What are the obstacles related to being transparent in AI-assisted governmental decision-making? An ethnographic study

Essen, Etiënne van

Essen, E.D. van
s1856685
Master thesis Public Administration: International and European Governance
Digitalisation and transparency in public organisations
Dr. A.R. Ingrams
27449 Words

~ What are the obstacles related to being transparent in AI-assisted governmental decision-making?~

An ethnographic study.

Preface

I hereby present to you my master thesis ‘What are the obstacles related to being transparent in AI-assisted governmental decision-making?’, a qualitative research paper based on timely interviews with experts from the sector on the promising technology of Artificial Intelligence. It has been written in order to fulfill the graduation requirements of the MSc Public Administration: International and European Governance. I was engaged in the writing of this thesis from September 2020 until June 2021.

During the research that started in September 2020, I simultaneously started with an internship at the Dutch Ministry of Infrastructure and Water Management up until April 2021. It was that opportunity that enabled me to conduct this research, I deeply cherish the support and freedom I enjoyed to do my research and to be able get in touch with people in the field. It hasn’t always been easy to graduate as well as work amidst a global pandemic, but the exposure to new experiences and the fun conversations I've had with like-minded people got me through it. Even if this was just mainly through a computer screen, it will undoubtedly be a period that I will remember all my life.

In particular, I would like to thank my supervisor Dr. Alex Ingrams for his superb guidance and unconditional enthusiasm for my research subject. I would also like to thank all of the respondents, without whose cooperation and interesting experiences I would not have been able to conduct this analysis.

To the colleagues with whom I have worked: I would like to thank you for the warm nest during my internship. I also benefitted from being able to discuss ideas and issues with the people in my immediate vicinity, thank you. And above all, I would like to thank my parents and family for their unconditional support during my graduation.

I hope you will enjoy your reading.

Best,

Etiënne van Essen

Leiden, May 13, 2021

Abstract

The increasing reliance on ICT within the public sector has changed the working ways of governmental bureaucracies from a paper reality to a digital one, and governments are eager to use new technologies for their business operations and reap its benefits just as the private sector does. Since technological advancement is driven by the private sector, and humans are increasingly accustomed to the speed and efficiency that technology brings, citizens are expecting governments to adapt and digitize as well. As such, an important trend that is being experimented with is the usage of self-learning algorithms, particularly Artificial Intelligence or AI. Since AI runs on data, it is only logical that an organization such as the government which holds an abundance of data would like to put this to use. Data that is collected might hold certain patterns, if you can find such patterns and assume that the near future will not be much different from when the data was collected, predictions can be made. However, AI systems are often deemed opaque and inscrutable, and this can collide with the judicial accountability that governments have towards their citizens in the form of transparency. Based on the assumption that the information that is used by AI i.e. data and algorithms, is not similar to documentary information that governments are accustomed to, there are added obstacles for governments to overcome in order to achieve the desired effects of transparency. The goal of this research is to explore the barriers to transparency in governmental usage of AI in decision-making by analyzing governmental motivation towards (non-) transparency and how the complex nature of AI relates to this. The question that stems from this is: *What are the obstacles related to being transparent in AI-assisted governmental decision-making?* In the study, a comparison is made between the obstacles to transparency for documentary information and the obstacles that experts encounter in practice related to AI, a contribution follows. Based on the literature, it is hypothesized that governments are limited by privacy and safety issues, lack of expertise, cooperation and inadequate disclosure. The results show that the obstacles are more nuanced and an addition to the theory is appropriate. The most important findings being: that data and algorithms should not be treated as documentary information; the importance of the policy domain in determinant for the degree of transparency; that lack of cooperation causes multiple obstacles to transparency such as self-censoring, accountability issues, superficial debate, false promises, inability to explain and ill-suited systems; that more information disclosure isn't always better; and that the public sector should rethink their overreliance on private sector business models. All these obstacles can be associated to losing sight of the fundamental function of government, serving citizens.

Table of Contents

Preface……………………………………………………………………………………ii

Abstract……………………………………………………………………………….…iii

Table of contents…………………………………………………………………………iv

1. Introduction……………………………………………………………………………1

1.1 Relevance of the research……………………….……………………………………..3

1.2 Research goal…………………………………………………………………….…5

2. Literature Review………………………………………………………………………6

2.1 Government, information and digital transformation………………………………6

2.2 Decision-making, the policy cycle and AI……………………………………….…9

2.3 Explaining the technology……………………………………………………...10

2.3.1 Algorithms…………………………………………………………………11

2.3.2 Artificial intelligence…………………………………………………………12

2.3.3 Machine learning……………………………………………………...14

2.2.3.1 Supervised learning……………………………………………16

2.2.3.2 Unsupervised learning…………………………………………17

2.2.3.3 Reinforcement learning………………………………………..18

2.3.4 Interpretability and human agency………………………………………19

2.4 Defining transparency…………………………………………………………..22

2.4.1 Data and information……………………………………..……...24

2.4.2 Achieving transparency…………………………………………….....25

2.5 Public sector obstacles to transparency…………………………………..…..27

2.5.1 Associated challenges for AI………………………………………….29

2.6 Hypotheses……………………………………………………..……..…..32

3. Research design……………………………………………………………………….35

3.1 Sampling……………………………………………………………..……36

3.2 Data collection method……………………………………………..…..39

3.3 Data analysis…………………………………………………….……..40

4. Results……………………………………………………………………………...41

4.1 Safety and privacy……………………………………………………42

4.2 Expertise and organizational cooperation……………………………....45

4.3 Inefficient disclosing…………………………………………………48

5. Discussion…………………………………………………………………………50

5.1 Governments are limited by privacy and safety issues to not be transparent in AI assisted decision-making………..……………………………………………51

5.2 Governments are impeded in being transparent in AI-assisted decision-making due to lack of expertise and cooperation…………………………………….. 54

5.3 Governments are impeding transparency by ineffectively disclosing information about AI-assisted decision making………………………………61

5.4 Limitations of the research…………………………………………......66

5.5 Recommendations for future research……………………………………..66

6. Conclusion…………………………………………………………………………68

7. Bibliography……………………………………………………………………..70

8. Appendices…………………………………………………………………………84

8.1 Appendix I interview transcripts…………………………………..…84

8.2 Appendix II interview guide……………………………………………...84

8.3 Appendix III codebook…………………………………………………86

1. Introduction

Over the past years a new technological development has become increasingly intertwined with our daily lives, Artificial Intelligence (AI). It is being used in medical diagnosis, logistics, navigation systems, weather predictions et cetera. Following the definition of the European Commision, AI "refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals" (European Commission, 2018). In an organizational context, algorithmic systems or AI systems, are extensively being used in the private sector in order to achieve efficiency gains and automate business processes. Due to its premised advantages, the public sector also wants a piece of the pie. Governments harbour a lot of data and information that substantiates their decision-making and ultimately their policies. It is therefore no surprise that algorithmic systems that need data to operate are increasingly being used as a tool within governmental decision-making processes. However, governments must be wary in its usage since AI systems may not always offer explanations for its outcome that are in line with the acquainted judicial and social expectations. Therefore, there is a "growing concern that the traditional frameworks for implementing transparency and accountability may not suffice as mechanisms of governance" (Koene et al., 2019, p. 1).

As opposed to the traditional statistics that decision-makers are accustomed to, AI is often referred to as 'a black box' where the preceding steps to reach an outcome are so complex that it cannot be explained. It is particularly the complexity and opaqueness of the technology that instigate these transparency issues. This apparent lack of transparency also amplifies issues related to accountability since no transparency also implies no recourse to explanation and method to ascertain particular faults or be compensated (Koene et al., 2019). There is a lot of literature on the challenges of AI application in the public sector: on ethical and social impact (Quraishi et al., 2017; Coeckelbergh, 2020), implementation of AI in governmental processes (Thierer et al., 2017), and legal issues related to responsibility and privacy (Scherer, 2015; Coglianese & Lehr, 2019). Therefore, it is apparent that attempting to use AI system in governmental decision-making is a challenge on its own. However, this research is focused on barriers to transparency since practice has proven that governments are still searching how to be transparent about AI systems, and being non-transparent often results in more disadvantages than advantages.

For example, SyRI or *Systeem Risico Indicatie* – which originated in 2014 – is an IT-system that was being used by the Dutch government to prevent and combat fraud and abuse.

This occurs by combining certain data from participating governmental institutions to determine if an individual is fraudulent or not (Rechtbank Den Haag, 2020). Based on a risk model that has 17 predetermined indicators, potentially fraudulent hits are exposed (Rechtbank Den Haag 2020). A fraudulent notification would entail that a person would be flagged for further official investigation (Rechtbank Den Haag, 2020), with all its consequences. On the 5th of February 2020, the Court of the Hague judged that the Dutch government should refrain from using the SyRI system that is being used to detect fraud on matters such as benefits, allowances and taxes (Rechtbank Den Haag, 2020). This judgement rests on the fact that the system conflicts with the European Convention on Human Rights (ECHR) (Rechtbank Den Haag, 2020). What weighed heavily in the Court's decree were the "fundamental principles underlying the protection of data", one important aspect being the 'transparency principle' that was determined by the European Commission (Rechtbank Den Haag, 2020). Since the State gave no insight into what indicators determined that there is a heightened risk for fraud, what algorithms are being used or how they work (Rechtbank Den Haag, 2020), the system was not transparent. In this case, the State chose not to be transparent about their system, however, it resulted in the abolishment of the systems usage. The State's reason for non-disclosure of SyRI's inner workings, was that citizens could adapt their behavior to prevent suspicion based on the released information (Rechtbank Den Haag, 2020). Even though the reason seems fair, it leaves suspicion whether there weren't other reasons for the State to withhold information.

Nevertheless, the importance of transparency is enhanced due to the possible discriminating effects of AI systems. SyRI was only being used in certain problematic neighborhoods meaning that there could be a risk that links were being made based on lower social status or a migration background (Rechtbank Den Haag, 2020). The lack of transparency means that there is no way to see if these risks are sufficiently safeguarded for. The fact that the Court of the Hague eventually decided that the usage of the SyRI system is unlawful has shown the necessity of transparency and verifiability of algorithm usage within governments and explanation towards your constituents.

The recalcitrance of government to disclose the information about what indicators were used to come to a certain decision is the point of interest for this thesis. Even though transparency of internal governmental organizational processes rests on the 'right to know' that is made explicit in Article 19 of the Universal Declaration of Human Rights (UN General Assembly, 1948), and disclosing governmental information is considered a precondition for good governance (Grimmelikhuijsen et al, 2013), it is often the case that governments refrain from disclosing information. Governments face obstacles to being transparent, either willingly or unwillingly.

This thesis means to focus on exploring the obstacles that governments encounter when attempting to be transparent whilst using AI in their decision-making processes. If governments want to use AI whilst ensuring fairness towards their citizens in the form of accountability, a first step in the process is ensuring transparency. Therefore, the researcher will explore the motivation of governmental behavior and how new technology potentially affects this. Thus, combining the interactions of disclosing government information versus transparency, and the challenges that exist in implementing AI in the public sector. The relevance of the research will now be outlined.

1.1 Relevance of the research

A system such as SyRI falls in line with ‘smart’ systems and algorithm usage. AI is a container concept of a lot of technological tools that can help transform data into information and uses algorithms to operate. AI systems particularly utilize ML algorithms which generates the ‘learning’ component of the technology, referring to its ‘intelligence’ (Coeckelbergh, 2020), these concepts will be outlined later on.

As was just briefly touched upon in the introduction, using learning algorithms in governmental decision-making can contain transparency related challenges. AI is made by humans, and just like humans make mistakes, sometimes AI – as a human product – can make mistakes. For example, a system can increase a certain bias towards societal minorities, this was potentially the case with the SyRI system (since SyRI was only used in problematic neighborhoods). AI related mistakes can often seem innocent like when Amazon’s virtual home assistant Alexa – that is AI driven – started a party at the owner's home without him being present eventually forcing the police to break down the door to stop the party (Olschewski, 2017). However, once mistakes adversely affect someone's life such as the denial of financial resources, the urgency for proper governance of these systems arises. Besides, mistakes can take on a more tragic form such as when Uber’s self-driving taxi killed a pedestrian, and the passenger was eventually charged with negligent homicide (Cellan-Jones, 2020). Considering that transparency lies at the heart of accountability (Naurin, 2002), when a system does make a mistake, it should be explainable in order to identify what went wrong and who is to be held accountable.

Unfortunately, the private sector innovates and the public sector reacts (Brown & Toze, 2017). Fundamental discussions about new technologies are therefore held too late and consequently laws and regulations often fall behind. Since AI is here to stay and is continuing to play a larger role in our lives, it is essential to think about how to govern this technology so governments can protect their citizens from the harm it can bring.

In the SyRI case, the AI system was given a lot of information from different participating governmental institutions and produced a prediction that classified an individual as being fraudulent or not. However, the Dutch government chose not to be transparent by concealing the information of SyRI's inner workings. In practice, a government can have many reasons for being non-transparent towards its citizens e.g. lacking resources or concerns for public safety (Pasquier & Villeneuve, 2007). Therefore, as a government there are certain barriers to being transparent.

There is extensive literature on organizational barriers to transparency that concerns the access to information, these studies entail transparency as a right, thus the ability of an individual to request information (in the form of documents) about a governmental action (Pasquier & Villeneuve, 2007). However, the nature of this information has changed from analog to digital and leads to changes in how daily work is performed and cases are handled by civil servants. The dawning of AI as a tool that can not only combine lots of information but also use it to analyze and support decision-making, adds a whole new dimension to the way governments use information. This also adds a new dimension to the explanatory demand of transparency itself, change in information usage evidently means change in the way information requests should be answered. Governments should not only explain how they reached an outcome but also explain the system that supported that decision. However, explaining the system can have its complications.

There are several big scholarly debates about transparency and governmental usage of algorithms. Some concerning if algorithmic usage in the public sector can meet the legal demands for transparency (Coglianese & Lehr, 2019). Others concerning if transparency is the solution to explaining and governing algorithm at all (Annany & Crawford, 2018). The issues that are being addressed in these debates revolve around the opaqueness and inscrutability of this technological tool and its relation to the long tradition of judicial accountability that the government has in the form of transparency towards its citizens. Concerns regarding responsible deployment of algorithmic tools by governmental bodies are increasingly receiving scholarly attention since governments cannot dawdle and should have "a response to the private economy's growing reliance on machine learning" (Coglianese & Lehr, 2019, p. 13).

However, the literature on using AI in the public sector is a rather young field (Wirtz et al., 2019). The field falls short in describing applications and challenges in the public sector. For example, understanding the impacts of AI on the workforce, organizational structures, economy, government, or society in general remains incomplete (Brynjolfsson & Mitchell, 2017; Faraj et al., 2018; Wirtz et al., 2019).

In addition, public administration literature on transparency is arguably lacking data and research. In the article by Grimmelikhuijsen & Meijer (2012) on if transparency will lead to trust, it is argued that “the empirical basis for both lines of argument is limited: both camps refer to anecdotal material rather than thorough empirical studies” (p. 138). Due to the fact that transparency literature is mostly normative, this research aims to contribute to the empirical work of public administration literature.

Even though, on the one hand there is vast literature on the organizational barriers of government to be transparent and, on the other hand, how the complexity of AI can make it difficult to be transparent and accountable, this research aims to combine these two strands of literature by contributing to the lacuna and identifying obstacles to transparency in AI applications in the public sector i.e. analyzing governmental motivation towards (non-) transparency and how the complex nature of AI relates to this.

1.2 Research Goal

The goal of this research is to explore the barriers to transparency in government usage of AI with an empirical foundation. The research will provide a unique snapshot of the obstacles that the Dutch government is facing related to transparency and new technology in the form of case related expert interviews. This research is not meant to be generalized across other cases, just as the experts interviewed were not meant to be representative of a population. On the contrary, the generalization to be made is related to theory rather than populations, the theoretical inferences stem from the collected qualitative data and determines this research’s generalization. It should be mentioned that another important goal is to implicate further research.

Since, analyzing governmental motivation towards (non-) transparency and how the complex nature of AI relates to this is the topic that will be focused on, the leading question that flows from this interest is:

<u>**What are the obstacles related to being transparent in AI-assisted governmental decision-making?**</u>

In order to structure the literature review, several sub questions were developed to lay the foundation for answering the main research question.

SQ 1: What preceded the need for the usage of AI in policymaking?

SQ 2: What is artificial intelligence?

SQ 3: What is transparency?

SQ 4: What obstacles do governments experience to being transparent?

SQ 5: What are the related obstacles to transparency for AI?

2. Literature Review

Before explaining what AI and transparency is and highlighting to what extent these two affect each other, some more general concepts related to the role of government in managing information and motives for using new technologies will be outlined to answer SQ 1. Afterwards, the technology will be outlined, answering SQ 2. Thirdly, the versatile concept of transparency will be outlined answering SQ 3. Thirdly, the barriers that governments face in achieving organizational transparency will be presented using a theoretical framework by Pasquier & Villeneuve (2007), this will answer SQ 4 and form the core of the hypotheses proposed. Lastly, SQ 5 will encompass the obstacles related to AI implementation in the public sector that coincide with the framework by Pasquier & Villeneuve (2007). This will lay the foundation for answering the main research question

2.1 Government, information and digital transformation

The working ways of governmental bureaucracies have changed during the course of history from a paper reality to a digital one. This is notable since a primary task of government is to manage information, it is “along with money and people, a core resource of public administration” (Brown & Toze, 2017, p. 582). The dawning of the information age has changed the way in which governments can manage and utilize their information. However, changing environments also leads to changing expectations. Therefore, governments are increasingly “changing their mode of operation to improve public service delivery, be more efficient and effective […] and achieve objectives such as increased transparency, interoperability, or citizen satisfaction” (Mergel, Edelmann, & Haug, 2019, p. 1) by adapting to this technological change.

Since, the process of technological development is very volatile and primarily driven in and by the private sector (Brown & Toze, 2017) governments have to deal with a permanent state of uncertainty. In addition, as Mergel et al. (2019) argue, since digitalization is happening outside of government this also means that citizens' expectations towards government are increasingly demanding. The "inherently unstable nature of the digital environment poses major challenges in terms of investments, management and governance of information, entailing a never-ending process of change management" (Brown & Toze, 2017, p. 584).

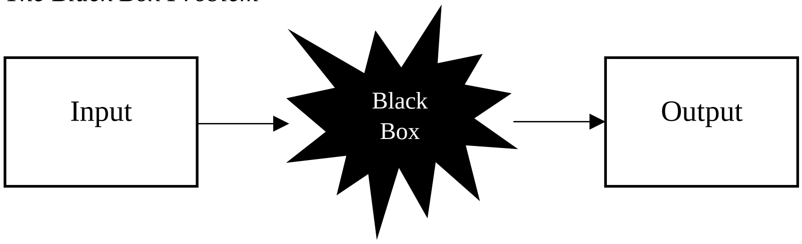
However, it is notable that "terms like digitization, digitalization, or digital transformation are used interchangeably in the literature" (Mergel et al., 2019, p. 1). Even though concepts related to digital transformation are used exhaustively, it is important to stipulate a concrete definition. Mergel et al. (2019) attempt to create a shared, empirically grounded definition for digital transformation by conducting expert interviews that highlights its comprehensive facets. It is more than just a mere alternation from analog to digital. The authors argue that "digital transformation is a holistic effort to revise core processes and services of government beyond traditional digitization efforts. It evolves along a continuum of transition from analog to digital to a full stack review of policies, current processes, and user needs and results in a complete revision of the existing and the creation of new digital services. The outcome of digital transformation efforts focuses among others on the satisfaction of user needs, new forms of service delivery, and the expansion of the user base" (Mergel et al., 2019, p. 12). Digital transformation permeates your entire organization.

An important trend of digital transformation that the public sector is experiencing is the usage of algorithms, in particular artificial intelligence. De Sousa et al. (2019) point out that "investment in new AI-based technologies has been one of the critical strategies of the public sector at various levels of government in several countries around the world" (p. 1). As previously mentioned, "AI can be defined as intelligence displayed or simulated by code (algorithms) or machines" (Coeckelbergh, 2020, p. 64), it "refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals" (European Commission, 2018).

However, even though there is an increase in the usage of algorithms in decision-making through its implementation in public systems such as policing and transport, this also creates a greater demand for algorithmic transparency (Diakopoulos, 2016; Pasquale, 2015). Unfortunately, this demand for transparency finds its origin in the citizens' increasing mistrust towards government and consequently, transparency is argued to be the remedy to this mistrust (Grimmelikhuijsen et al., 2013). If we would define transparency as "[...] the availability of information about an organization or actor that allows external actors to monitor the internal workings or performance of that organization." (Grimmelikhuijsen et al., 2013, p. 576), simply showing the AI algorithm that is used to execute an outcome would suffice. However, there are certain AI algorithms that are so complex that the generated outcome cannot be explained. This is what is called the 'black box' problem. It is because "AI algorithms suffer from opacity, […] it is difficult to get insight into their internal mechanism of work, especially Machine Learning (ML) algorithms" (Adadi & Barrada, 2018, p. 52138). This of course "further compounds the problem, because entrusting important decisions to a system that cannot explain itself presents obvious dangers" (Adadi & Barrada, 2018, p. 52138). Figure 1 depicts the black box problem.

Figure 1

The Black Box Problem



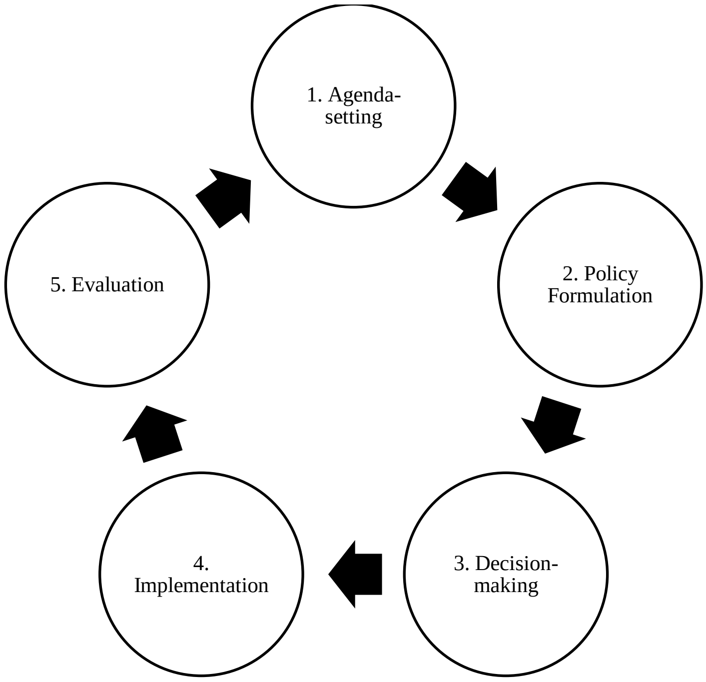
Thus, if a government would like to use AI there is a certain contradiction of interests. To satisfy its citizens a government has to adapt and digitize, if a government were to use algorithms there is a demand that these are transparent, but this might not always be possible. Governments who use AI to simultaneously digitize with its societal environment to meet citizens' demand for effectiveness might ultimately only generate more distrust in being unable to explain a policy outcome that was built on AI.

