, Philip, and Wlodek Rabinowicz. 2001. "Deliberative Democracy and the Discursive Dilemma." *Ridgeview Publishing Company* 11: 33.

Poel, Martijn, Eric T. Meyer, and Ralph Schroeder. 2018. "Big Data for Policymaking: Great Expectations, but with Limited Progress?: Big Data for Policymaking." *Policy & Internet* 10 (3): 347–67. https://doi.org/10.1002/poi3.176.

Rahwan, Iyad. 2018. "Society-in-the-Loop: Programming the Algorithmic Social Contract." *Ethics and Information Technology* 20 (1): 5–14. https://doi.org/10.1007/s10676-017-9430-8.

Rawls, John. 1971. "A Theory of Justice." In *A Theory of Justice*. Cambridge, Mass: The Belknap Press of Harvard University.

Richardson, Rashida, Jason M Schultz, and Kate Crawford. 2019. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *New York University Law Review* 94: 42.

Richardson, Rashida, Jason M Schultz, and Vincent M Southerland. n.d. "LITIGATING ALGORITHMS 2019 US REPORT:," 32.

Ripstein, Arthur. 2004. "Authority and Coercion." *Philosophy and Public Affairs* 32 (1): 2–35. https://doi.org/10.1111/j.1467-6486.2004.00003.x.

Rothwell, Janet, Zuhair Bandar, James O'Shea, and David McLean. 2006. "Silent Talker: A New Computer-Based System for the Analysis of Facial Cues to Deception." *Applied Cognitive Psychology* 20 (6): 757–77. https://doi.org/10.1002/acp.1204.

Sánchez-Monedero, Javier, and Lina Dencik. 2019. "The Politics of Deceptive Borders: 'Biomarkers of Deceit' and the Case of IBorderCtrl." *Cardiff University*, 17.

Scharpf, Fritz. 1999. *Governing in Europe: Effective and Democratic?* Oxford Scholarship Online

Schmidt, Vivien A. 2013. "Democracy and Legitimacy in the European Union Revisited: Input, Output and 'Throughput.'" *Political Studies* 61 (1): 2–22. https://doi.org/10.1111/j.1467-9248.2012.00962.x.

Schmidt, Vivien, and Matthew Wood. 2019. "Conceptualizing Throughput Legitimacy: Procedural Mechanisms of Accountability, Transparency, Inclusiveness and Openness in EU Governance." *Public Administration* 97 (4): 727–40. https://doi.org/10.1111/padm.12615.

Spielkamp, Matthias. 2019. "Automating Society Report." AW AlgorithmWatch gGmbH.

Stockmann, Daniela. 2018. "Toward Area-Smart Data Science: Critical Questions for Working With Big Data From China: Toward Area-Smart Data Science." *Policy & Internet* 10 (4): 393–414. https://doi.org/10.1002/poi3.192.

Tachelet, Marc. 2019. "European Commission: Article 7(2) of Regulation (EC) No 10492001 Application for Access to Documents Letter-to-R-Coluccini," 2019.

"The Project | IBorderCtrl." 2020. IBorderCtrl. 2020. https://www.iborderctrl.eu/The-project.

Waldman, Ari Ezra. 2019. "Power, Process, and Automated Decision-Making." *Fordham Law Review* 88: 21.

Weatherford, M. Stephen. 1992. "Measuring Political Legitimacy." *American Political Science Review* 86 (1): 149–66. https://doi.org/10.2307/1964021.

Weber, Max, Talcott Parsons, and AM Henderson. 1964. *Theory Of Social And Economic Organization*. A Free Press Paperback. London, England: Collier Macmillan Publishers.

Weston, Anthony. 2000. *A Rulebook for Arguments*. 3rd ed. Indianapolis: Hackett Pub. Co.

Whittaker, Meredith, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Myers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. 2018. "AI Now Report 2018." New York: AI Now.

Yampolskiy, Roman V. 2019. "Unexplainability and Incomprehensibility of Artificial Intelligence." *University of Louisville*, 14.

Zarsky, Tal Z. 2011. "Governmental Data Mining and Its Alternatives." *Penn State Law Review* 116: 46.

Appendix

Anonymous, and Daniel Cohen. 2020. Public Sector Anonymous Interview Google Meet. (Transcript available on request)

Brown, Jared, and Daniel Cohen. 2020. Senior Advisor for Government Affairs Jared Brown (FLI) Interview. Skype (Transcript available on request)

Lannquist, Yolanda, and Daniel Cohen. 2020. Head of Research and Advisory Yolanda Lannquist (Future Society) Interview. In-person (Transcript Available on request)

Special annex: Thesis Report

Thesis Report

Central European University / Institut Barcelona Estudis Internacionals

AI Assistants for Government Decision Makers: Threats to Democratic Accountability and Authority

Daniel Cohen
Erasmus Mundus Masters in Public Policy

Thesis Supervision

Prof Simon Rippon
Associate Professor of Philosophy, Central European University

Prof Aina Gallego Dobón
Associate Professor, Institut Barcelona Estudis Internacionals

Student number: 104602
Email: cohen_daniel@spp.ceu.edu
Number of words: 5,500

Submitted: August 30, 2019

Table of Contents

1.0 - Introduction

Big data, machine learning (ML) and artificial intelligence (AI) serve as an effective tool for policy makers to make informed decisions very efficiently through the collection of vast amounts of data and algorithms which can calculate outcomes at levels which exceed human cognitive capacities. However, this transition towards algorithmic policy making is a threat to democratic legitimacy and accountability to citizens. Decision makers' use of AI assistants enters the debate of democratic legitimacy through an instrumentalist lens in which too much emphasis is placed on the efficient outcome of an algorithm as opposed to the democratic procedure which achieves it. Cognitive outsourcing to AI assistants removes the autonomy of individual decision makers and will make them less accountable for the decisions that they make. There is concern that the algorithms, which form the foundation of AI assistants, lack transparency and contain a 'black box' in which humans cannot understand how results are produced. Without an understanding of the ways in which AI assistants make decisions, policy makers cannot be held accountable by the voters who grant them authority within a democracy.

