The Adoption of Artificial Intelligence In the Dutch Public Sector: How is the adoption of artificial intelligence affecting bureaucratic discretion?

Frederiks, Hannah

THE ADOPTION OF ARTIFICIAL INTELLIGENCE IN THE DUTCH PUBLIC SECTOR

Author: Hannah R. Frederiks

Supervisor: Dr. Amandine Lerusse

MSc Public Administration: International and European Governance

Faculty of Governance and Global Affairs Leiden University

Word Count: 14,869

Date: January 9th, 2022

Table Of Contents

Acknowledgements

The completion of this thesis could not have been possible without the support and guidance of my supervisor, Dr. Amandine Lerusse. I would also like to thank my parents and brother for always encouraging and championing me.

TABLE OF ABBREVIATIONS

ABBREVIATION	DESCRIPTION
AGI	Artificial General Intelligence
AI	Artificial Intelligence
ANI	Artificial Narrow Intelligence
AS	Actuarial Science
ASI	Artificial Super Intelligence
IBM	International Business Machines Corporation
ICT	Information Communication Technology
IT	Information Technology
ML	Machine Learning
NGO	Non-Governmental Organizations
PA	Public Administration
PPP	Public, Private Partnership
SSL	Semi-Supervised Learning
R&D	Research and Development

SECTION ONE | INTRODUCTION

Perhaps unbeknownst to us, ‘artificial intelligence’ ('AI') has permeated our society, transforming our day-to-day lives. We encounter AI technology every day- using facial recognition to unlock our phones, outsourcing questions to digital assistants like Apple’s Siri, and even parking our vehicles. AI technologies are creating exciting opportunities across all industries, but the public sector has yet to fully exploit AI's power and potential.

In the public sector, AI can improve decision-making, save government costs, and increase public worker’s access to resources and information (Dhasarathy et al., 2020). In essence, the adoption of AI is transforming administrative decision-making by delegating more tasks to AI systems than to humans (Bullock, 2019). The intersection of decision-making and AI is called algorithmic decision-making, which aims to either partially or fully substitute human analysis or determine a course of action (Busuioc, 2020). Consequently, the transition from normative administrative decision- making to algorithmic decision-making can affect public workers' discretionary power, the scope of work, restructure the workforce, and lastly, create or eliminate jobs (ibid., 2020). As for the citizen, algorithmic decision-making and AI systems offer greater consistency and assurance that citizens are assessed against an equal evaluative yardstick (ibid., 2020). However, the detection of algorithmic bias has implicated the uptake of AI in the public sector because it rejects AI as a neutral expertise (Misuraca and van Noordt, 2020).

It’s not the first time the public sector has seen these changes; the emergence of digital technology gave rise to less face-to-face interaction between public workers and citizens; instead, ‘information communication technology’ (‘ICT’) forged the connection between parties (Bovens and Zouridis, 2002). In such a setting, the role of public workers became increasingly more obsolete, as computers could more efficiently and effectively carry out their tasks (ibid., 2002). Boven and Zouridis (2002) seminal work explored the digital transformation of public administration through three transitory periods: ‘from street-level bureaucracy to screen-level bureaucracy to system-level bureaucracy.’ Central to their discussion was how the discretionary space in public organizations has been affected by digitization (van Eck et al. 2018).

Their work builds a necessary foundation that allows this study to further examine the relationship between AI and bureaucratic discretion within the Dutch public sector.

Today, there is an increasing tendency for an AI system to coordinate the dispensation of benefits, application entry, and issuing permits rather than a physical public worker (Bovens and Zouridis, 2002). The growing success of AI processing repetitive and routine tasks has prompted the engagement of AI at more mid to top-level management in executive public agencies (Zouridis et al. 2019). In these cases, the public sector is looking at incorporating self-learning AI systems. Self-learning AI systems assume that the decision systems no longer collaborate with data analysts; instead, they have complete autonomy and discretion in adjusting the decision-making algorithms (ibid., 2019).

The development of artificial intelligence has changed the nature of decision-making in bureaucracies, subsequently transforming the integral structure of public administrations and how they operate. AI takes on a similar role as ICT applications, aiming to improve and replace human discretion in bureaucracies (Bovens and Zouridis 2002; Fountain 2001; Busch and Henriksen 2018; Young et al. 2019; Zouridis et al., 2020). What remains unclear, is how bureaucracies are engaging with AI and what level of 'autonomous control' (National Research Council, 1999), is being allotted to them. In 2018, Busch and Henriksen aimed to clarify this by consulting various empirical studies and listing a typology of technologies ranging from the telephone to automated systems and arranged them by description, context (executive public agencies), and usage. The authors selected cases explicitly referenced in their sample of scholarly articles, thus omitting several real-time examples of ICT or AI directly replacing human discretion.

Since the publication of Busch and Henriksen (2018) and the accelerated growth of AI, scholars are presenting frameworks to assess the impact of artificial intelligence on bureaucracies (Bullock 2019; Young et al. 2019; Bullock and Kim, 2020; De Boer and Raaphorst, 2021). These theoretical frameworks are instrumental in foreshadowing how discretionary spaces in bureaucracies will evolve and help guide national AI strategies and regulations on artificial intelligence (i.e., EU Artificial Intelligence Act.). Despite their contributions, the extent to which AI replaces human discretion in bureaucracies is seldom explicitly discussed as they address more narrow research puzzles. In the case of De Boer and Raaphorst (2021), their research puzzle addresses how automation affects street-level bureaucrats' style of enforcement. Young et al. (2020) construct two theoretical frameworks: one, presented to public managers to offer guidance in implementing AI applications, and second, how artificial and human discretion compares to improving governance capacity. Bullock (2019) broadly examines how bureaucracy and governance are vulnerable to change in light of AI's impact on discretion. Lastly, Bullock and Kim (2020) propose a condition where AI systems become fully functioning bureaucrats ('artificial bureaucrats') and discuss its potential consequences on multi-agent bureaucratic systems. Aside from De Boer and Raaphorst (2021), none of the aforementioned scholars centre their research on a specific country; instead, they pool cases consistent with their discussions and findings. Chen and Salem's (2021) systematic literature review based on 26 articles and research agenda on the 'implications of the use of artificial intelligence in public governance', stipulated the necessity for more country-specific studies to broaden the field of comparative research (p. 15).

To uncover the extent to which AI replaces human discretion in bureaucracies, this study elects to examine the Netherlands and determines the extent of AI usage by the level of analysis (micro, meso, macro) in government. This study stands apart, as it will combine theory, current cases of AI, and context, to map how AI is evolving within the Dutch public sector. Examining the following research question:

How is the adoption of artificial intelligence affecting bureaucratic discretion?

This research question aims to build on previous scholarship about AI's impact on discretion and government practices (Boven and Zouridis, 2002, Busch and Henriksen, 2018; Bullock, 2019; Fountain, 2001; Young et al., 2019; 2020; Chen and Salem, 2021).

Unlike the United States, automated processes were not standardized until 1999, and the optimization of artificial processes was only recently introduced in the Dutch public sector (Rijksoverheid, 2020). The Netherlands is a key candidate for a wide range of AI adoption in the public sector. According to the Dutch government's '2019 strategic action report for AI,' the Netherlands has a suitable profile to coalesce AI in the public sector given its "high-quality connectivity, strong foundation for public, private partnership (PPP) and world-class research" (Netherlands, 2019, p. 7). Despite the depicted profile, the Dutch government's approach to AI has been soft; the current algorithms complete simple tasks and exclude autonomous and unsupervised algorithms (Netherlands Court of Audit, 2021). The lack of immersive AI systems in the public sector stems from the vacillation of introducing new technology, transparency issues, and staff's lack of expertise or familiarity with using AI. Most significantly, the public's concern about digital rights and data protection is restricting its entry. To this end, the audit highlighted that the use of AI algorithms has been limited to "operational management processes or provide services, such as the automated sending of letters and the initial selection of benefit applications" (Netherlands Court of Audit, 2021). This study promulgates whether it is appropriate to assume that sending letters and selecting benefit applications is equivalent and bona fide low discretion task or 'simple tasks' (ibid., 2021).

To determine these classifications, this study builds on the framework of Young et al. (2019) to tabulate tasks by the degree of discretion between low and high within three groups of cases. The first group are cases that utilize AI to identify social security fraud; the second group are cases that use AI for predictive policing and risk modelling; and the third, a close look at the city of Amsterdam identified as an AI hot spot for diverse AI applications. These cases will be further examined and will seek to test the proposed assumptions drawn forward by this study. Once the cases are classified by the degree of discretion, we determine the level of AI related reform in the Netherlands by building off Young's et al. (2019) 'matrix of task analysis by level of analysis and degree of discretion'.

This thesis is divided into six sections. The first section, the introduction, highlights the research question and identifies the significance of the research study. Section two expounds on all critical terms that are germane to the study. Section three, the literature review, examines existing scholarship relevant to the relationship of AI on bureaucracies and discretion. Section four presents the theoretical and conceptual framework aimed at uncovering the assumptions drawn forward. Section five provides an analysis of the findings within the selected AI cases in the Netherlands. Lastly, section six presents the conclusions of the study.

SECTION TWO | DEFINITIONS

2.1 Discretion

Discretion is a term that takes on a variety of meanings and applications. Every day public workers are required to make decisions that either affects the citizen or the management and processes of the organization. In the context of public administration, discretion can be assumed as the autonomy of a public worker to make decisions that are supported by the rule of law but not constitutionally enshrined (Cooper, 2000, p.300). Whereas, in the discipline of administrative law, administrators exercise discretion when navigating grey areas of the law that inadequately detail decision protocols (Otenyo, 2006, p. 180). Both interpretations of discretion pull at a common thread: public workers may need to use their judgment to solve problems in contexts of ambiguous situations. The extent a public worker has the freedom to exercise their judgment depends on the "specific context and the factors that give rise to [the] freedom in that context (Lipsky, 2010, p.2). This study proceeds by using Lipsky's (2010) definition of discretion, as cited above.

2.2 Artificial Intelligence

Artificial Intelligence is becoming increasingly more ubiquitous and transforming our everyday lives (Neti, 2016). Valle-Cruz et al. (2019) suggests that AI can't have a set definition because the rapid advancement of technology determines the evolutionary process of AI. There is a degree of truism to Valle-Cruz's statement, considering we have yet to unlock the full potential of AI. Nonetheless, for the purpose of this paper, it is critical we develop a comprehensive understanding of AI:

> *'To proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.' (McCarthy, 1955)*

McCarthy's (1955) theoretical conception of AI is relatively analogous to our current understanding of AI. McCarthy explains that AI is an 'intelligent machine' developed through a crossdisciplinary collaboration of engineering and science. The intelligent machine operates like a computer, but "AI does not have to confine itself to methods that are biologically observable" (McCarthy 1955, p.2). According to McCarthy, there are two strands of AI research: first, 'biological,' which implies a computer's efforts to emulate human cognition and physiology, given that AI's conviction is that 'humans are intelligent.' Second, 'phenomenal,' which assumes that that AI is problem-solving orientated. AI harnesses' facts and issues "that the world presents to the accomplishments of objectives" (McCarthy, 1955, p. 12). In a similar vein to McCarthy (1955), Poole et al. (2010) interpret AI as the study of 'intelligent computational agents.' The authors go one step further than McCarthy and examine the denotation of 'artificial' and 'intelligence' through the concept of an 'agent' (Poole et al. 2010). They identify an agent broadly as: sentient life-forms, animate objects and institutions. The authors are primarily concerned with how an 'agent' act in varying circumstances and environments. They contend that 'agents,' unlike 'computational agents' (Poole et al. 2010), can only observe and act in an environment, within a fixed period. Conversely, 'computational agents' act and make decisions by processing information that has been inputted or organically computed in a computer's 'hardware'. Poole et al. (2010) assert that AI is an experimental machine that utilizes intelligent behavior in order to solve problems and perform actions, previously perceived as only 'theoretical possibilities.' As we begin to learn more about artificial intelligence scholarship has identified three types of AI. To simply our understanding, AI scholarship has differentiated three types of AI: artificial narrow intelligence (ANI), artificial general intelligence (AGI) and artificial super intelligence (ASI). Each type of AI has a different set of goals and characteristics that will be outlined below.