2.2 Decision-making, the policy cycle and AI

The usage of AI within governmental organizations can aid in policy- and decision-making. At the heart of the decision-making process lies the policy cycle, a generic model which illustrates a process or lifecycle of policy implementation and policy decisions (Höchtl et al., 2016). There are multiple variations on the policy cycle, but following the *Handbook of Public Policy Analysis* it consists of five stages; agenda-setting, policy formulation, decision-making, implementation and evaluation (Jann & Wegrich, 2007), a depiction is shown in Figure 2.

Figure 2

The traditional policy cycle.



This research is primarily focused on the decision-making stage where "expressed problems, proposals, and demands are transformed into government programs" (Jann & Wegrich, 2007, p. 48), and the formal decision to take on the policy is made.

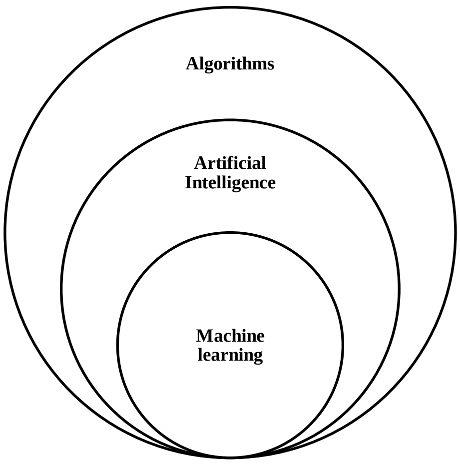
The usage of (big) data and AI can aid in the decision-making process the same way as scientific research and superior knowledge by experts can. Just like research and knowledge are used as advice, data can be used as evidence to support a decision. However, data "shifts the traditional knowledge used to inform policy by combining both objective and subjective measures of need and by increasing granularity of evidence to the level of the individual" (Craglia et al., 2020, p. 98). Providing insights into the needs of the individual for which the policy is created is thus a major advantage that data can provide. It is argued that the usage of data (and AI to process the data) can result in shorter feedback loops for policy evaluation (Craglia et al., 2020). For example, AI can aid by providing predictive analytics and scenario techniques (Höchtl et al., 2016), shortening the evaluation and overcoming the often lengthy implementation phase. Note that AI and its data analysis are used here in a complementary manner, systems making decisions autonomously are not the focus of this research. The next chapter on explaining the technology will show used cases on how AI can be of use in the public sector in general and will explain what the technology of AI encompasses. Firstly, the concept of ‘algorithms’ from which an AI is built will be outlined. Secondly, the concept of AI will be outlined accompanied by some public sector examples to better grasp the technology and its applications. Finally, ML which is a research field that is occupied with creating and using algorithms that can ‘learn’ and improve, the key mechanism that makes AI ‘intelligent’.

2.3 Explaining the technology

Artificial Intelligence has become a buzzword to increase you organizations efficiency and unburden humans of doing repetitive and sometimes dangerous work. It can help in medical care and is said to solve societal problems (Coeckelbergh, 2020). AI is increasingly deployed as a tool to aid humans in fields such as; recruitment, policing and autonomous driving. However, there are genuine fears that shelter in these examples just mentioned. For recruitment, AI is capable to analyze facial expressions and vocal tone through a camera and microphone during a job interview, with the justified fear of hollowing out human interaction (Buranyi, 2019). For policing, AI algorithms can forecast in which area of a city crime is more likely to occur, however, the result of this can be that certain specified racial or socioeconomic groups become scrutinized by police surveillance (Coeckelbergh, 2020). The premise to use AI is that it provides us with objective decisions as opposed to intuitive decisions. The intentions to use the technology are benevolent, but the problems that arise are often unintended consequences of its usage. AI will probably not take over the world or start to eradicate mankind just yet, but these unintended consequences need to be accounted for. In order to debunk the myths of Terminator robots and apocalyptic fiction, lets outline what AI actually is. However, first it should be mentioned that Artificial Intelligence is a container concept. As such, different components will be outlined to gain a better understanding of its workings. It is helpful to think of AI as a Russian doll in order to understand its context. As depicted in Figure 3, it is conspicuous to start by explaining what algorithms are

Figure 3

The relationship between Algorithms, AI and Machine Learning.



since they lay the foundation for AI. Secondly, the origin of the concept of AI is outlined followed by examples, and finally, the key component of ML which makes AI ‘intelligent’, is explained. Firstly, algorithms will now be outlined.

2.3.1 Algorithms

Following the definition of the Cambridge dictionary, algorithms are "a set of mathematical instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem" (Algorithm, n.d.). In essence, this means that algorithms are about doing something in a specific way following some kind of steps, a tool that can be used to help in solving particular problem (Louridas, 2020). It can be carried out by machines, people or nature, it is essentially a process that transforms information (Denning, 2007). An algorithm should include specific features: “its steps should be put into a *sequence* [...] steps may describe a *selection* that determines which steps to follow [...] steps can be put into a *loop* or *iteration*, where they are executed repeatedly” (Louridas, p. 19). These steps are called control structures.

It is inevitable to give a certain mathematical example to illustrate its workings. One of the most popular examples being Euclid’s algorithm. As we will see, this particular algorithm includes all the before mentioned control structures.

Without getting too clouded in mathematical terms, the necessity to mention this algorithm is due to its relevance for computer science as a whole. Euclid’s algorithm laid the foundation for measuring the efficiency of an algorithm i.e. stimulating the ideas of measuring how many steps it takes to get to a certain outcome. This is ultimately related to “the amount of resources it requires to run[, such as] time, how long it takes, and space, how much storage it requires in terms of computer memory” (Louridas, 2020, p. 31) and is called ‘algorithmic efficiency’.

Euclid’s algorithm is a tool to find the greatest common divisor – the greatest number that divides both without a remainder – of two integer numbers. As Louridas (2020) explains, it is a division which is repeated until it makes no sense to repeat it, $a = q \times b + r$, and it consists of two steps.

1. To find the greatest common divisor of a and b, perform the division of a by b. This will give us a quotient and remainder. If the remainder r is equal to 0, then we stop, and the greatest common divisor of a and b is b (Louridas, 2020).

2. Otherwise, we go back to step 1, but this time *b* will be the new *a* and *r* will be the new *b*. Or in other words, we go back to step 1, setting *a* equal to *b* and *b* equal to *r* (Louridas, 2020).

Table 1

Example of Euclid's Algorithm.

	$a = q \times b + r$		
a	*q*	*b*	*r*
178	4	42	10
42	4	10	2
10	5	**2**	0

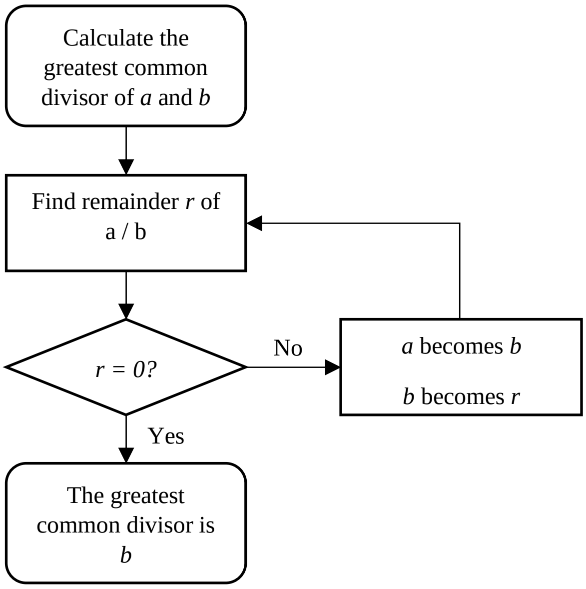
If for example, we try to find the greatest common divisor of 178 and 42 the algorithm should be filled in as follows. *a* = 178, the biggest number, *b* = 42, the smaller number, *q* = the amount of which *b* fits within *a* (the *quotient*), *r* = the *remainder*. If we fill in the steps the result will look like in Table 1.

The greatest common divisor of 178 and 42 is 2. If we would visualize this in a flowchart as if it were programmed, it would like Figure 4.

This tool which Louridas (2020) applicably stylized, follows all three basic control mechanisms which an algorithm should include: sequence, selection, loop/iteration. The next section will outline the origins of the concept of AI.

Figure 4

A stylized Euclides' algorithm. Louridas (2020).



2.3.2 Artificial intelligence

The infamous Dartmouth workshop that took place in Hanover 1956 is argued to be the starting point of contemporary AI where John McCarthy first coined the term, embracing digital machines and the simulation of human intelligence (Coeckelbergh, 2020). As mentioned earlier, "AI can be defined as intelligence displayed or simulated by code (algorithms) or machines" (Coeckelbergh, 2020, p. 64), it "refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals” (European Commission, 2018). It is a field which aims to make computers do things that the human mind can do (Boden, 2016). Psychological skills “such as perception, association, prediction, planning, motor control" are all part of the “richly structured space of diverse information-processing capacities" (Boden, 2016, p. 1) which constitutes AI. As Coeckelbergh (2020) puts it, “AI can be defined both as a science and as a technology" (p. 67), overlapping with fields such as cognitive science, psychology, data science and neuroscience. As a technology it is developed to aid in completing tasks. It is a tool that is created by human beings that appears to be intelligent by analyzing its environment and data and executing tasks with a certain autonomy (Boden, 2016).

It is helpful to give a few examples of what AI as a technology can look like. Most of the time it is part of another technological system such as in robots, algorithms and machines (Coeckelbergh, 2020). AI can “take the form of software running on the web (e.g. chatbots, search engines, image analysis), but AI can also be embedded in hardware devices such as robots, cars, or “internet of things” applications” (Coeckelbergh, 2020, p. 69). Several examples of AI applications that can specifically be applied in the public sector are: AI-Based Knowledge Management Software; AI Process Automation Systems; Virtual Agents; Predictive Analytics & Data Visualization; Recommendation Systems; and Speech Analytics (Wirtz et al., 2019), these applications are further specified in Table 2.

AI software can’t run without a physical infrastructure and hardware (Coeckelberg, 2020). It is therefore important to note that what lies at the heart of what determines the ‘intelligence’ of an AI is software: “an *algorithm* or a combination of algorithms” (Coeckelbergh, 2020, p. 70). Now the final and inner layer of the doll will be explained, ML which is a research field that occupies itself with using and creating algorithms that can make computers ‘learn’ and seem ‘intelligent’.

Table 2

Overview of several AI examples in the public sector.

AI Application	AI Functionality	Use Cases
AI-Based Knowledge Management Software	To efficiently systemize knowledge i.e. gather, sort, transform, share and record.	Clinical documentation that is supplemented by AI (Lin et al., 2018)
AI Process Automation Systems	Automating repetitive governmental tasks	Increased efficiency for processing immigration application forms (Chun, 2008)
Virtual Agents	Software that can interact with humans and perform tasks	Chatbot that assists refugees fill out and search documents (Mehr, 2017)
Predictive Analytics & Data Visualization	Statistical analysis based on quantitative data, can process big data to generate prescriptive and predictive analysis	Predicting ground water levels (Kouziokas et al., 2017)
Recommendation Systems	A system that can filter information	E-service provision for personalized information for employees (Cortés-Cediel et al., 2017)
Speech Analytics	Ability to understand and respond to natural language	Medical work assistance with voice to text transcription (Collier et al., 2017)

Notes: *These examples were succinctly enumerated by Wirtz et al. (2019).*

2.3.3 Machine learning

Before explaining what ML entails, the development that led to its necessity should be outlined first. Due to the global technological advancements there is a huge amount of data being produced by all the computerized machines that we own. At first, this was just a by-product that was being stored, but its potential utility has changed it into a resource (Alpaydin, 2016). This produced data during the last three decades resulted in ‘big data’, a concept coined and popularized by computer scientist John Mashey around 1990 (Lohr, 2013). Big data are “very large sets of data that are produced by people using the internet, and that can only be stored, understood, and used with the help of special tools and methods” (Big Data, n.d.).

The enormous heap of data that was being generated by the increase of digital devices sparked interest in its usage and “with this question, the whole direction of computing is reversed [...] data was passive [and now] data starts to drive the operation; it is not the programmers anymore but the data itself that defines what to do next" (Alpaydin, 2016, p. 11). Humans are obsessed with predicting, fantasizing about the bigger questions in life and what will happen next; will I find my true love?; will I ever achieve my dreams?; who will win next election?; what stocks should I invest in next? Data can aid in predicting.

If you look at human behavior it is – most of the time – not completely random. In a supermarket, the products a customer buys can be deemed complimentary or related to a certain season, analyzing such patterns in behavior is where data can be of use (Alpaydin, 2016). Data that is collected might hold certain patterns, if you can find such patterns and assume that the near future will not be much different from when the data was collected, predictions can be made (Alpaydin, 2016).

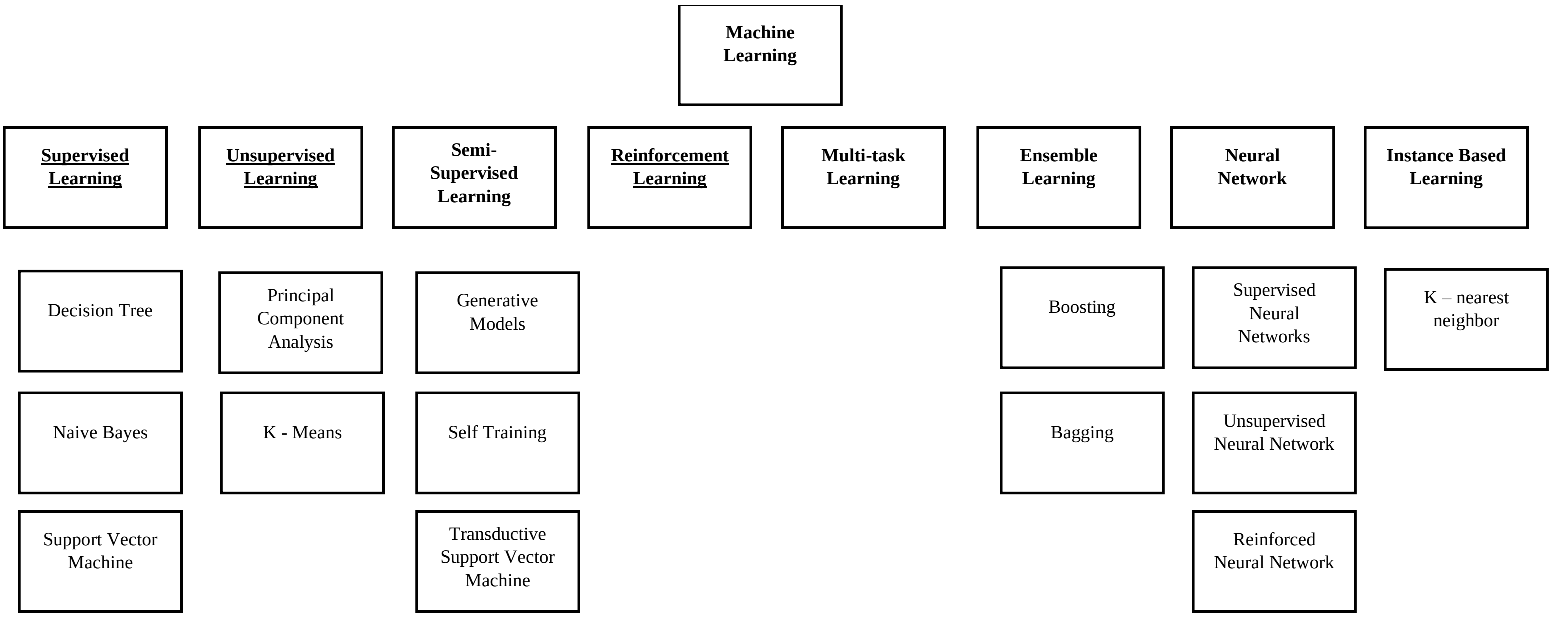
ML is an advanced research field within data science that is concerned with techniques that give a computer the ability to learn without them being programmed to. With this technique, certain patterns and relations can be found within big data sets. The challenge for ML is to generate a program that 'fits' the given data (Alpaydin, 2016). It entails the endeavor to choose an algorithm from a whole set of algorithms that explains the relationships between features in a dataset best and encode them as a computer program to generate a suitable model (Kelleher, 2019). Mapping the relationship between an input and output are often called 'functions' in the field of mathematics (Kelleher, 2019). Researchers tend to speak of 'functions', however, "in ML the concepts of function and model are so closely related that the distinction is often skipped over and the terms may even be used interchangeably" (Kelleher, 2019, p. 13). To sum up, an algorithm is a predefined process that a computer can follow that identifies patterns in data. A pattern that shows a relationship between a certain input and that will always return the same output is called a function. Discovering and learning functions from data whilst using algorithms to do so is the goal of ML (Kelleher, 2019). Encoding the best function into a computer program is then called a model.

It is important to mention that a ML program differs from an ordinary computer program in the sense that an ordinary program requires different values to be assigned to the programs parameters for it to do different things, a learning algorithm adjusts its parameters by optimizing a performance criterion defined on the data (Alpaydin, 2016). In other words, the learning algorithm updates its parameters (values which the model uses to make a prediction) gradually after it receives input to ultimately improve its performance. Thus, 'learning' by adjusting.

Dependent on the task at hand and the data it uses, a learning algorithm can differ substantially. As Jordan & Mitchell (2015) argue: "a diverse array of machine-learning algorithms has been developed to cover the wide variety of data and problem types exhibited across different machine-learning problems" (p. 255). Dey (2016) shows a useful summary of the different ML models and its learning algorithms in his article that are shown in Figure 5. This is of course but a grasp of a field that is constantly developing. It is out of the scope of this thesis to discuss all the methods that Dey (2016) mentions in his article. The next section will briefly touch upon the most common ML methods. ML algorithms can do many things; pattern recognition, learn associations between instances and even take autonomous action (Alpaydin, 2016). The founding paradigms of ML: supervised learning; unsupervised learning; and reinforcement learning will be outlined.

Figure 5

Types of learning (Dey, 2016, p. 1174).



***Notes**: The underlined models will be explained.*

2.3.3.1 Supervised learning

Supervised learning is a type of ML method where the input and output are known and the ultimate goal is to learn a mapping from input to output (Alpaydin, 2016). A key attribute of supervised learning is that it involves human agency in the sense that a supervisor administers correct values and the parameters are updated to create an output that can get as close as possible to the desired one (Alpaydin, 2016). Input and output are known and a human can provide the desired output.

Another important aspect of supervised learning is the way in which the model is trained to become more accurate. In supervised learning there is a dataset that is being used to train it, ultimately being able to detect patterns, relationships and yield good results when presented with new data (Rouse, 2020). As Alpaydin (2016) appropriately argues, a student that is only capable of solving the assigned exercises has not mastered the subject, the goal is to get a general understanding from the exercises so similar new questions about a same topic can be solved, this is no different for supervised ML.

Thus, an important goal when using supervised ML is pattern recognition under ‘supervision’ and recognizing the pattern as accurately as possible due to training. Supervised learning generally creates two types of results; regression and classification (Rouse, 2020). This can be used for a variety of things: text recognition, natural language processing, face recognition, speech recognition, sales forecasting, financial performance comparison, predictive analysis, pricing etc. (Rouse, 2020; Alpaydin, 2016), as long as it is known what the result should be.

Using supervised learning can result in accurate predictions but can be very costly both in money and time since you’ll need a supervisor and the training data for training, and it can be complex to compute (Rouse, 2020). In order to train a model ‘labelled data’ is needed, labelling (structuring) data is costly and therefore often outsourced. In addition, training the model is a process of trial and error. Further, supervised ML is quite limited in its applications and does not provide unknown information about the data (Joy, 2020a). However, supervised learning is a simple process to understand even though it is difficult to compute (Joy, 2020a).

2.3.3.2 Unsupervised Learning

In unsupervised learning, there is no predefined output, and thus no supervisor to train the model, there is only the input data (Alpaydin, 2016). The goal of unsupervised learning is to find regularities in the input by unleashing an algorithm on the data, there are no correct output values (Rouse, 2020; Alpaydin, 2016).

Unsupervised learning is primarily applied for two reasons; clustering and association (Rouse, 2020). Clustering entails the uncovering of groups in data, association entails attempting to find rules that ‘predict’ the data e.g. “people who buy X are also likely to buy Y” (Alpaydin, 2016, p. 118). These predictions can be developed further into recommendation systems, example being the ones used in online advertisement every day (Alpaydin, 2016). Clustering differs from classification used in supervised learning in the sense that the groups to be formed are yet unknown. An example being the classification of news articles, sometimes classifying them in predetermined categories such as sports or politics is oversimplified and not sufficiently informative as some articles can be about e.g. climate change technologies (Rouse, 2020), generalizing it under either technology or climate would be insufficient.

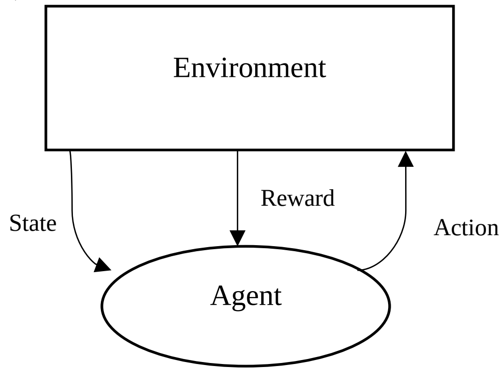
Unsupervised learning is often less complex to compute but harder to understand. It takes place in real time but is also less accurate since the desired result is unknown (Joy, 2020a). It also does not allow to estimate or map the results of new data, and results can vary quite a bit in the presence of outliers (Caballé et al., 2020). However, the fact that there is little to no human agency makes it easier to implement since it requires less human resources.

2.3.3.3 Reinforcement Learning

Reinforcement learning is a type of ML that allows for a ‘agent’ that is being placed in an ‘environment’ which alters its ‘state’ to learn due to trial and error (Bhatt, 2019). In reinforcement learning the agent “”[learns] with a critic”, as opposed to the learning with a teacher that we have in supervised learning” (Alpaydin, 2016, p. 127). Instead of feedback that is presented in the supervised model as a correct set of actions for a task, reinforcement learning works with rewards and punishments indicating positive or negative behavior (Bhatt, 2019). It is important to mention that unlike the two previous methods there is no external feature that provides training data, the agent creates its own data whilst experimenting with different actions in the environment and receiving feedback in the form of a reward (Alpaydin, 2016). Figure 6 depicts reinforcement learning.

Figure 6

Basic setting for reinforcement learning where the agent interacts with its environment (Alpaydin, 2016, p. 126).



The goal of reinforcement learning differs from unsupervised learning. For reinforcement learning the goal is to find a best action model that would generate the highest cumulative reward for the agent (Bhatt, 2019; Alpaydin, 2016).

One of the best way to understand this ML method is by looking at its first applications which is mainly in gameplay and robotics. For example chess, it has an agent (the player), an environment (the playboard), a volatile state (that changes after each move) and a reward (of winning the game). Scientists have been developing programs that can beat humans in games for a while, a famous example was in 1997 where IBM’s DeepBlue beat world chess champion Garry Kasparov (Harding & Barden, 2011). After machines were capable of beating humans, scientists and programmers stumbled into an arms race attempting to create the best program. In 2017, the AI called AlphaZero created by Google beat the champion chess program, Stockfish 8, in a 100-game match up after teaching itself in four hours (Gibbs, 2017). AlphaZero has even been generalized to learn other games such as Shogi

(Japanese chess) and Go, a complex Chinese boardgame (Gibbs, 2017).