Appropriate policies and technological improvements could ensure that algorithms will be ethical by design to ensure bias mitigation and a reasonable level of transparency which allows for the accountability of decisions made while not compromising the need for some opacity so as not to jeopardize the system. Furthermore, it is important to determine the level of human oversight required over an AI assistant for policymaking, in order to further ensure democratic accountability to the electorate.

2.0 - Hypothesis

The research will demonstrate that there is a strong relationship between the transition towards AI assistants for decision makers and its impacts on democratic legitimacy and authority. Data-driven algorithms for policy design could reinforce current beliefs and may not be the only approach to improving the policy-making processes. This could confine policy options of what is possible given the available data. The obsession with data driven and evidence-based policy making ties to the pure instrumentalist view of democratic authority, which leaves democratic procedures at the wayside. The use of AI assistants for governing decision makers will likely indicate a crossroads that requires a critique of our passive attitude to give up society's autonomy for deliberative democracy in the face of smart machines. Unless proper actions are not taken to ensure the human oversight for accountable decision making, understandable algorithms with regard to transparency and biases as well as an emphasis on maintaining the autonomy of decision-makers, society risks forfeiting instrumental democratic values. The thesis will propose ways in which current policies can ensure autonomous human oversight remains with regards to new systems of big data driven and algorithmic policy-making.

3.0 - Literature Review

3.1 Artificial Intelligence and its Usability in Government

3.1.1 - Big Data, Algorithms and AI

A general understanding of AI is necessary in order to delve deeper into the benefits it can provide as a tool to public sector decision makers. Russell and Norvig have laid out the most basic description of AI in which it is broken down into eight definitions and split into four categories: think like a human, act like a human, think rationally and act rationally (Russell and Norvig, 2010). More recently, academics have honed in on the latter part of the definition pertaining to rationality which states that AI is a system "that acts so as to achieve the best outcome, or when there is uncertainty, the best expected outcome" (Russell and Norvig, 2010; Danaher, 2016).

Over the last decade the developments of new technologies have led to an explosion of AI systems throughout our society. Big data refers to the "volume, velocity, variety and complexity" (Desouza and Jacob, 2014) of data which is collected throughout our society. Our actions throughout our digital surroundings are recorded at unprecedented levels and the limits of what data is collected and retained appears unbounded, in that its limits are unclear (Ibid.). With Big Data, datasets are capable of storing much larger quantities and algorithmic processes are capable of deciphering connections between complex, varied and massive amounts of data at unprecedented rates.

Zarsky (2011) explains the process of data mining as "the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data" (Zarsky, 2011). Beyond the collection of these large data pools, which is impressive in and of itself, mathematical algorithms are used to extract the information and draw connections between specific data points that would otherwise go unnoticed by the human eye. When we refer to algorithms in the context of AI and Big Data, we are not referring simply to the mathematical construct of an algorithm, but rather "the implementation and interaction of one or more algorithms in a particular program, software or information system" (Mittelstadt et al, 2016). These algorithms are applied to Big Data datasets in which data mining can occur with both descriptive and predictive functions. This is to say that a descriptive function would allow for an algorithm to search through a given data set and explain what has already happened, such as going through financial records in order to detect fraud (Danaher, 2016). Data mining takes a greater leap when it begins to use algorithms to make predictions about future outcomes based on historical data, this is called predictive analytics. An example of this could be recent use cases of predictive policing, credit, insurance and employment screening (Zarsky, 2011; Mittelstadt, 2016).

Machine Learning (ML) is a Pandora's box in regard to predictive algorithms applied to Big Data. ML is the technology which allows for the machine to go beyond the original tasks set out in the codes that instruct it how to act. Van Otterlo (2013) describes the process of ML as "any methodology and set of techniques that can employ data to come up with novel patterns and knowledge and generate models that can be used for effective predictions about data". The implication here is that ML goes beyond the original programmed functions of the algorithm to gather more information and recognize new patterns to make more accurate predictions about uncertain future outcomes.

3.1.2. - AI Assistants

AI assistants have been generously welcomed into our lives in the form of digital tools which help us "search, plan, message, schedule and so on" (Danaher, 2018). AI assistants are commonplace in our daily lives as we interact with our cell phones and web services that can help us discover music, purchase items, and tailor our digital existence to our personal preferences (Gal, 2016). Danaher sets out a proper definition of personal AI assistants that I will refer to throughout this paper, in which they are defined as "any computer-coded software/program that can act in a goal directed manner … that can set some target output and can select among a range of options that optimizes (according to specific metrics) for that output" (Danaher, 2018).

The advanced algorithms which form AI Assistants are based on coded decision trees that will assign various weights to different decision-making inputs which set the parameters for which an optimal recommendation can be made (Gal, 2016). As the AI assistant employs ML, the algorithm "self-adjusts based on its own analyses of data previously encountered, freeing the algorithm from predefined preferences" (Ibid). The overall aim of the AI assistant is to personalize and facilitate a decision-making process for the user, based on previous data which demonstrates what the most probable best option would be.