Artificial Narrow Intelligence, ANI also referred to as weak AI- is a type of artificial intelligence that outperforms humans when the task is limited in scope and very specific. In the case of ANI, the outcome of results or knowledge gained does not transfer to other domains or tasks. Currently, all applications and cases of AI use artificial narrow intelligence. Some practical examples include digital voice assistants, chatbots, autonomous driving and predictive analytics (Marr, 2021).

Artificial General Intelligence, AGI, also called strong AI, is a type of artificial intelligence that reflects human cognition, as it can self-learn and reason with its operating environment. Subsequently, it can store previously gathered knowledge and apply it to different contexts and settings, much like a human. There are no practical examples of artificial general intelligence, and it strictly remains a theorized possibility for AI development (Marr, 2021).

Artificial Super Intelligence, ASI also called super AI, is a type of artificial intelligence that exceeds human cognition in all capacities and capabilities (Bostrom, 2016; Tegmark, 2017; Russel, 2019). ASI would have the power to resolve complex problems and issues beyond mathematical equations and exhibits an ability to reason with consciousness or emotion (Marr, 2020).

The emerging variations- ANI, AGI and ASI- illustrate artificial intelligence's growing development and sophistication. Consequently, a subset of artificial intelligence that has received wide attention recently is 'machine learning' ('ML') (Gavrilova, 2020). ML is characterized as a search problem, and its problem-solving skills become more refined with experience (Mitchell, 1997). And to complete the comprehensive overview of artificial systems, deep learning is the subset of ML that uses multi-layer neural networks that uncover hidden patterns from large amounts of data, e.g., number plate identification (Costa, 2019). Many of the AI cases featured in this study will draw on applications that utilize machine learning tools. Notably, unlike AI, ML does not imitate human behavior and cognition; instead, it seeks to learn and utilize data "without being programmed explicitly" (Javapoint, 2022). Examples of machine learning usage and application include but are not limited to probabilistic inferences, speech recognition systems, video surveillance, phishing malware detection etc. (Mitchell, 1997; Techlabs, 2021).

There are four common ML approaches: I) supervised learning, II) unsupervised learning, III) semi-supervised and IV) reinforcement learning (Edwards, 2018). Each approach aims to parse data by using an algorithm, consequently the program can generate predictions or identify observations (ibid., 2018).

Supervised learning is characterized as a bottom-up approach because the data does not have a formal structure and is information-finding. The main objective of the artificial system that operates by supervised learning is to map “between the input and the output and predict the output of the system given new inputs" (Liu et al., 2012, p. 2). The artificial system uses data that is labelled in order to train the algorithm to yield specific outcomes (IMB, 2021). Akinsola (2017) chronicles the processing sequence of supervised learning: first, the algorithm is provided with a sample dataset (input) from a large sample size; second, the algorithm is taught to classify and group data by established parameters; third, the trained algorithm can apply different input data to identify patterns or predict relevant outcomes. Lastly, the data is withdrawn and applied. According to Wu et al., (2020) the reliability of results is contingent on the quality of labelled data because that determines the validity and reliability of the data output.

Unsupervised learning, more commonly known as self-learning, is identified as a top-down approach because it aims to identify hidden structures within the data (Testolin et al., 2016; Wu et al., 2020). Unsupervised learning consists of "clustering, artificial neural networks and dimensionality reduction" (Wu et al., 2021, p. 4). The methods employed are generative because unsupervised learning operationalizes data to allot them in specific categories and dimensions without human interference (ibid., 2021, p.4). This learning model does not make use of labelled data and must unilaterally discern relevant patterns and relationships between the data (ibid., 2021). Interestingly, unsupervised learning cannot determine casual relations within the data, only extrapolate relations (ibid., 2020, p. 4). Extrapolating data is understood as approximating the relationship of data "beyond the original observation range" (ibid., 2020, p. 4). The consequence of examining data outside of the observation range poses potential validity concerns and erroneous conclusions. To control this, running human interference in the data analysis stage can mitigate adverse or unwanted effects; however, the learning model would no longer be classified as unsupervised.

Semi-supervised learning ('SSL') falls between supervised and unsupervised learning. In order to perform predictions, SLL utilizes labelled and unlabeled data. An SSL approach is elected when there is a paucity of labelled data or when gathering data is a long-drawn out process (Chapelle, 2019) SLL carries out predictions through either transductive or inductive learning. Transduction builds a predictive model using a labelled training set and an unlabeled test set to perform projections. Transductive learning cannot make predictions with data that wasn't originally in the training set. Conversely, inductive learning uses its inputted labelled training set to build a predictive model. Consequently, the inductive predictive model can use its pre-existing training set to label for unlabeled data (ibid., 2019).

Reinforcement learning can be understood as a feedback mechanism between an agent (AI system) and the environment. The feedback mechanism functions as a trial-and-error system, whereby the environment shares information with the agent to achieve complicated goals. The algorithm built into the agent refines its skills and functions through a reward system. A positive or negative signal distinguishes the algorithm's feedback on the output data. If the signal is positive, the agent's skill is encouraged to be continued or repeated, whereas a negative signal is discouraged. Unlike supervised learning, reinforcement learning cannot assume if its action is correct, until the action has been executed. For example, the reinforcement learning chatbot is increasingly used in the public sector. Government chatbots provide quick access to public data, respond to frequent FAQs and can operate 24/7 (Streebo, 2021). Bots are a boon for government services, as they can accomplish routine tasks at "significantly lower costs" and "allow staff to focus on other complex initiatives" (Streebo, 2021). The Chatbot's response depends on the data that it's been fed, its established parameters, and what it has been programmed to accomplish. The chatbot becomes more sophisticated by the frequency of users and by updating its system design.

Penultimately, the **black box** is an AI system typically seen in machine learning algorithms, i.e., support vector machines and deep neural networks (Bathaee, 2018, p.892). Unlike other AI systems, a black box only makes the input and output of data perceptible; it excludes the user from discovering how decisions are made throughout. The black box is regarded as uninterpretable because the algorithm draws connections between variables that are not obvious or observable by humans (Rudin et al., 2019). The lack of transparency on the AI's thought process is considered problematic because the user fails to understand how the algorithm makes its decisions. Moreover, the user cannot surmise the programmer's intentions and whether their bias bears any influence on the conclusions generated (Bathaee, 2018, p. 893). Conversely, the advantages of using black-box models are that no one person is responsible for erroneous predictions or conclusions that the program generates. In the context of the public sector, the utility of a black box model juxtaposes the highly regarded intrinsic values of accountability and transparency. Governments will draw on evidence-based decision-making and include various stakeholders to account for multiple perspectives to maintain public trust in the public sector. Black box models may be operationalized to reflect different perspectives and specific values, but it cannot explicate how it derives to its conclusions. Nevertheless, the adoption of black-box models in the public sector has been growing, albeit that black box system caters towards less transparency. Adjusting the AI model to be more transparent is considered a 'design trade-off' (Bathaee, 2018); the model would have to decrease in size and apply narrow AI. Such structural changes would potentially alter the performance and intentional function of the AI model (Bathaee, 2018). Imposing measures to regulate transparency in AI systems ultimately deter technological innovation and reduces the algorithm's potential adroitness. Thus, governments expecting advanced AI systems, yet intensifying guidelines on transparency is incongruent for AI development (ibid., 2018). As a result, governments may feel more partial towards white-box models.

Lastly, the differentiation between white and black box models is unambiguously black and white. **White box** models assume the concept of interpretability, whereby the behavioural process of the algorithm and how it arises to its conclusions is apparent (Hulstaert, 2019). The advantages of using the white box model are that it's "easier to explain and interpret" and has "simpler computation" (Hulstaert, 2019). At the same time, the results have less accuracy, and the model is not "capable of modelling the inherent complexity of the data set" (Hulstaert, 2019).

2.3 A Diagrammatic Representation of AI and ML Systems

To summarize, figures one and two below taken from Dechesne et al. (2019) and serve as a visual representation in addition to the discussion above, describing the internal processes of AI and ML systems at its most basic form. The purpose of the extensive overview of all AI branches is necessary to understand the method of operationalization revealed later in chapter five.

Figure I. Artificial Intelligence

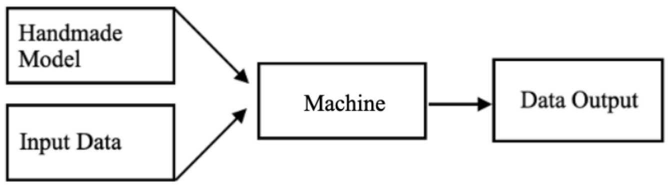
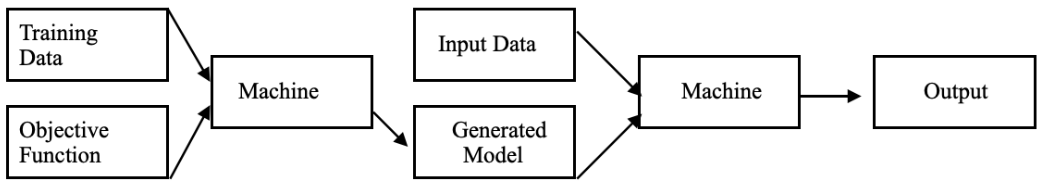


Figure II. Machine Learning



Note. From AI & Ethics at the Police, Dechesne et al., 2019, p. 3. Copyright 2019 by Leiden University and TU Delft.

SECTION THREE | LITERATURE REVIEW

The relationship between ICT and bureaucracy has received wide attention in the field of public administration (Bovens & Zouridis 2002; Cordella & Tempini, 2015; Wenger & Wilkins, 2009; Zouridis et al., 2020; Selten & Meijer, 2021). PA scholars, to great depth, have explored this relationship and its varying implications on bureaucratic accountability, transparency, and discretion (Barth & Arnold 1999; Bullock, 2019; Bovens & Zouridis 2002; Cummings, 2006; Sandor, 2012; Tummers & Bekker 2014; Calo & Citron, 2017; Busch & Henriksen 2018; Lennox & Payne, 2020; Busuioc, 2017, 2020). These studies provide invaluable insights and groundwork to predict and analyze the budding relationship between artificial intelligence and bureaucracies. However, the density of literature has created confusion surrounding tech phenomena within governmental bureaucracies, to an extent where scholars falsely interchange between ICT, IT, digitization, computerization, technology, automation, artificial intelligence, autonomous intelligent agents, and assign one of these terms as a hypernym. Many of these terms are conceptually similar but yield different nuances (De Boer and Raaphorst, 2021). For example, Gaynor (2020) suggests that distinguishing between AI and automation is becoming increasingly blurred. This is because automation is often confused with AI, yet automation excludes humans entirely and purely works on repetitive tasks based on set instructions and rules (De Boer and Raaphorst, 2021) (Gaynor, 2020). Recent evidence from International Business Machines Corporation ('IBM') contradicts this and cites four automation categories, of which the last is classified as AI automation (IBM Cloud Education, 2021). As such, this study identifies artificial intelligence and automation as interchangeable terms, given that IBM identifies AI as the most complex level of automation. By doing so, we avoid excluding useful literature related to the interaction of artificial intelligence or automation in governmental bureaucracies.

The literature review is split into two parts: examining the relationship of AI on bureaucracies and discretion. This allows us to better our understanding of how the adoption of artificial intelligence is changing the structure of bureaucracies and the nature of discretion.