The important takeaway is that what made AlphaZero different from its competitors is that it had no human input, only the basic rules of chess and it learned itself the rest by playing itself repeatedly with self-reinforced knowledge (Gibbs, 2017). Just like unsupervised learning, reinforcement learning doesn’t require human agency. Reinforcement learning can be used to create a perfect model for a particular problem, corrects its own errors that occurred during the training process and can be deployed when the only way to gain information about an environment is to interact with it (Joy, 2020b). However, it requires a lot of computation power due to its repetitive nature, and it is hard to determine what actions led to the reward and should get the credit otherwise known as the ‘credit assignment problem’ (Joy, 2020b; Sutton, 1985). Finally, since it is limited to solving a particular problem, its self-correcting and learning properties are often combined with other ML techniques (Joy, 2020b). The interpretability of this method differs depending on the difficulty of the models used. A hierarchical decision-tree (like Euclids’ algorithm in Figure 4) that schematically shows how the agent reached a single reward is easier to understand than a constantly changing environment with multiple rewards (e.g. self-driving cars).

To conclude, supervised learning that can help predict outcomes, unsupervised learning that can generate new insights that humans haven’t thought of before and autonomous reinforcement learning are all associated with the psychological skills such as perception, association, prediction, planning, motor control that are part of the information-processing capacities mentioned by Boden (2016) that are inherent to human intelligence. They laid the foundation for what is nowadays understood as Artificial Intelligence, and are the three prominent paradigms for ML. Each of these three ML methods and the methods mentioned by Dey (2016) in Figure 5 have cross relatives, entailing (un)supervised or reinforcing characteristics.

2.3.4 Interpretability and human agency

In order to achieve transparent machine learning models, it is important for them to be interpretable. As will be argued in the next section, interpretability and understanding is key in order to achieve meaningful transparency. As such, transparency can be expressed as interpretability.

Sometimes solely knowing the ‘what’ and not the ‘why’ suffices when using ML, e.g. when an outcome has no significant consequences or when you don’t want to enable malicious people or programs to manipulate the system (Molnar, 2019).

On the other hand, using ML in a governmental context usually has significant consequences e.g. denying people resources. As Molnar (2019) argues, the more a decision made by a machine affects a person's life, the more the necessity for the machines explanation of its behavior. In addition, Shen (2020) argues that interpretability provides contestability, i.e. the ability for a duped individual to make an appeal to a machine driven decision. Interpretability is essential in order to protect citizens from erroneous machine made decisions. As Doshi-Velez & Kim (2017) argue, interpretable ML models enable humans to check for certain traits such as fairness, ensuring no bias or discriminating effects underlie the decisions made; privacy, ensuring information is sufficiently safeguarded; and generally fosters trust in and public support for ML systems.

The easiest way to achieve interpretability of ML models is to use algorithms that generate interpretable models themselves due to its simplicity, however, this limits the user in using certain types of models which often have lower predictive utility (Molnar, 2019). In reality, ML models are combined in order to increase their predictive utility e.g. unsupervised reinforcement learning. In order to determine what is needed for each of these models to make them interpretable or transparent can be quite difficult. This is due to the tradeoff between 'model interpretability versus model performance'. Interpretability versus performance states that as model's accuracy increases so does the model's complexity, at the cost of interpretability (Brownlee, 2020). Evidently, the most complex models often yield the most accurate results.

However, apart from the degree of complexity, assuming that a ML model will not provide an explanation itself, human agency is needed in the decision-making processes of these models in order to make them interpretable. This is also advocated within the European Commission's *Ethics guidelines for trustworthy AI*, a report that was generated by a High-Level Expert Group on Artificial Intelligence (2019). Human oversight is deemed a key requirement for the usage of AI since it ensures that an AI system does not cause adverse effects and cannot undermine human autonomy (European Commission, 2019). The type of human oversight differs per system, this resulted in three types of governance mechanisms: human-in-the-loop (HITL); human-on-the-loop (HOTL) and human-in-command (HIC) (European Commission, 2019).

HITL refers to a systems capability for human intervention within every decision process of the system (European Commission, 2019). This seems most suitable for the Supervised Learning model since it is requires intensive human interaction with a data scientist training the model.

HOTL refers to the capability for monitoring the system's operations and the ability for human intervention throughout the design cycle (European Commission, 2019). This seems most suitable for the Unsupervised Learning model since the monitoring aspect provides the autonomy that the model requires and emphasizes the importance of the design choice for a suitable algorithm prior to its deployment.

HIC refers to overseeing the overall activity of an AI system and the ability to decide whether and how to use the system in a particular situation (European Commission, 2019). This governance mechanism seems most suitable for Reinforcement Learning since the autonomous nature of the model and the ability to master a single task emphasizes the importance of subsequently deciding whether the system should be deployed at all.

Table 3

Supervised Learning, Unsupervised Learning and Reinforcement Learning.

	Supervised Learning	**Unsupervised Learning**	**Reinforcement Learning**
Goal	Task Driven (predicting the next value)	Data Driven (finding structure)	Learn from mistakes (achieving the cumulatively highest reward)
Interpretability	Simple	Difficult	Differs
Uses	Regression and Classification	Clustering and Association	Gameplay and Robotics
Transparency relevance*	Human-in-the-loop	Human-on-the-loop	Human-in-command
Pros	Can predict accurate results, is simple to understand	Can be used on unlabeled data, easier to compute, little human agency	Little human agency, corrects own errors, can master a specific task
Cons	Time/money consuming, does not provide new insights from data, can be difficult to compute	Is hard to understand the process, does not provide accurate results	Hard to determine what actions led to the reward and should get the credit, requires lot computing power

***Notes:** *Transparency relevance refers to what type of human interaction is needed in order to make the ML models transparent.*

The choices made in assigning governance mechanisms to different ML models are debatable, especially since the models are combined in practice. However, in order to give an impression of the type of governance needed in order to make a model transparent, these assignments were based on the characteristics of the ML models. A summary of the mentioned methods are shown in Table 3 above. Now that the technology has been outlined it is important to define the contested concept of transparency.

2.4 Defining Transparency

Transparency has become a recurrent concept that promises to enhance public trust in government, it “shows whether goals and promises are being fulfilled and whether decisions are made in a prudent manner” (Grimmelikhuijsen et al., 2013, p. 579). It is often related to ethical concerns and is considered to be inherent to a fair society and a legitimate government. The Cambridge Dictionary describes it as “the characteristic of being easy to see through” (Transparency, n.d.), but this description is too minimalistic for this inexhaustible concept.

The postmodernist school of thought argues that words are more than just mere speech, they can become metaphors that contain symbols of ideas due to the way words can adopt meaning through its usage (Weick, 1995; Yanow, 2003). As such, words such as ‘transparency’ can have paradoxical features, on the one hand transferring the notion of openness whilst on the other hand generating the presence of secrecy (Stone, 2012). On the other hand, transparency is often considered inherent to better governance, but it is not a goal in itself, rather a means to achieve other goals (Grimmelikhuijsen et al, 2013).

Transparency is a recurrent concept in the public administration literature. In her article, Ball (2009) outlines the definition of transparency through its usage in international relations literature, public policy literature and literature on American politics and public administration. This resulted in three metaphors for the word transparency. Firstly, transparency as a norm of behavior to battle corruption, it is indirect in the sense that if citizens have information governance improves (Ball, 2009). Secondly, transparency as open government, where open decision-making is similar to transparency and accessibility to governmental information is a determinant. Lastly, transparency as an element of good policy, where it is integrated in the policy-making process and exerted by the policy officers making transparency a continuum (Ball, 2009).

Whilst describing a normative concept such as transparency, it helps to give a few examples of what transparency measures could look like. Apart from the definition of the concept used in the literature, Grimmelikhuijsen et al. (2013) concretize transparency by looking at the decision-making processes of government. Grimmelikhuijsen et al. (2013) refer to separate events of the processes: 1) decision-making transparency, openness about the rationale of and steps taken to reach a decision 2) policy transparency, disclosure of the adopted measures, implementation and implications of the policy itself and 3) policy outcome transparency, the provision of information regarding the effects of the policy measures.

In practice, sometimes governments and corporations choose to be voluntarily transparent. If there is no incentive to be transparent, governments can intervene with targeted policies that compel transparency (Diakopaulos, 2014). This is called ‘targeted transparency’, Weil et al. (2013) outline targeted transparency – which differs from the open-government policies – as using new scientific evidence in relation to public risks in order to achieve a specific goal, e.g. showing calories in restaurants in order to reduce obesity. Other examples can be automobile safety tests and restaurant inspections (Diakopaulos, 2014). These examples show ways for government to compel transparency from non-governmental organisations and are usually initiated when there is certain “missing information that might have bearing on public safety, the quality of the services provided to the public, or issues of discrimination or corruption” (Diakopaulos, 2014, p. 11).

Evidently, government can also compel transparency on itself. Examples of governmental self-disclosure are often anchored in legislation. Important examples are the American Freedom of Information Act (FOIA), or the similar Dutch *Wet Openbaarheid van Bestuur* (WOB), both laws for which government has to provide documents to constituents on request (Cogianese & Lehr, 2019; Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, n.d.). In addition, governmental self-disclosure can also take the form of ad hoc research committees that are put into power by the legislative branch following a certain disaster, accident, failed project or disturbance. Since the legislative branch does not always have the time and resources to conduct their own research, and the results have to remain politically unbiased, this is done by external assigned committees (*Externe Onderzoeken*, n.d.).

On the contrary, a more passive form of governmental self-disclosure that does not require citizens to actively rely on legislation is the ‘open government doctrine’. Open government refers to the right of disclosure and access to government information in order to have effective public oversight, and its goals are substantiated by the political assumption that information motivates and empowers citizens to exercise choice (Lathrop & Ruma, 2010; Dahlberg, 2011). The principles that constitute the open government doctrine entail: “opening public sector information and data and enabling citizens and entrepreneurs to access government-held data in a uniform way; opening government processes and operations to the public; explaining decisions and actions to the citizens, acting on requirements expected for the task and accepting responsibility for failure; engaging citizens in decision making; [and] enabling cooperation across different levels of government, between the government and private institutions and between the government and the citizens” (Veljković et al., 2014, p.

279). The Dutch government also engages in open government initiatives in their *Actieplan Open Overheid*. These initiatives have the similar ambition to improve access to government information, be accountable to society and promote government's cooperation with society (Ministerie van Algemene Zaken, 2017). Essentially, looking at the literature on transparency as well as the examples mentioned, it becomes apparent that what transparency essentially means in a citizen-to-government relation is the disclosure of and access to governmental information.

2.4.1 Data and information

Apart from the semantics of the concept of transparency it can take on many forms, and in the classical sense it consists of governmental documents which contain the desired information that was requested by citizens. These documents can then again take on many forms such as "reports, notes, minutes of meetings, e-mail, even unwritten documents such as telephone conversations" (Pasquier & Villeneuve, 2007, p. 149). It is important to make a distinction between the classical documented information and the information that is being used in AI. As was previously mentioned, AI needs data to operate, this data becomes input for a ML algorithm which then produces output data. Creating information from data also means that simultaneously information can be turned into data if you quantify it and store it on a computer. Data is an abstraction of a real-world entity for which the data describes its attributes (Kelleher & Tierney, 2018). For example, the entity of a person can have the following attributes: age, height, weight, nationality, gender etc. This example describes one entity, but data used for ML algorithms comes in greater volumes or 'datasets' relating to a variety of entities. An example of a real governmental dataset that describes individuals receiving social benefits and the duration of these benefits is shown in Table 4.

Table 4

Sample of a dataset from the Dutch Statistical Office describing individuals receiving social benefits and the duration of these benefits.

ID	Gender	Age	Migration background	Duration of benefits	Time periods	Persons receiving benefits
0	T001038	10000	T001040	T001066	2008KW01	317080
1	T001038	10000	T001040	T001066	2008KW02	312230
2	T001038	10000	T001040	T001066	2008KW03	305350
3	T001038	10000	T001040	T001066	2008KW04	304190

Source: Retrieved from:
https://opendata.cbs.nl/statline/portal.html?_la=nl&_catalog=CBS&tableId=82663NED&_theme=36 on 07-03-2021.

Referring back to the Cambridge definition of 'Big Data' it is a "very large sets of data that are produced by people using the internet, and that can only be stored, understood, and used with the help of special tools and methods" (Big Data, n.d.). It is particularly the latter part of the definition that is of importance here, simply having data without knowledge of the tools and methods to help understand it makes it fundamentally different than documentary information. This is no different for ML algorithms that are coded into software, without knowledge of the coding 'language' it becomes nearly impossible to be able to understand it.

Literature from the knowledge management discipline on the 'data-information-knowledge-wisdom hierarchy', further amplifies that data is not the same as (documented) information. The DIKW hierarchy is a prominent model in the field of knowledge and information management and is often target of debate where the challenge is to understand and explain how data is transformed into information, information into knowledge and knowledge into wisdom (Rowley, 2006). The model depicts a hierarchical pyramid which rests on the assumption "that data can be used to create information, information can be used to create knowledge, and knowledge can be used to create wisdom" (Rowley, 2006, p. 164). In her article, Rowley (2006) analyzes how data, information, knowledge and wisdom are defined in 16 scholarly textbooks in order to extract common definitions. Data is nothing more than just symbols that lack meaning or value, it is unorganized, unprocessed and generally useless if you don't know what is means (Rowley, 2006). Table 4 is an example. On the contrary, information is data with context and meaning, data that is processed for a specific purpose making it meaningful, valuable and useful (Rowley, 2006). An example of data being the fact that it's raining outside whereas information would offer enriched data including that the temperature has dropped, the air became more humid and it has started raining at 5 pm.

Now that it has been outlined that data and information are not self-explanatory similar to each other, the next section will describe what is necessary in order to reach meaningful transparency.

2.4.2 Achieving transparency

In order for (governmental) transparency to be meaningful, the information that is being presented needs to be understood by the recipient (Moss & Coleman, 2014). Solely publishing governmental transparent data is not enough. Swartz (2010) argues that in order for transparency to be useful, three actions have to occur: governments have to provide the right transparent data; published data must be correct; and citizens must understand the data that was published. All these conditions have to be met or else governmental transparency fails.

However, the latter condition is often not met due to several reasons. In communication terms, these reasons can be accounted to both the 'sender' (government) as the 'receiver' (citizen). As for the sender, too much data that governments publish are in their 'raw' form, i.e. too detailed and technical. The data being published is simply too difficult to understand as opposed to properly summarized data. Often, this can be inherent to the design of e-government services and the way in which information is presented throughout these services (Jaeger & Bartot, 2010).

Secondly, apart from the form in which information is presented, the quantity is also crucial in achieving meaningful transparency through understanding. Ultimately, transparency is about communicating information and data, meaning that the recipient of the information has to be able to understand it. Too much information can occlude information due to information overload and over exposure resulting in the inability for citizens to understand the 'raw' and excessive data (Annany & Crawford, 2018; Heald, 2012).

As for the receiver, problems of comprehending transparent data for a citizen can also be several. The inability for a citizen to comprehend the data presented can be due to: unfamiliarity with the legal framework in which the state operates or the processes take place; lack of depth with the subject at hand; lack of sufficient expertise to be able to determine what is deemed important; and lack of background knowledge of the policy or its outcome (Bannister & Connolly, 2011).

Worthy (2015) identifies that open data's underlying goal of increased information transmission towards citizens is subject to a fundamental problem, the "misunderstanding about data – [that] they are not 'power' by themselves: they require narrative and explanation" (p. 797). It is important to mention that open data and open government are concepts that remain puzzling due to its inconclusive usage and blurring of the distinction between the politics of open government and the technology of open data (Worthy, 2015). However, the need for context and explanation for the 'raw' data that is being disclosed seems indispensable. In earlier research on the new Transparency Agenda in the UK, governmental web statistics showed the public had almost no interest in disclosed transparent data (Worthy, 2013). In addition, "the data lacked context and comparability, with inconsistencies even for basic information such as dates" (Worthy, 2015, p. 797).

However, it is debatable whether it should be the government's responsibility to provide explanation and context. Margetts (2006) argues that that the responsibility lies with the citizens to understand the transparent data, not with the public authority. Preparing information in bite-sized formats for citizens to understand is unquestionably a challenge that governments face in order to be transparent.

2.5 Public sector obstacles to transparency

Since the concept of transparency is now outlined and the goal of this research is to explain the barriers to transparency, a theoretical framework by Pasquier & Villeneuve (2007) will now be presented in order to structure the arguments that will eventually be the core of the hypotheses proposed. Pasquier & Villeneuve (2007) argue that access to information through laws underlie the transparency of documents i.e. the opportunity to request documents that hold the desired information through judicial means. This entails transparency of documentary information.

However, even though there is general notion that documentary transparency is a self-evident right, governments can still – willingly or unwillingly – be recalcitrant in sharing information. Based on document analysis such as legal decision, official reports stipulating a lack of transparency despite regulations, and research work and publications that describe obstruction of transparency, a framework that describes the behavior used whilst denying transparency is created by Pasquier & Villeneuve (2007). Five types of impediment to transparency in information are presented in Table 5.

Table 5

Types of impediment to transparency in information (Pasquier & Villeneuve, 2007, p. 152).

	Not subjected	**Subjected**			
	Legal	Illegal	Legal		
	Non-Transparency	Averted transparency	Obstructed transparency	Strained transparency	Maximized transparency
Description	The concept of transparency does not apply. Transparency is only voluntary	The organizations directly disobey the law (refuse to participate)	Obstructions to transparency through using provisions of the law	Inability to cope with transparency due to an absence of resources or misunder-standing of information	Behaviour intended to forestall possible demands by making all the information available
Justification	'It's not necessary'	'The file doesn't exist'	'It wouldn't be responsible'	'We don't have the resources'	'It's simpler and less costly'

Firstly, there is non-transparency, this is typically the case when the whole of the organization or some activities are absolved from the accountability of information disclosure (Pasquier & Villeneuve, 2007). This can be the case when: 1) organizations are – at their own inquiry – excluded from the scope of law to give information, such as the General Intelligence and Security Services, 2) new corporate bodies are created that carry out certain public duties e.g. foundations or audit organizations, and 3) public services are outsourced to subcontractors, as a consequence information might be concealed to protect commercial confidentiality (Pasquier & Villeneuve, 2007).

Secondly, averted transparency, concerns the situation when an organization that is formally subject to the law but nonetheless willingly and illegally obstructs access to information (Pasquier & Villeneuve, 2007). A deliberate example of such behavior is the destruction or concealing of documents, a non-deliberate example poor document management that make it impossible to extract valuable information (Pasquier & Villeneuve, 2007). A second method is to politicize – or deliberately assign a political character to – a certain document, the problem being that there is no method to verify the validity of this decision (Pasquier & Villeneuve, 2007). In a third case, due to the danger of public disclosure of information, there tends to exist an organizational culture in which it is preferred to exchange crucial information and decisions orally rather than report them in a text, this evidently results in information loss (Pasquier & Villeneuve, 2007).

Thirdly, obstructed transparency, using legal means to obstruct access to information e.g. self-censorship or using the law (Pasquier & Villeneuve, 2007). This is a method that is used quite often by governments since it is perfectly legal to deny a request for information. A reason for governments to engage in such practices is the protection of privacy or information that was received anonymously (Pasquier & Villeneuve, 2007). Another reason can be state security and terrorism, terrorist events in the past have installed anti-terrorist laws that tend to increase power of government agents and simultaneously limits access to information (Mendel, 2003). This corresponds to what scholars often call the ‘state of exception’, when the constitutional order is at stake it can lead to a suspension of rights (such as transparency) by concentrating power to the executive (de Wilde, 2010). A final example related to obstructed transparency is the interdependence of states, since economic decisions are taken in consultation at the international level, a member state can nullify information requests (Roberts, 2001).

Fourthly, strained transparency refers to the willingly or unwillingly limiting of access to information due to unfamiliarity with the documents or lack of resources to process the demand (Pasquier & Villeneuve, 2007). Reasons mentioned by Pasquier & Villeneuve (2007) are poor information management or the financial cost of access to information. However, apart from the costs to process the demand of an information request, 'resources' can also be expressed in expertise, or as Howlett (2015) argues 'policy analytical capacity', referring to the "individual level analytical skill (competences) and resources (capabilities) […] on the individual level and specifically on the ability of individuals working in public policy organizations to produce sound analysis to inform their policy-making activities" (pp. 173 & 174). The argument being that an individual does not have the skills and resources to process a transparency demand.

Lastly, maximized transparency, a (hypothetical) situation where the organization discloses all information available and the public need not request it (Pasquier & Villeneuve, 2007). Even though this is often deemed the goal that is pursued by lawmakers, making all the information available can be a barrier to access information (Pasquier & Villeneuve, 2007). Giving too much information and failing to organize it might occlude important information, without proper management all transparency becomes futile (Pasquier & Villeneuve, 2007). As argued earlier, disclosed data by governments is often presented in its 'raw' form making it difficult for citizens to understand. Nevertheless, the open government doctrine underlies the choice to engage in maximized transparent behavior by governments. Besides open government's idealistic arguments to improve access to government information, be accountable to society and promote government's cooperation with society, showing all information would also be simpler and less costly by forestalling potential transparency demands.

2.5.1 Associated challenges for AI

Finally, in order to evaluate governmental motivation towards (non-) transparency and how the complex nature of AI relates to this, challenges of AI within the public sector that share common ground with Pasquier & Villeneuve's (2007) framework will be discussed in this section before formulating the final hypotheses. However, it should be clear that what is to be discussed entails the entirety of an algorithmic system, this entails the data the system processes as well as the steps the system uses to process the information, the algorithms.

Firstly, in relation to obstructed transparency, obstacles that stem from privacy and safety concerns are not uncommon for AI. The challenges are mostly related to technical and legal issues. The technical challenges mainly coincide with security, ensuring that an AI system can perform secure and its impact can be managed, this also includes cybersecurity precautions to ensure that the system is safe from external manipulation and the data from theft (Boyd & Wilson, 2017; Holdren & Smith, 2016). In addition, security issues are not limited to information security but also security in general, preventing that a system's malfunctioning physically harms humans e.g. a malicious hacker initializing the flooding of an autonomous dam.

The legal issues are mainly related to the privacy concerns that the technology potentially instigates. Essentially, privacy entails the right and privilege as an individual to have certain control over how your personal data is used and collected (Jain et al., 2016). However, in the public sector context, a lot of personal information is being collected during citizen centered services rendered by government and 'control' over the use of this information is often excluded. In addition, transparency of a governmental AI system that uses personal information for its insights thus also potentially jeopardizes the privacy of citizens. One of the most pressing contemporary privacy issues is the personal identification of personal information after it is disclosed on the internet (Porambage et al., 2016). This should be considered when attempting to be transparent about a governmental AI system.

Secondly, in relation to strained transparency, an obstacle that stems from unfamiliarity with the information and organizational capacity to process the demand for transparency are also not uncommon for AI. The challenges related to the unfamiliarity with the information correspond favorably with the complexity of AI systems and the expertise needed to work with AI. Given the possibilities of AI applications for process optimization, businesses are hunting for professionals to consolidate their ambitions, as a result there is a global shortage of AI experts. A survey of 400 senior executives from various industries in the public and private sector across eight major markets revealed that talent and skill is one of the major strategic challenges in the contemporary scenario of 2018 (The Economist Intelligence Unit, 2018). In 2019, Marr estimated that there is a need for 1 million AI experts worldwide but the available amount of experts were approximately 300.000. However, a more recent article by Chinn et al. (2020) reveal that Europe alone is experiencing an astonishing shortage of 8.6 million people in the public sector with necessary skills to implement the e-government initiatives that the EU have launched and aspire for 2023. The shortage in the public sector is further amplified given that the sector cannot financially compete. Proper AI expertise is considered a luxury that only hedge funds and tech giants can afford (The Economist, 2020).