The contribution which I seek to make in the literature is situated in the expansive shift of personal AI assistants from the private sector, that is for commercial uses, to the usage of AI assistants for public sector decision makers. Mittelstadt et al (2016) and Danaher (2016) have also set out along this path of highlighting the usage of AI assistants in the public sphere and attempted to discuss some of the potential issues associated with this level of cognitive outsourcing, and I hope to contribute further to the questions of democratic accountability, transparency in algorithms and human oversight.

3.1.3 - Big Data Governance

Applications of Big Data and AI are widely used throughout the private sector and the slow-moving nature of the public sector has made the emergence of AI products for evidence-based decision making a more recent development. Data governance is not new per se however, the applications of AI and Big Data extend its abilities to make more accurate predictions as a result of data mining and its predictive capacities. Giest (2016) describes data culture in government as "the capacity of both individual civil servants as well as the organizations as a whole to collect, merge and utilize big data and the institutional structure supporting this through training civil servants and open data initiatives". The vast data which is collected throughout government domains such as tax systems, social programs and health records can now be digitized and used by decision makers when formulating policies in education, economics, health and social welfare (Giest, 2016). At the government level, the pairing of artificial intelligence and the analysis of big data allows algorithms to make more accurate decisions (Gal, 2016).

Digital era governance emerges as a technological opportunity for governments to transition towards a more service oriented and accountable actor to its citizenry (Giest, 2016). Some of the initial ways in which these machines will be able to support government have been outlined by Engin and Treleavan (2019) among others with advancements that include: Government Data Facilities, Internet of Things (IoT), Big Data, Behavioural/Predictive Analytics and Blockchain Technologies. From these large pools of data and metadata, behavioural and predictive analytics are capable of synthesizing this information and uncovering hidden patterns, unknown correlations and personal preferences, something which would be impossible for a human (Engin and Treleavan, 2019).

Technical rationality as proposed by Max Weber (Bannister et al 2012) follows in the belief that government should act along scientific lines and make decisions. For example, the UK Government Transformation Strategy has set out plans for the better use of data for policy making in which they have indicated the importance of easy data-driven decisions for the benefit of citizens (Government Transformation Strategy, 2017). The strategy explains how, in the near future, governments will use predictive analytics to "to anticipate demand for services or policy changes and to prepare to meet citizens' and businesses' changing needs, informing government's decisions on what services to offer and how they should work" (Ibid.).

3.2 Democratic Authority and Legitimacy

3.2.1. - Value of Democracy

The concept of a democratic way of life for a society is not limited to elections or equal participation, rather it is the combination of instrumental and non-instrumental values of democracy. The democratic way of life permeates through a society at three levels: as a membership organization, a mode of government and a culture (Anderson, 2009). Anderson joins a larger body of scholars, notably John Stuart Mill and John Dewey, to argue that this democratic way of life is "justified as a matter of justice" (Anderson, 2009). I will provide a brief description of the different levels of the democratic way of life before focusing my particular critique with some of the aspects of democracy as a mode of government, particularly with relation to how citizens give authority to decision makers. Contextual understanding of the interplay between the forms of democratic life will later demonstrate how threats to one aspect can impact another.

A membership organization connotes the equality of all citizens of the community. The implication is such that no individual in society should preside above another. If one does not have such autonomy, there is no just reason for being compliant with a coercive law placed upon you (Pagallo, 2012). Autonomy of the individual as part of the membership, Kant claims, "is the independence from being constrained by another's choice" (Ibid.). This is important for each individual's equal claim to raise their opinion in society. Democracy ultimately pertains to citizens' equal opportunity to advance their interests.

Anderson (2009) describes the mode of government as a set of governing institutions that include: "a universal franchise, periodic elections, representative public officials, a free press and the rule of law". Each of these are important institutions which represent methods built to ensure accountability for the equal interests of citizens. The ways in which coercive decisions are placed upon citizens through these governing institutions leads us into the discussion of deliberative democracy in contrast to a system of majority rule. Collective decisions, in contrast to majority rule, drive moral outcomes (Christiano et al, 2018). Dewey describes collective decision making as "the exercise of practical intelligence in discovering and implementing collective solutions to shared problems, which is the basic function of community life" (Anderson, 2009). For collective society to validate the authority of decision makers, citizens must be autonomous actors, or as Rawls would say a "self-originating source of claims" (Ibid.). It is the collection of these claims which enhances citizens' autonomy to participate and uncover solutions to their shared problems.

Democratic institutions such as governing bodies or their constituents are an embodiment of a democratic culture. A democratic culture allows for individuals to deliberate, speak openly, critique and craft new solutions to their shared problems. Coercive decisions over society by decision makers are temporary and require feedback, furthermore, they are subject to revision (Anderson, 2009). Civil society represents a democratic culture wholeheartedly in that it provides an environment for citizens to deliberate freely on matters of public concern, serving as an intermediary between the private lives of individuals and the state (Ibid.).

3.2.2. – Proceduralism and Instrumentalism

The pure instrumentalist view of authority refers to those who believe that democratic decisions are justified in so far as the quality of their outcomes. In this regard, instrumentalism judges its justification of a decision on whether society chooses to abide by the coercive regulation or law. The authority of a decision is justified if citizens choose to follow it regardless of the democratic procedures by which this decision was created. The pure instrumentalist believes that a procedure only achieves legitimacy "solely in virtue of its consequence" (Danaher, 2015). How could evidence gathered by inhumane treatment of subjects lead to a legitimate and authoritative democratic decision?