3.1 ICT to AI- The Changing Structure of Bureaucracies

Arguably, the overarching purpose of a bureaucracy is to either regulate, administer, or implement decisions (Raaphorst, 2017). Scholars remain split on whether ICT has advanced the decision-making process, quality of service provisions, and work practice. Lipsky (2010) purports that machines cannot replace street-level bureaucrats because their work demands a level of discretion and judgment, that machines cannot replicate (p. 161). The shortcomings of Lipsky's analysis are that he fails to acknowledge that machines (automated systems) can act as decision aid tools and support bureaucrat's responsibilities and duties. Instead, Lipsky draws on cases where the dispensation of discretion was entrenched in the automated system and would directly affect the quality of people's lives, i.e., unemployment assessments and referrals (p.224).

A number of authors indicate that the spectrum of automation in decision-making tasks varies between partial or full automation (Bullock, 2019; De Boer, 2021; Young et al., 2019). The level of automation determines how much work is executed by a person or computer. Therefore, the more tasks bureaucracies assign to automated systems result in a change of public workers' competencies and functions, particularly their discretionary oversight.

The concept of transference of tasks from human to computer was significantly explored by Boven and Zouridis (2002). Boven and Zouridis (2002) analysis on the effect of ICT on the organizational structure of public agencies, determined that ICT would supplant administrative tasks, typically characterized as repetitive or routine. These administrative tasks were commonly carried out by 'street-level bureaucracies' such as: "tax departments, social security agencies and agencies that collect fines" (Boven and Zouridis, 2002). The advent of computers and ICT transformed street-level bureaucracies into 'screen level bureaucracies,' whereby public workers became less likely to interact with citizens face-to-face, instead carried out their tasks through a technological interface. The transition between street to screen level bureaucracies has altered the level of discretion exercised by public workers, as their skills emerge as less necessary. The authors predict that with increased advancement of ICT, the likelihood of communication networks and information systems becoming automated is high and inevitably excludes public workers entirely. The complete exclusion of public workers is referred to as a 'system-level bureaucracy,' whereby most tasks and decisions are automated (Boven and Zouridis, 2002). In sum, Boven and Zouridis (2002) argue that bureaucratic reform is driven by the incorporation of ICT and subsequently, changes the role and working practices of public workers. The inherent drawback of their analysis is they consider all new technologies to absorb a degree of discretion from human operators, rather than assuming some technologies playing a non-discretionary role and operating as support tools. Factoring this in, Boven and Zouridis's assessment of the paradigm shift resulting from ICT is specious because not all ICT applications pose a risk for the legitimacy and working functions at the street level (Garson, 2007, p. 119).

In contrast, to Boven and Zouridis (2002) scholars such as Kreamer and King (2006), Keld Pedersen (2018) and Noris and Reddick (2013), express that ICT has not changed the structure of bureaucracies, but rather enforced existing 'administrative and political arrangements' (Kreamer and King, 2006). The authors develop this assertion by drawing on the role of managerial executives and their intent on achieving organizational goals and objectives. They identify ICT as an opportunity to; efficiently pool information and data together, exercise greater judgement in decision-making, and exhibit legitimacy by utilizing a variety of resources (Kreamer and King, 2006, p. 6). As cited by Norris and Reddick (2013), these changes effectively build on previous structures and processes in the organization but do not indicate a transformation in the nature of the work itself (Kreamer and King, 2006). The authors contend that government operations and methods are only enhanced by ICT but are not "bold and innovative moves to reform public agencies" (Kreamer and King, 2006, p.9). Arguably, this assertion neglects to consider that bureaucratic reform cannot happen all at once but rather in stages and inter-departmental.

Similarly, applications of AI systems are cautiously incorporated at different stages and to specific departments within the public sector. This is because AI developers and government leadership wants to monitor how AI decision-making compares to human bureaucrats in improving organizational goals and outcomes. Most obvious is AI's ability to exceed a human's pace and efficiency in completing specific tasks. However, in the most transformative scenario, AI systems would be fully autonomous, cognisant and exercise discretion as situations evolve (Bath and Arnold, 1999, p. 335). Bath and Arnold (1999) stipulate that AI systems should fulfill three objectives in the public sector context in the scenario mentioned above. First, the AI system must be receptive and replicate diverging values and goals exhibited by relevant stakeholders. So instead of eliminating AI bias, the algorithms should reveal different nuances and 'opinions' that commonly appear in standard decision-making procedures. Second, the AI system should account for changes in the socio-economic and political landscape because it implicates how problems are addressed and decisions made. Lastly, the AI system should be independent and *feel* accountable to provide information when it is demanded. The authors effectively describe an ASI system, which is hardly a plausible reality in context to the current operative capacity of these technologies.

However, Bath and Arnold (1999) are not the only authors to tantalize the idea of an artificial system that assumes the role and tasks of public workers. Bullock and Kim's (2020) study aims to fill the gap on how artificial and human bureaucrats would band together to achieve the outcomes and goals of their governing institution. Bullock and Kim (2020) conceive the concept of an *individual artificial bureaucrats*, which depicts AI as an autonomous agent with capabilities to exercise discretion and make decisions within a multiagent system (p.31). The authors find that there is insufficient differentiation between the types of AI, i.e., supervised, and unsupervised learning- and how that implicates the decision-making protocol and the co-working between human and artificial agents. They stress that the addition or subtraction of human involvement in AI applications can potentially alter organizations' internal processes and structures, which might negatively affect the credibility of service and information output (ibid., 2020, p. 31). To prevent a legitimacy or democracy deficit in organizations, the authors stipulate a necessity for strict rules and regulations to sustain a symbiotic relationship between artificial and human bureaucrats that can mutually serve a common objective. The deficit to this solution, omitted by Bullock and Kim (2020), is that strict regulations can curb R&D and innovation and may ultimately derail the co-working arrangements between AI and human agents.

Nevertheless, both Bath and Arnold's (1999) and Bullock and Kim's (2020) foreshadowing of how the bureaucracies will be affected by AI has faceted a space for PA scholars to develop theoretical frameworks to predict how AI decision-making compares to human bureaucrats in improving organizational goals and outcomes (Busch and Henriksen 2018; Young et al. 2019; Bullock 2019; Saxena et al., 2021).

3.2 Artificial Intelligence and Discretion

Depending on the field of study, the concept of discretion can have varying definitions and interpretations. In the field of public administration, several scholars contend that discretion assumes a bureaucrat's freedom to enforce or make decisions within set parameters in a particular situation (Hupe & Hill, 2007; Lipsky, 2010; Tummers & Bekkers, 2014). Today, cases, where bureaucrats exercise discretion directly with citizens, are increasingly less likely, given that ICT and automated systems are replacing them (Boven and Zouridis 2002; Buffat, 2015; Boven et al. 2020). Whether the full or partial replacement of bureaucrats with technology is considered a positive or negative consequence is arguably moot, given the trajectory of digitization in the public sector.

One of the earlier works examining the implications of automation on bureaucratic discretion was by Aurelien Buffat. Buffat (2015) indicates that there are two theses about the effects of automated systems on bureaucratic discretion. The first is the curtailment thesis. Buffat (2015) discussed the 'curtailing effects' of automation on front-line policy discretion, effectively hindering and changing the scope of street-level bureaucrats' jobs, to an extent they may become obsolete or disappear. The opposing thesis was coined as the 'enablement effect': suggests that automation can bolster the competencies of bureaucrats and facilitate the distribution of information among citizens. The latter is consistent with the current trajectory of AI uptake in the public sector, wherein AI systems are employed as decision-aid tools to improve human decision-making (Busuioc, 2020, Selten & Meijer, 2021). This is illustrated by Wenger and Wilkins's (2008) empirical study that examined how the emergence of automation increased the opportunity for women to obtain unemployment insurance (UI). In this case, women who entered a UI agency faced more discrimination than those submitting their requests over the phone. Wenger and Wilkins's (2008) study highlights that in this case, enabling automation improves human decision-making by addressing the implicit bias of agents. This study signals one critical point missing in Buffat's discussion: how enabling automation increases agents' engagement and accountability in their work. In addition to the analysis, the Buffat (2015) identified several gaps related to the intersection of automation and the public sector. Factors such as the variety of technologies and their permissibility, the application and utility of the technology, and the contextual factors that inform technology uptake are considered as gaps requiring further research. The potential guardrails such as privacy compliance and regulation for future technologies imposed by governments are unaddressed in Buffat's assessment. These are important considerations because they affect the uptake of prospective automated applications and the decision-making systems within these institutions.

The relationship between AI and discretion has materialized to an extent where AI systems can make decisions with and without the supervision of a public manager or bureaucrats. As a result, researchers have conceived terms to describe an automated system exercising discretion. Busch and Henriksen (2018) defined "the use of computerized routines and analyses to influence or replace human judgements" (ibid et al., 2018, p. 4) as *'digital discretion.'* Young et al. (2019) took on a similar interpretation, but for cases where "artificial intelligence is used to augment or automate the exercise of administrative discretion" (ibid et al., 2019, p. 303) and defined this as *'artificial discretion'*. Bullock and Kim (2020) take one step further and consider the actualization of *'artificial bureaucrats'* understood as "AI agents [using] artificial discretion to make and execute decisions" (ibid et al., 2019, p. 30). The overarching consensus among these concepts is the propensity for less human judgment in decision-making in the context of bureaucracies.

Public managers determine the extent of artificial intelligence co-working with bureaucrats. Managers must decide which AI systems can or should replace work tasks and assignments. At the moment, most bureaucracies have established implicit boundaries, detailing that the adoption of AI is solely for achieving efficiency and truncating redundant tasks (Bullock and Kim, 2020). Governments have more resistance in adopting AI for tasks that determine a citizen's quality of life or are responsible for making high-stakes decisions.

SECTION FOUR | THEORETICAL FRAMEWORK

To ascertain how AI adoption affects bureaucratic discretion, this study builds on Young et al.'s (2019) theoretical framework that introduces a guideline for public managers to determine where and what [which] tasks can be replaced or supported by artificial intelligence. The 'where' referred to as the 'level of analysis' underscores the contextual factors that impact how tasks are completed. The 'what' refers to the tasks most appropriate for AI use, which is discerned by the 'degree of discretion' required. Young et al. (2019) sets the grounds for examining discretion affixed to AI in the context of administrative decision-making, which is directly in conjunction with the premise of this study. The rationale for drawing on Young et al.'s (2019) framework is on the basis that they operationalize the concept of discretion qualitatively and explicitly and examine how contextual factors might inform bureaucratic discretion.

Within the role of administrative decision-making, tasks vary both in scope and by the degree of discretion- ranging from low, medium, and high. Low discretion tasks are characterized as tedious and redundant, such as gathering and sorting data (ibid., 2019; Busch and Henriksen, 2018). Typically, tasks that bear no significant outcomes are ideal for automation, as it saves administrative costs, and human agents can be engaged with jobs that require more prudence (Young et al., 2019). An example of a low discretion task in the public sector is tax reporting. Tax reporting is considered a repetitive task because it relies on numerical data (i.e., income, expenses etc.) and abides by formalized rules, wherein the decision assessment across the board is uniform (Busch and Henriksen, 2018, p. 20). Medium discretion tasks are understood as cases that exhibit an inconsistent judgment among human agents, this is because the tasks consist of insufficient data, or are poorly structured (Young, 2019; Busch and Henriksen, 2018). High discretion tasks are coupled in a context of significant uncertainty/ambiguity and coupled by little data or rich data that inhibits relationship discovery (ibid., 2019). For example, Young et al. (2019) suggest that weather scientists exercise high discretion tasks because weather systems have dense multidimensional data sets with significant accuracy. Yet, forecasting is problematic given that weather systems are highly volatile. To sum up, Young et al.'s (2019) classification of tasks by the degree of discretion aims to evaluate whether AI is suitable to replace human judgment. By doing so, the authors introduce the construct of 'artificial discretion' and specify its potential use by drawing on tasks ranging in discretion, outlined below in framework one.