The challenges related to the organizational capacity associated to process the demand for transparency are mainly related to the proper management of information, in the context of

AI, the capacity to be able to manage and use the data on which AI would operate. Based on interviews with public officials from 11 organizations in the Dutch public sector, Klievink et al. (2017) identified that a major uncertainty regarding the usage of big data was whether the organization was sufficiently mature for big data usages. Drawing on literature on organizational maturity and by analyzing e-government growth stage models, Klievink et al. (2017) conceptualize organizational maturity as indicating “how far organizations have developed towards a state in which they collaborate better with other public organizations (and their IT) and provide more citizen-oriented services and demand-driven policies” (p. 273). Thus, the degree to which information sharing is possible amongst public organizations determines the degree to which it can properly utilize data. However, what can seriously hamper the interoperability of governmental organizations is the lack of architecture interoperability. Based on expert interviews on the barriers to e-government implantation, Lam (2005) highlighted that integrating systems which had previously existed as “islands of IT” poses a serious challenge in digitizing government. It is the “use of different technology platforms, use of proprietary technologies, the “closed” design of existing applications, absence of application interfaces and differences in development (programming) frameworks” (p. 519) that hampers IT collaboration.

Lastly, in relation to maximized transparency, an obstacle that primarily stems from either providing too much information or not making the information understandable enough is not uncommon in the AI literature. Literature from the arts and humanities discipline advocate ‘people-centered design’ or ‘human-centered design’ where the consideration of humans should be the starting point for AI development, ultimately designing solutions to cater to real people’s needs (Dwivedy et al., 2019). Riedl (2018) argues that when designing AI systems, the ‘nonexpert human’ and his needs have to be considered in the eventual interaction. The nonexpert will most likely not engage in seeking a detailed inspection of an autonomous system and its inner workings but will probably seek compensation for a perceived failure, a ‘remedy’ (Riedl, 2018). Prior to this remedy is an explanation, “explanations are post-hoc descriptions of how a system came to a given conclusion or behavior” (Riedl, 2018, p. 35), this is considered an important step towards transparency as well as accountability of AI systems. Alby & Flyverbom (2019) concur the notion that transparency is more than just information disclosure in order to verify good organizational conduct, transparency also implies a certain process of social interaction that contradicts the premise that more information is always better.

2.6 Hypotheses

Now that the sub questions: *What preceded the need for the usage of AI in policymaking?; What is artificial intelligence?; What is transparency?; What obstacles do governments experience to being transparent? What obstacles relate to governments being transparent about AI?* are answered and the foundation for the hypotheses have been laid, three hypotheses are proposed based on the literature presented. These three hypotheses stem from the framework presented in Table 5 and the additional obstacles known to AI and the public sector. The prior two types of transparency, non-transparency and averted transparency are deliberately excluded for this research. For non-transparency, this is because it would be redundant to analyze organizations to which the concept of transparency does not legally apply. The question why transparency does not apply on certain organizations is another scholarly debate on its own and out of the scope for this thesis. As for averted transparency, it is highly unlikely that the interviewees will admit to deliberately concealing information and disobeying the law.

Privacy and security related concerns

One of the main obligations that a government has towards it citizens is the protection of lives and properties. This right is anchored in the European Convention of Human Rights and includes the right to privacy mentioned in Article 8. To prevent disclosure of personal information and data, similar privacy laws have been created. Obstructed transparency entails using legal means to obstruct access to information, the justification being that it wouldn't be responsible to disclose information (Pasquier & Villeneuve, 2007). The state appeals to security and privacy related legislation in order to prevent being transparent about their internal governmental processes. Sharing this information could either be hazardous for security reasons or could potentially harm the privacy of civil servants whose information is being disclosed. A certain dichotomy underlies this hypothesis, the individual's need for security and privacy versus society's need for governmental transparency.

In the SyRI case presented in the introduction, the State chose not to disclose the inner workings of the system because citizens could adapt their behavior to prevent suspicion based on the released information. This would put citizens in the position to 'game the system' i.e. exploit, manipulate or undermine the system in order to gain an advantage over others. Disclosing the system potentially presents future (financial) security hazards and could promote fraud. The SyRI system was operational in the social domain, managing personal information in order to highlight fraud. However, security issues aren't limited to the social domain, a malfunctioning of an AI system that is operative in the physical domain e.g. that automatically operates emergency lanes on highways, can physically harm humans and result in casualties.

In addition, governments cannot disclose information due to privacy related concerns of the people involved in the process. This could be the civil servants themselves such as Pasquier & Villeneuve (2007) argue, but in an AI context could also be the citizens who are subject to the process i.e. the people whose data is being processed. As mentioned by Porombage et al. (2016), the most pressing contemporary privacy issues is the personal identification of personal information after it is disclosed on the internet. Disclosing personal data for the sake of transparency could consequently present privacy issues. Therefore, security and privacy related issues can prevent the government from being transparent.

> **H_1**: ***Governments are limited by privacy and safety issues to not be transparent in AI-assisted decision-making.***

Lack of expertise and cooperation

This hypothesis refers to strained transparency, the willingly or unwillingly limiting of access to information due to unfamiliarity with the documents or lack of resources to process the demand (Pasquier & Villeneuve, 2007). The technology that drives AI is quite complex and to use AI you need data for it to operate on. There are certain AI algorithms that are so complex that the generated outcome cannot be explained by its own creators. This is what is called the 'black box' problem. It is because "AI algorithms suffer from opacity, […] it is difficult to get insight into their internal mechanism of work, especially Machine Learning (ML) algorithms" (Adadi & Barrada, 2018, p. 52138). Given that AI is complex, in order to gain familiarity with the technology and the information that it creates you need the proper in-house knowledge and expertise. However, the hunt for AI experts by the private sector and the consequential shortage of experts poses a major obstacle for governmental self-sustaining AI innovation. In addition, the financial means necessary to compete with the headhunters from hedge funds and tech giants is inherently an issue of expertise. However, the argument is not that every civil servant that uses the AI tool should be a renowned data scientist or AI expert. Rather, in the context of using AI in policymaking, individual knowledge needed to comprehend the consequences of algorithms and AI in policymaking might also be of importance.

The lack of resources to process the demand can be expressed in the capacity to be able to manage and use the data on which AI would operate. This data is usually versatile in nature since data is often used from different suppliers, e.g. an AI application that is currently being used by the Dutch government to fight undermining crime uses information from 11 organizations (ICTU, 2017). It is therefore essential for organizations to be able to cooperate given that the input for AI requires this. As argued by Klievink et al. (2017), the degree to which information sharing is possible amongst public organizations determines the degree to which an organization is mature enough to exploit the perks of data. However, cooperation can be a major obstacle for governments since the systems that harbor data were previously considered 'islands of IT' and are now expected to cooperate with each other (Lam, 2005). Therefore, it is the lack of in-house expertise and cooperation that poses an obstacle for governments to be transparent about AI.

> **H_2:** ***Governments are impeded in being transparent in AI-assisted decision-making due to lack of expertise and cooperation.***

Ineffective disclosure

Maximized transparency, a situation where the organization discloses all information available and the public need not request it (Pasquier & Villeneuve, 2007), might not be as hypothetical as was being argued. Governments are answering civil society's call for transparency in the form of open government. Open government is multilevel, from municipalities to ministries. Examples of open government initiatives are municipalities that are engaging in 'open data' initiatives by disclosing data about the public sphere, national government that release its parliamentary papers and the 'open source' development method. Open source means disclosing the source code of, for example, a website, program or app making it freely available and possible for anyone to read, modify and distribute the source code (Ministerie van Algemene Zaken, 2017). Open data referring to the disclosure of data that the government uses for their analyzes. Open source and open data are *de facto* transparency enhancing initiatives. Since software embedded algorithms constitute AI, disclosing the source code of the software and the ML algorithm that runs the AI system seems like a positive development since transparency initiatives are considered an epitome of good governance. The disclosure of governmental information would lead to its divulgation and an informed society that can assess governmental legitimacy and take action accordingly (Albu & Flyberbom, 2016).

However, as ironic as it may seem, this hypothesis argues that one form of transparency can be an obstacle to another form. This has to do with the fact that the way in which the 'raw' information about the operation of the system is disclosed cannot effectively be understood by the group for which it is meant, the citizens. This hypothesis argues that the way in which governments disclose their data about AI is inefficient in that citizens have trouble comprehending the information. This hypothesis falls in line with Swartz (2010) and Moss & Coleman's (2014) argument that in order to achieve meaningful transparency, the citizen has to be able to understand the information that is disclosed. It is therefore important for governments to ask themselves who they're attempting to be transparent for and consider their needs when they develop AI systems and engage in transparency projects. The nonexpert human who is subject to an AI system will most likely not need detailed information about the system but would like the government to provide a remedy in the form of an explanation when the system commits errors (Riedl, 2018). Since data is not the same as information (Rowley, 2006), it requires narrative and explanation and this is lacking in the way in which 'open government' initiatives pushes governments to disclose AI related information in the form of open software and open data. As a consequence, transparency is not achieved due to the failing approach in which governments disclose information about AI.

> **H_3:** ***Governments are impeding transparency by ineffectively disclosing information about AI-assisted decision-making.***

3. Research design

This thesis means to focus on exploring the way that transparency can be deemed an obstacle in using AI in governmental decision-making. The way in which governments disclose information has shifted from analog documents to digital data and source code. Since AI has slowly but surely entered governmental decision- and policy-making, it also changed the way that governments are using information. Governments have their reasons to conceal information, the question is how new technology applies to these reasons. Past examples have shown that transparency can be an obstacle in using AI in governmental decision-making, finding out from experts in the field just how AI can impede governmental information, and how it can be an obstacle to its implementation is the goal of this research. For this research, it will be tested to what extent the organizational barriers to transparency proposed by Pasquier & Villeneuve (2007) apply to AI augmented decision-making within the Dutch government by asking the interviewees questions about those particular barriers. Finally, the research method used to answer these questions also creates the opportunity to gauge potential solutions that could lay the foundation for future research.

During the writing of this thesis, the researcher had the opportunity to engage in expert interviews due to an internship at the Dutch Ministry of Infrastructure and Water

Management. By asking superiors about this topic, they eventually redirected the researcher to the right people. In this study, 9 semi-structured expert interviews are conducted with distinctive public servants who work in the field of AI at the Dutch government. An interview guide has been created beforehand and can be found in the Appendix. To ensure that the information required would be obtained, several main research questions were developed. Transcriptions were made and later sent to the interviewees for adjustments and remarks. The data obtained was subsequently refuted against the theory that underlies the hypotheses proposed and formed the foundation for the discussion section.

Based on this design, several hypotheses were proposed, the main research question remained: *What are the obstacles related to being transparent in AI augmented governmental decision-making?*

The hypotheses proposed were:

> H_1: *Governments are limited by privacy and safety issues to not be transparent in AI-assisted decision-making.*
>
> H_2: *Governments are impeded in being transparent in AI-assisted decision-making due to lack of expertise and cooperation.*
>
> H_3: *Governments are impeding transparency by ineffectively disclosing information about AI-assisted decision-making.*

3.1 Sampling

The research method is qualitative and there has been relied on purposive research sampling. In purposive research sampling, the object of research is not selected randomly, on the contrary, the goal is to sample participant and cases that are relevant to the research questions proposed (Bryman, 2016). The reasoning behind the choice for purposive sampling was self-evident. Firstly, the research question heavily suggested that the unit of analysis would be government, sampling randomly would not be practical. Secondly, the fact that the internship made it easier to connect with experts in the field also meant that generalizing the data and research to a wider population seemed unworkable. Besides, purposive sampling is a non-probability sample where there cannot be generalized to a population or other cases (Bryman, 2016). But most importantly, because the goal of the research is to explore the motivation of governmental behavior, and speaking with experts in the field seemed the suitable way. In addition, as Bryman (2016) argues, “researchers basing their investigations on qualitative interviewing […] typically want to ensure that they gain access to as wide a range of individuals relevant to their research questions as possible, so that many different participant perspectives and ranges of activity are the focus of attention" (p. 408).

The choice of interviewing ministerial officials within government rests on a mix of three sampling approaches. Firstly, it is opportunistic sampling, the opportunity to already be part of the organization through an internship has led to an opportunity to connect with relevant participants on the researcher's own initiative. Secondly, it also partly rests on typical case sampling, since the general topic for this research was 'Digitalisation and transparency in public organisations', it was an obvious choice to sample participants in government since this exemplifies the dimension of interest. Finally, the snowball sampling approach also played a great part in finding the appropriate participants. Snowball sampling is an approach where a group of people relevant to the research group propose other participants who have characteristics or experience relevant to the research (Bryman, 2016). A group member from the project group for the Ministry of Infrastructure and Water Management – where the researcher was part of as an intern – held close ties with colleagues who are engaged in AI policy. Thus, introduced the researcher to potentially interesting interviewees who are in some way involved in creating, implementing or inspecting AI policy. From there on, new participants were introduced based on personal recommendations. The participants are divided over the three traditional tasks of government: creation of policy, implementing policy and monitoring policy compliance.

The literature states that it is difficult to determine a minimum sample size in qualitative research since it can differ substantially (Guest et al., 2006; Mason, 2010). Rather than looking at the size, this research will focus on the depth and achieving theoretical saturation i.e. that new data no longer suggests new theoretical insights (Bryman, 2016). In order to have provided sufficient theoretical saturation for the research, the participants would differ substantially in the task they were assigned in government. The premise being that differing policy tasks with differing responsibilities and portfolios would increase trustworthiness through the rich accounts of the details of a culture. Table 6 shows the composition of the sample. As shown, the sample is quite diverse since experts from both the 'social' domain (governing humans) as well as the 'physical' domain (governing assets) were interviewed scattered across the three governmental tasks of policy, inspection and implementation.

Table 6

Composition of the sample.

Respondents	Organization
Policy Officer A	Ministry of Justice and Safety
Policy Officer B	Ministry of Infrastructure and Water Management
Policy Officer C	Ministry of Infrastructure and Water Management
Policy Officer D	Ministry of Infrastructure and Water Management
Policy Officer E	Ministry of Infrastructure and Water Management
Policy Inspection Officer A	Ministry of Infrastructure and Water Management
Policy Inspection Officer B	Ministry of Justice and Safety
Policy Implementation Officer A	Ministry of Justice and Safety
Policy Implementation Officer B	Rijkswaterstaat

Each of the respondents in the sample are in some way engaged in the subject of AI, either in the creation of AI related policy or working on AI related projects. In order to adhere to the request of the respondents to remain anonymous, their work will be discussed in a general manner. As for project related expertise, for example, an interviewee was working on an AI project that uses ML-driven ‘smart camera’s’ for image recognition to automate incident management and rush-hour lane operations for traffic control centers in the Netherlands. Another interviewee was working on a project that intends to bundle data from as much transport and mobility services as possible in order to provide tailor-made traveling from door to door as quickly, sustainable and worry-free as possible. All collected data would be analyzed using ML in order to find patterns and gain insights about how humans move across the country. Another interviewee was engaged in the creation of an organizational framework to assess relevant judicial and ethical aspects of AI to support managerial decision-making, an ‘AI impact assessment’.

As for policy related expertise, for example, an interviewee was part of a specialized AI policy group that was primarily concerned with the sensible application of AI technology in one's own task performance, identifying and mitigating threats emanating from AI and protecting citizens from malicious AI. Another interviewee was engaged in the development of a new framework for the supervision and surveillance of algorithm usage within government processes, a new form of supervision that requires a new framework. Another interviewee was part of a governmental innovation hub and was occupied with the quality and ethics of the AI-driven data tools they are developing. Finally, two interviewees were doing dual doctoral research on AI related topics.

It is fair to say that the sample is quite diverse, each interviewee provided their own unique experiences and expertise. Any project or policy issue that concerns the usage of AI-driven tools at some point encounter issues related to transparency. What are the considerations that experts in the field have to make in order to be transparent? It is expected that their experiences and opinions can provide insights into the transparency related obstacles that they face in their day-to-day work in attempting to use AI tools for decision-making.

3.2 Data collection method

The primary method for gathering data will be by conducting qualitative interviews, more precisely semi-structured interviews. This method allowed for the interviewees to reflect on their own actions and experiences and provided the interviewer with leeway in how to respond. The flexibility of the research method allowed for in-depth questions that gave the interviewees the opportunity to explain their thoughts resulting in rich data. In preparation, an interview guide was created which can be found in the Appendix. In preparing the interview guide, several important elements were taken into consideration; not to make the questions to specific, using comprehensible language, not ask leading questions and remember to ask general information in order to contextualize the answers given.

Since the main goal of the research is to find reasoning behind certain behavior, using qualitative structured interviewing seemed the suitable method. The semi-structured format provides the opportunity to give insight into what the interviewees see as important when AI is being used in decision-making. To find insights into how the topic of AI and transparency affects their day-to-day work and study their experiences and problems potentially means departing from the interview guide and letting the interviewees be able to share their thoughts. The choice for semi-structured qualitative interviewing therefore seemed suitable for the collection of relevant data. Whereas going off topic during an interview is considered a nuisance in quantitative research, it is encouraged in qualitative research (Bryman, 2016). The initial tradeoff was that quantitative research would not provide me with the rich data that qualitative research could provide.

It is important to mention that during the collection of the data the topic of interest for this thesis received a lot of negative media attention. Multiple affairs involving the use of algorithms in government processes has resulted in some respondents being more cautious. This led to the request that some respondents wished to remain anonymous or not be cited at all. Therefore, in order to respect the privacy wishes of the respondents, the researcher chose to anonymize them.

3.3 Data analysis

The global COVID-19 pandemic forced the researcher to conduct the interviews using digital means since working from home was the norm and impeded face-to-face contact. Using the internal governmental video calling tool Webex, the researcher could conduct real-time conversations with the respondents. The videorecording function of Webex was disabled due to privacy concerns, therefore the researcher recorded the conversation with a mobile recorder application after the respondents gave permission. After the interview, the researcher transcribed the conversation by reviewing the audio file and sent it to the respondents to validate what was said and give them the chance to make adjustments or enhancements. After receiving the validated transcript the researcher had to anonymize the data in order to respect the privacy preferences of the respondents.

Even though the usage of digital means for the interviews created greater scheduling flexibility and also saved the researcher time and money that travelling would've cost. The faltering internet connection at times and the lack of human face-to-face interaction to prompt respondents in a more personal way were seen as methodological limitations.

The subsequent qualitative data analysis that underlies this research is thematic analysis. Thematic analysis can be seen as a generic approach to qualitative data analysis that is flexible for deducing central themes from verbal data (Bryman, 2016). As a generic approach it follows several steps which the researcher has gone through: 1) reading through the samples, 2) coding the materials, 3) elaborate many of the codes into themes related to the hypotheses, 4) evaluate the order of codes or themes, 5) examine possible links and connections between concepts and/or how they vary, 6) write up insights into a compelling narrative about the data (Bryman, 2016). This analysis laid the foundation for the research results.

In order to keep track of important topics in the data collected, qualitative coding is used for the qualitative analysis of this research. Coding is "a generative process that focuses on a close reading of data in order to capture, as best as possible, participant assumptions, insights, and complex motivations" (Mihas & Odum Institute, 2019, p. 2). Codes are developed either deductive or inductive. Deductive being a top down approach where, based on theoretical and conceptual frameworks, you start with codes and find excerpts that fit the codes in the data or transcripts (Mihas & Odum Institute, 2019). Inductive being a ground up approach where codes and new ideas are derived from the data itself (Mihas & Odum Institute, 2019). This qualitative data was coded in Word.

This research combines the deductive and inductive approach since this corresponds to the research design. Based on the types of impediment to transparency in documentary information by Pasquier & Villeneuve (2007), several codes were deducted from the theoretical framework. As shown in the interview guide in Appendix II, several interview questions were particularly aimed at retrieving data on these topics. However, after becoming more familiar with the data, several other codes that seemed relevant were generated from the data itself. Since the aim of this research is to explore the barriers to transparency in government usage of AI, and initially make an empirical contribution to the existing theory, the inductive coding approach was used at a later stage in order to generate new ideas and concepts. The deductive codes that stem from contemporary theory and the inductive codes that stem from the data are then reassessed and merged to form the basis for the theoretical contribution in the discussion section. The developed codes are further specified in Appendix III, the Codebook.

4. Results

The results presented stem from nine expert interviews that were conducted. As was previously mentioned in the methods section, the respondents were divided over the three traditional tasks of government. They will be referred to as either a Policy Officer (PO), a Policy Implementation Officer (PIMO) or a Policy Inspection Officer (PINO). The consideration to refer to the respondents in a differentiated manner is based on the result that the context of the government domain in which the respondent works is important for the content of the information provided. Before presenting the results, it should be mentioned that the respondents in no way whatsoever speak on behalf of their organization, their opinions and experiences are their own.

When asked about transparency related obstacles in using AI in decision-making, the respondents tended to give somewhat matching answers, the more general questions were answered in a divergent manner. The interviews revealed transparency related obstacles that were far more nuanced than the hypotheses proposed. There was also a general agreement that transparency is of the utmost importance for a healthy and functional democratic society. This importance was substantiated with moral reasons. For example, PO A (Full interview transcripts can be found in Appendix I) argued that:

“what we develop as civil servants is funded by tax money [...] so naturally it should be for the citizen [therefore] you have a moral duty to do it publicly”

Before outlining the results, it should be mentioned that the respondents’ interpretation of the concept of AI quite differed, some pointed to the usage of algorithms in general, others to ML. For example, when asked what PINO A meant by AI the response was:

“I have turned a bit in that. From my study background I would say all computer systems that attempt to replicate a part of human intelligence but I think that what we often talk about now when it comes to AI in government it is indeed a form of Machine Learning”

As such, AI and ML are used interchangeably but essentially mean the same. Another overall noteworthy discovery was the importance of the government sector in relation to the organizational maturity of the usage of AI applications. There is a significant difference between the policy domain in contrast to the inspection and implementation domain. The results are ordered according to the hypotheses and will now be presented.

4.1 Safety and privacy

On the topic of safety and privacy, the overall tendency was that both safety as well as privacy can be an obstacle in being transparent about an AI-assisted process. However, what part of the system would be made transparent was determinative for what kind of issues it could provoke. In addition, there was a distinction between the physical domain and the social domain. The physical domain entails the management of assets, objects or anything non-human, where mismanagement does not directly result in an impact on an individual’s wellbeing. The social domain entails the governing of humans, where mismanagement does lead to a direct impact on an individual’s wellbeing. For the physical domain security is generally more of an issue than privacy, nobody is concerned with the privacy of physical assets (PO C, Appendix I). For the social domain both security and privacy can be an issue.

When the algorithm, or mechanism which runs the system would be made transparent it could instigate security issues. For example, when PO A was asked what arguments she heard in her environment to not be too transparent about an AI system because it could be unsafe she replied:

“Yes, what I sometimes hear is that a hacker can use it to look for a vulnerability”

Another recurrent argument not to be transparent for the sake of safety was facilitating the possibility for malicious individuals to undermine or manipulate the system also called ‘gaming the system’ (PIMO A; PINO A, Appendix I).