In contrast to this view, a pure proceduralist believes that constituents should be subject to coercive decisions so long as the procedure which creates these decisions are just (Christiano, 2014). Furthermore, these decisions should be drafted in such a way that each individual constituent is included equally, this would be fair to all participants (Danaher, 2015). The critique of the instrumentalist approach of achieving democratic authority emerges from the question on how one should respond if they believe a democratic decision to be unjust. If the procedure which achieved such a result was just and accepted in advance, then those constituents subject to such a rule should concede to its authority. However, as with the case of the instrumentalist, if only the procedure is taken into consideration, one might be able to justify an evidently bad decision "simply because it treated people with respect and allowed them some meaningful participation" (Ibid.).

The intersection of these approaches to the determination of democratic authority is where several theorists believe decisions can be inherently legitimate. Christiano (2014) states that "we value political institutions because they make justice in society possible, because they advance the common good. And citizens within the democratic process argue in favor or against proposals on the grounds that certain policies and laws are just or desirable and others are not". The nuance in discovering a common ground between a just outcome and a just procedure stems from society's ability to have deliberative dialogue in achieving the shared solution for the collective problem and advancing each citizen's interests equally.

The discussion of democratic values and legitimate authority stems from the usage of AI assistants by democratic decision makers. For several reasons which I will discuss, big data driven policy decisions apply an overly instrumental approach to democratic legitimacy. At times, foundational algorithms of AI assistants can include inherent biases that are reflective of gaps in the input data which may lead to incomplete information or in some cases exclude certain portions of the democratic membership organization. The AI 'black box' refers to the inability for an AI Assistant to explain its output conclusions to the its user. This alludes to issues of transparency and the relevant literature which speaks to accountability achieved through transparency. As part of understanding the mixed approach, in which the decision-making process can be deemed just by the equal members of society, there must be a certain level of understandability and transparency which constituents can deem to be lawful, justifiable and legitimate. Lastly, in reference to the democratic processes of accountability in which democratic citizens enables certain individuals to become decision makers to steer the ship of mutual claims, the literature which points to the loss of autonomy of individuals in their use of AI makes their ability to be held accountable for the legitimacy of their decisions contestable.

3.3 Threats of AI Assistance to Democratic Values

3.3.1 - Cognitive Outsourcing and Autonomy

The increased ability of machines to do things better than humans has led to a significant rise in automation, whereby previously human performed tasks are routinely performed by machines. We regard this form of automation as the outsourcing of mundane or physically demanding activities. In many instances this is not objectively something negative or a cause for concern. As AI assistants enter various parts of personal and political life to help humans make more accurate and efficient decisions, this outsourcing extends into what some call algorithmic cognitive outsourcing. Danaher (2018) defines algorithmic cognitive outsourcing as "the offloading of a cognitive task to a smart algorithm". The outsourcing of tasks is not something novel to human behaviour in that we may ask an accountant to report our finances, or a hairdresser to cut our hair. What then is the difference between algorithmic cognitive outsourcing to AI assistants? When is it unethical to outsource our actions and decisions to a human, or even an algorithm? These questions have been approached from the perspective of private AI assistants (Gal, 2016) and of public AI assistants (Danaher, 2016; Helbing et al, 2014)

Algorithmic cognitive outsourcing limits the range of decisions available to an individual by the assurance that an algorithm is trustworthy in the provision of an efficient or most desirable outcome. The coded algorithm produces a recommendation based on given data inputs and weighted values as well as machine learning to predict future wants or decisions. AI is celebrated as an objective and efficient method of achieving evidence based, rational decisions, yet we do not consciously outsource these actions rather we become complacent in their abilities. As Gal et al (2017) states in reference to consumer choices "[consumers] display a pattern of conduct similar to that seen in relation to online contracts: accepting the algorithmic choice as default, without delving into the details and checking whether an optimal choice was made". The reliance on an AI assistant becomes pervasive and the desire of the individual to ensure that the decision they take are in their own interest rather than predetermined by an algorithm will require some understandability in how decisions are made in the private and public sphere.

In the context of public AI assistants, the autonomy of the decision maker is important in order to remain accountable to the constituents that elected them. There is concern that algorithmic *consumers*, those who use AI assistants to make decisions, are distanced from their actual choice field (Gal et al, 2017). Decisions outside of the purview of what an algorithm deems to be efficient and accurate will not be considered because the algorithm would in turn not suggest it and the algorithmic consumer would not have the willpower to look outside of predetermined options. In essence, decision makers will be nudged into ways of approaching policy issues based on the algorithms calculated best option (Diakopoulos, 2016). With increased citizen data that will be collected throughout Smart Cities and citizens' expansive existence in the digital, public decision makers should be weary that AI assistants become more capable of making accurate decisions, but also increasing our trust in their ability and loss of responsibility to make important political decisions. Helbing et al. (2019) warn that the combination of expansive data governance structures paired with big data creates "big nudging", where decisions are completely reliant on algorithms that can shape the future behavior of society in whichever way is deemed to provide the social good.

3.3.2 - Transparency and Accountability

Democratic methods of accountability were previously highlighted. However, of particular importance to the accountability with regard to AI assistants is the lack of transparency in the decision-making process. Frank Pasquale, author of *The Black Box Society* (2015) most famously identifies concerns with regard to the lack of transparency in emerging technologies and the lack of regulatory frameworks in place to take on the tech giants (Facebook, Twitter, Apple, Amazon, Google, Microsoft). Pasquale points out that decisions that used to be made through human reflection are now automated in that "software encodes thousands of rules and instructions computed in a fraction of a second" (Pasquale, 2015). When individuals can be responsible for their own decisions this raises its own concerns, but when the predictive algorithms that shape behavior interferes with society at large, the consequences could be more drastic. In reference to the organizing algorithms that curate our usage of online applications Pasquale states that the tech giants employ opaque technologies which effectively leave "users in the dark as to exactly why an app, story or book is featured at a particular time or in a particular place" (Pasquale, 2015).