Framework I. Potential Use of Artificial Discretion for Tasks by Degree of Discretion

Low Discretion	**Medium Discretion**	**High Discretion**
Automation	Decision-support tool, predictive analytics	New data generation, reduction of data complexity, relationship discovery

Source: (Young et al. 2019, p. 4)

Framework one will serve as a guiding tool to identify the task characteristics of each selected case in this study, which subsequently lists and designates their degree of discretion as to either: low, medium, or high. Once the degree of discretion is established, we must discern the contextual level of analysis for each chosen case.

The level of analysis reveals the contextual factors influencing the process to which tasks are completed (Young, 2019, p. 5). There are three levels. At the micro-level street-level bureaucrats deal with citizens directly and use their discretionary authority to apply policies to individual cases. The meso-level is about the bureaucracy's internal organization, specifically concerning the administrative working procedures with which policies are implemented. Lastly, the macro-level addresses' the context in which policies are designed and the preparations made for their implementation (Busch and Henriksen, 2018; Young, 2019). Young et al. (2019) note that discretion becomes more constrained and interdependent when more actors participate in the working or decision-making process.

Framework II: Matrix of Task Analysis by Level of Analysis and Degree of Discretion

Degree of Discretion Level of Analysis	Low Discretion	Medium Discretion	High Discretion
Micro-Level	Data entry, issuing licenses or permits	Placing children in foster care, sentencing/parole	Emergency response
Meso-Level	Facilities operations	Hiring processes, performance management	Goal setting, strategic planning
Macro-Level	Statutory obligations	Policy formulation	Crisis response and management

Source: (Young et al. 2019, p. 6)

4.1 Interlinkage Between Context (Level of Analysis) and Task Discretion

Examining the context of a task is relevant because it influences the outcome and how it is observed, thus dictating degree of discretion required. If we consider the concept of decision-making and deconstruct it to its primal definition, it is an action that evaluates a multitude of possible options. Understanding why certain decisions are made or addressed is explained best by the circumstances or conditions that led up to the decision. Referring to framework two, above, an emergency can occur at both the micro-level and macro-level. However, the actions at the macro level consist of coordinating between groups to ensure that resources and information is distributed accordingly- this is typically the responsibility in high-level government. Conversely, the task requires an immediate response at the micro-level, often in reaction to a situational issue that centers around a group of individuals.

4.2 Assumptions

The following discusses and justifies why examining the degree of discretion and level of analysis is instrumental for answering the research question. To assess the impact of AI adoption on bureaucratic discretion, this study puts forward two assumptions to narrow the focus of the research question.

Assumption A: We expect more AI adoption for tasks characterized by low discretion.

Assumption B: We expect more AI adoption at the micro level.

If we trace back to earlier theoretical models that examined how computers could replace humans in the context of public administration, their approach is also grounded in establishing the association between automation and discretion (Bainbridge, 1987; Johannessen, 1994; Sheridan, 1992). Sheridan (1992) developed a spectrum that distributed the level of automation on a ten-point scale. At the bottom of the spectrum, 1- assumes low levels of automation; thus, discretionary tasks and decisions are designated to humans. On the higher end of the spectrum- 10 assumes high automation levels, suggesting that computers have more discretion in making decisions than humans. Both Sheridan (1992) and Johannessen (1994) purported that, executive decisions with a high impact will remain the responsibility of human agents to illustrate that the decision-making process remains a human-centred approach.

In a similar vein, Boven et al.'s (2002) work examining how bureaucracies are transforming due to ICT also considered where the effects would be most significant- street or system level. The authors contended that front-line policy workers would be the first affected by automation and their jobs/tasks obsolete (ibid., 2002, p. 276). The system-level, which refers to the organization's decision-making body, is expected to see less discretionary curtailment for two reasons- first, decision-making bodies are responsible for designing automated systems and determining the threshold those systems can exercise discretion (Boven et al. 2018, p.6). Second, the source of data inputs and the exchange of data between ministries or departments for automated systems is decided by agents in decision-making roles (ibid., 2018, p. 6). It becomes clear that their perspective has not deviated from the 2002 to 2018 publication, but they highlight that "the ratio of automated decision making to people is changing," both at the street and system-level (ibid., 2018. p.16).

Bullock's (2019), approach to examining how the scope of bureaucratic discretion is affected by AI across bureaucracies is congruent with Boven et al. 2002 and 2018 work. Bullock applies two indicators, the level of uncertainty and complexity of tasks (high or low), in context to the level of analysis- street level and system level bureaucracies. The author promulgates that artificial intelligence is a candidate for tasks considered less complex and more routine, which seldom deviates from standard procedures. Also, tasks with greater interpretability and fewer unknowns (uncertainties) are considered prime for AI adoption.

These assumptions overlap closely with Young et al.'s (2019) differentiation between discretion across tasks, mentioned earlier, where less complex tasks are most suitable for AI because the outcome of the action is predictable.

By that same token, this study suspects that the inception of AI is a new procedure for decision-making in public administration. Therefore, it is most likely to emerge in procedures characteristic of low risk, low discretion, and simple task procedures. Once the organizational executive body identifies the procedures as successful or satisfactory, it is feasible to see AI implemented for tasks with high discretion, high impact, and high risk. The same ratiocination applies for the level of analysis, whereby AI is anticipated at the micro-level because the consequences of algorithmic decision-making per one case are less severe than for many. Moreover, adjusting how decisions are made at the micro-level is more malleable because the scope discretion for front-line workers is more. This study finds that the assumptions are consistent with the current AI transformation in the public sector, wherein chatbots and virtual assistants are employed as the first touchpoint between citizens and government (Council of Europe, 2021, p. 7).

SECTION FIVE | METHODOLOGICAL APPROACH

5.1 Case Selection

To test the assumptions mentioned above, a qualitative case study method is carried out to gather all enabled AI cases in the Dutch public sector, the degree of discretion exhibited by each case, and determining which level of government AI is being deployed. A qualitative approach is deemed most useful in this design because it emphasizes describing and interpreting the relationship rather than proving it. The criteria for selecting cases were based on three conditions; first, the case must be situated in the Netherlands; second, the case explicitly references the utility of AI; and third, AI was used in the context of bureaucratic decision-making. Discovering cases were achieved by researching the internet and reading journal articles that examined or alluded to AI applications in the context of public administration in the Netherlands (Bruxvoort and Keulen, 2021; Boven and Zouridis 2020; Meuwese, 2020; De Boer and Raaphorst; 2021). Furthermore, this study systematically searched for AI cases at every administrative level: local, regional, and national to avoid the unintentional omission of any cases. As a provision, web searches were also conducted in Dutch to prevent excluding original Dutch sources; additionally, not all government texts have been translated to English.

Incidentally, over the course of collecting AI cases, many sources mistook cases of actuarial science (AS) for artificial intelligence, when in fact, AS applies mathematics and statistical methods to make predictions. For example, the risk assessment tool- 'ProKid,' aimed at assessing the probability of prospective criminal conduct by children and young adults, was wrongfully assumed to operate on an AI algorithm (Ferris et al., 2021). The heterogeneity of interpretations and information regarding the type of AI for each case made the vetting procedure for case selection significantly challenging. To ensure the validity of the analysis presented in chapter 6, cases that produced an evaluation report were exclusively selected. The term ‘evaluation reports’ serves to encompass audit reports, inspection investigations, and in-depth assessments led by third-party organizations (i.e., TNO, academic institutions), internal oversight committees, NGOs (i.e., Amnesty International, Fair Trials), or a Dutch executive or legislative body. The purpose of only including cases with an evaluation report is because they provide a more comprehensive overview of each case, discuss the challenges and opportunities of the algorithm, and detail the properties of the algorithm and how it functions. Moreover, cases with evaluation reports can be considered most representative of AI applications in the Netherlands, as the reports are made publicly available.

To this end, the cases selected for this study have been arranged into three categories; first, cases tackling social security fraud; second, case examples of predictive policing/modelling risk; and third, AI cases localized in Amsterdam that are not interrelated to a specific function.

5.2 AI Cases Tackling Social Security Fraud

I. The System Risk Indication (SyRI) was an instrument developed in 2003 to challenge social security fraud in the Netherlands. Initially a pilot study developed by the national steering committee for intervention (LSI) and later, in 2014 practically implemented by the Ministry for Social Affairs and Employment, the SyRI project aimed to detect social security fraud by pooling data on citizens' profiles from multiple executive government agencies (Schets, 2019, p. 7). The listed agencies are police, the public prosecutor's office, the Dutch tax authorities, Employee Insurance Agency (UVW), the Ministry for Social Affairs and Employment (SZW), municipalities, the National Insurance Institute (NSI), police, and Immigration Authority (IND). In 2008, agencies were authorized to use SyRI if their investigation challenged social security fraud. SyRI operated as a black-box algorithm, whereby the data sources, risk model, and citizens' information history (employment, education, property registration, financial statements, and loans) were concealed and ciphered (ibid., 2019, p. 7; Vervloesem, 2020). The amalgamation of data from the agencies mentioned above allowed the SyRI algorithm to carry out predictive risk profiling to identify individuals suspected of fraud. A citizen or company was considered high risk of fraud if they matched the preset risk criteria. Only after SyRI's analysis were high-risk individuals investigated by a human administrator, and the identity of the profile was revealed. Subsequently, administrators reassessed the citizen's eligibility for benefits, and in cases of fraud, the prosecutor's office was notified to initiate legal action (Alfter et al. 2019). SyRI's approach to combatting social security fraud fell under scrutiny in 2017. A group of NGOs challenged SyRI's legality and considered the system's data acquisition and sharing of citizens' personal information a potential privacy breach. In 2020 the Dutch high court suspended the use of SyRI across all municipal governments- Capelle aan den Ijssel, Eindhoven, Haarlem, and Rotterdam (NOS, 2020), for having violated Article VIII of the ECHR "right to respect for private and family life, home and correspondence." (Council of Europe/European Court of Human Rights, 2021).

II. The second examined case is the Toeslagenaffaire (Dutch childcare benefits scandal). After suspected exploitation of the Dutch child benefits scheme, in 2013, the Dutch government introduced an anti-fraud committee for childcare benefits affairs, run by the Tax and Customs Administration (Belastingdienst). The committee challenged fraud within the child benefits system by using a self-learning AI algorithm that learnt to identify false claims. The algorithm was taught to distinguish between correct and incorrect applications using a series of indicators, for example, the algorithm flagged claimants made by citizens with dual nationality as high risk for committing fraud (Dutch Data Protection Authority, 2018, p. 14). Other indicators included: income, credit score, nationality, the distance between a person's house and their daycare center (ibid., 2018, p. 23). These indicators created the systemized model for risk, which the algorithm used to filter applications for fraud. Each indicator in the model was assessed on a metric from zero to one, where zero is the lowest risk score, and one is the highest. If the application scored highly on several indicators, the claim would receive a high-risk score and then be passed for manual verification (ibid., 2018). Employers only received the applications risk score during verification, consequently concurring with the algorithm's assessment. The implications of the AI systemized risk model were exposed in the 2018 investigation report highlighting a disproportionate influence dual nationalism had on an application's risk score. Thus, many benefit allowances were discontinued, denied, or had to repay debts/fines. As such, the AI system caused the ruination of 26,000 parents in the past decade and mistook them as fraudsters (Brenninkmeijer and ten Seldam, 2021). In 2021 the Parliamentary Interrogation Committee on Childcare Allowance published a damning report, 'unknown injustice', exposing the discriminatory system aimed at singling out potential benefit fraud (Rijksoverheid, 2021).