However, in all cases, the respondents argued that non-transparency is never an option that can be substantiated by safety and or privacy concerns. There is always a certain degree of transparency possible especially when it concerns human lives. Respondents argued that as a government you have a different dependency position towards your citizens, a government is not a business and people don’t choose to have their information used by government (PINO A; PIMO A, Appendix I). As PINO A argues:

> *“you don’t have to necessarily put your entire model online in order to provide insight”*

A positive practical example concerning the usage of a crime prediction system by the police that is made transparent was mentioned by PINO B:

> *“a large part of how the system works can be found online. What type of data sources it uses, not the exact data […] what kind of output it generates and how is dealt with that output […] you can get a clear image as a citizen what happens in the system without the risk of criminals ‘gaming the system’”*

However, there are certain liabilities such as when the input data on which the algorithm runs were to be made transparent it could instigate privacy issues. In the big data era it is difficult to anonymize data, when the government discloses information and removes personal data it does not automatically make it anonymous (PO A). PIMO A emphasizes this by arguing:

> *“[…] I think that there are so many possibilities in the field of data analysis in development and so much data is made available at the moment that it is very difficult to foresee the consequences of current publication […] they are irreversible”*

A final noteworthy discovery was finding that the degree to which the policy, inspection and implementation domain were engaged in the development of AI application substantially differed. This obstacle is also the main reason why it was so important to differentiate between the respondents and their given information. Overall, within the policy domain the topic of AI is less lively because the application of the technology and the potential benefits are less tangible. On the contrary, within the inspection and implementation domain the presupposed goals are better quantifiable, for example, who has a higher risk to be fraudulent or how can we optimally use our scarce assets? PO E for example argued that:

> *"The research question is quite clear: Who would be more likely to engage in fraudulent behavior? […] I think it is in any case easier with an inspection than with a policy directory where the question is: What would you like to use AI for at all?"*

These results highlights that the differentiation between the statutory tasks of government as well as the policy domain is of importance.

Culture of fear and the media

The culture of fear and the media is an obstacle to being transparent that was mentioned several times by the respondents and essentially referred to the privacy concerns of the public officials themselves (PO A; PO D; PO E; PINO B, Appendix I). Ironically, the fact that the researcher had to anonymize the transcripts due to the negative media attention that the research subject was receiving at the time of examination is an evident example of non-transparency.

There is a mutual sentiment of fear amongst the interviewee's (PO A) colleagues that instigates non-transparency, especially when it concerns sectors that generally receive negative media attention such as immigration. For example PO A responded that:

> *"You never hear about positive examples, even though there are plenty. This also creates a kind of fear like 'if we are going to be very transparent about this, we will be in the newspapers again with big headlines'"*

Apart from the lack of opacity that this fear generates, there is also the consequence of a negative spiral that this behavior invokes for governmental AI development in general. PINO B emphasizes this by mentioning that:

> *"I think that a lot of negative reporting has created a lot of fears […] As a result, there is now a potential stop in certain developments around the application of algorithms […]therefore there is less capacity to develop things properly […] so you're kind of in a negative spiral"*

Apart from fear there is also a kind of shame that people don't know what they're talking about when it comes to AI or new technologies in general (PO E, Appendix I). This lack of expertise and cooperation will now be presented.

4.2 Expertise and organizational cooperation

The results related to lack of expertise on the topic of AI to provide transparency showed some noteworthy dynamics. Generally, there is a lack of expertise on the topic of AI, this is no surprise since the technology is only until recently receiving a lot of attention and the government is gradually adapting. However, it not that there is a particular shortage on expert programmers, it is a rather distinctive expertise that seems to be lacking mentioned by PO E, the expertise to be able to assess whether the algorithms and data that are used to support certain decision-making coincide with the domain in which the system will be deployed. To what extent the measured reality and context do sufficient justice to reality.

PO E argued that, in general, the people working in government have a background in alpha sciences such as history and linguistics. She argues that:

> *"Generally, you have a lot of alphas within the government. So a lot of people have a completely different background meaning they are not trained in recognizing and explaining formulas, numbers, graphs to something that happens in the real world, making it more difficult to interpret what an algorithm predicts or gives as output"*

On the topic of organizational cooperation the respondents were the most talkative. The first argument related to organizational cooperation is the inadequate information architecture on which the daily business operations run. In some cases it is just simply not possible to provide transparency due to the presence of 'legacy systems' that were not built with the consideration to provide transparency, it would require a great deal of time and money to make that transparent (PIMO A). This is also an obstacle for developing AI applications in general. The experimental environment in which an AI tool is developed does not coincide with the present information architecture where it should supposedly be enrolled (PINO A).

However, the most recurrent obstacle – mentioned by seven out of nine respondents – to not be able to provide transparency due to the lack of cooperation amongst the different organizational components on how to be transparent. Each governmental department seems to be occupied with creating their own framework on how to provide transparency in their effort to properly govern AI. All these frameworks should be harmonized in order to create a clear duty for transparency towards citizens (PIMO A). For example PIMO A responded:

> *[On transparency guidelines] "On the European level they're busy with a discussion about a Human Impact Assessment for algorithms […] But also from the Ministry, the*

General Audit Office, The National Audit Service, the Authority of Personal Data, The Telecom Agency, everyone is engaged with this topic"

The lack of cooperation is also reflected in the lack of agreement on what transparency should mean. For example, whereas the data scientists think in terms of 'technical transparency' (PO C), the bureaucrats express it in terms of explainability (PINO B). This engagement in technical transparency is also reflected in the lack of a common policy how to share AI related information. For example, PIMO A argued:

"[…] so that is an important obstacle, do we technically have the platforms to publish all this information at all? […] The data on data.overheid […] is purely a website for publishing with no substantive check by the web administrator […] you have few standards in that area"

This lack of cooperation also trickles down to the organizational dynamics. In general, the clear division between policy, inspection and implementation also presents transparency issues. Since the policy domain creates the rules, the inspection checks its compliance and the implementation executes the policy, the lack of AI awareness in the policy domain (as mentioned in the prior results) impedes effective problem solving. In addition, even though the chain of policy is thought out well, in practice there is a certain pillarization that ensures that the problems are not known to each other. For example PIMO A responded:

"The person who receives the complaint of a citizen is, in most cases, a policy officer. This person knows nothing of the system that produced the error. So, the system administrator can fully understand the problem, but the recipient of the complaint doesn't transfer it to the system administrator. This way the problem won't be resolved, and it's a fundamental problem that has to change. It is truly because of the separation between implementation and policy"

Also, there was a general agreement amongst the respondents that anybody who is can be deemed a critical actor in the development of AI systems e.g. a politician, a manager, or bureaucrat should at least have a general understanding of how the system works. However, it is not expected that they should be able to explain it from front to end. The results show that there are crucial mismatches of knowledge of each other's affairs on AI between political executives and the bureaucrats, managerial executives and data scientists and bureaucrats and data scientists. These mismatches each present obstacles to transparency for their own reasons.

Firstly, the mismatch of knowledge between political executives and the bureaucracy. The parliament has the duty to inspect the government and its bureaucrats. However, when politicians ask parliamentary questions about AI and transparency, due to a lack of knowledge in each other's affairs on the behalf of the politicians (both in government as well as in parliament), nuance is missing in the debate concerning this topic (PIMO A). PIMO A accentuates this issue by arguing:

> *"There are a lot of questions being asked by parliament about transparency and AI […] but the knowledge level from the First and Second Chamber about IT is just very low in general […] and that is reflected in the questions. It results in questions such as 'to what extent does this system look like Judge Dredd?' to which the Minister answers 'this system does not look like Judge Dredd' period. That way you don't have the necessary discussion!"*

This lack knowledge of each other's affairs between political executives and the bureaucracy is anchored in the dichotomy of politics and its administration. The bureaucrat's expertise is slowly but surely improving due to the hiring of AI specialists and organization wide schooling initiatives. But when asked why the politician's expertise can't be improved the same way as the bureaucrat's, a respondent replied that is not possible due to the strict separation of politics and bureaucracy and the risk of conflict of interest (PIMO A). This lack of awareness of the affairs of the bureaucracy also leads to politicians making promises to solve IT-related (transparency) issues which are often not feasible, resulting in disappointment (PO E).

Secondly, the mismatch of knowledge between bureaucrats who will use the system in their work and data scientists. The discrepancy in knowledge of affairs between the end users of the system and the people who programmed the system can be an obstacle to being transparent. This is due to the users incapability to properly explain their AI-assisted decision other than arguing the system advised them to. In addition, a respondent mentioned that poor interaction between data scientists and bureaucrats can result in opaque systems that are ill-suited for the context they're deployed in (PO B).

Thirdly, the mismatch of knowledge between managerial executives and data scientists. This gap in knowledge is inherent to the hierarchical nature of government. The more you ascend the hierarchical ladder, the bigger this gap in AI related knowledge becomes, this is however no surprise since executives are not hired to know the exact details of their subordinates (PINO A).

PO E argued that the problem of the knowledge gap between contemporary managerial executives and newly hired experts is not easily solved due to insufficient hiring during the last financial crisis. There seems to be a certain mismatch between the people in management who try to keep up with the technological trends and the newly hired people who seem to have figured out the problem. For example PO E argued:

> *" [...] the problem is that you bring in a lot of new people who have a lot more knowledge but a lot less experience and are put in a position with lesser influence"*

It seems that the government is hiring young experts with AI related knowledge but no administrative experience who are unable to generate meaningful change within the organization due to their minor rank. These newly hired experts also seem to be having a hard time realizing change due to a certain lack of awareness amongst older bureaucrats with lesser AI related knowledge but greater administrative experience (PO E).

In any case, it is essential that the managerial top understands what is happening with these systems before they take the responsibility to put such a system into practice, if not it can result in accountability issues (PINO B). PINO B highlights this by arguing:

> *"[...] is working on ethical guidelines for data scientists which I find very strange [...] everyone within your organization should act ethical, not just data scientists [...] I think it is a symptom of the lack of the manager's expertise who carries the responsibility, this way responsibility is delegated downwards when it should be at the top. A manager should be able to see 'yes we built this system [...] that should meet certain ethical standards [...] if not we cannot use it'"*

4.3 Inefficient disclosing

On the topic of inefficient disclosing as an obstacle to being transparent, all respondents agreed that simply disclosing an AI algorithm by engaging in open source nor disclosing the data that feeds a system by engaging in open data is sufficient when attempting to be transparent as a government. PINO B for example responded:

> *"You want to be transparent in a way that the citizen can understand what's happening inside government. And that doesn't always mean that you – for example with an algorithm – disclose the source code. Because there are maybe a few thousand individuals in the Netherlands who can read that and enjoy reading it, but for everyone else it is useless"*

The contemporary way in which information about AI systems is disclosed is not achieving its goal of making sure that citizens understand what is going on in government. Simply disclosing this raw information is what respondents called 'technical transparency', but the context (usage, goal or result) of the data or algorithm is of greater importance for citizens (PO A; PO D; PIMO A; PINO B). Respondents claimed that in general, government should invest in proactive transparency rather than reactive transparency where disclosure only follows after problems arise (PO D; PINO B). For example, PO D responded:

> "*How can we make sure that it [transparency] truly contributes and we don't get stressed afterwards when we have to justify ourselves? How can you be more transparent on the front? How can we ensure that we're a reliable and diligent government?*"

In addition, reactive transparency doesn't promote governmental trustworthiness. If documents and information are requested afterwards, in order to protect the civil servants and their personal opinions, their information is painted black within the documents, this can result in entire pages painted black which is not conducive to governmental image (PO D).

An important nuance that was mentioned by PINO A was the difference between information and source code. As opposed to the painted documents PINO A argued that:

> *"[…] what you always see are those painted document, but if you would do that with code… I see a great operational problem how we would need to communicate that to the citizen in a way in which it is of use to them."*

Another noteworthy obstacle was the premise that the public sector is similar to the private sector in how efforts are made to realize AI-assisted innovation. As a government you have a different position towards your citizens than a company would have. A business, as an organization created to earn money and keep shareholders pleased, has to take less account of citizen interests. However, the methods used in achieving innovation are similar to the ones used in the private sector and that presents obstacles. Consequently, a translation is made from a model that exists to achieve financial gain to the efficiency of governing citizens (PO E). As a result, citizens can be taken less taken into account in the design process of these applications due to the failure to recognize the complexity of the administrative context. This presents certain issues that are reflected in the anecdotal example by PINO B:

> *He [an inspector] said: "well, it is put together nicely, but something really needs to be done about the sensitivity and robustness of your model". However, the inspectors*

didn't say "you should stop using that algorithm". They just said this is really something you should do something about. The analysts [data scientists] understood it immediately and they also wanted to do something about it, but the managers didn't understand it at all. They had no clue what sensitivity and robustness are. The managers eventually thought, well the traffic light is green so we just continue. And so the issues were not prioritized to be treated by the analysts even though they knew it was important.

This result shows that a focus on the result of the project causes them to erroneously prioritize the outcome of an AI system rather than what calculations led to the outcome or how to make it transparent (PINO B). For example, they cannot take the responsibility to prioritize certain miscalculations because they can't determine its earnestness due to unfamiliarity with the subject (PINO B). As long as the development can continue and the deadlines can be met, certain flaws in an algorithmic model aren't prioritized.

5. Discussion

This research aimed to contribute to the lacuna of identifying obstacles to transparency in AI applications in the public sector i.e. analyzing governmental motivation towards (non-) transparency and how the complex nature of AI relates to this. The results of the qualitative interviews has shown that the obstacles to being transparent about AI-assisted decision-making are much more nuanced than the obstacles proposed in the literature. Whereas the disclosure of documentary information has its known obstacles described by Pasquier & Villeneuve (2007), the challenges associated with the disclosure of AI related information such as algorithms and data on which the system operates are in most cases fundamentally different.

The most unexpected findings were that what part of the system would be made transparent is determinative for the type of issues it would instigate as well as that the policy domain is also of importance in determining how to be transparent. The inner workings of the system presents safety issues, the (input) data can present privacy issues. However both are never reasons to be non-transparent, there is always some degree of transparency possible. Negative media attention can demotivate development resulting in less hiring of experts who would need to solve these issues which then results in less capacity for proper development making it a matter of time before another algorithmic related scandal becomes published by the media. Lack of cooperation on how to be transparent seems to impede a clear duty of transparency towards citizens and 'legacy systems' are often not technically capable to provide transparency. The discrepancy of knowledge of each other's' affairs amongst executives, data scientists and bureaucrats highlighted six obstacles to transparency and even though government is engaging in proactive transparency initiatives, it misses the mark thinking 'technical transparency' is sufficient.

Thus, the obstacles to being transparent that stem from the usage of AI are somewhat different than decision-making that relies on documentary information and doesn't use AI. The uptake for this argument and any argument that follows lies in the fundamental difference of the information that is to be made transparent. Data, as argued by Rowley (2006), is nothing more than just symbols lacking meaning or value and can be considered useless if you don't know what is means. In addition to data, algorithms are just that to any laymen without the proper knowledge of programming or AI. It seems that AI and the particular usage of a kind of 'information' that can only be understood as 'information' by a handful of experts presents a whole set of other obstacles to be transparent as a government than documentary information does when used in decision-making. This puts a strain on governments, as a novel way to use information also means that it requires a novel way to remain transparent and diligent as a government. Governments should not only explain the information that constituted an outcome but also explain the system that supported and generated that decision. Being transparent as a government by disclosing reports, notes, minutes of meetings, or e-mails won't require the effort as being transparent about data and algorithms that are fundamental to AI. The importance of a well-considered commitment to being transparent as a government when using artificial intelligence should not be underestimated.

The results will now be presented against findings from other studies, presenting commonalities and differences for each hypothesis based on the type of information. The findings will be evaluated based on the known obstacles for documentary information and the discovered obstacles for AI and its related digital information. This will ultimately lead to implications for further research and a conclusion.

5.1 Governments are limited by privacy and safety issues to not be transparent in AI assisted decision-making

Multi-sidedness of digital information

The results show that governments are limited by privacy and safety issues to not be transparent about AI assisted decision-making. The line of argument for the security issues involved with the disclosure of documentary information coincides with that of the disclosure of the inner workings of an AI system, its algorithms. Pasquier & Villeneuve (2007) argue that security challenges are mostly related to national security e.g. terrorism, whereas the disclosure of the algorithms of an AI system makes it prone to attack by a malicious hacker (Shokri et al., 2019). It is debatable whether a hacker can be considered a terrorist and vice versa, but whether it is the refrainment from disclosure of an AI algorithm or a highly classified document, (national) security can be a reason to do so. However, whereas Pasquire & Villeneuve (2007) argue that disclosure of documentary information also instigates privacy issues, this is only primarily true for the input data and not so much a concern for the disclosure of algorithms. The finding that what part of the system would be made transparent is determinative for the type of issues it would instigate is a first example that shows that the one-sidedness of information that is so self-evident in traditional decision-making does not apply to AI. It is therefore relevant what part of the system is to be made transparent.

The importance of the policy domain

Additionally, another result that advocates the multi-sidedness of digital information used in AI is that of the importance of the domain in which AI is used as being determinative for whether privacy or security is an issue. The social domain instigating both security and privacy issues, the physical domain mostly security issues since the privacy of physical assets aren't prioritized over that of human beings. Therefore, privacy and security should not be treated as a similar issue that requires the same treatment when disclosing information about a decision-making process.

The difference between the social and physical domain is a clear example of the fact that transparency isn't a one-size-fits-all solution to increase governmental legitimacy. In her article, de Fine Licht (2014) argues that the transparency has different effects in different policy areas. This difference is based on the theory of taboo trade-offs that argues that disclosing decisions that are related to human life and well-being versus monetary considerations are ethically difficult to process for people (de Fine Licht, 2014). Therefore, disclosing information about political decisions that involve a trade-off between human life or well-being and money (e.g. setting health care priorities) triggers negative feelings towards decision makers and reduces political legitimacy (de Fine Licht, 2014). On the contrary, policy areas that typically handle routine trade-offs (e.g. asset management and money), are less likely to cause negative reactions (de Fine Licht, 2014). The results by de Fine Licht (2014) show that "transparency effects can be conditioned by the type of policy area and, more specifically by the type of trade-offs typically carried out within these areas" (p. 367).

The results have shown that it isn't desirable to be completely non-transparent about decision-making as this will only generate distrust in government and can result in the abolishment of a system as with SyRI. However, there is always some degree of transparency possible, the question is what degree of transparency is desirable in what policy area. Based on the literature and the results, what type of transparency is required in order to achieve public acceptance of political decision-making differs. When it concerns the social domain, where taboo trade-offs are more likely to occur, transparency about solely the justification for the decision is the best choice. Limited transparency is beneficial in the policy areas where taboo trade-offs occur (de Fine Licht, 2014). When it concerns the physical domain, where routine trade-offs are likely to occur, there is the choice for either solely the justification or a thick description of the decision-making process. There is no tendency that transparency of any degree has a negative effect within the policy areas involving routine trade-offs (de Fine Licht, 2014). Even though Pasquier & Villeneuve (2007) argue that security and privacy are reasons for government to be non-transparent, both the results from this research as the results from de Fine Licht (2014) contradict this, both are never a reason to be completely non-transparent about a governmental process or outcome. To sum up, there is always a certain level of transparency possible i.e. telling from where you obtain the information but not precisely what information you use that can provide some benefits of transparency while avoiding its disadvantages. In addition, it is important to bear attention to the policy domain where transparency is to be provided.

Reputation and the media

Interestingly, the privacy issues that Pasquire & Villeneuve (2007) mention apply to the privacy of civil servants, whereas the results show that the disclosure of governmental data emphasizes the privacy of citizens. Interviewees were concerned that in the big data era it is difficult to anonymize data and with good reason. Löfgren & Webster (2014) argue that repurposing and reuse of data sets can result in the (re)identification and profiling of individual citizens at a later stage. The constant developments in data analytics makes it hard to foresee the consequences of contemporary disclosure of data since publication is irreversible. It takes a skilled data scientist only but a few data sources of personal information in order to re-identify individuals in Big Data sets (Jain et al., 2016). Not to mention how data brokers within the private sector combine different datasets in order to generate profiles of individuals which are sold to the highest bidder in order to be used in targeted advertising (Pasquale, 2015).

However, it is not that the privacy of civil servants is insignificant. Rather, the results show that there is a genuine fear for the media and negative publication of IT projects in general. It is not as much an issue of privacy since the civil servants are generally protected by legislation but rather an issue of reputation. This result was particularly surprising since so many interviewees highlighted it as an concern, the fact that the results had to be anonymized strongly projects this fear for the media. However, apart from the finding that the media plays a significant role in the recalcitrance of governmental transparency, negative media attention has even more dire consequences. The Dutch government seems to have become entangled in a negative spiral when it comes to developing AI applications, and the media is to blame for this. Negative media attention demotivates AI development resulting in less hiring of experts who would need to solve these issues which then results in less expertise and capacity for proper development making it a matter of time before another algorithmic related scandal becomes published by the media. Not only is this detrimental for governmental reputation, this reputation can act as a deterrent for the experts that governments so desperately need. Therefore, the role that the media plays in the digitalization effort that the public sector experiences shouldn't be overlooked. This need for expertise is a nice mnemonic for the following hypothesis on expertise and cooperation.

5.2 Governments are impeded in being transparent in AI-assisted decision-making due to lack of expertise and cooperation

Lack of expertise

Both the literature as well as the results show that there is a general lack of AI experts that needs to be addressed. In order to better grasp what skills are needed, Chinn et al. (2020) determined a set of skills across three dimensions in assessing the European skill gap. A distinction is made between technological skills, digital citizenship skills and classical skills, the latter not being of interest. The most relevant technological and digital citizenship skills for this research encompass: being able to moderate between technology experts and nonexperts who are involved in a project; and collaborate effectively during projects irrespective of different disciplines and cultures (Chinn et al., 2020). It is the former skill that coincides with the results as being the most important skill missing at the moment called 'tech translation'. The results show that the primary workforce within government and public administration in general are alumni from alpha sciences such as history and linguistics, therefore they lack the proper background to be able to recognize and explain digital information such as data and algorithms. One interviewee argued that there is not just a shortage in expert programmers or policy domain experts but another certain kind of expertise. The expertise to be able to assess whether the algorithms and data that are used to support certain decision-making coincides with the domain in which the system will be deployed. Experts who can assess to what extent the measured reality and policy context do sufficient justice to practice, individuals who are able to critically moderate between experts and nonexperts during AI development seem to be lacking.

This line of argument coincides with the philosophical theory of ontology and anthropomorphism which Hawley (2019) highlights in his article *Challenges for an Ontology of Artificial Intelligence*. Ontology is a branch of the philosophical school of thought that is primarily concerned with the concepts of existence, being, becoming and reality, arguing that "things act in accordance to they are *i.e.*, their ontology" (Hawley, 2019, p. 2). A concept that is associated with the ontological debate of understanding new phenomena is anthropomorphism, the tendency to ascribe human characteristics traits and attributes to non-human things (Hawley, 2019). Anthropomorphism is a hotly debated topic within the field of AI since AI entails the attribution of something human like 'intelligence' to something non-human as computers and algorithms. It is a human trait that in order to make sense of our surroundings we assign meaning to them, it is "the "hammer" we try to apply to many "nails"" (Hawley, 2019, p. 10). However, anthropomorphism also has a downside referred to as 'dehumanization', or 'objectification' an ontological error that denies the personhood and human value of 'beings' and replaces them thus regarding humans as mere things (Hawley, 2019). This is a particular problem that potentially arises when you quantify reality and try to model human behavior for the sake of recommendation systems (Hawley, 2019). Drawing on Hawley (2019), the argument that should be made here is that when you objectify and model human behavior into data and algorithms it is an imminent risk that you dehumanize people and no longer take the human dimension into account. There is particularly need for guardians who assess whether the correct 'hammer' is used and whether all 'nails' are identified through intensive moderation between the experts and the nonexperts. How this skill should be specified and whether it yields results requires further research.