The lack of regulations for ensuring ethical algorithms has significant effects on democratic institutions. Diakopolous (2016) argues that algorithms are increasingly "exercising power over individuals or policies in a way that in some cases (hidden government watch lists) lacks any accountability whatsoever". The increased influence of algorithms over decisions requires that the users can understand what values are being taken into consideration into the provision of suggestions. In the context of a music suggestion from Spotify or online purchase, this may seem fairly innocuous because one assumes that the recommendation derives from previous listens and similar users' preferences and has no serious drawbacks, however decisions by AI assistants in the public sphere demand another level of transparency and accountability.

Transparency is necessary in the determination of just processes and maintaining accountability for decisions. Annany and Crawford (2016) highlight that the ability to observe can be understood as "a diagnostic for ethical actions, as observers with more access to the facts describing a system will be better able to judge whether a system is working as intended and what changes are required" (Annany and Crawford, 2016). Citizens are more capable of determining the authority of decisions when decision making processes are more visible. For example, if society was unaware that a politician had external motivations for making a decision, or that it would have beneficial implications to their own wellbeing then society would be incapable of assessing the injustice and need for change. Furthermore, the notion of transparency and observation allows for the public acceptability of a decision and public trust in institutions (de Fine Licht, 2011). Transparency is a symbol of good government, it builds trust and can enable a system of accountability.

On the other hand, transparency can also be harmful. For example, when companies want to hide trade secrets, they require a certain level of opacity so as not to lose their competitive edge. Algorithms are similar in this regard, if they were to be fully transparent it would be difficult to prevent someone gaming the system (Diakopoulos, 2016). If someone were to understand the weighted values of an algorithm that assesses the level of welfare a person should receive, or which tax bracket they should be taxed under, then individuals could alter their behavior in order to rig the system in their favour. Annany and Crawford (2016) proceed to outline some limitations of transparency, pertaining to algorithms, which include: the technical limitations of deep learning, machine learning and the inability to explain in a humanly logical sense how a decision was achieved, the flaw of visibility in contrast to understandability as well as the overabundance of variables that render algorithms illegible to experts themselves.

3.3.3 - Data inputs and Bias

Many critics have raised of AI decision making have raised their concerns in regards to modern applications of AI which have led to biased decision making in predictive policing and facial recognition applications. Books like Weapons of Math Destruction by Cathy O'Neil have led the movement in criticizing the usage of AI and ML with regard to impactful decisions on society. Algorithms include the human influence of their creators due to the "criteria choices, optimization functions and training data" (Diakopoulos, 2016). Algorithms that thus claim to be completely objective and rational are also then subject to the human flaws in their creators. Rather, algorithms make normative decisions rather than objective or rational ones as a result of their weighted values in the codes (Mittelstadt, 2016, Hao, 2019).

One of the more notable cases was Amazon's facial recognition which produced gender and racial bias (Miller, 2019). This was also the case in police forces which attempted to use algorithms for predictive policing software which directed police to target minority communities and had discriminatory consequences for minority individuals (Brantingham et al. 2018). There are significant limitations to the usage of algorithms for predictive policing, in that relying on poor data can result in systemic bias and data censoring which can lead to discriminatory practices (Perry, 2013).

On the other hand, some argue that machines are less biased than humans. Miller describes some examples in which algorithms were given tasks normally intended for humans. First, a 2002 study by economists on automated underwriting systems demonstrated that the systems more accurately produced results compared to manual underwriters, which also resulted in "higher borrower approval rates, especially for underserved applicants" (Gates et. al 2002). The next case points to a study on a job-screening algorithm at a software company in which the algorithm "favored non-traditional candidates" in particular those without top ranking universities on their CV (Miller, 2018). A study at the National Bureau of Economic Research analyzes the way in which a behavioral/predictive algorithm could replace the job of a judge that makes bail decisions. The study found that the algorithm could achieve significantly more equitable decisions with "jailing rate reductions of up to 41% with no increase in crime rates" (Kleinberg et al, 2017). The report goes on to suggest that the jailing rate reductions can be achieved "while simultaneously reducing racial disparities" (Ibid.).

Cave et al (2018) suggests that ensuring that algorithms and new advanced AIs achieve an ethical alignment to a human level acceptable standard or what one would find minimally acceptable from a human would be trivial. The paper claims “human decision-makers often fall below the standards we expect of them, whether through accident, malice or failures of reasoning.” Should we expect better than this from an algorithm? The study goes on to demonstrate that machine ethicists believe that it may be possible that automated moral reason could improve the ethical alignment of human and machine decision-making because it would demonstrate more consistent ethical commitments and beliefs and offer ways by which they can be followed more accurately.

4. 0 Methodology

4.1 - Ethics of Risky New Technologies

The aim of this paper will be to demonstrate how the risks to accountability and democratic legitimacy associated with the increased use of AI assistance as tools for public decision makers can be mitigated. Wolff (2019) sets out a framework by which ethical implications can be considered when assessing the implementation of what he calls "risky new technologies". He argues that the introduction of "very innovative technology takes us into uncharted territory and [this] means that it is very difficult to make a well-grounded assessment of the possible outcomes and their probabilities" (Wolff, 2019). The future capabilities of AI are advancing at such a rate that their long-term implications are difficult to currently assess. This is a case of radical uncertainty (Wolff, 2019) that necessitates a different approach in order to provide suggestions on how to approach a new technology with cautious optimism.