III. In 2018, the Municipality of Nissewaard implemented a fraud detection algorithm, engineered by Totta Data La. Prior to the implementation of the algorithm, government employees assessed the validity of social welfare claims in two ways; first, through periodical checks, whereby residents filed their social information annually, and second, through investigating reports or tips provided by residents indicating potential fraud (TNO, 2021, p. 11). This traditional approach was deemed neither cost-effective nor timesaving; thus, the Municipality of Nissewaard employed an AI 'supervised-learning' algorithm to detect potential welfare fraud by using various risk indicators (ibid., 2021, p. 13). The algorithm established the risk indicators by discovering patterns among claims that indicated an increased risk of fraud (ibid., 2021, p. 13). Subsequently, those indicators produced a risk model for claim comparison. As such, the algorithm gave each claim a risk score (ibid., 2021, p. 8). A claim with a very high-risk score would be sent for secondary screening by a supervisor. Other claims would be processed without scrutiny; however, the enforcement and issuance of fines were done manually. The Municipality of Nissewaard discontinued using the algorithm after TNO's (the Netherlands organization for applied scientific research) due to ethical considerations, the AI's poor audibility, and the algorithm's heterogenous assessments when one code was run multiple times (ibid., 2021, p. 21).

5.3 AI Case Examples of Predictive Policing and Risk Monitoring

I. The Crime Anticipation System (CAS), a predictive policing program used across different cities in the Netherlands: Amsterdam, Enschede, Hoorn, the Hague and Groningen Noord since 2017, aimed at identifying areas in cities with a high risk for crime (Oosterloo and van Schie, 2018, p. 4). CAS makes its predictions by data mining from three statistical sources: the Basisvoorziening Informatie (basic information provision, BVI), the Central Bureau of Statistics (CBS) and the

Municipal Administration (BAG) (ibid., 2018, p. 3). From each source, specific data is observed to curate fifteen specific indicators per zip code area, i.e., number of inhabitants, the average size of household, average age, number of foreigners (discontinued in 2017), property value, social benefits etc. Subsequently, CAS's artificial neural network predicts the location, time, date, and statistical probability of a crime occurring. Every two weeks, CAS graphically visualizes high-risk zones on city maps, using spatiotemporal data analysis, which assists police officers in deploying surveillance teams in those areas (ibid., 2018, p. 3). CAS is one of the few AI systems still operating in the Netherlands; out of 167 police branches, 110 use CAS.

II. Roermond Sensing Project, a predictive policing tool carried out by the Dutch police and used in the Municipality of Roermond since 2019. The Sensing Project uses real-time footage of cars driving near the luxury outlet center to collect data on the vehicle model, license plate, number of passengers, and the vehicle's travelling direction. Subsequently, the algorithm presents a risk score using the indicators above to determine a potentially relevant target at risk for committing petty crimes. The algorithm is an example of supervised learning, whereby the output is evaluated by manual verification, in this case, a policeman.

5.4 Case Examples of AI City Operations

I. The Automated Parking Control System deployed by the Municipality of Amsterdam aims at ensuring, that vehicle owners have a permit for parking or have paid for the short-term parking fees (Gemeente Amsterdam, 2021a). The control system is operated by a city-enforcement vehicle that drives around scanning license plates and simultaneously runs the plates through the National Parking Registry to authenticate if the car is permitted to park in that spot (ibid., 2021a). If the vehicle is illegally parked, the AI- Parking Control System flags the case for manual inspection and follow-up (ibid., 2021a).

II. In the city of Amsterdam, there has been a history of landlords and temporary tenants violating the city rental code. The purpose of the rental code is to avoid city overcrowding and manage visitor pressure. Violations include but are not limited to rental contracts that supersede thirty nights. Violations are recorded in the form of a report often provided by neighbors or concerned citizens, subsequently passed onto the Department of Surveillance and Enforcement for investigation (Gemeente Amsterdam, 2021b). However, given the high volume of reports in July 2020, the municipality of Amsterdam introduced a supervised deep learning algorithm aimed at expediting the process (ibid., 2021b). The algorithm calculates the likelihood of housing fraud on a reported address using data from the past five years on illegal housing cases (ibid., 2021b).

III. Maintenance issues, congestion, and general obstructions are persisting challenges in a dense city like Amsterdam. To combat these issues, the Municipality of Amsterdam introduced an online reporting system operated by a supervised machine learning algorithm that extracts keywords from the citizen's incident reports and forwards them to the relevant department in the Municipality (Gemeente Amsterdam., 2021c). A dataset of 500,000 reports trained the supervised learning algorithm to allot text responses to a category correctly (ibid., 2021c). Before this system, citizens frequently classified the issue category incorrectly, causing a slower response rate by the department to mediate the problem (ibid., 2021c).

5.5 Summary of Cases

Case	Sector	Purpose	Operational Status
1. SyRI	Ministry for Social Affairs and Employment (SZW)	Detecting Social Security Fraud	Discontinued
2. Toeslagenaffaire	Ministry Tax and Customs Administration	Detecting Social Security Fraud	Discontinued
3. Municipality of Nissewaard's fraud detection algorithm	Nissewaard Municipality	Detecting Social Security Fraud	Discontinued
4. Crime Anticipation System (CAS)	Law Enforcement, & Ministry of Justice and Security	Predictive policing to anticipate crimes	Active
5. Sensing Project-Roermond Police	Law Enforcement, Municipality of Roermond	Predictive policing to predict petty crimes	Active
6. Automated Parking Control in Amsterdam	Economic Services Department, Municipality Amsterdam	Identifying illegal parking	Active
7. Rubbish Collection	City Management Municipality Amsterdam	Identifying irregularities in cities (i.e., garbage)	Active
8. Illegal holiday Rental Housing Risk	Amsterdam Municipality, Department of Surveillance and Enforcement	Detecting Rental Housing Violations	Active

5.6 Method of Data Analysis

Ordinarily, case selection for case-study analysis is accomplished by examining cases that are most similar or most different (Gerring, 2008). However, this study has chosen not to take this approach because it aims to provide a comprehensive overview of all current and past AI cases employed by the Dutch public sector. Notably, the summary list above is not exhaustive; certain cases of AI have been excluded because they were classified as pilot projects, or they lacked sufficient information detailing its purpose and type of AI used.

A meta-analysis method that generates a word cloud by selected key phrases is eliminated, given the variety of terms used to describe a one-collect concept. This approach is feasible for a concept such as AI that has a broader representation by the number of software-based techniques: "artificial neural networks, evolutionary computation (consisting of genetic algorithms, evolutionary strategies, and genetic programming), fuzzy logic, intelligent systems, multi-agent systems, natural language, expert systems, learning classifier systems, automatic learning, and deep learning (Valle-Cruz et al. 2019, p. 93). However, the concept of discretion cannot be distinguished by key terms or phrases as it takes on a variety of meanings and is understood by its context.

Therefore, to determine whether the adoption of artificial intelligence is affecting bureaucratic discretion in the Netherlands, this study utilizes the two proposed assumptions to approximate the changes in bureaucratic discretion- (1) we expect more AI adoption for tasks characterized by low discretion and (2) we expect more AI adoption at the micro level. With regards to assumption one, the degree of discretion in each case is ascertained by identifying the type of AI used. If the case documents that the type of AI used is either unsupervised learning (self-learning) or black box, then the tasks carried out are considered low discretion; this is because the aforementioned AI systems make no use of human involvement or supervision. This suggests that the agency has entrusted the AI system to operate independently and process information automatically. Conversely, if the case documents that the type of AI used is either supervised or semi-supervised, then tasks carried out are considered high discretion; this is because the aforementioned AI system makes use of human involvement or supervision. In cases where the AI system requires human involvement, or oversight, it suggests that the agency does not fully trust the system to exercise discretion independently. This idea goes hand in hand with Young et al. (2019) who suggests that systems whereby relationships must be extrapolated from data are considered as high discretion. Relationship discovery requires surmising the influence of extraneous factors or variables on independent and dependent variables, and whether an autonomous AI system can mimic a cognitive approach for relationship discovery remains unknown or in development (Delua, 2021). Till then, supervised, or semi-supervised learning is most used for classification problems or calculating regression models that addresses the relationship between the independent and dependent variable (ibid., 2021).

To this end, the degree of task discretion per case is revealed by matching the written details (quotational evidence for the type of AI used) of each case with the matrix shown below. This approach is considered as a qualitive method of documentation analysis.

5.7 Matrix for Operationalizing Discretion

	Low Discretion	**High Discretion**
Task Discretion	Simple, repetitive and redundant tasks, typically fit for automation (Young et al., 2019; Busch and Henriksen, 2018)	Decision-support tool, predictive analytics, new data generation, reduction of data complexity, relationship discovery. (Young et al., 2019).
Operationalization	A system that runs tasks without human involvement and executes decisions without any supervision. (Busch and Henriksen, 2018)	A system that runs tasks with human involvement and executes decisions with supervision. (Busch and Henriksen, 2018)
AI Classification	Unsupervised learning (self-learning, deep learning) and black box	Supervised and semi-supervised learning

5.8 Matrix for Operationalizing Level of Analysis

	Micro- Level	**Macro- Level**
Operationalization	The AI system deals with citizens directly and /or supports bureaucrats' discretionary authority to apply policies to individual cases (Busch and Henriksen, 2018; Young et al., 2019).	The AI system assist in designing and arranging the preparation made for the implementation or enforcement of policies (Busch and Henriksen, 2018; Young et al. 2019).

This study differentiates between the micro and macro level by assessing the intention of the AI system and who or what it directly affects. The tasks of the AI system that directly deal with individual cases are considered as micro-level. Conversely, AI systems identified as macro-level are responsible for how policies/rules or laws are designed and disseminated across the organization.

In sum, for both assumption A (discretion) and B (level of analysis), the results will be chronicled in a table and totalled together to discern whether the initial assumption is valid- there is more AI adoption at the micro-level and tasks characterized by low discretion. If the table reveals more cases are used for high discretion tasks and employed at the macro level, we accept the assumption alternative.

SECTION SIX | ANALYSIS

In this section, the degree of task discretion and level of analysis is revealed by matching the written details of each case with the matrix that identifies the characteristics of tasks classified as low vs high discretion and micro vs macro contexts. Please refer to matrix 5.7 and 5.8 in the method of analysis on how this is methodically addressed and deduced.

6.1 AI Systems Tackling Social Fraud

In 2015, the Dutch government cited that it had taken steps directed at curtailing fraud by providing enhanced training programs for employees in the respective competent authorities (European Commission, 2015). However, during the case discovery process, it became clear that the Dutch Government considered this as insufficient and instead, dedicated a significant number of resources to challenge social security fraud both regionally and nationally. In all three examined cases- SyRI, Toeslagenaffaire and TNO Nissewaard the AI fraud detection system was considered unlawful or a violation because it encroached on citizens' rights to privacy. As a result, the AI systems in the cases mentioned above were suspended because NGOs and private citizens had collected sufficient evidence indicating discriminatory profiling, including but not limited to various indicators: income, registered address, marital status, and concealment of assets. Of the three cases, SyRI and Toeslagenaffaire classified as low discretion because the programming of the AI system excluded human interference or supervision, and TNO Nissewaard's algorithm classified as high discretion because the program requires human interference or oversight. We arrive at this conclusion by the evidence indicating that the system operated using unsupervised learning (self-learning) or supervised learning. With regards to the context in which these cases are situated, the AI systems of SyRI and the Toeslagenaffaire are employed at the macro level and TNO Nissewaard at the micro level. These findings are derived by drawing parallels between the description of the case and the level of analysis matrix.

Unlike the Toeslagenaffaire or TNO Nissewaard's algorithm, SyRI's algorithm has not been explicitly defined by the State because they feared citizens will learn to circumvent the system if they choose to implement something similar in the future. The District Court of the Hague criticized SyRI's enigmatic decision-making process; to an extent, they pseudonymized SyRI as a Blackbox (NJCM et al. v The Dutch State, 2020, para. 6.53). Although the Ministry of Social Affairs and the Dutch tax authorities denied that SyRI operated on deep learning or self-learning algorithms, there was no method to verify that SyRI used simple decision trees either (NJCM et al. v The Dutch State, para. 6.47). In response, the court indicated that-

> *"The SyRI legislation does provide scope for the development and application of a risk model using "deep learning" and data mining, and for the development of risk profiles (NJCM et al. v The Dutch State, 2020, para. 6.63).*

Moreover, the Court concurred with the Advisory Division in their assessment-

> *"That the application of SyRI "is in line" with 'deep learning' and self-learning systems" (NJCM et al. v The Dutch State, 2020, para. 6.65).*

Given SyRI's lack of transparency, the Court could not ascertain unequivocally that the system was indeed deep learning or self-learning in its final judgment. As for the Dutch government, they failed to invalidate the suppositions of the court; a multitude of reasons could explain this; one, fear of public scrutiny by the lack of human oversight in the attempt to detect fraud; or second, the government foresees the use deep learning and self-learning algorithms in the future, thus safeguarding the legitimacy of AI measures. Considering this, this study agrees with Court's initial opinion/assessment of SyRI as referenced above.

Regarding SyRI's level of analysis, we look at its principal objective- it is a system that amalgamates data about citizens shared between government departments to detect fraud with taxes, social benefits, and allowances (Rijksoverheid, 2020). SyRI is only operable if a department elects to use the system; thus, it is a tool that facilitates the enforcement of policies prescribed by the Dutch government. To this end, referring to matrix 5.8, it is clear that SyRI ascribes to the macro-level.

In the case of the Toeslagenaffaire, discerning the level of discretion by the type of AI was less convoluted because the respective evaluation reports (Dutch Data Protection Authority and Amnesty

International) detailed the internal design of the risk classification model that was used for monitoring childcare benefits.

> *"The risk classification model is a 'self-learning model' that is trained with examples of correct and incorrect application."* [1]

Notably, the risk classification model is one of several tools the Dutch tax authorities used for fraud monitoring in childcare benefits. Applications with a very high risk-score- established by the risk classification model, were passed on for manual verification, but inspectors hardly refuted the conclusion because there was no justification explaining the risk score (Amnesty International, 2021, p. 16). And so, given that the crux of the fraud monitoring relied on the risk classification model operated by a self-learning algorithm, the classification of task discretion for the system's task is low.

As for the level of analysis, the self-learning risk classification model used by the Dutch tax authorities did more than support the discretionary authority of administrative agents. In most cases, the risk model systematically and independently monitored applications for childcare benefits, and only profiles with a high-risk score for fraud were passed on for manual inspection. Given this fact, and consulting with matrix 5.8, we eliminate that the case attributes to the micro-level because the system was an instrument for prioritizing fraudulent applications necessary for enforcement or fines- in this matter. Thus, the tasks exercised by the self-learning risk classification model took place in the macro-level context.

The fraud detection system instated by the Municipality of Nissewaard operated on a supervised learning method, the corollary being that it ran its tasks with human supervision.

> *"The algorithm developed by TDL works according to a 'supervised learning' method, whereby to train this method a dataset is used with social assistance recipients in which abuse, or improper*

[1] Original Dutch text reads: "Het risico-classificatiemodel is een zelflerend model dat wordt getraind met voorbeelden van juiste en onjuiste aanvragen" (Dutch Data Protection Authority, 2018, p. 14).

use has previously been established (also called 'the positives') and data from all other social assistance recipients (these are treated as 'the others')." [2]

TNO's algorithm operates in a context whereby it co-works with the administrative agent to assess individual cases for benefits fraud, thus classifying it as a task that works at the micro-level. Both TNO and the municipality considered it necessary to involve a human agent in the process to oversee that the supervised learning method avoided demographics-based targeting and repeating profiles the system had previously investigated (TNO, 2021, p. 13). It is clear from TNO's evaluation report that the AI system is considered a decision-support tool and a data reduction tool that sieves through thousands of applications and prioritizes profiles that indicate an increased risk of fraud. Referring to matrix 5.7, the classification of discretion for the system's task is high because the AI tool requires human involvement and aligns with the high discretion indicators formulated by Young et al. (2019).

Arguably, of the three cases, Nissewaard's fraud detection system can be considered most logical/ appropriate, given that decisions determining if a citizen is suspected of fraud and abusing the welfare system should fall in the domain of human discretion. The implications of an AI system wrongfully classifying applications with a high-risk score based on false or discriminatory indicators are detrimental to citizens' quality of life. This very fact is consistent with the events that transpired of the Toeslagenaffaire algorithm. [3]

[2] Original Dutch text reads: "Het door TDL ontwikkelde algoritme werkt volgens een 'supervised learning' methode, waarbij voor de training van deze methode een dataset wordt gebruikt met bijstandsontvangers waar eerder misbruik of oneigenlijk gebruik is vastgesteld (ook wel genoemd: 'de positieven') en gegevens van alle andere bijstandsontvangers (deze worden behandeld als 'de overigen')" (TNO, 2021, p. 13).

[3] Approximately 26,000 parents in the Netherlands were falsely accused of child-benefit fraud. Affected families suffered from crippling financial issues ranging from debt, unemployment, and destabilizing family dynamics (Amnesty International, 2021, p. 5).

6.2 AI Systems for Predictive Policing and Risk Monitoring

The growth of AI adoption in law enforcement has become ostensible, particularly in the domain of predictive policing and risk monitoring. AI predictive policing has improved the markers determining when and where a crime is likely to occur and who might be a victim or complicit in a crime. The latter falls in the domain of risk monitoring, which often links data between government branches to discern the propensity of criminal activity by a citizen. In 2020, Ferdinand Grapperhaus, the Minister of Justice and Security, provided an extensive overview to parliament on AI police experiments with mass surveillance. The parliamentary inquisition was set off in light of concerns that the current applications of AI used by the police lead to (un)intentional discrimination and profiling. The Dutch government remains firm that this is strictly not the case and purports that risk assessment models are only used as support tools but not as actionable intelligence for predicting crime or identifying criminals. The extent to which the Dutch government has adopted or experimented with AI tools (that we know of) is limited to two cases: CAS and the Remote Sensing Project in Roermond. Both cases are classified as low discretion because the internal mechanism of the AI system operates using unsupervised learning which excludes human oversight. In the first case, CAS-

> *"CAS uses a neural network; a machine learning algorithm that learns to recognize and adapt to patterns in data as a result of self-correction in iterative cycles"*
> *(Mutsaers and van Nuenen, 2020, p.6).*

As pointed out in the reference above, CAS uses a neural network, a programming technique used in unsupervised machine learning to detect where crimes are likely to occur in a city. The algorithm relies on other analytical methods and big data sources to identify hot-crime zones; however, the utility of a human agent or knowledge is not required. The quote cites that CAS operates in iterative cycles, suggesting that tasks are both redundant and repetitive. If we recall matrix 5.7, we note that low discretion tasks are considered repetitive and unvaried (Young et al., 2019). By this logic, CAS classifies as low task discretion. In the case of the Remote Sensing Project-

> *"The police use various data for the analyses, including ANPR [(Automatic Number Plate Recognition) using deep learning]- cameras65 and brand-model color recognition cameras.66 The data are processed in a 'points system', or risk model. The sensors are programmed in such a way that they link observations in the street scene to the risk profile".*[4]

Notably, the risk model used in the Remote sensing project does make use of AI tools. However, the data obtained by which the risk model generates hits on vehicles with a greater propensity for criminality does. The cameras mounted around the city of Roermond are interlinked with an ANPR software that uses deep learning, an unsupervised machine learning technique. Thus, while the Dutch government claims that the remote sensing project does not use artificial intelligence, the police's data acquisition methods are dependent on artificial intelligence tools[5]. Given the risk model's dependency on ANPR to detect crime, we maintain that the case classifies as a low discretion task because ANPR sets off the action for the operation of the risk model and does this without human involvement (this abides to matrix 5.7). The uncertainties and contradictory messages communicated by the Dutch government on the type of AI used indicate a lack of transparency on the utility of technologies in law enforcement.

The context in which these cases operate is different; the AI system of CAS is employed at the Macro level and the Remote Sensing Project at the micro level. CAS's classification is considered macro-level because the objective of CAS is to predict crime hot spots which aids the police in organizing the

[4] Original Dutch text reads: *"De politie maakt voor de analyses gebruik van verschillende gegevens, waaronder ANPR-camera's65 en merk-model-kleurherkenning-camera's.66 De gegevens worden verwerkt in een 'puntensysteem', ofwel een risicomodel. De sensoren zijn zo geprogrammeerd dat ze observaties in het straatbeeld koppelen aan het risicoprofiel*".

[5] Minister Grapperhaus remarks to the parliamentary inquiry on predictive policing expresses that *"the police currently have only a limited number of applications of AI in the sense that systems exhibit intelligent behaviour and can, to a greater or lesser extent, learn and take actions independently"* (Tweede Kamer, 2020, p. 5).

deployment of surveillance teams in those areas. To put it simply, it is an instrument that facilities the working process of police and their ability to enforce policies at the street level. Conversely, the Remote Sensing project is considered micro-level because the AI system directly influences a police officer's discretionary authority and whether they follow up on a vehicle with a high-risk score. The intelligence obtained by the AI system is immediately actionable and implicates the citizen who is unaware they have been flagged.

6.3 City Operations via AI

In September 2020, the Municipality of Amsterdam launched an AI index that provides residents and citizens with a comprehensive overview of active AI systems that optimize city services. Currently, three AI systems are listed on the index that targets parking control, holiday rental fraud, and issues in public spaces. The AI systems are under the management of the municipality but are operated by a third party, and this serves as an indication of growing private-public partnership in AI. In contrast to the aforementioned cases outlined in sub-section 6.1 and 6.2, these AI systems are human-centred, meaning that they rely on the input or feedback of citizens for the tasks to operate or be initiated.

Amsterdam is a densely populated city with limited space, and the municipality has imposed a threshold on the number of vehicles allowed to park in the city (Gemeente Amsterdam, 2021b). Before the AI parking control system, local police were responsible for checking permits and issuing fines for illegally parked cars. The introduction of the AI-automated parking control system has arguably taken over the discretionary authority of police officers responsible for parking enforcement. In terms of the task discretion classification- referring to Young et al. (2019), they consider that dispensing penalties and fines is a low discretion task because it is a menial and repetitive process with a fixed outcome. This rhetoric does not change because the AI system takes up this task. The AI parking control system detects violations through:

> *ANPR (Automatic Number Plate Recognition [using deep learning])- camera scans license plates and translates the images into data." (Gemeente Amsterdam, 2021b)*

Matrix 5.7 indicates that AI systems that operate on deep learning execute decisions without supervision; hence, the AI parking control system, which runs on deep learning, is entrusted to exercise discretion autonomously. Thus, the task discretion classification for the AI parking control system is low. The context in which this task performs is at the macro level. It assists the institution of law enforcement in enforcing policies and is not considered a decision support tool that aids police officers in determining if a violation has or has not occurred.

Similar to Amsterdam's lack of parking availability, there is also a tight capacity for available holiday housing. To better anticipate a potential fraudulent holiday rental, the municipality introduced an AI system to support surveillance and enforcement officers to prioritize cases and expedite the processing of reports, detailing suspicions of illegal holiday rentals. The algorithm of the AI system uses-

> *"A random forest regression (…) to calculate the probability of housing fraud, [next] SHAP is used to calculate which features have resulted in a high or low suspicion of housing fraud" (Gemeente Amsterdam, 2021a).*

Random forest regression is a supervised machine learning technique that establishes parameters in this case (high or low suspicion in fraud) to predict the initial default problem (risk of rental fraud). (Gramegna and Giudici, 2021). AI systems that use supervised learning techniques include human supervision, indicating that the task discretion is high. This conclusion aligns with Young et al.'s (2019) assessment of high discretion tasks, which are regarded as decision support tools for predictive analytics and decision making, which is consistent with the objective in this case. Regarding the analysis level, the algorithm operates in a context whereby it co-works with the surveillance and enforcement officers to assess individual cases for rental fraud, thus classifying it as a task that works at the micro-level.