The second skill entails 'collaboration' to collaborate effectively during projects irrespective of different disciplines and cultures (Chinn et al., 2020). And the results show that it is not a phenomenon that cannot be ascribed to a single person but to an entire organization.

Lack of cooperation and awareness

A distinction can be made between lack of cooperation on the operational level, on the strategic level as well as on the organizational level. Firstly, as for the operational cooperation, drawing back on the literature by Klievink et al. (2017), the degree to which information sharing is possible amongst public organizations determines the degree to which it can properly utilize data and thus AI. However, interviewees responded that there are certain concerns about the inadequate information architecture on which the daily operations run. There are still legacy systems present that are not built with the consideration to provide transparency let alone communicate with other systems.

Secondly, on the strategic level, the results show that each governmental department seems to be occupied with creating their own framework in how to provide transparency in their effort to generate ethical AI clearly lacking strategic cooperation. Dawes et al. (2009) speak of public sector knowledge networks or PSKNs and its challenges, PSKNs being information systems that serve as communication tools and data resources in order to address public needs that no single organization can handle. Dawes et al. (2009) argues that an important aspect of developing a knowledge network is social interaction over time in order to make sure there is a shared understanding of the standard definitions and certain concepts. This is particularly important for contested concepts such as transparency, however, the results show that there is no common understanding about what transparency should look like, data scientists referring to technical transparency whereas bureaucrats are expecting explainability. This implies that there is a lack of cooperation on the strategic level. This ultimately generates negative consequences for transparency since the lack of cooperation on the strategic levels also results in lack of a unified approach on how to be transparent which will be assessed in the final hypothesis.

Thirdly, on the organizational level, the results show a discrepancy in cooperation, not only on the general topic of AI but also related to the degree that crucial actors involved in the creation of AI systems aren't aware of each other's affairs. This may be the most surprising result since this discrepancy in cooperation generates multiple issues. This lack of cooperation can be ascribed on two levels, the organizational level and the individual level.

Firstly, on the organizational level, the results show that there is also a certain pillarization amongst the different tasks of government. Even though the chain of policy is thought out well, the separation of governmental tasks ensures that certain problems are not know to each other. Dawes et al. (2009) concur this as a barrier, arguing that the sharp lines of authority, different rolls and functions as well as competing missions all impede effective knowledge sharing. Löfgren & Webster (2014) speak of 'silos' of policies and services that generates problems for digital governance in general. Both authors pointing to the lack of cooperation that is so essential in digitizing governments in general. Frankly, unless these apparent differences become known to each other they cannot be reconciled (Dawes et al., 2009), it is high time that government agencies start communicating better with each other if they wish to properly innovate government. The way that the pillarization of government agency's negatively affect cooperation and transparency is definitely a topic worth exploring. However, the most contributing finding is the lack of cooperation, and particularly awareness on the individual level that will now be outlined.

Dichotomy of politicians and administrators

A lack of awareness of affairs between the political executives and their administration leads to superficial debate about the technology on how to realize transparency as well as false promises to enhance transparency that are often not feasible. This lack of awareness of each other's affairs coincides with the theory of the 'politics-administration dichotomy' that argues that public administration is and should remain distinct from politics (Demir & Nyhan, 2008). The reason for this separation of spheres is that "politics is a process by which disagreements and conflicts are worked out [and] ends with laws and policies through legislation. The purpose of politics is to provide political guidance to public administration" (Demir & Nyhan, 2008, p. 82). On the contrary, public administration has the task to translate "value choices into concrete results […] public administrators apply special knowledge and skills called expertise. The purpose of public administrators is to provide neutral competence to the policy process" (Demir & Nyhan, 2008, p. 82). This separation of political guidance and neutral competence is mechanism to prevent political corruption (Demir & Nyhan, 2008) as well as prevent conflict of interest as one interviewee argued.

Nevertheless, even though the separation of politics and policy is well intended, it is questionable if politicians are capable to provide 'political guidance' to the administration on the subject of AI and digitalization in general. Demir & Nyhan (2008) argue that the two ways in which politicians exercise political guidance is through policy leadership and legislative oversight. Policy leadership, in the sense of policy management, entails the ability to exercise strategic tasks related to setting goals and priorities and provide guidance for the development implementation of policy strategies (Demir & Nyhan, 2008). However, the results clearly show that the lack of knowledge on behalf of the politicians results in superficial debate, about the technology as well as about how to make something transparent. This is particularly troublesome since the results also show that the way in which information is disclosed is essential in order for citizens to understand it. It can therefore be argued that it is difficult to provide adequate policy leadership on topics they don't understand sufficiently.

The other competence related to proper political guidance is 'legislative oversight', "[ensuring] that policy implementation proceeds in conformity with legislative intentions and instructions" (Demir & Nyhan, 2008, p. 85). However, the results show that the oversight that politicians have seems below par due to the complexity of the subject if AI and digitization in general. This is reflected through the fact that the feasibility of the solutions promised aren't realistic, for example the organizational overhaul that a completely new system requires. This further compounds the problem since it only leads to disappointment and doesn't lead to actual solutions to become more transparent and properly use AI or any other technology.

Therefore, the complex topic of digitization and AI provides reasons to rethink the relationship between politics and the public administration. In order for politicians to provide political guidance and set the task for public administration it seems essential that both parties understand the topics sufficiently. What measures are possible to equalize the knowledge of politicians to that of administrators whilst keeping the spheres separated seems an interesting topic. However, it is questionable whether this is feasible at all and whether politicians are eager to undergo such endeavors. Perhaps a better choice is to institutionalize the responsibility of digitization by assigning a Minister of Digital Affairs.

Data scientists versus domain experts

The lack of awareness of affairs between data scientists and bureaucrats consequently leads to bureaucrats being unable to explain an outcome and results in ill-suited systems due to poor interaction. This lack of cooperation between the bureaucrats who will eventually use the system in their business operations and the data scientists who built the AI system is not uncommon in the literature. In his article *Data Scientists Aren't Domain Experts,* Viaene (2013) expertly puts his finger on the sore spot highlighting the presence of a 'wall' between data scientists and 'domain' experts. Domain experts are the individuals who have worked in their domain for years and years and who've accumulated their knowledge through experience (Viaene, 2013). In the contemporary setting, there is a lack of cooperation and dialogue between the domain experts and the data scientists, the data scientists being expected to produce new insights based on a dataset, fleeting ideas and limited business knowledge (Viaene, 2013). This presents obvious issues since "without any further clarification of a particular idea and its business context, data scientists are likely to misinterpret (and thus misrepresent) the idea, leading to the data science exercise's failure" (Viaene, 2013, p. 13). In addition, it is primarily during conversation with the domain experts that data scientists can uncover underlying assumptions of business ideas and bias in the data (Viaene, 2013). Thus, the discrepancy of knowledge between the data scientists and the bureaucrats will most likely lead to ill-suited systems since data scientists are decoupled from the organizational processes and therefore lack contextual knowledge and lived experiences.

It can be argued that much can be gained by including bureaucrats and perhaps citizens who are subject to the system in the design process. An example being the study by Frey et al. (2020) where the involvement of marginalized communities that are subject to algorithmic analysis as 'domain experts' leads to a better understanding of context and culture. This knowledge is then incorporated in the development process of data scientists leading to the production of algorithms with a positive social impact that otherwise would include erroneous assumptions and implications about these marginalized communities (Frey et al., 2020). Therefore, more intensive collaboration between domain experts and data scientists is expected to lead to fruitful results and can even lead to domain experts becoming data science advocates for the organization, adjusting expectations and showing value in use (Viaene, 2013).

Finally, since the bureaucrats are the end users of the system that will be incorporated in their work processes, but lack the knowledge to explain the analysis that advised them to do so, the discrepancy in knowledge between the bureaucrats and the data scientists leads to inability of bureaucrats to explain, let alone be transparent, about their decision. It is a difficult problem to determine how much knowledge of the system the end user should have, whose responsibility it should be to provide this knowledge and if it is even possible to train this aged audience the determinants of an AI system. However, it is expected that the degree of knowledge that bureaucrats have of the system inherently relates to the degree of transparency that can be provided about any decision-making that occurs.

Hierarchy and idea generation

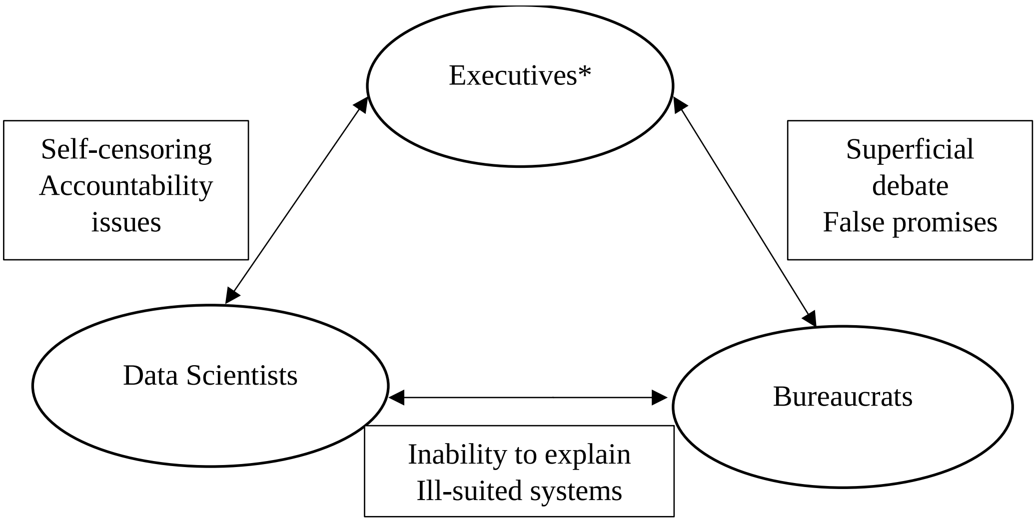
A lack of awareness of affairs between administrative executives and their data scientists impedes effective solutions for transparency issues due to the hierarchical culture and can lead to accountability issues. The lack of AI expertise amongst administrative executives as opposed to the data scientists shows how the hierarchical culture of government impedes transparency. This impediment to transparency is due to the limitations that hierarchy instigates in idea selection (Keum & See, 2015). The results show that newly hired experts experience resistance and are unable to generate meaningful change within the organization due to their minor rank. This is no surprise since “greater hierarchy of authority reduces the generation of expression of ideas that might be deemed too risky or may not conform the preferences of one’s supervisor or other higher-ups” (Keum & See, 2015, p. 657). Therefore, in order to provide meaningful innovation, C-level executives should create an environment where collisions of ideas can happen, encouraging dialogue (Viaene, 2013).

In addition, the lack of understanding of AI related topics amongst executives and wretched cooperation between executives and data scientists leads to accountability issues. The results show that currently there are ethical guidelines for data scientists in development, virtually burdening the data scientists when they develop an ethically transgressing system. An interviewee argued that this is a direct result from the lack of expertise that executives have on this topic as they are delegating accountability to the data scientists, not wanting to take responsibility for something they cannot fully understand. However, it is very questionable whether data scientists should be held accountable for the societal impact their systems produce, especially if their liability is the results of a lack of expertise of their superior. However, as another interviewee responded, executives are not hired to know the exact details of their subordinates’ work. It is a hotly debated topic, specialists versus generalists and the breadth and depth of knowledge that executive managers should master (Ferreira & Sah, 2012). Nonetheless, in the context of AI, accountability and expertise thus presents itself as a topic worth exploring.

The finding that there seems to be a general lack of awareness in each other’s affairs between important actors involved in the development of AI causes potential issues for transparency. Figure 7 depicts these obstacles to transparency.

Figure 7

The obstacles to transparency due to mismatch in knowledge of other actors' affairs



**The relationship between executives and bureaucrats entails political executives (politicians), the relationship between executives and data scientists entails managerial executives (C-level managers).*

5.3 Governments are impeding transparency by ineffectively disclosing information about AI-assisted decision-making.

Lastly, the paradoxical hypothesis that one form of transparency leads to the impediment of actual transparency. In general, the results contradict the maximized transparent situation of Pasquier & Villeneuve (2007) where the organization discloses all information available and the public need not request it by showing that even after disclosure they most likely still need to request a certain explanation because it is not understandable for the public. Governments are increasingly engaging in open government initiatives, expecting that it will increase governmental legitimacy by generating an informed society that can assess governmental legitimacy and take action accordingly (Albu & Flyberbom, 2016). In their research, Albu & Flyverbom (2016) attempt to conceptualize transparency based on different streams of research on the topic. The results are two paradigmatic positions on the topic of transparency; the verifiability approach and the performativity approach (Albu & Flyverbom, 2016). The former, verifiability approach, focuses on the way in which information is disclosed in order to 'verify' a certain state of affairs, the assumption being that "by making more information available we can regulate behavior and improve organizational and societal affairs through processes of verification" (Albu & Flyverbom, 2016, p. 281). This approach underlies the assumptions of the open government doctrine, that more information disclosure is better, verifying that it is conducting good governance. On the contrary, the performativity approach contradicts the assumption that more information is always better, emphasizing "the complexity of communication and interpretation processes and [focusing] on the complications and paradoxes generated by transparency projects" (Albu & Flyverbom, 2016, p. 281).

As argued earlier, data is nothing more than just symbols lacking meaning or value and can be considered useless if you don't know what is means (Rowley, 2006). In addition to data, algorithms are just that to any laymen without the proper knowledge of programming or AI. As such, the results show that governments fundamentally misses the mark when they engage in verifiability driven transparency, the assumption that more information is always better e.g. 'open source' and 'open data' initiatives. Consequently, as argued by Albu & Flyverbom (2016) "the focus on verifiability takes it for granted that those involved in the mediation and reception of disclosures are always willing and able to process, digest, and interpret the information" (p. 286). It appears that the target audience for which the endeavors of transparency are done are completely forgotten in the process. Governments have to acknowledge the performativity approach to transparency, paying more attention to the complexity of communication and interpretation processes (Albu & Flyberbom) that are inherent to transparency projects.

In their article, *Transparent to whom? No algorithmic accountability without a critical audience*, Kemper & Kolkman (2019) further highlight the erroneous assumption of the verifiability approach to transparency that would supposedly increase governmental legitimacy by generating an informed society. Kemper & Kolkman (2019) argue that transparency might induce a less critical attitude towards a product but doesn't necessarily lead to a better product, transparency doesn't ensure a critical evaluation of an algorithmic model or system. An important aspect of the critical evaluation of government is an apparent critical audience. And since AI models are so complex that even experts can't always explain the exact operations of these models, it is likely that "measures of transparency are at risk of remaining empty signifiers if no critical and unbiased engagement emerges from their installment" (Kemper & Kolkman, 2019, p. 2092). Based on these arguments it is of the utmost importance that the value of transparency can't be unseen from its practicalities and eventual engagement, an emphasis on context.

Private sector business models

The results show that the primary reason to always be transparent as a government as opposed to the private sector is the different dependency position that a government has towards citizens than a business. Anything that is develop by civil servants is essentially funded by tax money and should there be disclosed to citizens, in addition, citizens cannot choose whether they want their information to be used by government or not. However, governments are increasingly copying innovation models from the private sector hence increasing the likelihood that performance indices and efficiency will trump public values. This problem was raised early on by Clark & Newman (1997) who spoke of the 'managerial state' where a managerial approach to societal issues would instigate perverse effects that contribute to a process of alienation of government from society. Therefore this finding potentially accounts to this deficit in consideration of people in general. This is amplified by Löfgren & Webster (2014) who argue that because the private sector serves as an inspiration and benchmark for the public sector, "the rationale and functioning of the public sector, including the safeguarding of core public values is usually ignored in exchange for prospects of enhanced efficiency and customer-satisfaction" (p. 3). Further, the results reflect that a managerial executive is assessed on the basis of the results that he or she achieves, primarily fixated on the outcome. An interviewee concurred this by an anecdotal example where a flaw of an algorithmic system was pointed out to a manager, it was rather a tip to consider and had no consequences for the system's development, the tip was not processed because the manager was mainly concerned with the fact that development could continue.

Whereas governments and public administrations were first primarily concerned with inputs and processes, this shifted to a focus on results and outputs due to a major reform that took place in the 1980s and is still dominant today, New Public Management or NPM (Bekkers et al., 2011). NPM has put "measurement and quantification, especially through the development of performance indicators and benchmarking systems" (Bekkers et al., 2011, p. 10) on a pedestal contrary to political and substantial values. This generates issues since innovation within the public sector substantially differs from the private sector. Innovation in the public sector is primarily concerned with achieving legitimacy whilst taking into account the institutional context in which these innovations emerge (Bekkers et al., 2011). This goes beyond the goals of the private sector that are primarily concerned with exploiting new markets and inventing new products and services for consumers (Bekkers et al., 2011). Even though one of the primary reasons for government to engage in the adoption of private sector business models is to better interact with private sector partners and mitigate barriers to (technological) innovation (Micheli et al., 2012), the previous arguments highlight that it can alienate them from society. Therefore, when engaging in AI development, a new balancing act between the needs and challenges of actual citizens and economic values has to occur. An emphasis on engagement and context perhaps requires a new culture. A culture where multi-disciplinary development teams receive a bigger role in the process of AI development and transparency projects, practicing ‘human-centered design’ that puts humans – or in this case citizens – at the epicenter of development. Especially in the public sector, where the government exists to provide services to its people.

In sum, this research has provided new insights into the way in which governments react to new technology that changes how daily work is performed and cases are handled by civil servants, AI. While previous research has focused on ways governments react to transparency demands for documented information, these results demonstrated that the transparency demand for digital information presents broader obstacles than the theory proposed. For the documentary transparency that concerns the disclosure of documents, the disclosed information can already be understood. However, for digital information this is not the case since data cannot be considered information. Without context, the disclosure of digital information related to AI can hardly be understood by someone who’s unfamiliar with the technology. Therefore, an update to the existing conceptual framework seems appropriate where a division is made between analog documentary information and digital information i.e. data and algorithms. The theoretical framework depicted in Table 5 by Pasquier & Villeneuve (2007) will therefore receive a contribution, the enhanced framework is shown in Table 7.

Table 7

Types of impediment to governmental transparency in documentary and digital information.

	Documentary information			Digital information		
	Obstructed transparency	Strained transparency	Maximized transparency	Multifaceted transparency	Scattered transparency	Alienated transparency
Description	Obstructions to transparency through using provisions of the law	Inability to cope with transparency due to an absence of resources or misunder-standing of information	Behaviour intended to forestall possible demands by making all the information available	Limitations based on privacy and security issues that are instigated by what part of the system is disclosed, differing policy areas, and fear for the media.	Limitations based on lack of ontological expertise, cooperation and presence of governmental pillarization.	Limitations based on inefficient disclosure due to an uncritical engagement by citizens and overreliance on private sector innovation models.
Justification	'It wouldn't be responsible'	'We don't have the resources'	'It's simpler and less costly'	'Due to technological advancements, some consequences are unforeseen, we should think about our reputation'	'The model isn't sufficiently suited for this context, we have our own task on which we should focus'	'If we disclose these data and algorithms we'll increase our legitimacy, it doesn't matter how we get there as long as we get there'

5.4 Limitations of the research

Obviously, this research method has its limitations in confirming hypotheses based on a small number of interviews. Firstly, the research is difficult to replicate. Since the method is quite unstructured and relies on ingenuity, replication is nearly impossible. In addition, the 'instrument' of collection and measurement is the researcher, therefore the focus of the observations are based on preferences (Bryman, 2016). The fact that the sampling approach is opportunistic further compounds the problem of replication. Since the interaction is face-to-face, the interviewees are likely to be influenced by the researcher's personal characteristics e.g. personality, gender, or professional position (intern).

Secondly, there is no objective measurement, the data consists of opinions. Apart from the problem just described relating to the researcher's determination of importance, there is always a personal relationship with the people studied (Bryman, 2016). This gives reason to doubt the opinion of the civil servants studied since they work from a certain professional work ethic. Speaking negatively about the organization might be a sensitive issue for them, there is an expectancy that the promise of confidentiality towards them might've mediated this problem.

Thirdly, external validity i.e. the extent to which research findings can be generalized across different social settings (Bryman, 2016) is little to none. However, the goal of the research was not meant to be generalized across other cases, just as the people interviewed were not meant to be representative of a population. On the contrary, the generalization was related to theory rather than populations, the theoretical inferences that were made from the collected qualitative data determined this research's generalization. To conclude, theoretical generalization and structuring further research were the goals to be pursued.

5.5 Recommendations for future research

Based on the limitations of the research and the discussion underlying the results, there are some key topics for potential future research. Firstly, for the hypothesis on the topic of privacy and safety, the research primarily focused on the disclosure of algorithms and input data, however an AI system has more facets on which light can be shed. Examples being the output of a system, or the compliance reports that preceded the development of the system. What types of issues the disclosure of other parts of an AI system would instigate can be fruitful. In addition, further emphasis on the importance of the policy domain in the differing needs for transparency can be fruitful, what degree of transparency is desirable in what policy area? But the most surprising finding is the importance of the media and its potential effect on innovation as a whole, to what extent does the media play a role in the recalcitrance of governmental transparency?

Secondly, on the hypothesis of expertise and cooperation, further research on what the moderation between experts and nonexperts of AI or any other technology essentially means is necessary, the expertise of ‘tech translation’ requires further specification. In addition, the extent to which the upgrading of legacy systems will improve operational cooperation and if it is truly a technical limitation can be interesting. In addition, on the separation between politics and administration, the results give reason to rethink the relationship between politics and administration. To what extent can politicians provide policy leadership to their administration on complex topics? What is needed to equalize the knowledge of politicians to that of administrators whilst keeping the spheres separated? On the topic of data scientists versus domain experts, how much knowledge of the system should the end user have? Whose responsibility should it be to provide this knowledge? Is it feasible to impart this knowledge on an aging workforce? Furthermore, on the topic of hierarchy and accountability, to what extent can a manager be accountable for something he doesn’t understand?

Thirdly, in relation the third hypothesis on inefficient disclosing. It is questionable whether transparency is of any use if it lacks a critical audience to evaluate the information disclosed. In addition, the reliance on private sector innovation models evokes a whole set of questions, to what extent can private sector innovation models be used without its negative consequences for the public sphere? Can models that are used for efficiency be used to quantify welfare?

Finally, recommendations for further research based on the limitations of this research. Since this research is of an exploratory nature, perhaps it generates far more questions than it initially answered. This shouldn’t particularly have to be a bad thing, but quantitative measurement would certainly adorn the qualitative findings since the qualitive measurements are not objective. Since the findings can’t be generalized across other governments, it could be fruitful to research how other national governments handle being transparent about AI-assisted decision-making. In addition, upscaling and further differentiating the sample could provide clearer data saturation, such as including the private sector or interviewing other levels of government such as street-level bureaucrats or provincial bureaucrats.