A restrictive attitude to new technologies could lead to a loss of something which would be of great benefit to society. The promises of AI assistants to create innovative and evidence-based policies should not be ignored though we should take precautions. Wolff states, "the trick will be to decide when precaution is necessary and when it is unnecessary or in other words, when the introduction of a new technology should be treated in a different way" (Wolff, 2019).

Bearing this in mind, I will apply the precautionary checklist for risky new technologies set forward by Wolff, in order to determine how we should approach the introduction of AI assistants into public sector decision making.

The checklist includes the following questions (Wolff, 2019):

1. Are the costs and risks of the new technology tolerable?
2. Does it have significant benefits?
3. Do these benefits solve important problems?
4. Could these problems be solved in some other less risky way?
5. What are the possible long-term economic consequences of introducing the technology?
6. What is the possible long-term political consequences of introducing the technology?

If we determine that the variables for AI assistants are too uncertain due to the incalculable benefits with further technological developments and the degree to which humans will remain present in the decision-making process, then there may very well be a risk that the negative drawbacks could overwhelm the benefits. This paper will attempt to set some points of limitation by highlighting certain areas of concern that should be considered with the increased development of AI assistants and their integration into the public-sector decision-making process. Understanding the potential risks associated with these innovations will allow the thesis to offer suggestions on how to mitigate risks to democratic legitimacy and the accountability of decision makers while not being overly cautious and impeding innovations that could potentially benefit society.

5.0 - Conclusions

In light of the literature with the application of this methodology, this paper will attempt to do a cost-benefit analysis as it pertains to the risks associated with AI assistants used by policy makers in order to better understand the policies necessary to regulate such an innovative new technology in which consequences are unclear and the benefits are unbounded. First, I will begin by outlining some of the areas in which AI assistants to policy makers can be of great benefit, through the usage of more efficient predictive analytics tools and cost-effective government spending that emerges out of the collection of big data paired with machine learning and artificial intelligence. Second, this paper will discuss the importance of a pluralistic approach in the debate surrounding instrumentalist and proceduralist views of democratic authority, particularly in relation to the use of AI assistants. I will highlight the risks associated with AI assistants that place too much emphasis on instrumental authority whereby authority is justified by the result, instead of the pluralistic approach that includes a level of proceduralism that retains the value of authority which stems from a legitimate and democratic process. This democratic process must include transparency, to the point of understandability, in the algorithms that shape AI assistants so that the electorate can hold decision makers accountable when a policy is deemed unjust. Third, I will discuss how decision makers' cognitive outsourcing to AI assistants diminishes their level of autonomy, and thus accountability to citizens. This is true unless elected officials are capable of understanding the ways in which AI assistants produce their results, ensuring that the data inputs used to produce results do not include biases that exceed a human level of bias. It is important to note here that humans often make biased decisions and thus the expectation from an AI assistant should be that it produces equally or less biased decisions than a human. Fourth, I will provide some preliminary results for preventative actions that policymakers can take to regulate the usage AI assistants. This will include a discussion of the need for transparency as well as opacity in algorithms. I will also argue for the sake of an 'in the loop' human oversight which implies that a human should always be in a position to be held accountable to the final decision. This stems from the unknowable risks associated with the further development of AI assistants which could legitimize concerns of superintelligence and algocracy.

Word Count: 5,500

6.0 - Bibliography

Annany, Mike and Crawford, Kate. “Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability.” *New Media and Society*, 2016, 17.

Anderson, Elizabeth. “Democracy: Instrumental vs. Non-Instrumental Value.” In *Contemporary Debates in Political Philosophy*, edited by Thomas Christiano and John Christman, 213–27. Oxford, UK: Wiley-Blackwell, 2009. https://doi.org/10.1002/9781444310399.ch12.

Atkinson, Robert D. “‘It’s Going to Kill Us!’ And Other Myths About the Future of Artificial Intelligence.” *INFORMATION TECHNOLOGY*, 2016, 50.

Atkinson, Robert D., Information Technology, and Innovation Foundation. “Will Smart Machines Be Less Biased Than Humans?” GE Reports, September 19, 2016. https://www.ge.com/reports/will-smart-machines-be-less-biased-than-humans/.

Bannister, Frank, and Diana Wilson. “O(Ver)-Government?: Emerging Technology, Citizen Autonomy and the Regulatory State.” *Information Polity* 16, no. 1 (May 4, 2011): 63–79. https://doi.org/10.3233/IP-2011-0225.

Bostrom, Nick. “Ethical Issues in Advanced Artificial Intelligence.” *Future of Humanity Institute*, 2003, 7.

Bostrom, Nick, Allan Dafoe, and Carrick Flynn. “Policy Desiderata for Superintelligent AI: A Vector Field Approach,” 2018, 29.

Brantingham, P. Jeffrey, Matthew Valasik, and George O. Mohler. “Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial.” *Statistics and Public Policy* 5, no. 1 (January 2018): 1–6. https://doi.org/10.1080/2330443X.2018.1438940.

Brill, Julie. “Scalable Approaches to Transparency and Accountability In,” 4, 2018.

Buchanan, Ben, and Taylor Miller. “Machine Learning for Policymakers.” *Machine Learning*, 2017, 58.

Cave, Stephen, Rune Nyrup, Karina Vold, and Adrian Weller. “Motivations and Risks of Machine Ethics.” *Proceedings of the IEEE* 107, no. 3 (March 2019): 562–74. https://doi.org/10.1109/JPROC.2018.2865996.

Christiano, Thomas. “The Authority of Democracy*.” *Journal of Political Philosophy* 12, no. 3 (September 2004): 266–90. https://doi.org/10.1111/j.1467-9760.2004.00200.x.