The last AI case centered on optimizing city operations in Amsterdam is an AI system that expedites the response time of relevant municipal departments for issues in public spaces. The algorithm of the system uses-

"A logistic regression (a machine-learning technique) of that combination of words is then used to determine which category is most likely to fit, and therefore which department within the municipality needs to act on the report" (Gemeente Amsterdam, 2021c).

Logistic regression is a supervised machine learning technique. To this end, using matrix 5.7, AI systems that operate using supervised learning are considered high discretion because it demands human oversight or involvement. This AI system is viewed as a decision support tool, given that 40% of assessments made by the algorithm are forwarded to an administrator at the Action Service Centre that manually verifies if the chosen category is representative of the report (Gemeente Amsterdam, 2021c). To this end, AI systems characterized as decision support tools are considered high discretion tasks as outlined in matrix 5.7. Regarding the level of analysis, the explanation is concomitant with the rental housing fraud case, whereby the algorithm works with an administrator to prepare respective departments with the necessary information to fulfill their duties.

6. 4 Summary of Findings

The results of the analysis indicate that five out of the eight cases are classified with low task discretion and three out of eight with high task discretion, thus assumption A- *we expect more AI adoption for tasks characterized by low discretion* is **true.** This conclusion was derived by assuming that a system that runs tasks without human involvement and executes decisions without any supervision would typically be tasks characterized as repetitive and straightforward. Whether or not the mechanisms the AI system runs its tasks are considered complex or not, is irrelevant. If the organization has approved that a task can be carried out by an unsupervised, self-learning or deep learning machine algorithm that excludes human oversight, it indicates a lack of necessity for human judgement or discretion. In principle, human discretion is necessary for trivial tasks with uncertain outcomes or that the consequences can significantly affect citizens quality of life or the functioning of an organization (Young et al., 2019; Busch and Henriksen, 2018). By this logic, it would be expected that cases that challenge issues such as social benefits fraud

(SyRI) or detecting suspicious individuals (sensing project Roermond) would not allocate complete discretionary authority to an artificial system. However, the analysis of the cases pointed to the contrary.

Furthermore, in all five cases that are classified as low task discretion, the density of information and data the AI algorithms process is beyond what a human agent(s) could administer. It becomes clear that the Dutch government prioritizes efficiency over accountability irrespective of how the outcome may implicate individuals. To this end, while the analysis is consistent with assumption A, establishing a judgment on how bureaucratic discretion is transforming because of AI in the Dutch government could be argued as unpredictable and arbitrary.

Regarding the analysis level, the results indicate that five out of the eight cases were employed at the macro level and three out of eight at the micro; thus, assumption B- *we expect more AI adoption at the micro-level* is **false**. Instead, the analysis identified that the Dutch public sector employed AI systems that assisted in designing and arranging the preparation made for the implementation or enforcement of policies (Busch and Henriksen, 2018; Young et al., 2019). This study considered the rejection of the assumption to be unlikely given that AI is a new phenomenon in the public sector; thus, the proclivity for AI to establish in areas of decision-making systems is less likely. It can be understood that decision-making systems are complex given the interdependencies across departments and therefore demand to be adaptive and open to change. By this logic, the public sector's inclination to move in tandem with the private sphere and optimize AI systems to achieve best practices is not inconceivable. It has consequences on bureaucratic discretion because previously, at that level, public workers established their command actions, and now the opposite effect is seen, and public workers rely on the AI system’s input to follow up on the deliverable.

Summary of Findings

Type of AI		**Black Box**	**Unsupervised (Self-learning, Deep Learning)**	**Supervised Learning**	**Semi-Supervised Learning**	**Task Discretion Classification**	**Level of Analysis**
Cases							
I.	SyRI	**X**	**X**			Low Discretion	Macro
II.	Toeslagenaffaire		**X**			Low Discretion	Macro
III.	Municipality of Nissewaard's fraud detection algorithm			**X**		High Discretion	Micro
IV.	Crime Anticipation System (CAS)		**X**			Low Discretion	Macro
V.	Sensing Project-Roermond			**X**		High Discretion	Micro
VI.	Automated Parking Control in Amsterdam		**X**			Low Discretion	Micro
VII.	AI Reporting Issues in Public Spaces			**X**		High Discretion	Macro
VIII.	Illegal holiday Rental Housing Risk		**X**			Low Discretion	Micro
					Total	Low Task Discretion: 5/8 High Task Discretion: 3/8	Micro Level 4/8 Macro Level: 5/8

SECTION SEVEN | LIMITATIONS, SUGGESTED RESEARCH AND CONCLUSION

The discussion of AI in the public sector is becoming increasingly relevant as international institutions ratify AI data regulations and acts. The academic community has played a profound role in orienting government for best methods and practices to implement AI without causing significant shockwaves within governments and for service recipients. The findings in this study have shown that despite the limited number of AI applications in the Netherlands, there is sufficient evidence indicating that the scope of human discretion in the public sector is reducing, especially for low discretion tasks and at macro level.

The most notable limitation of the research presented in this study is the uncertainty on whether the identified cases are representative of all AI applications currently used in the Dutch public sector. While the case selection method aimed to ensure that the searches conducted on the internet were sufficiently broad and specific to flag AI hits, the certainty that all AI applications would be discoverable and listed is unlikely. To control for this, the study could have had prompted an inquiry with varying ministries and departments to procure a more accurate overview. However, due to the restricted time allotted for the completion of this study, this approach was not considered constructive. Moreover, the deficit to the latter approach is the notion that governments are reluctant to disclose the use of AI, as the regulations and laws governing the permissibility of AI are in development. Evident from the analysis, government spokespersons (i.e., Minister of Justice and Security) responsible for communicating the trajectory and use of AI, are intentionally ambiguous, and use terms such as automation and self-sufficient systems to disguise the utility of artificial intelligence. The speculated rationale for this approach, is based on Holland's history of subjective algorithms that have (in)advertently targeted marginalized groups.

Another observed limitation is that the parameters of the case selection method were too narrow. Had we extended the range of permissible AI cases, the study could have listed potential pilot-project that the government is currently vetting and considering. For instance, there is an exploratory research project into the use of prediction models for the purpose of risk-based oversight and inspection of primary education facilities. The research project was carried out by the Education Inspectorate of the Netherlands Ministry of Education, Culture and Science, in close cooperation with research scientists of the Amsterdam Free University and used data on primary education in the Netherlands over the period 2011 – 2018. The reports indicates that these prediction models can be useful for the assessment of less quantifiable quality areas, such as the teaching process, quality care and ambition and the corporate culture at school. This new technology has the potential of identifying risk areas that otherwise would have remained blind spots and thereby help to identify schools at risk that subsequently should be priorities for inspections. Including this case would have added more variety to the list of discovered cases, especially as its purpose deviates significantly from the other, i.e., detecting fraud, risk modelling, and city management.

This study has identified two principal avenues for future research specific to interlinkages between AI, bureaucracies, and discretion. The first recommendation is to urge scholars to examine the needs of end-users. We refer to end-users as public workers who engage with an AI system that aids or contributes to their decision-making process. There has been a lack of literature citing what areas and tasks public workers would like to implement or adopt AI tools. It is understood that institutions approach AI as a cost-saving and productivity mechanism, but rarely is the discourse disclosed on the motivations for AI task replacement and why. By identifying the desires and needs of public workers through a survey-questionnaire method, scholars can better infer an informal timeline for AI uptake. Furthermore, if it becomes apparent that public workers like or dislike the implementation of AI, it will illustrate their willingness to co-work with future AI-systems.

Next, this study recommends allocating closer attention to the role of program analysts and developers responsible for the operation of AI applications within the public sector. This small sum of individuals has the potential to dictate the structure and working methods of both front-line workers and managers implicitly or explicitly. Examining how much authority and discretion is allotted to these individuals is another explanation for the changing scope of discretion in bureaucracies resulting from AI.

The conclusion of this study indicates that bureaucratic discretion is transforming as a result of AI in the Dutch public sector and will continue to do so. How bureaucratic discretion is changing was assessed by examining the type of AI in conjunction with the two indicators borrowed from Young et al. (2019)-

task discretion and the level of analysis. The identified AI cases served as a snapshot of what can be further expected and in which ministries and departments AI applications will be more prevalent. We can expect that emerging AI data regulations by the European Union will focus on how, when, and where AI capabilities are permissible; thus, again shaping the transformation of bureaucratic discretion.

***Blank Page ***

REFERENCES

Alfter, B., Muller-Eiselt, R., & Spielkamp, M. (2019). Automating Society 2019. AlgorithmWatch. Retrieved 20 December 2021, from https://algorithmwatch.org/en/automating-society-2019/.

Amnesty International. (2020). We Sense Trouble, Automated Discrimination and Mass Surveillance in Predictive Policing in The Netherlands [PDF] (pp. 4-43). Amnesty International Ltd. Retrieved 20 December 2021, from https://www.amnesty.org/en/wp-content/uploads/2021/05/EUR3529712020E NGLISH.pdf.

Amnesty International. (2021). Xenophobic Machines Discrimination Through Unregulated Use Of Algorithms In The Dutch Childcare Benefits Scandal [Pdf] (1st ed., pp. 8-39). Amnesty International. Retrieved 5 January 2022, from https://www.amnesty.nl/content/uploads/2021/10/20211014_FINAL_Xenophobic-Machines.pdf?x42580.

Barth, T. J., & Arnold, E. (1999). Artificial Intelligence and Administrative Discretion. American Review of Public Administration, 29(4), 332–351. https://doi.org/10.1177/02750749922064463

Bathaee, Y. (2018). The artificial intelligence black box and the failure of intent and causation. Harvard Journal of Law & Technology, 31(2), 889.

Bazuin, A., Boesman, T., Catsburg, J., Van Ginneken, C., van Ham, M., Koenis, M., & Schelfhout, T. (2021). Rapport Werkgroep Reflectie Toeslagenaffaire Rechtbanken Recht Vinden Bij De Rechtbank (pp. 6-63). The Hague: The judiciary, Landelijk Overleg Vakinhoud Bestuursrech.

Beuken, A. (2021). HCC - Wat is een algoritmeregister en waarom is het belangrijk?. Hcc.nl. Retrieved 28 December 2021, from https://www.hcc.nl/kennis/vraag-van-de-week/4871-wat-is-een-algoritmeregister-en-waarom-is-het-belangrijk.

Boost.ai. (2019). Boost.ai - Automation for the people: using conversational AI to benefit the public sector. Boost.ai. Retrieved 19 December 2021, from https://www.boost.ai/case- studies/automation-for-the-people-using-conversational-ai-to-benefit-the-public- sector.

Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. Public Administration Review, 62(2), 174-184. https://doi.org/10.1111/0033-3352.00168

Brenninkmeijer, A., & ten Seldam, B. (2021). The Dutch benefits scandal: a cautionary tale for algorithmic enforcement – EU Law Enforcement. Eulawenforcement.com. Retrieved 20 December 2021, from https://eulawenforcement.com/?p=7941.

Buffat, A. (2015). Street-Level Bureaucracy and E-Government. Public Management Review, 17(1), 149–161. https://doi.org/10.1080/14719037.2013.771699

Bullock, J. B. (2019). Artificial intelligence, discretion, and bureaucracy. American Review of Public Administration, 49(7), 751-761. https://doi.org/10.1177/0275074019856123

Busch, P. A., & Henriksen, H. Z. (2018). Digital discretion: A systematic literature review of ICT and street-level discretion. Information Polity, 23(1), 3-28. https://doi.org/10.3233/IP-170050

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. Public Administration Review, 81(5), 825-836. https://doi.org/10.1111/puar.13293

Chapelle, O., Scholkopf, B., & Zien, A. (2019). Semi-supervised learning. The MIT Press.

Cooper, P. J. (2006). Public law and public administration. Wadsworth Publishing Company.

Cordella, Antonio and Tempini, Niccolò (2015) E-government and organizational change: reappraising the role of ICT and bureaucracy in public service delivery. Government Information Quarterly, 32 (3). pp. 279-286. ISSN 0740-624X

Costa, C. (2019). What Is Machine Learning & Deep Learning?. Retrieved 30 December 2021, from https://medium.com/@harish_6956/what-is-machine-learning-deep-learning-7788604004da de Boer, N., & Raaphorst, N. (2021). Automation and discretion: Explaining the effect of automation on how street-level bureaucrats enforce. Public Management Review, ahead-of-print(ahead-of-print), 1-21. https://doi.org/10.1080/14719037.2021.1937684

De Boer, N., & Raaphorst, N. (2021). Automation and discretion: explaining the effect of automation on how street-level bureaucrats enforce. Public Management Review, 1–21. https://doi.org/10.1080/14719037.2021.1937684

Dechesne, F., Zardiashvili, L., Dignum, V., & Bieger, J. (2019). AI & Ethics at the Police: Towards Responsible use of Artificial Intelligence in the Dutch Police [PDF] (1st ed., pp. 1-28). Leiden University Center for Law and Digital Technologies (eLaw) and TU Delft Institute of Design for Values. Retrieved 20 December 2021,fromhttps://www.universiteitleiden.nl/binaries/content/assets/rechtsgeleerdheid/instituut-voor- metajuridica/artificiele-intelligentie-en-ethiek-bij-de-politie/ai-and-ethics-at-the-police-towards-responsible-use-of-artificial-intelligence-at-the-dutch-police-2019.pdf.

Delua, J. (2021). Supervised vs. Unsupervised Learning: What's the Difference? Ibm.com. Retrieved 22 December 2021, from https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning.

Dhasarathy, A., Jain, S., & Khan, N. (2020). When governments turn to AI: Algorithms, trade-offs, and trust [PDF] (1st ed., pp. 1-7). McKinsey & Company. Retrieved 19 December 2021, from https://www.mckinsey.com/industries/public-and-social-sector/our-insights/when-governments-turn-to-ai-algorithms-trade-offs-and-trust.

Dutch Data Protection Authority. (2018). Belastingdienst/ Toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag (pp. 3-58). The Hague: Autoriteit Persoonsgegevens.

European Commission. (2015). Fraud and error in the field of social security coordination (pp. 7-58). Brussels: European Commission, Directorate-General for Employment, Social Affairs and Inclusion.

Fountain, J. E. (2001). Building the virtual state: Information technology and institutional change. Brookings Institution Press.

Garson, G. D., InfoSci-Books - Copyright 2007, & Books24x7, I. (2007). Modern public information technology systems: Issues and challenges. IGI Pub.

Gramegna, A., & Giudici, P. (2021). *SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk*. Frontiers. Retrieved 9 January 2022, from https://www.frontiersin.org/articles/10.3389/frai.2021.752558/full.

Gemeente Amsterdam. (2021). Automated parking control. City of Amsterdam Algorithm Register. Retrieved 30 December 2021, from https://algoritmeregister.amsterdam.nl/en/automated-parking-control/.

Gemeente Amsterdam. (2021b). Illegal holiday rental housing risk. City of Amsterdam Algorithm Register. Retrieved 28 December 2021, from https://algoritmeregister.amsterdam.nl/en/illegal-holiday-rental- housing-risk/.

Gemeente Amsterdam. (2021c). Reporting issues in public space. City of Amsterdam Algorithm Register. Retrieved 30 December 2021, from https://algoritmeregister.amsterdam.nl/en/reporting-issues-in-public- space.

Gerring, J. (2008). Case selection for Case-Study analysis: Qualitative and quantitative techniques. In H. E. Brady, D. Collier & J. M. Box-Steffensmeier (Eds.), The oxford handbook of political methodology (1st ed.,). Oxford. University Press. https://doi.org/10.1093/oxfordhb/9780199286546.003.0028

Hoetker, G., & Mahony, J. T. (2002). Building the Virtual State: Information Technology and Institutional Change. The Academy of Management Review, 27(4), 619–622. https://doi.org/10.5465/AMR.2002.7566114

Hulstaert, L. (2019). Black-box vs. white-box models. Towards Data Science. Retrieved 19 December 2021, from https://towardsdatascience.com/machine-learning-interpretability-techniques-662c723454f3.

JavaPoint. (2022). Difference between Artificial intelligence and Machine learning - Javatpoint. www.javatpoint.com. Retrieved 5 January 2022, from https://www.javatpoint.com/difference-between-artificial-intelligence-and-machine-learning.

Johannsen, G. (2014). Integrated systems engineering. Pergamon.

Lipsky M (1980) Street –level bureaucracy: dilemmas of the individual in public services. Russell Sage Foundation, New York

Lipsky, M., & Project Muse University Press eBooks. (2010). Street-level bureaucracy: Dilemmas of the individual in public services (30th anniversary expand ed.). Russell Sage Foundation.

Liu, Qiong & Wu, Ying. (2012). Supervised Learning. 10.1007/978-1-4419-1428-6_451.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence: August 31, 1955. The AI Magazine, 27(4), 12.

Meuwese, A. (2020). Regulating algorithmic decision-making one case at the time. Case notes on: District Court of The Hague, 5/02/20, ECLI:NL: RBDHA:2020:865 (NJCM vs the Netherlands (SyRI)). European Review of Digital Administration & Law, 1(1), 209-211.

Misuraca, G., & van Noordt, C. (2020). Overview of the use and impact of AI in public services in the EU, [PDF] (pp. 2-83). Publications Office of the European Union. Retrieved 19 December 2021, from https://publications.jrc.ec.europa.eu/repository/bitstream/JRC120399/jrc120399_misuraca-ai-watch_public-services_30062020_def.pdf.

Mitchell, T. M. (1997). Machine Learning. New York: McGraw-Hill. ISBN: 978-0-07-042807-2

Mutsaers, Paul & van Nuenen, Tom. (2020). Predictively policed: The Dutch CAS case and its forerunners. 10.13140/RG.2.2.10960.74243.

NJCM et al. v The Dutch State (2020) The Hague District Court ECLI: NL: RBDHA:2020:1878 (SyRI). Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878 [6.63].

Oosterloo, S., & van Schie, G. (2018). The Politics and Biases of the "Crime Anticipation System" of the Dutch Police. CEUR Workshop Proceedings, 2103, 30.

Politie. (2021). Crime Anticipation System. studylib.net. Retrieved 20 December 2021, from https://studylib.net/doc/5800826/crime-anticipation-system.

Raaphorst, N. (2017). Uncertainty in Bureaucracy: Toward a Sociological Understanding of Frontline Decision Making: Onzekerheid in bureaucratie : naar een sociologisch begrip van besluitvorming door contactambtenaren. Erasmus University Rotterdam.

Rijksoverheid. (2020). Staat niet in hoger beroep tegen vonnis rechter inzake SyRI. Rijksoverheid. Retrieved 7 January 2022, from https://www.rijksoverheid.nl/actueel/nieuws/2020/04/23/staat-niet-in-hoger-beroep-tegen-vonnis-rechter-inzake-syri.

Rijksoverheid. (2021). Kabinetsreactie op het rapport 'Ongekend onrecht'. Rijksoverheid. Retrieved 20 December 2021, from https://www-rijksoverheid-nl.translate.goog/actueel/nieuws/2021/01/15/kabinetsreactie-op-het-rapport-ongekend-onrecht.

Roosen, M. (2020). What SyRI can teach us about technical solutions for societal challenges. Globaldatajustice.org. Retrieved 5 January 2022, from https://globaldatajustice.org/2020-02-20-roosen-syri/.

Saxena, D., Badillo-Urquiola, K., Wisniewski, P. J., & Guha, S. (2021). A framework of high-stakes algorithmic decision-making for the public sector developed through a case study of child-welfare. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1-41. https://doi.org/10.1145/3476089

Schets, F. (2019). Should I be unconsciously afraid of myself? The (dis)proportionate use of risk profiling practices by public authorities to combat social security fraud [PDF] (pp. 5-49). Tilburg Institute for Law, Technology and Society (TILT). Retrieved 20 December 2021, from http://arno.uvt.nl/show.cgi?fid=150066.

Selten, F., & Meijer, A. (2021). Managing Algorithms for Public Value. International Journal of Public Administration in the Digital Age, 8(1), 1–16. https://doi.org/10.4018/IJPADA.20210101.oa9

Sheridan, T. B. (1992). Telerobotics, automation, and human supervisory control. The MIT Press.

Streebo. (2021). Chatbot for Government. Streebo inc. Retrieved 19 December 2021, from https://www.streebo.com/chatbot-for-government.

Techlabs, M. (2021). How Beneficial is Detecting Fraud Using Machine Learning?. Medium. Retrieved 5 January 2022, from https://medium.com/mlearning-ai/how-beneficial-is-detecting-fraud-using-machine- learning-b00089d84930.

TNO. (2021). Technische evaluatie van het AI-algoritme dat de Gemeente Nissewaard inzet voor de opsporing van misbruik en oneigenlijk gebruik van bijstandsuitkeringen [Ebook] (1st ed., pp. 1-44). TNO. Retrieved 6 January 2022, from http://TNO-eindrapportage-van-de-Nissewaard-casus-22-juni-2021-v1.0.pdf.

Tummers, L., & Bekkers, V. (2014). Policy Implementation, Street-level Bureaucracy, and the Importance of Discretion. Public Management Review, 16(4), 527–547. https://doi.org/10.1080/14719037.2013.841978

Tweede Kamer der Staten-Generaal. (2020). *Rechtsbescherming over het gebruik van discriminerende algoritmes en massasurveillance door de politie*(pp. 1-6). Den Haag: Tweede Kamer.

Valle-Cruz, D. (2019). Public value of e-government services through emerging technologies. The International Journal of Public Sector Management, 32(5), 530–545. https://doi.org/10.1108/IJPSM-03-2018-0072

van Bekkum, M., & Borgesius, F. Z. (2021). Digital welfare fraud detection and the Dutch SyRI judgment. European Journal of Social Security, 23(4), 323–340. https://doi.org/10.1177/13882627211031257

Vaughn, J., & Otenyo, E. E. (2007). Managerial discretion in government decision making: Beyond the street level. Jones and Bartlett Publishers.

Wenger, J. B., & Wilkins, V. M. (2008). At the Discretion of Rogue Agents: How Automation Improves Women's Outcomes in Unemployment Insurance. Journal of Public Administration Research and Theory, 19(2), 313–333. https://doi.org/10.1093/jopart/mum044

Wirtz, B. W., & Müller, W. M. (2019). An integrated artificial intelligence framework for public management. Public Management Review, 21(7), 1076–1100. https://doi.org/10.1080/14719037.2018.1549268

Wu X., Liu X., Zhou Y. (2021) Review of Unsupervised Learning Techniques. In: Jia Y., Zhang W., Fu Y., Yu Z., Zheng S. (eds) Proceedings of 2021 Chinese Intelligent Systems Conference. Lecture Notes in Electrical Engineering, vol 804. Springer, Singapore. https://doi.org/10.1007/978- 981-16-6324-6_59

Young, M. M., Bullock, J. B., & Lecy, J. D. (2019). Artificial discretion as a tool of governance: A framework for understanding the impact of artificial intelligence on public administration. Perspectives on Public Management and Governance, https://doi.org/10.1093/ppmgov/gvz014

Zouridis, S., Eck, B. M. A. van, Bovens, M., & Evans T., H. P. (2019). Automated discretion. Discretion and the Quest for Controlled Freedom.

Zuiderwijk, A., Chen, Y., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. Government Information Quarterly, 38(3), 101577. https://doi.org/10.1016/j.giq.2021.101577