6 Conclusion

This exploratory research aimed to contribute to the lacuna of identifying obstacles related to being transparent in AI-assisted governmental decision-making. As such, the study seeks to make a contribution by providing sufficient tools for further research on the topic of AI in governmental decision-making. Before coming to the conclusion, it should be said that this research solely focused on obstacles whereas there are numerous opportunities to being transparent in AI-assisted governmental decision-making. These obstacles must be read in opportunities at a time when technology continues to amaze us all and the future remains ever so exciting and unpredictable.

AI as a technological tool holds great opportunities for the public sector, however, governments have to overcome certain barriers to be transparent about its usage in decision-making. Transforming government in such a way to address obstacles related to being transparent in AI-assisted governmental decision-making is going to be a major challenge in the coming years since they're not objective in nature but most of all subjective and cultural. Time will tell whether governments can keep up the pace with the ever so rapid changes of AI technology to satisfy and protect its citizens. Continuous small scale experimenting in order show the world that it is not all trial and error and break the negative stigma that IT and government has whilst being transparent should definitely be prioritized.

In attempting to answer the research question: what are the obstacles related to being transparent in AI-assisted government decision-making? Several obstacles have been found. Since data and algorithms are not the same as documentary information it will require a greater effort for governments to engage in proactive transparency rather than reactive transparency. Engaging in reactive transparency, attempting to be transparent, about AI in the decision-making presents additional barriers than documentary information would instigate.

The barriers are the following, the results show that it is important to determine prior to disclosure what part of the AI assisted decision-making should be made transparent as different parts instigate different issues. In addition, this highlights that the digital information used by AI cannot be treated as uniform as documentary information. Also, whereas the disclosure of documentary information has put an emphasis on the privacy of civil servants, the disclosure of (personal) data can therewithal jeopardize the privacy of citizens. However, it is not that the privacy of civil servants is jeopardized, rather, it's the reputation of the organization and the significance of the media that poses the real issue. Nevertheless, the government has to step up and break the spiral by showing some merit and good innovative examples if it wishes to attract talent to achieve their ambitions.

Even though there is an apparent lack of AI experts globally, it is wrong to assume that more programmers and data scientists are the experts needed to fill the expertise gap within government. On the contrary, experts who can provide ‘tech translation’, who are able to critically moderate between experts and nonexperts of AI and act as a guardian to prevent dehumanization in the development of AI is lacking.

In addition, there is a lack of cooperation on the operational level that is reflected in the inadequate information architecture of government and the presence of legacy systems. Further, there is a lack of cooperation on the strategic level that is reflected in the fragmented initiatives and lack of common understanding of concepts. Additionally, there is a lack of cooperation on the organizational level due to the pillarization of governmental tasks, rolls and agenda's ensuring that problems aren't known to each other. Lastly, related to the organizational sphere, there is a lack of cooperation or rather a lack of awareness amongst critical actors in the AI development process leading to transparency issues.

It appears that governments engage in transparency as verifiability, expecting to generate legitimacy by providing their critical citizenry with the information to assess this legitimacy. However, this approach doesn’t acknowledge the complexities that are involved in communicating disclosures, taking for granted that anyone can understand it. In addition, citizens aren’t capable to engage in unbiased and critical engagement with the information provided given its complexity. It is therefore an erroneous assumption by government that disclosing input data and algorithms will lead to increased government legitimacy. It appears that the target audience for which the endeavors of transparency are done are completely forgotten in the process. A potential reason for this alienation from society is the indiscriminate copying of management techniques from the private sector. This results in public values being ignored in exchange for prospects of enhanced efficiency and output. It is high time that the government stops being occupied with mitigating barriers with the private sector and starts focusing on the consequential alienation from society. Thus, when it concerns being transparent about AI in contemporary decision-making, it seems that the public sector has to refocus on their essential task, serving citizens.

7. Bibliography

Abbott, A. (2014). *The system of professions: An essay on the division of expert labor*. University of Chicago press.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138-52160.

Albu, O.B., & Flyverbom, M. (2016). Organizational Transparency: Conceptualizations, Conditions, and Consequences. *Business & Society, 58*(2), 268-297.

Algorithm. (n.d.) in *Cambridge Dictionary.* https://dictionary.cambridge.org/dictionary/english/algorithm

Alpaydin, E. (2016). *Machine learning* (MIT Press Essential knowledge series). Cambridge: MIT Press.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society*, *20*(3), 973-989.

Ball, C. (2009). What Is Transparency? *Public Integrity, 11*(4), 293-308.

Bannister, F., & Connolly, R. (2011). The trouble with transparency: a critical review of openness in e-government. *Policy & Internet*, *3*(1), 1-30.

Bekkers, V., Edelenbos, J., & Steijn, B. (2011). Linking innovation to the public sector: Contexts, concepts and challenges. In *Innovation in the Public Sector* (pp. 3-32). Palgrave Macmillan, London.

Bhatt, S. (2019). Reinforcement Learning 101 - Towards Data Science. Retrieved from https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292

Big Data. (n.d.) in *Cambridge Dictionary.* https://dictionary.cambridge.org/dictionary/english/big-data

Boden, M. A. (2016). *AI: Its nature and future*. Oxford University Press.

Bosque, G., del Campo, I., & Echanobe, J. (2014). Fuzzy systems, neural networks and neuro-fuzzy systems: A vision on their hardware implementation and platforms over two decades. *Engineering Applications of Artificial Intelligence, 32*(1), 283–331.

Boyd, M., & Wilson, N. (2017). Rapid developments in artificial intelligence: how might the New Zealand government respond?. *Policy Quarterly*, *13*(4).

Bryman, A. (2016). *Social research methods*. Oxford University Press.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530-1534.

Brown, D. C., & Toze, S. (2017). Information Governance in Digitized Public Administration. Canadian Public Administration, 60(4), 581-604.

Brownlee, J. (2020, August 15). *Model Prediction Accuracy Versus Interpretation in Machine Learning*. Machine Learning Mastery. https://machinelearningmastery.com/model-prediction-versus-interpretation-in-machine-learning/

Buranyi, S. (2019). “Dehumanising, impenetrable, frustrating”: the grim reality of job hunting in the age of AI. Retrieved from https://www.theguardian.com/inequality/2018/mar/04/dehumanising-impenetrable-frustrating-the-grim-reality-of-job-hunting-in-the-age-of-ai

Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 2053951715622512.

Caballé, N. C., Castillo-Sequera, J. L., Gómez-Pulido, J. A., Gómez-Pulido, J. M., & Polo-Luque, M. L. (2020). Machine learning applied to diagnosis of human diseases: A systematic review. *Applied Sciences, 10*(15), 5135.

Cellan-Jones, R. (2020). Uber’s self-driving operator charged over fatal crash. Retrieved from https://www.bbc.com/news/technology-54175359

Chinn, D., Hieronimus, S., Kirchherr, J., & Klier, J. (2020, December 4). *The future is now: Closing the skills gap in Europe's public sector*. McKinsey & Company. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/the-future-is-now-closing-the-skills-gap-in-europes-public-sector

Chun, A. H. W. (2008). An AI Framework for the Automatic Assessment of e-Government Forms. *AI Magazine*, *29*(1), 52.

Clarke, J., & Newman, J. (1997). *The managerial state: Power, politics and ideology in the remaking of social welfare*. Sage.

Coeckelbergh, M. (2020). *AI Ethics*. MIT Press.

Coglianese, C., & Lehr, D. (2019). Transparency and algorithmic governance. *Admin. L. Rev.*, *71*, 1.

Collier, M., Fu, R. & Yin, L. (2017). Artificial intelligence: Healthcare's new nervous system. Edited by Accenture. Retrieved January 14, 2021, from: https://www.accenture.com/t20170418T023052Z__w__/au-en/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf

Cortés, U., & Sanchez-Marre, M. (1999). Binding environmental sciences and artificial intelligence in environmental modelling & software. *Environmental Modelling and Software*, *14*(5), 335-337.

Cortés-Cediel, M. E., Cantador, I., & Gil, O. (2017, August). Recommender systems for e-governance in smart cities: State of the art and research opportunities. In *Proceedings of the international workshop on recommender systems for citizens* (pp. 1-6).

Council of Europe., & Council of Europe. (1952). *The European convention on human rights*. Strasbourg: Directorate of Information.

Craglia, M., Hradec, J., & Troussard, X. (2020). The Big Data and Artificial Intelligence: Opportunities and Challenges to Modernise the Policy Cycle. *Science for Policy Handbook*, 96-103.

Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society*, *20*(1), 88-104.

Dahlberg, L. (2011). Re-constructing digital democracy: An outline of four 'positions'. *New media & society*, *13*(6), 855-872.

Dawes, S. S., Cresswell, A. M., & Pardo, T. A. (2009). From "need to know" to "need to share": Tangled problems, information boundaries, and the building of public sector knowledge networks. *Public administration review*, *69*(3), 392-402.

de Fine Licht, J. (2014). Policy area as a potential moderator of transparency effects: An experiment. *Public Administration Review*, *74*(3), 361-371.

de Sousa, W. G., de Melo, E. R. P., Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly*, *36*(4), 1-14.

de Wilde, M. (2010). Locke and the state of exception: Towards a modern understanding of emergency government. *European Constitutional Law Review*, *6*(2), 249.

Demir, T., & Nyhan, R. C. (2008). The politics–administration dichotomy: An empirical search for correspondence between theory and practice. *Public Administration Review*, *68*(1), 81-96.

Denning, P. J. (2007). Computing is a natural science. *Communications of the ACM*, *50*(7), 13-18.

Dey, A. (2016). Machine learning algorithms: a review. *International Journal of Computer Science and Information Technologies*, *7*(3), 1174-1179.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, *59*(2), 56-62.

Diakopoulos, N. (2014). Algorithmic accountability reporting: On the investigation of black boxes. Columbia Journalism School.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *Retrieved from: https://arxiv.org/abs/1702.08608*

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 101994.

European Commission. (2016). Guidelines on Transparency. Luxembourg: Office for Official Publications of the European Communities. Retrieved from: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227

European Commission. (2018). *Artificial Intelligence for Europe*. Brussels. Retrieved from: https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe.

European Commission. (2019, April 8). *Ethics guidelines for trustworthy AI*. Shaping Europe's Digital Future - European Commission. Retrieved from: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Externe onderzoeken. (n.d.). Parlement.com. Retrieved from: https://www.parlement.com/id/vibrc7hwl2z0/externe_onderzoeken

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, *28*(1), 62-70.

Ferber, J. (1999). *Multi-agent systems: An introduction to distributed artificial intelligence.*

Boston, MA: Addison-Wesley Longman.

Ferreira, D., & Sah, R. K. (2012). Who gets to the top? Generalists versus specialists in managerial organizations. *The RAND journal of economics*, *43*(4), 577-601.

Frey, W. R., Patton, D. U., Gaskell, M. B., & McGregor, K. A. (2020). Artificial intelligence and inclusion: Formerly gang-involved youth as domain experts for analyzing unstructured twitter data. *Social Science Computer Review*, *38*(1), 42-56.

Gibbs, S. (2017). AlphaZero AI beats champion chess program after teaching itself in four hours. Retrieved from https://www.theguardian.com/technology/2017/dec/07/alphazero-google-deepmind-ai-beats-champion-program-teaching-itself-to-play-four-hours

Grimmelikhuijsen, S., Porumbescu, G., Hong, B., & Im, T. (2013). The effect of transparency on trust in government: A cross-national comparative experiment. *Public administration review*, *73*(4), 575-586.

Grimmelikhuijsen, S., & Meijer, A. J. (2014). Effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment. *Journal of Public Administration Research and Theory*, *24*(1), 137-157

Griggs, T., & Wakabayashi, D. (2018, July 30). *How a Self-Driving Uber Killed a Pedestrian in Arizona*. The New York Times Company. https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field methods*, *18*(1), 59-82.

Harding, L., & Barden, L. (2011). From the archive, 12 May 1997: Deep Blue win a giant step for computerkind. Retrieved from https://www.theguardian.com/theguardian/2011/may/12/deep-blue-beats-kasparov-1997

Hawley, S. H. (2019). Challenges for an Ontology of Artificial Intelligence. *arXiv preprint arXiv:1903.03171*.

Heald, D. (2012). Why is transparency about public expenditure so elusive? *International Review of Administrative Sciences, 78*(1), 30-49. doi:10.1177/00208523|1429931

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1-2), 147-169.

Holdren, J. P., & Smith, M. (2016). Preparing for the future of artificial intelligence. *Executive Office of the President of the United States, National Science and Technology Council, Committee on Technology*.

Howlett, M. (2015). Policy analytical capacity: The supply and demand for policy analysis in government. *Policy & Society, 34*(3-4), 173-182.

ICTU. (2017, July 29). Ondermijnende criminaliteit beter in vizier met data analyse. Retrieved from https://www.ictu.nl/nieuws/ondermijnende-criminaliteit-beter-in-vizier-met-data-analyse#

Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, 371-376. doi:10.1016/j.giq.2010.05.003

Jain, P., Gyanchandani, M., & Khare, N. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data*, *3*(1), 1-25.

Jann, W., & Wegrich, K. (2007). Theories of the policy cycle. *Handbook of Public Policy Analysis: Theory, politics, and methods*, *125*, 43-62.

Jordan, M. I, & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science (American Association for the Advancement of Science), 349*(6245), 255-260.

Joy, A. (2020a). *Pros And Cons Of Supervised Machine Learning*. Pythonista Planet. https://www.pythonistaplanet.com/pros-and-cons-of-supervised-machine-learning/

Joy, A. (2020b). *Pros And Cons Of Reinforcement Learning*. Pythonista Planet. https://www.pythonistaplanet.com/pros-and-cons-of-reinforcement-learning/

Kelleher, J. D. (2019). *Deep learning*. (MIT Press Essential knowledge series). Cambridge: MIT Press.

Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.

Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, *22*(14), 2081-2096.

Keum, D. D., & See, K. E. (2017). The influence of hierarchy on idea generation and selection in the innovation process. *Organization Science*, *28*(4), 653-669.

Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information systems frontiers*, *19*(2), 267-283.

Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). A governance framework for algorithmic accountability and transparency. Brussels: European Parliamentary Research Service.

Kouziokas, G. N., Chatzigeorgiou, A., & Perakis, K. (2017). Artificial intelligence and regression analysis in predicting ground water levels in public administration. *European Water Publications*, *57*, 361-366.

Krenker, A., Bešter, J., & Kos, A. (2011). Introduction to the artificial neural networks. *European Journal of Gastroenterology & Hepatology*, *19*(12), 1046–1054.

Lam, W. (2005). Barriers to e-government integration. *Journal of Enterprise Information Management*.

Lathrop, D., & Ruma, L. (2010). *Open government: Collaboration, transparency, and participation in practice*. " O'Reilly Media, Inc.".

Liddy, E. D. (2001). Natural language processing. *Encyclopedia of library and information Science* (2nd ed.). New York, NY: CRC Press Marcel Dekker.

Lin, S. Y., Shanafelt, T. D. & Asch, S. M. (2018). Reimagining Clinical Documentation With Artificial Intelligence. *Mayo Clinic Proceedings*, *93*(5), 563-565.

Löfgren, K., & Webster, C. W. R. (2020). The value of Big Data in government: The case of 'smart cities'. *Big Data & Society*, *7*(1), 2053951720912775.

Lohr, S. (2013) "The Origins of 'Big Data': An Etymological Detective Story" *The New York Times*, February 1 2013 [Online]. Available at: https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/

Louridas, P. (2020). *Algorithms*. MIT Press.

Man, K. F., Tang, K. S., & Kwong, S. (1996). Genetic algorithms: Concepts and applications. *IEEE Transactions on Industrial Electronics*, *43*(5), 519–534.

Marr, B. (2019, March 8). *The AI Skills Crisis And How To Close The Gap*. Forbes. https://www.forbes.com/sites/bernardmarr/2018/06/25/the-ai-skills-crisis-and-how-to-close-the-gap/?sh=530574e931f3

Mason, M. (2010). Sample Size and Saturation in PhD Studies Using Qualitative Interviews. *Forum: Qualitative Social Research*, *11*(3), article 8, Retrieved from: www.qualitative-research.net/index.php/fqs/article/view/1428

McGonagle, T. (2017). "Fake news" False fears or real concerns?. *Netherlands Quarterly of Human Rights*, *35*(4), 203-209.

Mehr, H. (2017, August). *Artificial Intelligence for Citizen Services and Government*. Ash Center for Democratic Governance and Innovation, Harvard Kennedy School. https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. Government Information Quarterly, 36(4), 1-16.

Micheli, P., Schoeman, M., Baxter, D., & Goffin, K. (2012). New business models for public-sector innovation: Successful technological innovation for government. *Research-Technology Management*, *55*(5), 51-57.

Ministerie van Algemene Zaken. (2017, December 1). Open overheid. Retrieved from https://www.rijksoverheid.nl/onderwerpen/digitale-overheid/open-overheid

Molnar, C. (2018). A guide for making black box models explainable. *Retrieved from: https://christophm.github.io/interpretable-ml-book*.

Moss, G., & Coleman, S. (2014). Deliberative manoeuvres in the digital darkness: e-democracy policy in the UK. *The British Journal of Politics and International Relations*, *16*(3), 410-427.

Naurin, D. (2002). *Taking transparency seriously*. Brighton, UK: Sussex European Institute.

McJohn, S. M. (1998). Artificial legal intelligence. *Harvard Journal of Law & Technology*, *1*(12), 241–261.

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. The Penguin Press: USA.

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

Pasquier, M., & Villeneuve, J. P. (2007). Organizational barriers to transparency: A typology and analysis of organizational behaviour tending to prevent or restrict access to information. *International Review of Administrative Sciences*, *73*(1), 147-162.

Razavian, N., Knoll, F., & Geras, K. J. (2020, February). Artificial intelligence explained for Nonexperts. In *Seminars in musculoskeletal radiology* (Vol. 24, No. 1, p. 3). NIH Public Access.

Resnick, M., Berg, R., & Eisenberg, M. (2000). Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *The Journal of the Learning Sciences*, *9*(1), 7-30.

Mendel , T. (2003). National Security vs Openness: An Overview and Status Report on the Johannesburg Principles, *National Security and Open Government: Striking the Right Balance*. New York: The Maxwell School of Syracuse University.

Mihas, P., & Odum Institute. (2019). Learn to build a codebook for a generic qualitative study. In *SAGE Research Methods Datasets Part 2*. SAGE Publications, Ltd. https://www.doi.org/10.4135/9781526496058

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (n.d.). *Openbaarheid van overheidsinformatie*. Wet openbaarheid van bestuur (Wob) | Rijksoverheid.nl. https://www.rijksoverheid.nl/onderwerpen/wet-openbaarheid-van-bestuur-wob/openbaarheid-van-overheidsinformatie

Olschewski, M. (2017). A German Alexa owner returned home to find his Amazon device had started a "party" at 2am, leading to police breaking down his door. Retrieved from https://www.businessinsider.com/amazon-alexa-started-party-2am-police-broke-down-door-2017-11?international=true&r=US&IR=T

Porambage, P., Ylianttila, M., Schmitt, C., Kumar, P., Gurtov, A., & Vasilakos, A. V. (2016). The quest for privacy in the internet of things. *IEEE Cloud Computing*, *3*(2), 36-45.

Quraishi, F. F., Wajid, S. A., & Dhiman, P. (2017). Social and ethical impact of artificial intelligence on public: A case study of university students. *International Journal of Scientific Research in Science, Engineering and Technology*, *3*(8), 463-467.

Rechtbank Den Haag. (2020, February 5th). SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens/C-09-550982-HA ZA 18-388. Retrieved from: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:865

Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies*, *1*(1), 33-36.

Roberts, A. (2001). Structural pluralism and the right to information. *The University of Toronto Law Journal*, *51*(3), 243-271.

Rouse, M. (2020, July 8). Supervised Learning. Retrieved from https://searchenterpriseai.techtarget.com/definition/supervised-learning

Rowley, J. (2006). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science*, *33*(2), 163-180.

Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harv. JL & Tech.*, *29*, 353.

Shen, O. (2020, November 30). *Interpretability in Machine Learning: An Overview*. The Gradient. https://thegradient.pub/interpretability-in-ml-a-broad-overview/

Shokri, R., Strobel, M., & Zick, Y. (2019). On the Privacy Risks of Model Explanations. *arXiv preprint arXiv:1907.00164*.

Smola, A. J., & Vishwanathan, S. V. N. (2008). *Introduction to machine learning* (1st ed.). Cambridge University Press.

Stone, D. (2012). *Policy paradox: The art of political decision making* (Third ed.). New York, NY: W.W. Norton.

Sutton, R. S. (1985). Temporal Credit Assignment in Reinforcement Learning.

Swartz, A. (2010). When is transparency useful? In Lathrop, D., & Ruma, L. *Open government: Collaboration, transparency, and participation in practice* (pp. pp. 267-274) O'Reilly Media, Inc.

The Economist. (2020, June 10). *Businesses are finding AI hard to adopt*. https://www.economist.com/technology-quarterly/2020/06/11/businesses-are-finding-ai-hard-to-adopt

The Economist Intelligence Unit. (2018, July). *INTELLIGENT ECONOMIES: AI's transformation of industries and society*. https://eiuperspectives.economist.com/technology-innovation/intelligent-economies-ais-transformation-industries-and-society

Thierer, A. D., Castillo O'Sullivan, A., & Russell, R. (2017). Artificial intelligence and public policy. *Mercatus Research Paper*.

Transparency. (n.d.) in *Cambridge Dictionary.* https://dictionary.cambridge.org/dictionary/english/transparency

UN General Assembly. (1948). *Universal Declaration of Human Rights*. Retrieved from: https://www.un.org/en/universal-declaration-human-rights/

Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, *31*(2), 278-290.

Viaene, S. (2013). Data scientists aren't domain experts. *IT Professional*, *15*(6), 12-17.

Weick, K. E. (1995). *Sensemaking in organizations*. Sage Publications (Thousand Oaks).

Weil, D., Graham, M., & Fung, A. (2013). Targeting transparency. *Science*, *340*(6139), 1410-1411.

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—applications and challenges. *International Journal of Public Administration*, *42*(7), 596-615.

Worthy, B. (2013). *David Cameron's Transparency Revolution? The Impact of Open Data in the UK* [Paper]. Available at SSRN: https://ssrn.com/abstract=2361428 or http://dx.doi.org/10.2139/ssrn.2361428

Worthy, B. (2015). The impact of open data in the UK: Complex, unpredictable, and political. *Public Administration*, *93*(3), 788-805.

Yanow, D. (2003). *Constructing Race and Ethnicity in America*. Florence: Routledge.

8. Appendices

8.1 Appendix I interview transcripts

The interview transcripts can be requested by contacting the researcher at: e.d.van.essen@umail.leidenuniv.nl

8.2 Appendix II interview guide

1. **Introductie**

- Welkom en bedankt voor het interview
- Persoonlijke introductie van mijzelf
- Vragen om introductie van de geïnterviewde
- Uitleg over mijn onderzoek en het nut van dit interview
- Bevestiging van vertrouwelijkheid
- Vragen om het gesprek op te nemen

2. **Algemene vragen**

- Wat vindt u van transparantie in data/AI gedreven besluitvorming?
- In hoeverre vindt u het transparant zijn als overheid van belang?

3. **Privacy en veiligheid**

- Leg uit wat je bedoelt met 'obstructed transparency'
- In hoeverre is het zo dat veiligheidsoverwegingen een obstakel zijn om informatie over een AI systeem niet te delen?
- In hoeverre is privacywetgeving een obstakel in het delen van informatie over algoritmen?
- Hoe denk je over de balans tussen transparant zijn om te tonen dat je de veiligheid en privacy waarborgt, en té transparant zijn wat mogelijk veiligeids- en privacyrisico's met zich meebrengt?

4. **Organisatorische middelen en expertise**

- Leg uit wat je bedoelt met 'strained transparency'
- Wat is het belang van het hebben van de juiste systeemgeralateerde kennis?
- Zouden de mensen die er uiteindelijk mee moeten werken deze kennis moeten hebben? Waarom?

- Uitleggen kost geld, in hoeverre zijn financiële middelen een obstakel?

5. **Open overheid en open source**

- Leg uit wat je bedoelt met 'maximized transparency'
- In hoeverre zijn deze open overheid initiatieven van toepassing op het gebruik van AI? Voorbeelden?
- Ben je van mening dat je als overheid voldoende transparant bent als je de broncode van een systeem deelt?
- Wie zou de doelgroep zijn die je hiermee bereikt?
- Is het doel van transparantie bereikt op de manier zoals dat nu gebeurd?
- Wie zou verantwoordelijk moeten zijn voor die uitlegbaarheid?

- <u>Zijn er nog andere factoren die nog genoemd moeten worden als het gaat over obstakels van het toepassen van AI in de besluitvorming?</u>

6. **Afsluiting**

- Vraag om opmerkingen of aanvullende informatie
- Bedanken

7. **Opvolging**

- Stuur het transcript

8.3 Appendix III codebook

Table 8

Theory-driven codes

Code	Description	Example
Privacy	Expert mentions that privacy in the broadest sense can be a reason for non-disclosure of AI related information.	"Het is best wel precair, daar moet je goed over nadenken. Dus ik kan mij voorstellen dat het met open data soms een issue kan zijn. Wij doen ook bij [redacted] ook soms een DPIA ook al is het open data. Dan doen we een privacy impact assessment terwijl het in principe anoniem zou moeten zijn maar ja dat is het dan niet altijd" (PO A).
Security	Expert mentions that (state) security can be a reason for non-disclosure of AI related information.	"Het gaat natuurlijk wel deels om als het gegevens zijn die ervoor zorgen dat je makkelijker kan inbreken als hacker … ik kan mij voorstellen bij de stormvloedkering dat je niet zomaar wilt delen hoe dat natuurlijk werkt" (PO D).
Organizational resources	Expert mentions that lack of organizational resources i.e. poor information management to process transparency demand is reason of non-disclosure of AI related information.	"Ja maar het is ook een infrastructuurprobleem. ICT bij de overheid dat is ook niet helemaal je-van-het. Er werken ontzettend veel getalenteerde ICTers bij de overheid dus ik vind het altijd 'IT bij de overheid is per definitie slecht' dat is zeker niet waar. Maar het is wel zo dat elke organisatie doet zijn eigen ding. Iedereen maakt zijn eigen systeempje of ze besteden dat uit, ik vind dat systemen slecht samenwerken" (PO A).
Expertise	Expert mentions that unfamiliarity with the documents/information is	"[…] een beslissing bevoegde ambtenaar die moet het systeem kunnen begrijpen anders krijg je een hele slechte kwaliteit

	reason of non-disclosure of AI related information.	van besluiten of dan weet hij niet wat de gevolgen zijn. Dus ja systeemkennis is echt super cruciaal" (PIMO A).
Transparency projects	Expert mentions that transparency projects that don't properly organize information fail to reach its goal of being legitimate.	"[…] je wilt transparant zijn op zo'n manier dat de burger snapt wat er gebeurt binnen de overheid. En dat betekent niet altijd dat je bijvoorbeeld in het geval een algoritme de source code openbaar maakt. Want er zijn misschien een paar duizend mensen in Nederland die dat goed kunnen lezen en die het ook heel leuk vinden om te lezen, maar de rest kan er helemaal niks mee" (PINO B).

Table 9

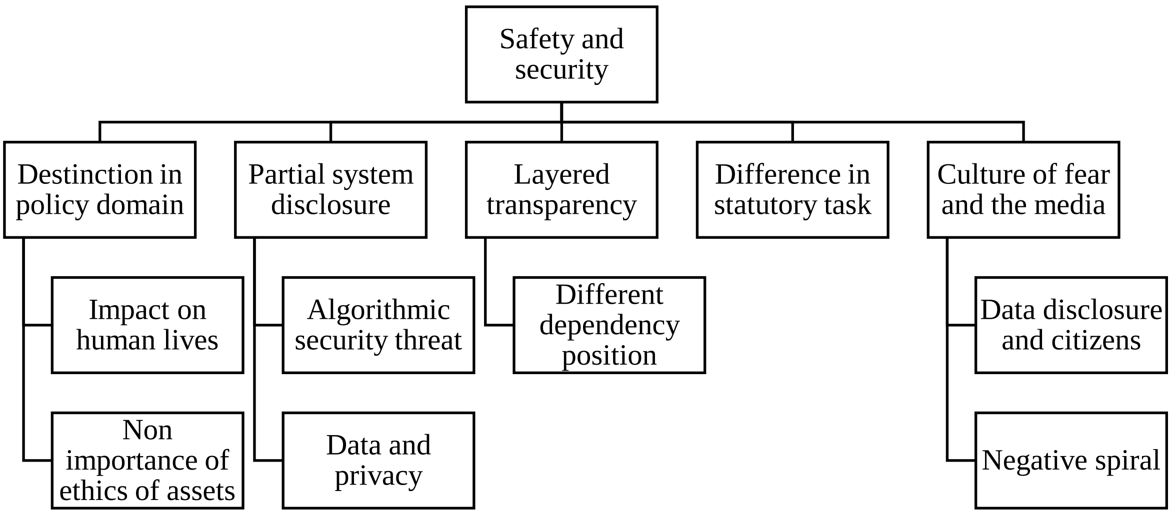
Data-driven codes merged with theory-driven codes per theme

Safety and Security		
Code	**Description**	**Example**
Distinction in policy domain	Expert mentions that policy domains differ.	"Ja ik denk dat dat een beetje verschilt per domein. Dus ik ga nu even vanuit de […] praten, als je het wilt kan ik dat ook via Defensie doen die kijken er net iets anders tegenaan" (PINO A). "Maar hoeveel energie een brug gebruikt heeft toch wel minder impact op mensen dan als je in het sociaal domein werkt. Dus dat betekent, bij ons komen mensenrechten bijvoorbeeld niet zo gauw in het geding, als je aan assetmanagement werkt" (PO C).
Impact on human lives	Expert mentions that what determines the difference in	"[Oké je noemt nu het verschil tussen het fysieke domein en het sociaal

	policy domain is the impact on human lives.	domein, in hoeverre vind jij daarin een verschil?] Dat is echt de directe impact op mensen" (PO C).
Non importance of ethics of assets	Expert mentions that ethical concerns are less prominent in the physical domain.	"bij ons komen mensenrechten bijvoorbeeld niet zo gauw in het geding, als je aan assetmanagement werkt […] ja ik maak wel eens het flauwe grapje: de privacy van een baksteen" (PO C).
Partial system disclosure	Expert mention that what part of system is disclosed is of importance.	"Alleen ik denk dat de werking van een algoritme uitleggen zeg maar dat hoeft niet per se de privacy in de weg te zitten"
Algorithmic security threat	Expert mentions that disclosure of algorithms instigates security issues.	"[Dus als het gaan om algoritmen dan is privacy niet echt een issue. Maar meer misschien veiligheid dag?] Ja ik denk het wel" (PO D).
Data and privacy	Expert mentions that data disclosure instigates privacy issues.	"Nou ja als je persoonsgegevens eruit haalt betekent dat nog niet meteen dat het anoniem is" (PO A)
Layered transparency	Expert mentions that there is always some possibility for being transparent.	"[…] er is altijd wel een mate van transparantie die wel kan en ik denk dat je daarin heel goed zou kunnen afwegen wat dan nuttig is voor de burger om te weten en wat daadwerkelijk je operatie schaadt" (PINO B).
Different dependency position	Expert mentions that a government has a different dependency position than a company.	"[…] te maken hebt met een afhankelijkheidspositie. De burger kiest er niet voor dat zijn gegevens door ons, door het ministerie worden verwerkt. En ja dan heb je denk ik een zwaardere transparantieplicht dan een

		bedrijf waar je voor kan kiezen en niet voor kan kiezen" (PIMO A).
Difference in statutory task	Expert mentions that separation of statutory tasks of government can present issues.	"[…] in een beleidscontext , efficiencyslagen wel worden geprobeerd te maken maar efficiency zoals het geldt bij een uitvoering als in een werkproces heeft echt een totaal andere dimensie" (PO B). "Dan wordt [het probleem] niet opgelost. En dat komt echt door die scheiding tussen uitvoering en beleid" (PIMO A).
Culture of fear and the media	Expert mentions the importance of the media and an apparent presence of fear of disclosure amongst civil servants.	"Positieve voorbeelden die hoor je eigenlijk niet, terwijl die er natuurlijk ook zijn. Waardoor er ook een soort angst ontstaat van 'als we hier heel transparant over gaan zijn dan komen we weer met grote kop in de kranten', dat soort sentiment heerst er ook wel" (PO A).
Data disclosure and citizens	Expert mentions that disclosure of data can have unforeseen consequences.	"[…] ik denk dat er zoveel mogelijkheden aankomen op het gebied van data analyse en zoveel data beschikbaar wordt gesteld dat je nu heel moeilijk de consequenties kan voorzien voor publicatie" (PIMO A).
Negative spiral	Expert mentions that negative media attention instigates a negative spiral for AI development in general.	"Want op dit moment wordt de rem erop gezet waardoor er minder capaciteit is om dingen goed te ontwikkelen. En wat jij net zegt, er is ook capaciteit nodig om dingen transparant te maken of in ieder geval een bepaalde uitleg te geven naar

		burgers waar ze op zitten te wachten. En daar is nu … ik denk ook niet dat daar de geldkraan voor wordt opengezet terwijl je dat denk ik wel al wilt. Je hebt denk ik binnen de hele overheid in alle divisies en regiones heb je mensen nodig die snappen wat dit soort systemen kunnen wat de risico’s zijn, technisch en dat vertalen naar socio-technisch en daar heb je gewoon capaciteit voor nodig en dat staat nu in een keer op de rem. Ja dus je bent eigenlijk een soort van in een negatieve spiraal beland” (PINO B).



Organizational cooperation and expertise		
Code	**Description**	**Example**
Ontological expertise	Expert mentions that there is a distinctive type of expertise lacking within government, assessing the	“En dat moet je dus ook doen als je met data werkt bij de overheid dat je … het is niet zo makkelijk om een werkelijkheid weer te geven met data […] er moeten meer mensen

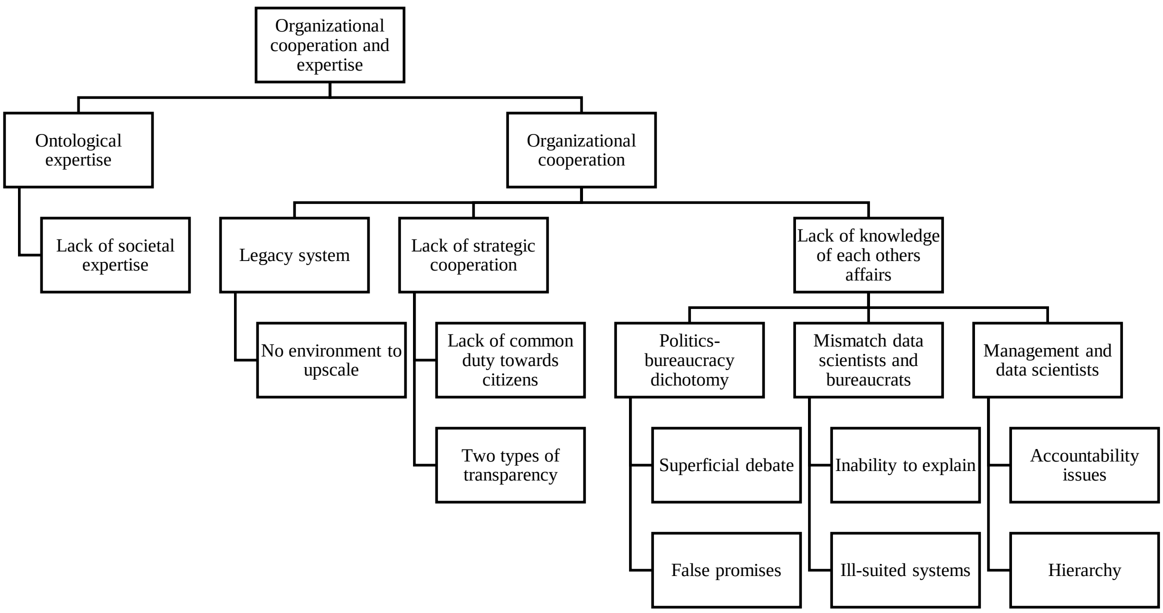
	exactness of data measurements.	binnenkomen die de vertaalslag duidelijk kunnen maken en die tegen de minister ook in zijn leefwereld kunnen vertellen 'oké dit is wat er is misgegaan'" (PO E).
Lack of societal expertise	Expert mentions that a lack of societal implications is present.	"[Zo'n] aanpak objectiveert door er algoritmes voor te schrijven maar vervolgens niet meer goed voor ogen hebt wat de impact is op de samenleving" (PO E).
Organizational cooperation	Experts mentioning that it is not lack of experts, rather cooperation.	"Er werken ontzettend veel getalenteerde ICTers bij de overheid dus ik vind het altijd 'IT bij de overheid is per definitie slecht' dat is zeker niet waar […] En misschien een obstakel die er nog kan zijn, is gebrek aan samenwerking tussen organisaties […] ik zie niet heel veel samenwerking op dat gebied, denk dat dat meer kan worden" (PO A).
Legacy systems	Expert mentions that there is still presence of legacy systems.	"Hebben wij de gegevensinfrastructuur om transparantie te bieden?' Nou heel vaak is dat niet zo want de systemen … in de tijd dat die werden ingericht, bijvoorbeeld in het gevangeniswezen daar zitten heel veel oudere legacy systemen. Die zijn niet ingericht met het oogpunt op transparantie dus dat kost gigantisch veel tijd en geld om transparantie te bieden" (PIMO A).
No environment to upscale	Expert mentions the obstacle of upscaling an	"[…] hebben we dan een experimenteeromgeving van maar dat is niet een omgeving waar inspecteurs

	application to be used in the organization.	bij kunnen of waar de rest van de organisatie bij kan […] dat is gewoon een heel technisch obstakel, hoe zorg je ervoor dat ze ook daadwerkelijk met de spullen kunnen werken die wij maken" (PINO A)
Lack of strategic cooperation	Expert mentions that there is a lack of strategic cooperation on how to be transparent.	´Ze zijn nu bezig met discussie over Human Impact Assessment rondom algoritmen. Maar ook wij vanuit ministerie, de algemene rekenkamer, de auditdienst rijk, de autoriteit persoonsgegevens, de agentschap telecom, iedereen die is hiermee bezig" (PIMO A).
Lack of common duty	Expert mentions that lack of cooperation results in lack of common duty on how to be transparent.	"Dus het is wel belangrijk dat dat wordt geharmoniseerd dat je een duidelijke plicht en niet tien verschillende toezichtkaders hebt" (PIMO A).
Two types of transparency	Expert mentions that there is either technical transparency or explainability.	"dat is ook het hele probleem van de discussie rondom transparantie dat we eigenlijk met containerbegrippen werken" (PIMO A). "Een data scientist die denkt in technische transparantie terwijl een beleidsambtenaar denkt in transparantie naar de burger toe en uitlegbaarheid" (PINO B).
Lack of knowledge of each other's affairs	Expert mentions that there is a general lack of knowledge of each other's affairs between governmental departments.	[…] degene die de klacht van de burger krijgt, dus de burgerbrieven is vaak een beleidsambtenaar. En die weet niks van het systeem waar de burger mee te maken heeft met die fout, dus diegene

		die dat systeem beheert die kan het helemaal begrijpen maar degene die die klacht krijgt die zet dat vaak niet door naar de beheerder van het systeem. Dus dan wordt het niet opgelost" (PIMO A).
Politics-bureaucracy dichotomy	Expert mentions that it is forbidden that politicians and bureaucrats are simultaneously schooled on a topic.	"En ik heb ook wel eens intern gevraagd van kunnen we niet een soort kennissessie organiseren om de politici te informeren van 'wat zijn wij überhaupt aan het doen'. Maar dat is heel lastig want dat kan politiek gezien niet want dan heb je opeens ambtenaren en politici bij elkaar en dat is dan problematisch" (PIMO A)
Superficial debate	Expert mentions that lack of knowledge between politicians and bureaucrats leads to superficial debate about the topic of AI and transparency.	"Ja er worden best wel veel Kamervragen gesteld ook over transparantie en AI. Maar niet dat het kennisniveau vanuit de eerste en tweede kamer rondom ICT is überhaupt gewoon heel laag. Volgens mij gaan ook bijna alle dossier houders van ICT en privacy ook weg dit jaar. En dat zie je terug aan de vragen. Dan krijgen wij vragen als 'in hoeverre lijkt dit op Judge Dredd?'. Ja dan kan de minister zeggen 'dit lijkt niet op Judge Dredd punt'. Maar dan heb je dus niet discussie waar het over zou moeten gaan!" (PIMO A).
False promises	Expert mentions that lack of knowledge of affairs between politicians and	"'Nee nieuw systeem', midden in de pandemie. En dat komt er dan ook waarschijnlijk niet, want volgens mij is

	bureaucrats leads to false promises and disappointment.	dat helemaal niet te doen. Ik weet niet of het er komt maar ik kan me best voorstellen dat het dan geroepen wordt en dan vervolgens niet lukt. En dan is er weer een teleurstelling” (PO E).
Mismatch data scientists and domain experts	Expert mentions that there is often a mismatch between the knowledge of a data scientist and the knowledge of a domain expert.	“Ik denk dat het vaak wordt onderschat hoe belangrijk domeinkennis is bij het ontwikkelen van een ML model […] het begint uiteindelijk wel bij het accepteren dat zowel domeinkennis als kennis over data science noem ik het even nu, ook samenkomen en dat ze open staan voor elkaars bevindingen en inzichten. Daar ontbreekt het nu het meest aan” (PO B).
Inability to explain	Expert mentions that there is a knowledge gap between the developers and end users of a system.	“Er is sowieso een kennis gap tussen de mensen die data science of AI maken en de mensen die het moeten gebruiken” (PINO A).
Ill-suited systems	Expert mentions that domain knowledge is essential in order to develop a suited system.	“Vaak wordt er dan gedacht ja we moeten mensen hebben die heel goed zijn in ML en dat model te kunnen optimaliseren. Maar je ziet dat iemand die daar goed in is vaak heel veel behoefte heeft aan iemand met goede domeinkennis om te kunnen toetsen/verifieren of het model echt goed werkt” (PO B).
Management and data scientists	Expert mentions that there is a fundamental mismatch between the knowledge of managers and data	“Dus je ziet dat het management niet de ernst snapt van bepaalde miscalculaties in zo’n systeem en die kunnen dus ook niet de juiste

	scientists on the subject of AI.	verantwoordelijkheid nemen" (PINO B).
Accountability issues	Expert mentions that it is wrong to delegate responsibility of a systems consequences to a data scientist.	"En ik denk dus dat het een symptoom is van het missen van de kennis bij leidinggevende die de verantwoordelijkheid wel heeft. Dus de verantwoordelijkheden die worden naar beneden gedelegeerd terwijl die eigenlijk gewoon bovenop moeten liggen" (PINO B).
Hierarchy	Expert mentions that hierarchy has a negative effect on idea generation hence impeding change.	"[…] het probleem is dat je heel veel nieuwe mensen binnenhaalt die wel veel meer kennis hebben maar wel veel minder ervaring, veel minder in een machtspositie zitten waar ze iets mee kunnen" (PO E).



Inefficient disclosing		
Code	**Description**	**Example**
Technical transparency	Expert mentions that government engages in technical transparency.	"Dat is puur een platform je kan het publiceren maar je hebt geen inhoudelijke check door degene die die

		website beheert […] de organisaties zijn zelf verantwoordelijk voor de kwaliteit. Je hebt dus weinig standaarden op dat gebied […] alleen broncode is niet genoeg er moeten … allerlei documentatie maar daarvoor zijn geen standaarden" (PIMO A).
Proactive transparency rather than reactive	Expert mentions that proactive transparency would be more beneficial than reactive transparency.	"En ik denk dat het ook als overheid ons kan helpen als we daar soms aan de voorkant beter over nadenken in plaats van dat er een WOB verzoek binnenkomt en dat we denken 'shit dit moeten we openbaar maken' […] En dan heb je van die zwarte strepen zeg maar, dat soort dingen dat zorgt natuurlijk ook voor veel wantrouwen. Dus ik denk aan die kant dat het aan je verantwoordingskant heel nuttig kan zijn om transparant te zijn." (PO D)
Information is different than source code	Expert mentions that information is different than source code.	"Want je hebt wel altijd van die zwartgelakte stukken maar als je dat bij code zou doen … ik zie gewoon heel erg een operationeel probleem hoe we dat in godsnaam bij een burger terecht moeten krijgen op een manier dat ze er ook iets mee kunnen" (PINO A).
Private sector innovation models	Expert mentions that government relies on private sector innovation models.	[…] wat mij opvalt is dat datagedreven beleid of datagedreven inspectie et cetera, het wordt heel erg gekopieerd vanuit het bedrijfsleven. Van we zetten een aantal KPI's op, dan gaan we dat besturen, et cetera. Maar binnen de overheid is dat een stuk complexer dan

		binnen een bedrijf dat stuurt op financieel gewin (PO E)
Not acknowledging social complexity	Expert mentions that private sector innovation models don’t acknowledge governmental complexity.	“Die vertaalslag kun je eigenlijk helemaal niet zo maken. Dus daarin zit ook een stuk dat is niet eens de expertise dat is gewoon … dat zijn vertaalvragen eigenlijk omdat de complexiteit niet onderkend wordt in het bestuurlijke versus het bedrijfsmatige denken” (PO E).
Prioritizing outcome rather than process	Expert mentions that outcome is being prioritized rather than the process towards it.	“[…] ik had laatst met een andere toezichthouder [redacted] en die had een keer aanbevelingen gegeven over de doorontwikkeling van een bepaald algoritme bij een uitvoeringsorganisatie in Nederland. Die zei van ‘nou ja het zit opzicht goed in elkaar maar de sensitiviteit en de robuustheid van je model daar moet echt iets aan gedaan worden’. Maar ze hadden niet gezegd van joh je moet stoppen met dat algoritme, ze hadden gewoon gezegd dit is echt iets waar je wat aan zou moeten doen. Die analisten die begrepen dat meteen en die wilden daar ook iets aan doen maar die managers die snapten er helemaal niks van. Die hadden geen flauw idee wat sensitiviteit en robuustheid is. En dachten van ja uiteindelijk staat het stoplicht op groen dus we gaan gewoon verder. En dat kreeg dus niet de prioriteit om aangepakt te worden door

die analisten ook al wisten zij wel dat
het belangrijk was” (PINO B).