Christiano, Tom. “Democracy.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2018. Metaphysics Research Lab, Stanford University, 2018. https://plato.stanford.edu/archives/fall2018/entries/democracy/.

Christman, John. “Autonomy in Moral and Political Philosophy.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2018. Metaphysics Research Lab, Stanford

University, 2018. https://plato.stanford.edu/archives/spr2018/entries/autonomy-moral/.

Citron, Danielle Keats, and Frank Pasquale. “The Scored Society: Due Process for Automated Predictions.” *WASHINGTON LAW REVIEW* 89 (2014): 33.

Cohen, Julie E. “The Regulatory State in the Information Age.” *Theoretical Inquiries in Law* 17, no. 2 (January 1, 2016). https://doi.org/10.1515/til-2016-0015.

D’Agostino, Fred. “Expertise, Democracy, and Applied Ethics.” *Journal of Applied Philosophy* 15, no. 1 (January 1998): 49–55. https://doi.org/10.1111/1468-5930.00072.

Danaher, John. “The Threat of Algocracy: Reality, Resistance and Accommodation.” *Philosophy & Technology* 29, no. 3 (September 2016): 245–68. https://doi.org/10.1007/s13347-015-0211-1.

Danaher, John. “Toward an Ethics of AI Assistants: An Initial Framework.” *Philosophy & Technology* 31, no. 4 (December 2018): 629–53. https://doi.org/10.1007/s13347-018-0317-3.

Datta, Anupam, Shayak Sen, and Yair Zick. “Algorithmic Transparency via Quantitative Input Influence:,” 2016, 20.

Diakopoulos, Nicholas. “Accountability in Algorithmic Decision Making.” *Communications of the ACM* 59, no. 2 (January 25, 2016): 56–62. https://doi.org/10.1145/2844110.

Dressel, Julia, and Hany Farid. “The Accuracy, Fairness, and Limits of Predicting Recidivism.” *Science Advances* 4, no. 1 (January 2018): eaao5580. https://doi.org/10.1126/sciadv.aao5580.

Engin, Zeynep, and Philip Treleaven. “Algorithmic Government: Automating Public Services and Supporting Civil Servants in Using Data Science Technologies.” *The Computer Journal* 62, no. 3 (March 1, 2019): 448–60. https://doi.org/10.1093/comjnl/bxy082.

Feenberg, Andrew. “Critical Theory of Technology: An Overview.” *Tailoring Biotechnologies* 1, no. 1 (2005): 10.

Feenberg, Andrew. “The Technocracy Thesis Revisited: On *the Critique of Power* ∗.” *Inquiry* 37, no. 1 (March 1994): 85–102. https://doi.org/10.1080/00201749408602341.

Fine Licht, Jenny de. “Do We Really Want to Know? The Potentially Negative Effect of Transparency in Decision Making on Perceived Legitimacy: Do We Really Want to Know? The Potentially Negative Effect of Transparency in Decision Making on Perceived Legitimacy.” *Scandinavian Political Studies* 34, no. 3 (September 2011): 183–201. https://doi.org/10.1111/j.1467-9477.2011.00268.x.

Gal, Michal S. “ALGORITHMIC CHALLENGES TO AUTONOMOUS CHOICE,” 2016, 40.

Gal, Michal S, and Niva Elkin-Koren. “ALGORITHMIC CONSUMERS” 30, no. 2 (2017): 45.

Gates, Susan Warton, Gail Perry, Vanessa and Zorn, Peter M. (2002) Automated underwriting in mortgage lending: Good news for the underserved?, Housing Policy Debate, 13:2, 369-391, DOI: 10.1080/10511482.2002.9521447

Giest, Sarah. “Big Data for Policymaking: Fad or Fasttrack?” *Policy Sciences* 50, no. 3 (September 2017): 367–82. https://doi.org/10.1007/s11077-017-9293-1.

Gigerenzer, Gerd. “On the Supposed Evidence for Libertarian Paternalism.” *Review of Philosophy and Psychology* 6, no. 3 (September 2015): 361–83. https://doi.org/10.1007/s13164-015-0248-1.

“Government Transformation Strategy.” GOV.UK, 2017. https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy.

Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. “When Will AI Exceed Human Performance? Evidence from AI Experts.” *ArXiv:1705.08807 [Cs]*, May 24, 2017. http://arxiv.org/abs/1705.08807.

Gutmann, Amy, and Dennis Thompson. “Ch. 6 The Constitution of Deliberative Democracy.Pdf.” In *Democracy and Disagreement*, 1996.

Hao, Karen. “This Is How AI Bias Really Happens—and Why It’s so Hard to Fix.” MIT Technology Review, 2019. https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/.

Helbing, Dirk, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V. Zicari, and Andrej Zwitter. “Will Democracy Survive Big Data and Artificial Intelligence?” In *Towards Digital Enlightenment*, edited by Dirk Helbing, 73–98. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-319-90869-4_7.

Helbing, Dirk, and Evangelos Pournaras. “Build Digital Democracy.” *Nature, Macmillan Publishers*, 2015.

Hurley, Dan. “Can an Algorithm Tell When Kids Are in Danger?” *The New York Times*, January 2, 2018, sec. Magazine. https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html.

Kleinberg, John, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. “Human Decisions and Machine Predictions.” *National Bureau of Economic Research*, 2017, 1–76.

Krakauer, David. “Will A.I. Harm Us? Better to Ask How We’ll Reckon With Our Hybrid Nature.” Nautilus, September 6, 2016. http://nautil.us/blog/will-ai-harm-us-better-to-ask-how-well-reckon-with-our-hybrid-nature.

La Fors, Karolina, Bart Custers, and Esther Keymolen. “Reassessing Values for Emerging Big Data Technologies: Integrating Design-Based and Application-Based Approaches.” *Ethics and Information Technology*, April 24, 2019. https://doi.org/10.1007/s10676-019-09503-4.

Li, Michael. “Addressing the Biases Plaguing Algorithms.” *Harvard Business Review*, May 13, 2019. https://hbr.org/2019/05/addressing-the-biases-plaguing-algorithms.

Liang, Fan, Vishnupriya Das, Nadiya Kostyuk, and Muzammil M. Hussain. “Constructing a Data-Driven Society: China’s Social Credit System as a State Surveillance Infrastructure: China’s Social Credit System as State Surveillance.” *Policy & Internet* 10, no. 4 (December 2018): 415–53. https://doi.org/10.1002/poi3.183.

Manheim, Karl, and Lyric Kaplan. “Artificial Intelligence: Risks to Privacy and Democracy” 21 (2019): 83.

Manyika, James, and Kevin Sneader. “AI, Automation, and the Future of Work: Ten Things to Solve for (Tech4Good) | McKinsey,” June 2018. https://www.mckinsey.com/featured-insights/future-of-work/ai-automation-and-the-future-of-work-ten-things-to-solve-for.

Mehr, Hila. “Artificial Intelligence for Citizen Services and Government,” 2017, 19.

Miller, Alex P. “Want Less Biased Decisions? Use Algorithms.” *Harvard Business Review*, 2018, 5.

Mittelstadt, Brent. “Auditing for Transparency in Content Personalization Systems,” 2016, 12.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. “The Ethics of Algorithms: Mapping the Debate.” *Big Data & Society* 3, no. 2 (December 2016): 205395171667967. https://doi.org/10.1177/2053951716679679.

Noorman, Merel. “Computing and Moral Responsibility.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2018. Metaphysics Research Lab, Stanford University, 2018. https://plato.stanford.edu/archives/spr2018/entries/computing-responsibility/.

“Owakah and Aswani - 2009 - Technocracy and Democracy The Challenges to Devel.Pdf,” n.d.

Owakah, Fea, and Rd Aswani. “Technocracy and Democracy: The Challenges to Development in Africa.” *Thought and Practice: A Journal of the Philosophical Association of Kenya* 1, no. 1 (September 28, 2009): 87–99. https://doi.org/10.4314/tp.v1i1.46308.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press, 2015.

Perry, Walt L. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Santa Monica, CA: RAND, 2013.

Poel, Martijn, Eric T. Meyer, and Ralph Schroeder. “Big Data for Policymaking: Great Expectations, but with Limited Progress?: Big Data for Policymaking.” *Policy & Internet* 10, no. 3 (September 2018): 347–67. https://doi.org/10.1002/poi3.176.

Singh, Varun, Ishan Srivastava, and Vishal Johri. “Big Data and the Opportunities and Challenges for Government Agencies” 5 (2014): 4.

Stockmann, Daniela. “Toward Area-Smart Data Science: Critical Questions for Working With Big Data From China: Toward Area-Smart Data Science.” *Policy & Internet* 10, no. 4 (December 2018): 393–414. https://doi.org/10.1002/poi3.192.

Viechnicki, Peter, and William Eggers. “How Much Time and Money Can AI Save Government?” Deloitte Center for Government Insights, 2017. https://www2.deloitte.com/content/dam/insights/us/articles/3834_How-much-time-

and-money-can-AI-save-government/DUP_How-much-time-and-money-can-AI-save-government.pdf.

Wal, Zeger van der, and Yan Yifei. “Could Robots Do Better than Our Current Leaders?” World Economic Forum, 2018. https://www.weforum.org/agenda/2018/10/could-robot-government-lead-better-current-politicians-ai/.

Weiser, Stephanie. “Requirements of Trustworthy AI.” Text. FUTURIUM - European Commission, April 8, 2019. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1.

Wolff, Jonathan. “Risky New Technologies.” In *Ethics and Public Policy: A Philosophical Inquiry*, 2nd Edition., 330. Routledge, 2019.

Yiu, Chris. “The Big Data Opportunity” *Policy Exchange* (Think Tank), 2012. https://policyexchange.org.uk/the-big-data-opportunity/.

Zara, Christopher. “FTC Chief Technologist Ashkan Soltani On Algorithmic Transparency And The Fight Against Biased Bots.” International Business Times, April 9, 2015. https://www.ibtimes.com/ftc-chief-technologist-ashkan-soltani-algorithmic-transparency-fight-against-biased-1876177.

Zarsky, Tal Z. “Governmental Data Mining and Its Alternatives.” *Penn State Law Review* 116, no. 2 (2011): 46.

7.0 Appendix

Table 1: Workplan and timetable for completion of the Masters Thesis

Deadline	Work to be completed
August 15, 2019	First full Draft of Thesis Report
August 30, 2019	Submission of Thesis Report
December 2019	Discussion with Professor Gallego on Empirical Approach to Masters Thesis
April 2020	Content Analysis for Masters Thesis
End of May 2020	Completion of analytical research methods
June 15, 2020	First full Draft of Masters Thesis
June 30, 2020	Submission of Masters Thesis
September 2020	Oral Defense of Masters Thesis

8.0 - Declaration of Authorship

I, the undersigned Daniel Cohen hereby declare that I am the sole author of this thesis report. To the best of my knowledge this thesis contains no material previously published by any other person except where due acknowledgement has been made. This thesis report contains no material which has been accepted as part of the requirements of any other academic degree or no non-degree program, in English or in any other language. This is a true copy of the thesis report, including final revisions.

Date: August 30, 2019

Name: Daniel Cohen

Signature: