



UNIVERSITÀ DEGLI STUDI DI MILANO



UNIVERSITÀ DEGLI STUDI DI TORINO

Ph.D. PROGRAM in
SOCIOLOGY AND METHODOLOGY OF SOCIAL RESEARCH
37th Cohort

Department of Social and Political Sciences

Constructing Automated Societies

Socio-cultural Determinants and Impacts of Automated Decision-making in Public Services

Doctoral dissertation by
Diletta Huyskes

Supervisor: Prof. Alessandro Gandini

Director of Doctoral Programme: Prof. Alessandra Rebughini

A.A. 2024-2025

Table of contents

Table of contents	3
Acknowledgments	5
Introduction	8
1. Literature Review & Theoretical Framework	17
1.1. Contemporary Approaches to Technology and Society	18
1.2. The Social Shaping of Technology	24
1.3. The Gendered Shaping of Technology	32
1.4. From Effects to Causes	39
2. An Ethnomethodology to Reveal the Social Construction of Algorithms	42
2.1. A Case Study Approach for Technology-in-Context	43
2.2. Observing Humans Behind Algorithms	47
2.3. Following the Actors	50
2.4. Researcher's Positioning in the Field	51
2.5. Data Collection	53
2.5.1. Interviews	54
2.5.2. Participant Observation	59
2.5.3. Document Analysis	61
2.6. Limitations	62
3. "Unprecedented Injustice": Network Determinism in the Dutch Digital Welfare State	64
3.1. Introduction	64
3.2. Theoretical Background	67
3.2.1. From Technological Determinism to Network Determinism	67
3.2.2. The Algorithmization of Social Services	71
3.2.3. The Cultural and Social Determinants of Algorithmic Practises	73
3.3. The Algorithmization of Social Benefits in the Netherlands	75
3.3.1. Constructing the Dutch Digital Welfare State	75
3.3.2. System for Risk Indication (SyRI)	78
3.3.3. The Childcare Benefits Scandal	82
3.3.4. The Case of Rotterdam	94
3.4. Networks, Determinants and the Amplification of Discrimination	100

3.5. Conclusion	104
4. The Machinery of Government: Bureaucracy, Automation and Institutional Black-Boxing	108
4.1. Introduction	108
4.2. Theoretical Background	110
4.2.1. The Machinery of Government	110
4.2.2. Blackboxing as a Practice	116
4.3. Transparency Barriers in Italian Algorithmic Governance	120
4.3.1. Legal barriers	122
4.3.2. Awareness and competence	124
4.3.3. Private interests	126
4.3.4. Governance and strategy	128
4.3.5. Accountability	129
4.3.6. Bureaucratic resistance	130
4.3.7. A Lesson from the Algorithm: the Buona Scuola Case	133
4.4. Behind Closed Doors: Measuring Technological and Institutional Black-Boxing	137
4.5. Conclusion	140
5. “The Algorithm Made Me Do It”: Automation, Culture and the Erosion of Accountability	143
5.1. Introduction	143
5.2. Theoretical background	145
5.2.1. Algorithmic Legitimacy	145
5.2.2. Algorithmic Accountability	148
5.3. Responsibility, shifted: “it’s not me, it’s the data!”	153
5.4. A Cross-Cultural Erosion of Accountability	165
5.5. Conclusion	169
6. Conclusions and future steps for research	172
Bibliography	179

Acknowledgments

Since 2018, during the third year of my undergraduate studies in Philosophy at the University of Milan, several events occurred that significantly influenced my academic and professional journey. If I were to start from there, I would have countless people to thank. That year, I took part in a seminar led by Prof. Andrea Guardo on free will and determinism and, from the very first session, I was captivated. At first glance, it had nothing to do with technology, but now, looking back, I realize that everything started there. Around the same time, other courses introduced me to theories on gender, technology, and the work of feminist scholars that have accompanied me ever since.

During this three-years research journey, I was guided by a quote by Donna Haraway in *Staying with the Trouble* (2016): «*Our task is to make trouble, to stir up potent responses to devastating events, as well as to settle troubled waters and rebuild quiet places*». That is what I felt in early 2021 when, just before applying for the PhD program in Sociology and Social Research Methods at NASP, the Dutch childcare benefits scandal - one of the main case studies of this dissertation - erupted. Like many others, I felt powerless in the face of the systemic discrimination made possible by a simple algorithm deployed by a public institution. From that moment on, I devoted much of my energy to understanding what had happened and shedding light on the processes and decisions behind it. This event, more than any other, pushed me to pursue this path, not only to spark critical responses but also to investigate how to rebuild spaces of reflection and action. If this was possible, I have to thank above anyone else my supervisor Alessandro Gandini, who has shown endless support, trust and understanding in me, this research and many other important choices in my life. Despite my descriptive tendencies and my philosophical background (!), he welcomed me into the Department of Social Sciences and never once stopped believing in me over these past three years. Many times along this long journey, I doubted whether I was capable of navigating a discipline and methodology I was unfamiliar with. I will always be grateful for your

encouragement and for always pushing me beyond my own expectations, and to do better. You were right. I was very scared of this adventure because of everything that is said about it, but you know how to make it an experience of true growth, responsibility and satisfaction. I was incredibly lucky to have you as my supervisor: if we could multiply you, we definitely would, because every doctoral student would deserve guidance like yours.

This dissertation is the result of a deeply held conviction, but also of the unwavering support of many. This work is, at its core, a collective effort sustained by many voices that often go unheard. The scholars whose work inspired me gave me the courage to explore uncharted territories and ask questions that may have seemed unorthodox. Among many, I want to thank Judy Wajcman and Andrew Feenberg, who directly shared their invaluable ideas with me. I will never forget that. I thank all the professors I had the privilege of learning from, my PhD colleagues - the 37th cohort of the Network for the Advancement of Social and Political Studies (NASP) - especially for their patience in explaining basics statistics to me in the first year, without which many of the arguments in this thesis would never have taken shape. A special thanks to Margherita, with whom I shared the initial part of this journey with great affection. Thanks to Maria, my friend and roommate in Milan, the best person to have by my side for all the care that was needed in the beginning of my PhD.

I owe a lot to all the people who materially contributed to this research: the individuals I interviewed in the Netherlands and Italy, as well as those who gave me invaluable advice off the record, whose voices resonate in these pages even if they are not formally cited. I am deeply grateful to the Data School at Utrecht University and to everyone working there, especially Prof. Mirko Schäfer and Iris Muis, for welcoming and supporting me as I explored new trajectories. I also thank my colleagues at Privacy Network, who trusted me to develop some of our projects on algorithmic transparency in public administrations. Many fragments of you are embedded in this thesis, which, despite being a solitary endeavor, still feels like a collective achievement.

Thank you to all those with whom I have shared debates, conferences, and symposiums, to everyone working in different ways to ensure a fairer technological future and a more responsible, thoughtful approach to digital innovation in public institutions. To my comrades and feminist peers, whether working on technology or not, with whom I have the privilege of exchanging invaluable ideas.

Thank you to my families, both by blood and by choice. To my parents and grandparents, for their immense love and for never doubting for a second about my path. Thanks to my mother, because you may not realize it, but more than any seminar, I think the reason I am here doing exactly what I do is thanks to you. To all my friends, both old and new, who continue to enchant me even when I think it won't happen anymore - your support is incredible. And finally, thank you, Alessandro, for your first piece of advice to undertake exactly this path. We have learned together that not everything real is rational.

Introduction

For almost two years now we have been experiencing and adapting to what has been described as a new revolution. Artificial intelligence (AI), and particularly its latest application, generative AI, has been described as such in the public debate, but also as an innovation that will lead to “human extinction” (Cheng 2023). Only in the last period, the impact of large language models (LLMs) and the architectures that allow us to interact with them has been described as a “threat to the entire humanity like pandemics” a “destruction for human civilization”, “something scary”, “a steam engine-like transformation” and beyond (Cheng 2023). Although many of these statements come from key industry leaders and several commentators have since explained the reasons and logic behind this kind of narrative as favoring their own business in Silicon Valley (Swan 2024), this rhetoric did not sound so distant and absurd. To compare AI to the steam engine is to entrust it with revolutionary, in the sense of transformative and paradigmatic, power for society. This is only the most recent version of it, as in the past few years many digital technologies or techniques involving algorithms - understood as computational processes (though not necessarily so) that follow a set of predefined rules to solve a given problem and that can vary in complexity and sophistication (Hildebrandt 2020) - have been described as revolutions, as in the case of big data (Mayer-Schönberger & Cukier 2013) and machine and deep learning (Sejnowski 2018).

In 1922, William Ogburn and Dorothy Thomas were writing down the question: “given the boat and the steam engine, isn’t the steam boat inevitable?”. Using the same rhetoric, in the essay *Are Inventions Inevitable?*, they argued that revolutionary inventions are simply inevitable events, because in the presence of the necessary constituent cultural elements, it is impossible to avoid a new technology. The two writers described it as an unstoppable process, taking the very

example of the steam engine and the steam boat. This idea is still very prevalent in our present, to the extent that we find it particularly easy to compare the words used in 1922 with those used today for AI. The logic underlying this rhetoric is common and crosses various historical stages, as it reflects the dominant paradigm for interpreting the relationship between technology and society, according to which innovation follows extremely linear processes of invention (Godin 2016) through which societies and the underlying relationships evolve and develop. This paradigm has been described as technological determinism, a theoretical perspective that assumes technology to be an autonomous force driving social change (Veblen 1899). This reductionist stance suggests that technological innovations inherently dictate the structure and behavior of social structures, relations and institutions (Winner 1980). Beginning in the 1970s, many authors in the social sciences - then ascribed to the broader field of Science and Technology Studies (STS) - developed a comprehensive scholarship aimed at refuting the deterministic approach to technological innovation. In particular, these authors focused on the social *construction* of any technology, and insisted on the study and analysis of all that is human and intentional about it (Pinch and Bijker 1984). As Ogborn and Thomas (1922) had already assumed, however, the importance of cultural elements for technological construction is undeniable and, even more, strongly inherent to it. From the design of the Victorian bicycle (Ibid.), or that of certain bridges (Winner 1980), ballistic missiles (MacKenzie 1995) or the microwave oven (Cockburn and Ormrod 1993), numerous empirical studies have reacted willing to open the technological box to social analysis and show that far from being inevitable and autonomous, every artifact is the result of precise choices, more or less conscious, carried out by different actors and their interactions, negotiations, conflicts. Why, at some point in history, does one technology impose itself on another possible trajectory, and what determines its realization (Noble 1978)? What are the controversies that influence and make possible the construction of that artifact (Callon 1991)? These are some of the questions that constructivism has been asking to show that no technology is

inevitable. Several illuminating evidences emerged from these analyses about all the factors that lead to the consolidation not only of an artifact, but also of its social consequences (Huyskes 2024). In this sense, the present and pressing statements in the contemporary public debate with respect to AI and its social impacts should lead to a new season of unraveling the human dynamics behind their design and their entry into the market and society, which are too often passed off as invisible and inconsistent aspects compared to technical and statistical correctives (see Du 2019). So many authors in recent years have been committed to this very purpose, in showing the human, social, and political roots of AI's large-scale impacts. For example, Kate Crawford in her latest book has extensively documented the environmental impacts of modern digital technologies, prior to generative AI, which are indistinguishable from the social impacts associated with digital colonialism and extractivist practices (2021).

The negative impacts and risks associated with automated practices, then, are extremely real. But one does not need to go as far as catastrophic scenarios such as the destruction of humanity and, more importantly, treat them as inevitable, in order to observe them. Other authors who have dealt with this extensively in recent years have pointed out, for example, how automated systems can perpetuate and even amplify existing social inequalities, reflecting the biases inherent in the data and the practises they are trained on (O'Neil 2016; Eubanks 2018; Barocas & Selbst 2019; Buolamwini & Gebru 2018). In general, modern scholars argue that these technologies amplify social inequalities and institutional biases, particularly when used in high-stakes decision-making environments that impact vulnerable populations (boyd and Crawford 2012; D'Ignazio and Klein 2020). In response to these issues, scholars have proposed both technical solutions and socio-technical frameworks to address the discriminatory impacts of automation. While technical solutions like fairness metrics and computational formulas have been developed to mitigate bias at the data or algorithm level (e.g., Bellamy et al. 2018; Lepri et al. 2018), other scholars argue for a more holistic approach.

In this thesis, one of the key theoretical frameworks employed is drawn from

feminist scholarship on technology. Gender and technology studies have long argued that technologies are not neutral, but rather embedded in social relations that reflect and reproduce gendered power dynamics (Harding 1986; Wajcman 1991, 2004; Haraway 1991; Cockburn & Ormrod 1993). Feminist studies of technology have looked more boldly at how technology can exclude certain social groups, particularly women, than classical constructivism. They have applied social sciences' ethnographic methodology to the study of networks of actors involved in specific instances of technological design, revealing their roles and weights in the shaping of artifacts and their consequences. Following this model and retrieving the important contributions of this literature, this thesis aims to trace how technologies, and specifically automated decision-making (ADM) systems, are deeply shaped by the cultural and institutional environments in which they are developed. Drawing on the works on the social construction of technology, the following chapters intend to look at how different cultures and specific human, social and political determinants influence the development and use of technology in unique ways. This reflects the main hypothesis behind this work, based on what Hughes defined as "technological style" (1987), which posits that different cultures influence the development of technology in unique ways. Hughes observed that national differences, such as organizational cultures, political structures, and economic policies, play a significant role in how technologies are designed, implemented, and used. These differences reflect underlying cultural values and priorities, with significant implications for how technology functions in practice. In the context of ADMs, AI and automation, these insights are particularly valuable for understanding how algorithmic systems can perpetuate inequalities, particularly those related to gender, race, and class. As Thouvenin et al. (2019) note, the very notion of "automated decision-making" is an oxymoron, as decision-making requires context, interpretation, and judgment, qualities that automation often suppresses (see Parasuraman & Riley 1997; Skitka et al. 1999; Cabitza et al. 2023). However, the contemporary literature on the impacts of automation on society, for example with regard to the amplification of existing discrimination and social

exclusion, has paid limited attention to the specific contexts and contingencies of technological implementation understood not as a product in itself, but as a project that is always situated and to be interpreted in conjunction with the specific cultural norms and social values and actors that influence it. Algorithms and AI systems used in increasingly more ways in order to automate certain processes in our lives - and I am not referring to online platforms, which belong to large, centralized companies and are part of a larger infrastructure, that of the Internet, which is more difficult to contextualize - which were previously in the hands of humans are particularly well suited for such research, as institutionalized within pre-existing, value-dense human practices. As Eubanks (2018) has shown in the case of some U.S. public services, these technologies rarely create new worldviews. However, almost none of these studies have worked to trace and unravel the specific dynamics that contextually influence the effects of technologies. While there has been extensive work on the impacts of automation and ADM systems, particularly on how they exacerbate social inequalities, there is a lack of in-depth empirical studies focusing on the social, cultural, and political determinants that shape these technologies in their design phases. Much of the contemporary research looks at the consequences of algorithms and automation, but fewer studies engage with what makes these technologies possible in the first place, examining the values and motivations behind their construction and contingent factors such as design choices, governance structures, and human decision-making that impact the final outcomes. As MacKenzie and Wajcman noted, «social scientists have tended to concentrate on the ‘effects’ of technology, on the ‘impact’ of technological change on society. This is a perfectly valid concern, but it leaves a prior, and perhaps more important, question unasked and therefore unanswered. What has shaped the technology that is having ‘effects’? What has caused and is causing the technological changes whose ‘impact’ we are experiencing?» (1995). This being the main purpose of this thesis, it will focus in particular on how automated processes are shaped and “caused” by following the actors who have an influence on them. Historically, the design and deployment of mainstream and

commercial technological systems have often excluded marginalized groups. In AI systems and automated processes, this exclusion is compounded by the fact that the algorithms themselves are often trained on data that reflect historical biases. This will be empirically observed in the context of essential services, particularly public services, which at a time when they are increasingly being automated, the question of how such procedures are shaped by cultural and institutional factors remains critically underexplored. Since history is strictly connected to culture, these biases cannot be analyzed in a vacuum.

This thesis will be structured as follows. After an account of the theoretical background from which this work originates and a review of the literature in Chapter 1, Chapter 2 will illustrate the methodology based on a case study approach, focused on a range of qualitative techniques and aimed at exploring the socio-political and cultural contexts that shape the algorithmization and automation of public services. Specifically, the research uses ethnographic methods to investigate the interactions between public actors and automated processes. The cases chosen, at the national level, are those of Italy and the Netherlands, due to their different approaches to the automation of public services (see DESI 2022). This approach allows for a detailed look at how these systems are used, incorporated, interpreted and understood within public institutions. This is further investigated through semi-structured interviews with key stakeholders involved in the design, implementation and supervision of ADMs in the Italian and Dutch contexts. Finally, it will involve the analysis of various documentary sources, examining government reports, relevant publicly available policy documents and legal frameworks, and also material obtained through generalized civic access requests.

In the context of ADM systems in Italy and the Netherlands, cultural values deeply influence how these technologies are adopted and perceived. Chapters 3 and 4 will explore these differences in the two empirical contexts, revealing how systems operate within specific national and institutional contexts. Chapter 3

focuses on the Dutch digital welfare system, which has been shaped by multiple cases intended to optimize fraud detection in welfare distribution and which disproportionately targeted vulnerable groups. The *Toeslagenaffaire* scandal serves as a critical case study, illustrating the role of institutional choices and the over-reliance on statistical correlations. This chapter also highlights how public backlash and the exposure of algorithmic discrimination led to policy reforms and calls for greater transparency in algorithmic governance. Chapter 4 shifts the focus to Italy, highlighting the bureaucratic opacity surrounding automation and its implementation in the country, where years ago an algorithmic system designed to assign teachers to schools resulted in widespread errors and dissatisfaction, showing how cultural and bureaucratic practices in Italy contribute to the black-boxing of algorithmic systems. Unlike in the Netherlands, where different scandals sparked public debate and reform, the Italian system has faced fewer challenges in terms of public accountability and transparency. Chapter 5 offers a cross-cultural analysis precisely of these topics, emphasizing how these systems lead to the erosion of responsibility across different cultural contexts. After strongly arguing for the need of contextual and contingent analyses to understand the causes and effects of automation and its social construction, the last empirical chapter discusses a result that should be understood in a wider cross-cultural perspective. It explores how public officials in both countries increasingly defer to automated systems, using them as a shield to avoid responsibility for decision-making and governance errors. The chapter examines the broader implications of this phenomenon, investigating how automation bias and the shift in accountability from human actors to technical systems undermine classical notions of legitimacy (Suchman 1995) and accountability (Bovens 2007). This cross-cultural analysis also draws attention to the differing levels of public scrutiny and reform efforts, illustrating how national contexts shape responses to the challenges posed by technologies. These national differences and similarities demonstrate that the design and functioning of automated processes are deeply shaped by social determinants and reflect the values, biases, and institutional

frameworks of the societies in which they are embedded, challenging the view that technology alone dictates social change. Through these case studies, this thesis aims to contribute to the literature by filling gaps in classical studies on the social construction of technology, especially with regard to the specific cultural and institutional determinants that influence modern high-impact technologies. Positioning itself at the intersection of technology, governance, and social justice, this work reconstructs several key constructivist works, believing that their core propositions are the most appropriate theoretical and methodological framework for analyzing modern digital and automated technologies. Indeed, the role of actors is perhaps even more relevant in the contexts discussed in this thesis, and it is thus crucial to reflect on the implications with respect to public accountability and values, understood as the benefits and outcomes that government institutions generate for society, beyond mere service delivery (Moore 1995). Studying the automation of public services, in fact, requires interacting with a broader body of studies concerning bureaucracy, public administration, and the role of institutions. In this, the key approach in the next few pages aims to re-center the attention on the people who work on or influence these algorithms, rather than on the algorithms themselves. One of the key arguments proposed in this regard will be that of institutional blackboxing, a core concept that describes the process by which bureaucratic systems - before algorithmic ones - become opaque, making it difficult to trace decision-making back to human or institutional actors. This blackboxing can occur through various barriers that shield the decision-making logic from scrutiny and take attention away from human beings. It aligns with broader concerns about algorithmic opacity and accountability, where decisions appear as neutral outputs of a system rather than contested political or social human choices. When algorithmic decision-making is framed as an unavoidable consequence of technological progress (in a deterministic stance), institutions may obscure their own role in shaping and deploying these technologies. Moreover, if decisions appear to be the inevitable result of complex networks rather than human choices, challenging their legitimacy becomes harder.

Moreover, by revealing the specific socio-political dynamics that shape automated decision-making systems in public institutions and social sensitive domains, the collected and discussed findings provide an important insight into the centrality of cultural differences in shaping the trajectories of automation and innovation, a timely reflection now that Europe and several states are regulating the use of different technologies and high-risk AI systems (European Commission 2024). This work, with common arguments in the social sciences, argues that even in the present, far from being inevitable and autonomous, every artifact is the result of precise choices and points of view that are more or less traceable, which it is our task to identify and increasingly highlight. Doing so can help unveil who is included and who continues to be excluded from innovation, as technological feminism did, and lead us to a systematic way of addressing the impacts of automation and algorithmization so that they can be critically evaluated and used to hold accountable those who continue to maintain the power to influence these technologies.

1. Literature Review & Theoretical Framework

The idea that technology and society are mutually constitutive (Wajcman 1991) is widely shared today. Although the view of technology as a stand-alone optimization process of social life that is independent of society still exists, the increase in social and cultural research driven by growing data gathering, computational power and technical sophistication shows that the bond between technology and society is strengthening (Chun 2017).

This was not always the case. In fact, the relationship between these two elements throughout history has gone through more complicated times, and in general there is little clarity about the different ways in which this relationship can be realized. Until the 1950s, technology was primarily conceptualized in its various economically driven transformations. During the industrial revolutions, technologies had played such a fundamental role that every social change was thought to be consequent and due to their introduction (Smith 1776). For some years it was taken for granted that they were the undisputed protagonists in the history of progress and chains of innovation: through linear models, initially prompted by scientific discoveries and then directly by the stimuli and demands of the market (Godin 2006), all social change hinged on technological adoption and its diffusion. Various research groups and scholars within Science and Technology Studies (STS) challenged this paradigm, noting that it normalized the relationship with technology and promoted uncritical adoption of it (Winner 1980). From these stances, an interesting theoretical and empirical tradition, full of nuances, contradictions and evolutions, has unfolded. The social studies on technology included various perspectives, including the social shaping of technology (SCOT) and actor-network theory (ANT) but also the various feminist contributions on gender and technology. It is one of the goals of this thesis to show how this vast academic and empirical production, although overlooked, has profoundly influenced the way we understand the relationship between technology and society

today and is still fundamental to further understanding it. When we ask the question about the neutrality of technology, for example, we do so thanks to the last two generations of scholars in these fields. How exactly has this literature unfolded in our time?

1.1. Contemporary Approaches to Technology and Society

In modern times, digital technologies are employed for the automation of crucial processes and decisions for people and society such as welfare, justice, education, labor and finance. As these uses are increasingly common, so is the existing literature on these practices and their relationship to society. Research dealing with automation in these contexts often discusses their technical features, spanning from artificial intelligence (AI) systems, data analysis, algorithms and statistical modeling.

We have been dealing with the automation of processes and the resulting human behavior for a long time (Friedman & Nissenbaum 1996; Stitka et al. 2000), but with the proliferation and collection of so-called big data (Mayer-Schönberger 2013) the prevalence of these practices has increased as new methods and formulas have emerged to identify relationships, networks, patterns and correlations between the enormous amount of data that a modern dataset can contain (boyd and Crawford 2012).

In recent years, the effects and impacts of these technologies have been addressed in many different ways, including those on the economy and labor (Kellogg et al. 2018), on education (Guilherme 2019), the impacts on policy and legislation (see e.g. in Europe the Artificial Intelligence Act, the Digital Services Act (DSA), the Digital Market Act (DMA), the Data Act and more¹), a normative perspective on what limits they should maintain (Mokander et al. 2021) and the impacts more closely related to society and culture. This thesis will mainly focus on the latter category and the observation of visible causes and consequences, e.g. the

¹ <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>

amplification of existing social inequalities or the decrease in equity due to the values embedded in these processes.

Literature focusing on these issues is also extensive. For example, many scholars turned on the datasets used to train the models (Barocas and Selbst 2016; Buolamwini and Gebru 2018; Suresh and Gutttag 2021; D'Ignazio and Klein 2020) as a cause for their potential discriminatory performance. Barocas and Selbst notably wrote that “an algorithm is only as good as the data it works with” (2016, 1) examining big data’s impact through the lens of American antidiscrimination law. O’Neil prominently examined the negative impacts on society of big data and algorithms and their power to perpetuate and exacerbate social inequalities (2016). She discussed negative impacts by observing some case studies in the areas of education, and specifically the use of value-added models (VAMs) for teacher evaluations, predictive policing in criminal justice, hiring algorithms in employment and credit scoring in finance. The author defined these tools as weapons of math destruction (WMDs), which operate without transparency, making it difficult for individuals to understand or challenge the decisions made about them. Both Jenna Burrell (2016) and Frank Pasquale (2015) also focused on the opacity and lack of transparency inherent in these processes, using the metaphor of the black box (as ANT scholars already introduced for technical artifacts) to explain the inscrutability of automated processes.

In addition to outlining impacts in real-life scenarios, some of the work also had a technical objective in highlighting the limitations of analyzing large amounts of data through statistical and AI-based methods. Buolamwini and Gebru’s prominent study sought to assess the accuracy and performance of AI-powered gender classification systems in the United States highlighting differences across various demographic groups, particularly focusing on intersectional characteristics like gender and skin type (2018). The study emphasized the importance of using diverse datasets in training AI systems to ensure fair and unbiased performance, since the error rates the authors measured were highest for darker-skinned females, paving the way for several researches that recounted the implications of automated

discrimination (Eubanks 2018; Benjamin 2019), particularly on intersectional identities (Crenshaw 1989).

When this scholarship began to emerge, another academic strand developed to propose mathematical solutions and interventions for solving bias and disparate impacts at a technical level, e.g. through different computational fairness and ethics formalisations (Lepri et al. 2018; Gajane, P., & Pechenizkiy 2017; Scantamburlo et al. 2018; Beretta et al. 2019). Other scholars, however, have also emphasized the need to address the discriminatory impacts of automation through data analysis as a socio-technical, political and cultural matter. For example, Selbst et al. criticized the prevailing approaches to fairness in machine learning, identifying several “abstraction traps” such as treating fairness as a purely technical problem without considering social and historical contexts (2019).

Many works are keen to contextualize the cultural and political bond of pre-existing practices that are being automated. The works of Ruha Benjamin and Virginia Eubanks, for example, go in this direction. Benjamin compares processes such as policing algorithms, facial recognition technology and predictive analytics in social services that disproportionately harm people of color to Jim Crow's late 19th and early 20th century racial segregation laws (writing about a “New Jim Code”). The author uses this comparison to highlight the potential segregative power of such tools and promote their development and use that actively dismantles systemic racism (Benjamin 2019).

The automation of complex, high-stake processes and decisions can have a significant impact in various contexts, but in some areas its effects may be particularly pronounced. For this reason, other scholars focused on specific organizational and empirical contexts to analyze social impacts such as the exclusion, unrecognition or systematic oppression of specific social groups. These include healthcare (McCradden et al. 2020; Lecher 2020), the criminal justice system (e.g. Angwin et al. 2016; Chouldechova 2017), employment and human resources (Hamilton 2018; Bogen 2019), education (Kolkman 2020) and public services (Chouldechova 2017; Richardson et al. 2019). Similarly to Benjamin,

Eubanks seeks to uncover how these technologies disproportionately affect low-income and marginalized communities, and more specifically explores how automated decision-making systems are used in public services and welfare programs in the United States. In addition to describing the social impacts of these projects and the consequences they had on those affected, Eubanks makes their political and cultural matrices explicit. She documents how decision-making can be obscured by automation, making it difficult to hold systems and administrators accountable for mistakes and injustice. Famously, she writes about a “feedback loop of injustice” to refer to automated processes that instead of alleviating poverty and inequality often exacerbate them by cutting off access to essential services and supports (Eubanks 2018).

Because of the profound transformative nature on different levels of these processes, philosophy is one of the disciplines that has been more concerned with these cases in recent years, to conceptualize and set ethical limits to these situations rather than to investigate them empirically. The ethics of technology, in fact, has ancient roots in philosophy (Horkheimer and Adorno 1947; Marcuse 1964; Habermas 1968) but its potential was revived in recent years precisely with datafication and the development of techniques for analyzing the enormous amount of data. These analytical techniques have to do with statistics, data science but also more traditional computer science (Floridi 2010; Johnson 2009; Weckert 2007; van den Hoven & Weckert 2008), and in particular automation and artificial intelligence (AI) (Nucci & Santoni de Sio 2016; Mittelstadt et al. 2016; Bostrom & Yudkowsky 2014; Wallach & Allen 2009, Coeckelbergh 2020b). The ethics of technologies, however, has quickly become a deeply interdisciplinary area of research, connecting philosophy, the social sciences, and law. In some cases, this literature has been addressed with acronyms such as FAccT (Ziosi et al. 2024), which denote shared principles such as Fairness, Accountability, and Transparency. In this case, traditional ethics are taken in the form of normative principles to analyze their applicability to the governance and design of machines and AI systems (Mittelstadt 2016; Ananny & Crawford 2018; Dignum 2018; Floridi 2019;

Lo Piano 2020; Floridi 2023). Elsewhere, some specific ethical and social implications of AI, such as the human-machine relationship (Rogers 2012), discrimination, and the automation of decision-making (Lepri et al. 2018), have been extensively addressed by this literature. This thesis, in particular, focuses on the latter two dimensions, which in recent years have been increasingly addressed in academia at the intersection with journalism (Angwin et al. 2016), regulation (OECD 2019; European Commission 2019) and civil society (AlgorithmWatch 2019).

Some of the studies presented so far have focused on the automation of public services and their delivery by public and governmental bodies with respect to their social consequences and impacts. This is precisely the empirical context that this research will analyze, given its relevance for social studies. I consider it to be particularly sensitive and worthy of a closer attention, since it is where decision-making carries a higher degree of responsibility in ensuring individuals and groups' public and/or essential services, and more generally the promotion of the common good (Broomfield e Reutter 2022). As already noted by O'Neil in the case of her WMDs (2016) and many other authors, in fact, one of the most important characteristics of these processes applied to public or essential services is their scale: they affect large numbers of people, amplifying their impact. Given their potential to embed existing biases and discrimination, high-responsibility decision-making contexts such as public organizations and administrations can lead to a cycle of marginalization and exclusion of certain groups. Moreover, the lack of transparency and accountability in algorithmic systems undermines democratic principles, as individuals often cannot contest or understand decisions that affect their lives, causing even greater problems for the public interest.

Several authors have analyzed the implications of automation for governments and public services. Some works focused on the consequences for human discretion (Young, Bullock & Lecy 2019; de Boer & Raaphorst 2021) and other challenges of integrating AI systems into public governance, such as accountability

(Meijer & Bovens 2003; Bovens 2007b; Meijer & Grimmelikhuijsen 2020; Cobbe et al. 2021; Cooper et al. 2022; Wieringa 2022). For example, Young, Bullock & Lecy differentiate between routine automation (simple, rule-based tasks) and non-routine automation (tasks requiring discretion and judgment), in order to consider the impacts based on the nature of tasks and highlight the potential benefits of AI for efficiency and decision-making while cautioning against risks related to transparency, accountability, and equity that erode public good.

Investigating this issue in public services necessarily means considering bureaucracy in its traditional features and how they respond to automation. In recent years, relevant scholarship on this emerged, e.g. Meijer et al. (2021) studied the “algorithmization” of bureaucratic organizations through ethnographic methods, focusing on the changes that occur at the organizational level and less on the study of the technological process, with a comparative case study about predictive policing in the cities of Amsterdam and Berlin. Drawing on psychology and public administration literatures, Alon-Barkat & Busuioc (2021) investigate automation bias as the overreliance on algorithmic advice and other possible biases as consequences of the use of such processes in bureaucratic contexts. Wihlborg, Larsson and Hedström investigated the relationship between civil servants, citizens and automated-decision making in Swedish public administrations, discussing the division of responsibilities (2016). In other words, the topic can be approached both by asking what happens to individuals and groups when public services are automated and what happens to organizations.

Although academic research on the social impact of automation has evolved enormously in recent years, including empirical studies that promote inside-the-black-box analysis (Pasquale 2015; Bonini and Gandini 2020; Buolamwini and Gebru 2018) and categorization of different possible harms (Buçinca et al. 2023), an organic field of research that assesses the actual impacts on social groups of automated processes based on mathematical models and artificial intelligence systems in relation to their root causes (such as their design

and governance) has never been fully developed. After presenting the evolving and more recent contemporary study of the relationship between technology and society, and before outlining what I consider to be the gaps in these literatures to which this research intends to contribute, I will briefly present the main classic reference literature in social studies on technology, which I believe can still greatly support and sustain research in this direction.

1.2. The Social Shaping of Technology

Starting from the 1970s, an extensive and diverse body of social research on and around technology gradually took shape. The starting point was to counter the dominant conception about technology, namely, that it was the driver of all social change. This idea, reflecting a typically reductionist approach, drew on technological determinism, which pointed to the assumed autonomous nature of technological development. Social scientist Thorstein Veblen was among the first to talk about technological determinism in the second half of the 19th century and to leave a theoretical legacy on these issues, and in particular on the idea that social changes, the passing of the eras, and major economic and political events are determined by technological revolutions (1899). According to a determinist perspective, technology is endemic to the development of each historical stage and as a productive force, based on its logic of efficiency, it determines social structure, cultural values and ideology, according to a self-managed and uncontrollable motion.

For a long time, technology was interpreted as instrumental to itself. Theodor W. Adorno and Max Horkheimer (1947) give ample evidence of this and trace it back to the logic of the Enlightenment. Technology was distant from society, parallel, not responding to real social goals. Karl Marx's political and economic philosophy contains many determinist elements in the way he presents and treats technology. At its core, there was the idea of the productive forces determining the conditions of society, and in particular of labor: the relationship between workers

and capitalists is articulated on the basis of these forces (1867). Technology, as a productive force, is part of the structure, which produces and influences the political, ideological and religious superstructure of society. Technological determinism, in retrospect, is emblematically identifiable in a quote by Karl Marx: “The hand mill leads to a society with the feudal lord; the steam mill to a society with the industrial capitalist” (Huyskes 2024; Shaw 1979). Although Marx never openly took a stand to defend a determinist logic (and indeed was very critical of technology), his arguments along with many others contributed to this rationale that shifted power to technologies as inevitable.

By the early 1970s, the sociology of science was a well-established field that questioned the objectivity of scientific knowledge, epistemology, and the different explanations each person might have of the same “facts” (e.g., Barnes 1974; Barnes & Edge 1982). On this basis, the sociology of technology, joining innovation studies and the history of technology, has emerged to open up technological artifacts to social analysis. Interestingly, in fact, technology was already considered as a black box by social scientists and economists, who studied it from the outside but never “from the inside” (Bijker and Pinch 1984, 404). In order to confront the determinism that had prevailed up to that time, Bijker and Pinch also questioned the linear model of innovation, that is, the classical interpretation of technological change that takes place through a series of linear and consequential steps driven initially by scientific research and oriented towards production (Godin 2006). According to this model, each technology proceeds solely on the basis of its economic success, in a smooth and rational manner, as if the world we find ourselves in were the perfect outcome of a predetermined and successful plan (Bijker and Pinch 1984). In general, these social approaches to technology have sought to highlight anomalies, changes in trajectory, demonstrating non-linearity and non-determinism, but adding a bold new piece to the argument that had remained implicit or undirected: every technological artifact is a social construction. This position is situated more broadly in the strand of social constructivism, a theory of knowledge that counters the rigidity and inevitability of

certain characteristics - such as gender, race, class, which are considered biological and immutable - by arguing rather that they are products of human interpretation shaped by cultural and historical contexts (Subramaniam 2010). The technological constructivism of Bijker and Pinch and many other authors (see MacKenzie and Wajcman 1999) is just one of the many approaches that the social sciences adopted toward technology in those years. Others were actor-network theory (ANT), critical theory that arose within the Frankfurt School (Horkheimer and Adorno 1947; Marcuse 1964; Habermas 1968), and philosophical and political theories on artifacts and innovation, as will be introduced. Among the main building blocks of this social inquiry, for example, was the seminal study on laboratories conducted by Latour and Woolgar (1979; Woolgar, 1982), from which constructivism drew inspiration for its empirical and observational approach.

Social constructivists studied everything that contributed to the establishment of a technology in the marketplace and society, and thus to its existence. Historian David Noble contributed greatly by simultaneously investigating a technology which “made it” and one that “did not make it” to show their symmetrical development (1978). Michel Callon, as a sociologist, decided to focus explicitly on technological controversies drawing on extensive case study of the electric vehicle in France to show the high level of negotiability of any choice and decision (1980; 1991). To investigate the social construction of technology means taking into account many more factors than just the external observation of what is there and what is not, and requires a particular internal look. These studies, in other words, have rejected the inevitability and autonomy previously entrusted to technology, refuting the self-generation paradigm. Instead, they worked to show that technology does not unidirectionally determine human action. Rather, there is always at least one human action that shapes technology in the first place. This is why motivating the existence of a technology as “the best” sounded partial and misleading: these authors were interested in unveiling the definition of the criteria and metrics that made it such (Huyskes 2024). To study technological development

and its influential factors, constructivists borrowed from the sociology of scientific knowledge an empirical methodology useful to focus on the problems, controversies, and failed uses. In deciding what is a problem or not, “a crucial role is played by the social groups concerned with the artifact, and by the meanings which those groups give to the artifact: a problem is only defined as such, when there is a social group for which it constitutes a ‘problem’” (Bijker and Pinch 1984, p. 414).

Bijker and Pinch’s proposed method of observation for studying the relationship between technology and society was that of interpretive flexibility. For example: if for some an artifact (or some of its components) constitutes a problem and for some others it does not, it means that that artifact possesses a certain degree of flexibility. This means that it does not have the same degree of appreciation, functioning and acceptability for everyone. In other words, interpretive flexibility consists of the way different social groups may conceive the same technology (Huyskes 2024).

FIGURE 7
Some Relevant Social Groups, Problems and Solutions in the Developmental
Process of the Penny-Farthing Bicycle

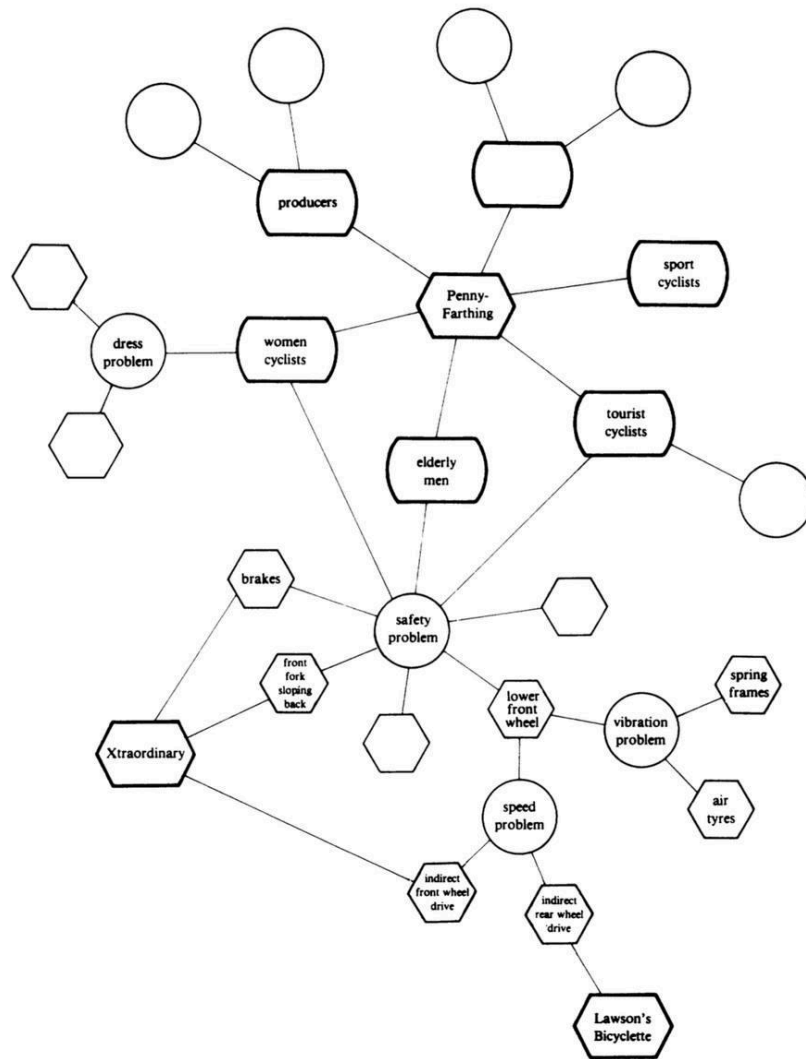


Figure 1: Interpretative flexibility scheme for the bicycle (Bijker and Pinch 1984)

Bijker and Pinch used interpretative flexibility, a formal method of inquiry, for the social construction of technology, and applied it to the example of the bicycle. In order to study different interpretations, the authors divided their observations into different relevant social groups. These groups may represent entire institutions or more spontaneous groups of individuals, the relevant factor being that they share the same interpretation or meaning attached to that technology (Bijker and Pinch 1984; Bijker, Hughes and Pinch 1987). Based on the problems identified

(and shown here in Figure 1, with respect to the specific bicycle example), there could be different solutions, with different levels of compromise, negotiation and optimization for different groups (e.g. women, elderly men, professional and sport cyclists). As Judy Wajcman also wrote, “technology is always[...] the result of conflicts and compromises, the outcomes of which depend primarily on the distribution of power and resources between different groups in society” (1991, 162).

What determined the transition from the historic Victorian bicycle, with its large front wheel, to the modern chain bicycle? The Penny Farthing, which became obsolete in the late 1880s, provided high speeds but was highly dangerous because the rubber-covered wheels did not provide stability and suffered from the vibrations of road bumps. For this reason, it was mainly used by wealthy young men, not as a real means of transportation in itself but mainly for recreational purposes. Other social groups were naturally excluded from its use (Huyskes 2024). For example, its design was unsuitable for women, as it did not accommodate dresses and skirts, but only trousers. The main reason for the initial modifications and redesign of the bicycle was its safety, which led to the development of air tires (Woodforde 1970). This innovation was initially ridiculed and dismissed as aesthetically unpleasant, especially by engineers, until its use demonstrated the increased usability for multiple social groups. Interpretive flexibility, in fact, applies not only to what different social groups mean by the same artifact, but also to how they are designed, because “there is not just one possible way, or one best way, of designing an artifact” (Bijker and Pinch 1984, 421). According to the constructivist study of technological development, which focuses on problems and their solutions instead of the linear, determinist model, the negotiation and interpretation phase, characterized by controversies and solutions, is followed by a stabilization phase, a “closure mechanism” in which the most relevant social groups “accept” a certain version of an artifact that then settles in the marketplace (Bijker and Pinch 1984, p. 429). This is what Thomas P. Hughes referred to as the “technological momentum”, that early stage in the life of a technology when it is

most susceptible to human intervention, since when it stabilizes more than naturally begins to influence and enter society, institutions and even cultural norms (1994).

As also constructivists pointed out, in addition to the consolidation of a methodology, more research was needed to study the cultural and social influence of some choices at the expense of others. It was necessary to look more closely at these social groups and the power dynamics between them. This is what other scholars focused on, shedding light on the more clearly political influences in the processes of artifact construction and development.

To claim that every technology incorporates specific social values, and is constructed with them, is to reject the idea that there are neutral technological processes. In line with the constructivist, non-linear model of innovation, Donald Mackenzie and Judy Wajcman have deeply criticized this approach, arguing that the only way to define a technology as neutral is to not know the purpose it will have to fulfill, if it has never been used before and especially if the people who created it had no idea about its intended purpose (1999). Mackenzie and Wajcman noted gaps in the social sciences, which despite calling it social “construction” have in fact mostly limited the reflection - albeit important - to study technological impacts and effects on society, without asking what made those technologies possible in the first place, i.e. their causes and determinants. This approach, they argued, tends to view technology as inevitable and to ask questions such as “how can society best adapt to changing technology?” (Mackenzie and Wajcman 1999, 3) instead of asking how to adapt the design of technology to the society to which it will apply. To advance the instances of interpretive flexibility, the authors also stated that “the same technology can have very different ‘effects’ in different situations.” But what are the social conditions that cause these different effects? The question about social shaping was the one they intended to ask, beginning with Langdon Winner’s influential paper on the politics of artifacts to show that some technologies can be designed as “gatekeepers”, for opening or blocking access to specific services, goods, or rights. Famously, Winner draws on the bridges built by architect Robert

Moses in Long Island, New York in the 1930s and 1940s as a case study. Moses' bridges, built in order to open up new possibilities, were the perfect example of a technology that, in its design, represents and includes specific political values (Winner 1980). In this example, the height of the bridge did not allow for the passage of public transport, specifically buses, which were too tall to pass. The political scientist concluded that this infrastructure was designed to have a particular social effect, that of excluding social groups that used public transportation, such as poor or Black people, from enjoying certain spaces, in this case Long Island. According to Winner, and others after him, racial discrimination influenced the (social) construction, as in many other cases of urban design.

Another strand in constructivist research on technology is that carried on from the Frankfurt School, focusing on politics and the transmission of values. For example, philosopher Andrew Feenberg, a student of Herbert Marcuse, believes that it is not the technology itself, but its design, that determines what human beings are and will become. Technological design, then, is "an ontological decision fraught with political consequences" (Feenberg 2010). With an argument grounded in the philosophy of technology, to which Martin Heidegger and Jacques Ellul, among others, had already contributed, Feenberg argued that starting from the Industrial Revolutions we have always conceived technology as an economic and commercial product, historically equated with progress in itself. This has distanced it from the empirical conditions of existence (Feenberg 1999). In other words, technological development has been assimilated with economical development and not necessarily social development. Feenberg spoke of a "technical code" to describe the realization of an interest or ideology in a technically consistent solution to a problem. Specifically, it is a criterion that selects among alternative technical projects those that are feasible. "Feasible", in this case, means technically achievable (Feenberg 2010).

Other scholars such as Bruno Latour, Michel Callon, Madeleine Akrich and John Law have developed actor-network theory (ANT) by taking a different path in studying the relationship between technology and society. The starting point, in

fact, was not to treat them as two separate spheres that influence each other (as constructivists theorized), but by analyzing what they have in common, namely the network (Huyskes 2024). Networks connect human beings with other entities, such as nonhuman artifacts, that mediate, normalize, impose, and make certain things possible, and because of this they acquire agency in the sociotechnical network. ANT has contributed substantially to the study of the role of design and designers, allowing it to move out of the purely humanistic study of technology and expanding the discourse to other audiences. Among the many examples used to untangle the morality of things, Bruno Latour gives that of artificial speed bumps, which force drivers to slow down in place of human policemen but performing the same role, and even more, since the threat is to damage the vehicle and the driver himself (Latour 1992, 1999).

Madeleine Akrich argued that “a large part of the work of innovators is that of ‘inscribing’ this vision of (or prediction about) the world in the technical content of the new object.” She refers to what follows this relationship as the “script” or “scenario”: like a film script, “technical objects define a framework of action together with the actors and the space in which they are supposed to act” (Akrich 1992, 208).

1.3. The Gendered Shaping of Technology

One of the most important empirical constructivist studies of technology for the way it traces the entire life trajectory of an artifact is Cynthia Cockburn and Susan Ormrod’s study of the microwave oven (1993). Their book, which grew out of a European project aimed at understanding how new technologies influence gender relations, was intended to trace the alliances between the actors involved in the construction of a consumer good that was already circulating, observing how they behaved, what they said, and showing the multidirectionality of each innovation or, in other words, the directions that are decided for each artifact and that influence its results.

References to the work of Bijker, Pinch and Hughes are explicit where they write that technology is culture, that every idea and artifact are social constructions resulting from negotiations among different actors and groups. But the main purpose of Cockburn and Ormrod's microwave study was to observe the impacts of a given technology on society (and specifically on gender relations) as directly dependent on the choices made in the design, production, and commercialization of that technology (Huyskes 2024). Cockburn and Ormrod spent considerable time at different Electro UK sites, one of the companies that designed and sold the microwaves, to conduct empirical observations. To study the effects of this technology with respect to gender relations, the authors firstly looked at them within the company, because they imagined they might already be reflected in the artifact. They were not wrong, because they first discovered that women did not occupy management or scientific positions, but rather the factories and stores that received them for selling. Interestingly, women also occupied another position: that of culinary consultants in the Test Kitchen, a laboratory where they were asked to give input on culinary knowledge, recipes, and test different models (Cockburn and Ormrod 1993). Another element that makes Cockburn and Ormrod's work important and worth mentioning in relation to this thesis is that they have followed the microwave throughout its life cycle, from the factory to the stores. It was precisely when they went to the stores that they discovered that the configuration completely changed there. If in the company, from the design but also from the sales perspective, it was taken for granted that the end user would be a woman, the final marketing of the product revealed the opposite: microwaves were initially sold as hi-tech "brown goods", such as televisions, video recorders, cameras, and electric shavers. Once it entered the market and users' homes this configuration did not prove successful. People realized that the microwave actually was a kitchen appliance, and they quickly forgot the identity of the technology that had been promised to them from the store shelf. At that point, they considered rethinking its placement, moving it among "white goods", alongside white appliances from refrigerators to hair dryers, that according to store managers naturally interests

women and families. In this case, then, in this product life chain, its characterization seems to have changed at least three times: the one originally attached to it by designers, the one given by retailers, and, finally, the interpretation that end users attributed to it. With the goal of tracing gender relations, this study tracked the division between domestic economists (the consultant-housewives) and engineers, who inscribed their choices in the product but also negotiated different values and needs. These included, for example, the specific needs of women as cooks (promoted in the Test Kitchen) but also the inclusion of high-tech features as designers desired. By opening the construction of artifacts to social analysis, these social and cultural values can be revealed, and this is where the importance of empirical studies such as that of the microwave oven lies. This is true for the constructivist literature on technology, and more generally for better understanding the social implications of the technologies we use. For example, other feminists focused on other domestic technologies, such as the washing machine, in order to understand how the progressive hyper-technologization of household tasks was conditioning care work in the home (in line with a slogan of the Italian feminist movement of the time that claimed “super-mechanized, but always fatigued” (see Vanek 1974, Cowan 1979; 1983). Some of them concluded that the advent of domestic technologies and the engineering of homes - rather than liberating women, as marketed - had reinforced gender roles, particularly caregiving roles intended for women, locking them into these oppressive patterns. Because of this, these artifacts seemed to be built more to meet the needs of designers than those of women (Wajcman 1991, Cockburn and Ormrod 1993. Berg and Lie 1993).

Studies focusing on the interaction between gender roles and technologies, besides fitting within the framework of constructivism, have carved out a field of research of their own. Gender and technology studies arose for a very specific reason. Some scholars who were already contributing in the field of Science and Technology Studies (STS) noticed that a fundamental gender lens was missing in critical reflections on the social construction of both science and technology. Crucially, they pointed out that in the texts and methodologies proposed by the

social shaping of technology approach (SCOT), the “shaping” failed to consider gender dynamics (Wajcman 1991; 2004; 2009, Gill & Grint 1995, Faulkner 2001). Dynamics that, along with others, constitute a broader and thicker power network within our societies. For example, based on the principle of flexibility, SCOT proposed to investigate the relevant social groups and their possible different interpretation of the same technology. These included those who propose it, design it, finance it, adopt it, and modify it (Pinch and Bijker, 1984; Bijker, Hughes and Pinch 1987). But what happens to those who do not fall within these groups? Women, in fact, normally fell only among those who used that technology (as in the case of domestic or reproductive technologies) but were typically excluded from decision-making, design and managerial positions, and thus from the crucial stages of construction (Wajcman 2004). Some scholars proposed to look more directly at how the exclusion of women - and gender relations in general - influenced the social shaping of technology. The extensive literature that can be reconstructed and that constituted the broad field of gender and technology studies looked at highly practical cases, addressing mainly (i) domestic technologies (ii) reproductive technologies (iii) military technologies, (iv) urban technologies, and (v) the Internet and the web. The former include the microwave oven study and other broader contributions on the impact of technologies in the home (Cockburn and Ormrod 1993). Much work was focused on reproductive technologies, characterized by both empirical and theoretical studies. Judy Wajcman referred to them as “technologies of life”, and noted how they were able to divide feminist movements and literature. In fact, reproductive technologies such as in vitro fertilization, egg donation, sex predetermination, the birth control pill, and the speculum (Stanworth 1987) represent the most controversial topic within feminism with respect to technology, and also where the distance between constructivism and determinism came back into focus (Wajcman 2004). On the one hand, were those who embraced and even encouraged the advent of such technologies as a revolution for women's lives, as they would grant them greater control and freedom over their bodies, and saw them in line with the struggles being fought for

abortion rights (e.g. Firestone 1970). On the other side, however, were the feminist scholars who saw reproductive technologies as yet another means for men to control women's bodies, arguing that it was always them who were creating, directing and giving purposes to these artifacts, turning women into what Gena Corea metaphorically called "The Mother Machine" (Corea 1985) (see also Rowland 1985, 1988; Mies 1987). Feminist scholarship also dealt with public, private, and urban space, and with what has represented a growing research in the last years too: the exclusion of women in the design of spaces (see Saegert 1980, Criado-Perez 2019), but also with military technologies as a masculine social construct to fulfill the "obsession with control" (Wajcman 1991; see Easlea 1983).

The relationship between technology and female reproduction also traces the more general relationship between technology and science. What divided feminism over these specific applications actually divided them primarily with respect to the distinction between nature and culture, or between what is considered by some to be naturally masculine or feminine and what is instead socially constructed. In the former case, gender binarism and essentialism were defended as the cultural basis of sexual difference, while in the latter, more space was left for more fluidity in identities. In this divide, technology played an opposite role: in the former case it was treated as a tool of control in the hands of men and of a culture that was already oppressing women, while in the latter, technology could facilitate their emancipation and liberation. For example, much ecofeminist literature (e.g., Merchant 1980) dismissed the technological innovation of the time as irremediably tied to a masculine standpoint, detached from nature and people. At the other end of the spectrum, political and optimistic expectations were placed in technology (e.g., Plant 1998). At the other extreme, they relied on political and optimistic expectations in technology (e.g., Plant 1997). This enthusiastic conceptualization of technology was already present, as mentioned, in its applications of domestic and reproductive artifacts. However, the idea of technology as a force to be reclaimed expanded furthest in the realm of computation. On this basis, specifically, the cyberfeminism of Donna Haraway was born to propose a positive synergy between

machines and humans with the metaphor of the cyborg (Haraway 1991). Haraway's goal was above all epistemological, to unhinge the supposed scientific objectivity and investigate specificity without plunging into relativism (Braidotti 1995). Haraway worked on a feminist epistemology that similarly to other authors proposed to question the objectivity of traditional scientific knowledge, emphasizing the importance of considering social and gender positions in the process of knowledge production. However, while Haraway focused on the broader social situation and the interconnectedness of nature and culture, feminism focused on science specifically looked at the role of gender in the scientific process (see Harding 1986). In line with Haraway, Plant stressed the importance of women's active participation in shaping and using digital technologies to ensure a more equitable and inclusive future. But her specific analytical framework was centered on the virtual space, where gender can be shed and identity can be redesigned and reinvented. Even more significantly, Plant extensively argued her thesis that - unlike what the dominant culture advanced and what feminists who were concerned with technology argued - machines are more like women than men (Plant 1998). This led Plant, in a way similar to that already unveiled by Haraway, to propose an alliance between women and machines, not because they are "naturally" similar, but because perceptions and expectations toward both are influenced by the same power structures (Plant 1998; 1996).²

Despite the extensive contribution each of these works has made to the broader literature on gender and technology, the division in feminism over the more pessimistic and the more utopian-optimistic interpretations has never really been solved. As Judy Wajcman - who has analyzed the evolution of theories of gender and technology from a constructivist perspective - has repeatedly pointed out, the feminist project in this sense is incomplete (Wajcman 2004). Much of the literature on gender and technology had focused on the exclusion of women, first from technological work and then from the design of artifacts, to expose how

² A very similar but more recent proposal is also found in Hester, H. (2018). *Xenofeminism*. Polity Press.

technologies were contributing to maintain the status quo and shape a patriarchal society. In doing so, they progressively focused more and more on technology (either as a tool of control or as a force for empowerment), attributing much power to it. In this way, from both essentialist and utopian perspectives, both views ended up falling back on determinist formulations, from which traditional studies of technology (despite their lacking on other grounds) had distanced. For this reason, Wajcman argued that the feminist project was “incomplete” (2004, 127), in the sense that it alternately fell into essentialism or forced optimism without finding a new ground. However, she proposed to use many of these important contributions, e.g., gender blindness and the focus on women’s exclusion, Haraway’s materialism, empirical inquiry, and contingency to bring attention back to the initial goals of constructivism (1991; 2004; 2006; 2009). The problem with the early studies was that while they focused on the inherently social and political construction of technology, they did not in fact act as if their methodologies were influenced by the social context. For example, they did not reflect on the preponderance of white, privileged, heterosexual men in their analyses, not only as investigators but also as investigated (Wajcman 2004, 126). Wajcman referred to the need to reframe interpretive flexibility (Bijker and Pinch, 1984) and other methods of inquiry that put the analysis of design processes at the center (e.g. in line with Cockburn and Ormrod, 1993), including in their exploration of controversies, limitations, and alternatives, to continue to show technology after technology that rarely the choices that significantly contribute to make them possible are inevitable.

1.4. From Effects to Causes

In gathering and reconstructing this literature, I aim to highlight two major aspects. First, how considerable is the conceptual and empirical work that has been undertaken since the 1980s to investigate the relationship between technology,

politics, and society, and how these frameworks are still highly relevant today. Second, among social studies, how crucial the contribution of gender and technology studies has been in outlining for the first time an academic inquiry into the different impacts that a technology can have on different social groups. In other words, on how different social groups can interpret and conceptualize the same technology differently, based on the antecedent values and determinants embedded in its design. All these works have unveiled the fundamental link between social construction, design choices, political and social determinants, networks of actors (such as the trade and sales chain), controversies and what follows. These were essential contributions in the direction of a real analysis of choices and impacts on certain categories of people that can still provide us with much guidance today, both methodologically and conceptually. However, the limitation that we can recognize in this literature today, in retrospect, is that it falls into the same trap denounced by constructivism: that of analyzing only one manifestation (albeit material and contingent) of power, in this case that exercised by men over women. While dealing with gender, many of these studies - even the least essentialist - have in fact dealt predominantly with women without empirically distinguishing between different women, practicing specificity, but addressing them only as a category. In addition to this, they did not interact with a political and social analysis that aimed to intersect different forms of discrimination (gender, social class, age, origin, religion...) and oppression, such as intersectionality in the legal field (Crenshaw 1989) or the matrix of oppression for social and political studies (Collins 2000).

Starting from the denunciation of gender blindness in social studies on science and technology, several other forms of blindness remain to be addressed. Although the goal of constructivist feminist research and action was explicit and situated in a cultural-historical moment, that of the 1970s and 1980s, in which instances were in line with second wave feminism political demands that started from women's liberation and the conquest of their rights (cf. Firestone 1970; Harding 1986), failed to deliver on its promises to propose a feminist epistemological alternative related

to the development and adoption of technologies with respect to different identities and social groups. One of the shortcomings found in the classical studies on gender and technology, e.g. by Wajcman, is that it was an incomplete project, in the sense that in its two opposing expressions it assigned “too much agency to new technologies and not enough to feminist politics” (2004, 127-128). Moreover, although the feminist reflection on technology has been taken up mainly by materialists, in some cases it has expanded empirical reflection without providing new practical proposals for structural intervention.

On the other hand, the contemporary literature on the social impact of technologies, as highlighted at the beginning of this chapter, has also worked in recent years to seek practical solutions to the problem of discrimination, for example by working on computational fairness and other statistical metrics (Bellamy et al. 2018; Dwork et al. 2012; Hardt et al. 2016). In addition to this, contemporary studies are also concerned with what is allowed by regulations and have a normative ethical approach, identifying different principles with which to direct the development and use of technologies (Mittelstadt et al. 2016; Coeckelbergh 2020b). However, the contemporary literature on the relationship between technology and society lacks an in-depth approach that opens the box to social analysis as much as possible and searches for the causes, determinants, and shaping factors related to impacts - such as the amplification of discrimination or exclusion - in construction practices (the interaction between actors, design choices, life cycle...). Only a few of the aforementioned studies have focused in any significant way on the causes, but have tended to concentrate only on the effects, consequences and impacts of algorithms on society (MacKenzie and Wajcman 1999). When they do, they normally take a principles-based approach, noting the lack of general oversight, accountability and fairness which, although necessary, does not look at contextualized and contingent factors. This is also evidenced by the great attention that studies which propose technical and quantitative corrections ex-post continue to receive, revealing a responsibility placed on the more technical aspects that are less directly related to human choices, such as data

collection and dataset construction, as if “a machine learning model merely reflects the algorithmic biases existing in the dataset and does not itself contribute to the harm” (Hooker 2021). Instead, as constructivism emphasizes, the search for the social, cultural and political determinants (Huyskes 2024) that led to the construction of the technology in a certain specific way is crucial to understanding its consequences and possibly modifying its design (MacKenzie and Wajcman 1999). It seems, however, that a reflection more akin to constructivism has been lost. Datafication and artificial intelligence have been widely studied from feminist and social justice perspectives in recent years (D'Ignazio & Klein 2020; Browne et al. 2023), but not with the empirical methods proposed by constructivist feminism on technology. In particular, I am interested in returning to the idea present in the early social studies on technology that, in order to identify the social, cultural and political traits inscribed in design, it is necessary to examine the different meanings and values attributed to it by different groups of actors (Bijker and Pinch 1984; Bijker, Hughes and Pinch 1987).

Given that feminist politics remains one of the main sources of contestation of inequalities and privileges in a world where it often seems that previously achieved gains can be easily lost (Wajcman 2004), it seems crucial to take up these important contributions and apply them beyond gender. I think that this literature and some of the main principles it has advanced through the ethnography of certain socio-technical systems is useful today to represent as many exclusions from the use of technology as possible and to trace the origins and consequences of these choices, which are to be sought again in design and in particular in how it relates - embodied by the people who deal with it - to the specific values and contingent culture of the context where that design and planning takes place.

This research attempts to make two contributions to the literature. First, I will propose to empirically investigate some negative impacts of automated decision-making processes in the specific realm of public services, expanding the reflection to contingent causes, cultural values, and human decisions with respect

to the design trajectories that influenced the final outcomes (Bijker and Pinch 1987; MacKenzie and Wajcman 1999; Wajcman 2004). More specifically, I am approaching socio-technical contexts wondering what would happen if, in observing the results, we trace back social and political intentions inscribed during design by innovators and scientists. Second, and most importantly, I believe that innovation, design and the interests that drive them, not just the consequences of technology on society, should be of primary concern for social studies, and I believe that the application of constructivist methodologies like interpretative flexibility and contingency (Haraway 1991) to the study of automated decision-making processes can help us in a broader understanding of these phenomena, especially in identifying methodologies for their prevention.

2. An Ethnomethodology to Reveal the Social Construction of Algorithms

This research aims to investigate the social shaping factors, i.e. determinants, and impacts of the algorithmization and automation of public services. More specifically, it observes some of their social effects in context, in order to trace their potential origin in design choices. The broader interest is to observe how cultural, social and political values influence the socio-technical construction of automated processes - including their governance, the actors which are involved in them and their choices - in public administrations and agencies.

This takes the form of a main research question and two sub-questions:

Q.R. 1: How do cultural, social and political values shape and influence decisions regarding the design and use of automation and algorithmization in public services?

S.Q. 1: To what extent can the design choices and social construction of automated decision-making (ADM) explain the negative impacts that subsequently occur?

S.Q. 2: What are the key actors influencing these choices of both technical and social construction, and generally bringing cultural, social, and political determinants into it?

These research questions will be addressed through a case study approach, considering the cultural, social, political factors which shaped specific cases. The approach is based on qualitative techniques, specifically qualitative document analysis, ethnographic observation, interviews and content analysis. The following paragraphs will detail and motivate the case study approach. Next, the choice of methods will be briefly explored, introducing data collection procedures, the positioning of the researcher in the field and general methodological limitations.

2.1. A Case Study Approach for Technology-in-Context

The nature of digital technologies is one of continuous becoming, thanks to the data streams used to train them with which there is a relationship of mutual influence (Gillespie 2012). The datasets chosen and used are intended to tell portions of reality, mostly historical samples that deal with specific contexts and should ideally serve the analysis to be carried out. Algorithms, AI systems and automated processes are socio-technical systems embedded in specific contexts and cultures. They are technologies, in Thomas P. Hughes' words, "creative activities, individual and collective, of craftsmen, mechanics, inventors, engineers, designers, and scientists" (2004, 11). In reality, the networks of actors that influence the construction of contemporary algorithms can be much more broad. For these reasons, in recognising their social and not just technical nature, it is important to study these technologies and their effects in their contexts, to analyze the specific surrounding causes that shaped them in a certain way (boyd 2010), as we can only grasp the responsibilities and roles behind each technology when analyzed in their

situatedness, partiality and locatedness, rather than transcendence (Haraway 1991). For example, Orlikowski's theory of technology-in-practice is particularly useful to this research approach, as it focuses the attention not primarily on the technology itself, but rather on what people - or in this case, public organizations - decide to do with it (2000). Similarly, as Christin points out in her analysis of the use of algorithms by web journalists and legal professionals, there is a lack of analysis in actual practices, while much of the debate focuses on the instruments and not their use (2017). This work, incorporating this perspective, intends to add the filter of context to analyze technologies not as such, but in the roles that people entrust to them and their contextual uses.

Algorithms often arise in practice in relation to other computational forms (the data, models, calculations and codes) or, in the case of public administrations, arise within already potentially codified institutional, administrative and bureaucratic processes that give them meaning and assign them a role. Potentially, the search for the cultural and contextual elements that lead to the birth and construction of an algorithmic or automated practice can be pursued *ad infinitum*. This is not the aim of this project, but rather to understand how the algorithm, or the mathematical model it then becomes, comes to play the role assigned to it. This role will take part in a larger system, in a particular network, with a specific mathematical configuration, specific chosen metrics to explain their performance according to the observed characteristics (Hull 1996) and above all, it will have particular effects based on all its specific characteristics. For all these reasons, the study of technology embedded in a culture, as put forward by this research, will rely on a case study approach. This will not leave aside considerations of continuity and certain cross-cutting factors common to the adoption of automated decision-making processes by public bodies, but rather it will be of greater interest to note even these similarities in the presence of very different dynamics present within single settings (Eisenhardt 1989). As reconstructed by Eisenhardt (1989), case study research has been common both in the study of specific technological settings, innovation projects and decision-making situations. This type of research

is not necessarily aimed - as in this case - at comparison, but rather at the possibility of gaining access to the specifics of different cases and situations within their real-life context, providing detailed insights.

The choice and sampling of cases in case study research is primarily based on theoretical (not statistical, see Glaser & Strauss 1967) sampling. In doing so, choosing “polar” and more extreme cases (Eisenhardt 1989) in order to ensure greater transparency and observability of results can be helpful (Pettigrew 1990), especially in the study of processes and phenomena that are scarcely investigated in the existing literature. Therefore, in order to select the socio-cultural contexts that would serve as cases for this study, I started from two theoretical factors: the already evident presence of events and projects related to the automation of public services and the ranking of the different European countries in the DESI index (Digital Economy and Society Index, 2022). The choice of restricting the field to European contexts depends on the fact that, in order not to make the value observation excessively broad and complex, I decided to limit it to a context that shares regulations (such as the General Data Protection Regulation and the Artificial Intelligence Act) and common principles (such as the European Commission’s Trustworthy AI framework published in 2019) on the use of automated processes in the decision-making context, while bringing together different countries and cultures with different adoption and implementation indices. Looking at these data, I chose Italy, one of the most underdeveloped countries in the automation and transparency of public services, and the Netherlands, ranked as one of the most automated and transparent governments in Europe (DESI Index, 2022). Out of 27 total countries, the Netherlands is ranked 3rd and Italy 18th, where 'the uptake of other key technologies such as big data and Artificial Intelligence is still quite limited' (DESI, 2022).

Nonetheless, both of these countries in recent years used two algorithms in their public services that contributed to a problematic impact on the lives of citizens, not only as individuals but especially as social groups. Namely, the Netherlands has been defined as the first European country to carry out mass

data-driven discrimination of citizens (Hadwick & Lan 2021) for a scandal that unfolded in 2021 related to the allocation of child benefits and alleged fraud in applications reviewed by an algorithm. Wrongfully reported fraud has led to bankruptcy and the loss of children for thousands of families. A similar although less tragic case occurred in Italy between 2016 and 2017, linked to the “Buona Scuola”, a major education reform that reorganized teachers allocation and mobilization with an algorithmic process. Errors in the output led to several rulings on the use of the automated practice. Unlike in the Netherlands, where a lot of media attention and several researches have already been placed on this scandal and its consequences, in Italy what happened did not trigger a public and research discourse on the construction, impacts and governance of these technologies on society, as DESI data also show. With respect to this digital agenda, unlike the Netherlands, Italy remains below the European average. Out of 27 total countries, the Netherlands ranks 3rd while Italy ranks 18th, and where 'the adoption of other key technologies such as big data and artificial intelligence is still rather limited'. (DESI Italy, 2022, p. 4). In particular, Italy ranks 19th in the EU in terms of digital public services, with a gap compared to the European average precisely in relation to the offer of digital public services to citizens (67 in Italy against 75 in the EU, DESI Italy, 2022). On the same topic, however, the Netherlands outperforms the European average by 10 points (85 in the Netherlands 75 in the EU, DESI Netherlands, 2022). Finally, with regard to governance, while the new government formed in Italy in 2023 dismissed the Ministry for Digital Transformation (replacing it with an undersecretary of State for the country's digital affairs), there is an increasing focus of Dutch governance on the digital transition (DESI Netherlands, 2022). Within the chosen cases, the Netherlands and Italy, specific scenarios, public administrations and bodies with different objectives and values were selected to allow for environmental and dimensional variation (Eisenhardt 1989).

2.2. Observing Humans Behind Algorithms

This research is based on various qualitative techniques. The rationale behind the choice to structure its design on exclusively qualitative methods is reasoned and motivated. While in fact involving algorithms, AI systems and other digital technologies with a quantitative functioning as the object of study, this thesis is primarily interested in observing how humans in particular contexts behave towards them, why and how they use them and what roles they attribute to them. Therefore, the choice has been to focus on the study of human actors, the organizations they are part of and their choices. This can be traced through a mix of different qualitative and ethnographic techniques, in particular interviews, participant observation and document analysis. This choice turned out to be the best based on the research objectives already anticipated, in particular the need to root and situate technological practices within their socio-cultural context and observe their interpretation (Cardano 2011) according to the groups of relevant actors identified. As Eisenhardt puts it, a diverse set of techniques is the most recommended choice for case studies (1989). This, in fact, provides freedom to make controlled changes during the data collection process, e.g. the addition of cases that emerge during collection. In this, the author reminds us, fieldnotes are particularly helpful, because they allow a continuous focus on what is happening in both observation and analysis (Ibid.).

The ethnographic approach to the study of algorithms has become widespread in recent years, especially for the study of cultural and organizational influence (Caliandro & Gandini 2016; Caliandro 2018; Bonini and Treré 2024). Christin (2020) provides a valuable review of the ethnographic literature on digital technologies. However, the majority of these studies focus on platforms and the algorithms that govern them, or on private contexts such as those in Silicon Valley where engineers are driven by capitalist logics (Christin 2020). Nonetheless, these same methods can be applied to the context of public organizations, to study the

values involved there (Orlikowski 2000). More importantly, I think this methodology is most akin to those introduced and pioneered by social constructivists for early artifact studies (Pinch and Bijker, 1984). By observing the interactions between the different actors and their responsibilities at play, one can capture elements that motivate or refer to the final effects of the technology they control and employ (MacKenzie and Wajcman, 1999) (cfr. Case Study Approach).

The STS literature has repeatedly shown that ethnographic methods are particularly suited to analyzing emerging practices in the development and evolution of scientific and technological artifacts (Latour and Woolgar 1986; Orlikowski 2000). Christin noted the need to include ethnographic methods in the study of algorithms to overcome algorithmic opacity by decentralizing the analysis: “instead of focusing on the ‘black boxes’ of algorithms, ethnographers can study how collectives of human and non-human actors emerge, solidify and evolve over time” (2020). This was the same technique proposed by Seaver with reference to information technologies: “A great deal of information about algorithmic systems is available to the critic who does not define her object of interest as that which is off limits or intentionally hidden” (2017, 7; see Wieringa 2023 for a practical adoption).

Similar to Christin (2017), this work adopts a multi-sited fieldwork approach, as it does not focus a priori on specific technologies, but rather on their social effects and to follow the work of the actors involved. This approach also finds affinities with the ethnomethodology inaugurated by Harold Garfinkel (1974) and used among others by Bruno Latour (2005) in the study of human and non-human actors, in order to study actors’ behavior towards the social world in a contextualized and localized way. In particular, the phenomenological framing of these approaches is useful in the observation of the more ‘naïve’, natural and unprogrammed human interaction vis-à-vis the automation of processes and their construction. In this, Alfred Schütz’s social phenomenology as a method of analysis oriented towards the observation of actors and their experiences that are

never identical, as they are always situated in different positions and motivations, finds several points of encounter with constructivist interpretative flexibility (Pinch and Bijker, 1984). Feminist epistemology as described by Donna Haraway (1985) and Sandra Harding (1986), with particular reference to the criticism of the objective point of view, also referred to as God's trick (Haraway 1985; D'Ignazio and Klein 2023) and attributed to the male, privileged, white gender, is greatly inspired by phenomenology and has much in common with ethnomethodology, in situating practices related to technology and its experiences, an approach that this thesis seeks to learn from and follow.

In order to transpose a constructivist analysis as close as possible to the one inaugurated in the 1980s with interpretative flexibility (Bijker, Hughes and Pinch 1987), therefore, in this research the actors and their organizations are followed, which in specific empirical cases play an important role in the development, governance, technical and political management, understanding and interpretation of certain ADM processes as socio-technical systems.

The key in this process is to study the enrollment (Callon 1986), the technical code (Feenberg 2010), the technological style (Hughes 1987) or the inscription (Akrich 1992) of social, human cultural values in technologies. In order to bring this constructivist and feminist theoretical approach as far as possible within the methodology guiding this research, (i) the case study approach was chosen, in order to be able to get to know each selected context as thoroughly and as much as possible as a cultural and value entity in its own right (Eisenhardt 1989) before generalizing it, and (ii) to support the selection of privileged participants through the snowball sampling technique, which in my opinion is the closest to the principle of "following the actors".

2.3. Following the Actors

The “snowball” or “chain” sampling, described by Noy as a constructivist and feminist hermeneutic, can generate a “unique type of social knowledge, an emergent, political and interactive knowledge” (2008, 328). Particularly useful for investigating organic social networks and the social dynamics of specific groups of actors, because it consists in the researcher accessing informants through contact information that is provided by other informants in a repetitive process where “informants refer the researcher to other informants, who are contacted by the researcher and then refer her or him to yet other informants, and so on. Hence the evolving ‘snowball’ effect, captured in a metaphor that touches on the central quality of this sampling procedure: its accumulative (diachronic and dynamic) dimension” (Noy 2008, 330).

It is relevant to note that Pinch and Bijker’s exhortation to open artifacts to social analysis (1984) in the case of automated decision-making processes and artificial intelligence systems is made more difficult than in the case of other technologies such as the bicycle, the microwave oven (Cockburn and Ormrod 1993) or ballistic missiles (MacKenzie 1993). This is because of the layered, complex and immaterial nature of software, which makes it easier to treat it as a black box (Pasquale 2015) where the only things that can be easily accessed are the inputs (in the case of the technologies of interest to this research, the data) and the outputs (the automated or semi-automated recommendations or decisions). Latour wrote back in 1999 to refuse to accept black-boxes as inevitable and that “scientific and technical work is made invisible by its own success. When a machine runs efficiently, when a matter of fact is settled, one needs to focus only on its inputs and outputs and not on its internal complexity. Thus, paradoxically, the more science and technology succeed, the more opaque and obscure they become” (Latour 1999a, 304). Efficiency and success, however, are terms that need to be questioned in order to understand exactly how they are measured (Feenberg 2010).

What this research proposes to do on a methodological level is to go beyond this assumption, to analyze AI-based processes beyond data (Mantelero 2022; Renieris 2022) - on which the FaccT and bias literature has focused heavily in

recent years (cf. Theoretical Framework, Contemporary approaches) - to see what is the role of other observable components (e.g. choice of parameters, weight of variables, analysis techniques, choice of model) in addition to the dataset construction, which can be limiting and disempowering, in shaping the outcomes (Purtrova and Huyskes 2024) and to shift the analysis “from final products to production, from 'cold' stable objects to 'warmer' and unstable ones [...] before the box closes and becomes black” (Latour 1987, 21).

2.4. Researcher’s Positioning in the Field

To contextualize at least part of the fieldwork, it is necessary to clarify my positioning as a researcher in this field. Since 2019, I have been part of an Italian non-profit association called Privacy Network, which deals with digital rights at several levels. From September 2020 to January 2024, in particular, I volunteered as the Head of the Advocacy & Policy team, and led several projects including the Automated Administration Observatory (OAA). Thanks to this role, I had access over the years to different contexts related to both national and European policy-making and governance strategies on the topics of technology, AI, ethics and fundamental rights. Thanks to my previous involvement in these themes, in early 2021 I closely witnessed the outbreak of a scandal in the Netherlands, named *Toeslagenaffaire*, related to the automation of controls to detect social fraud in the context of a particular public benefit. The Dutch public debate, the study of the context and some information shared by colleagues from civil society inspired me to direct this research and begin a study around similar cases in Italy. For instance, thanks to the work on the OAA, I was able to analyze several documents obtained through access requests under the Freedom of Information legislation (FOI) (for a detailed proposal on how to use these requests to facilitate social research, see Savage & Hyde 2012) sent to central public administrations suspected of using algorithms in the provision of public services. The purpose of the FOI procedure

for the OAA was to map the use of algorithms used by Italian central public administrations in order to ensure transparency for citizens, and proved to be useful for this dissertation, particularly for Chapter 5. Starting from these documents and responses received before and during the past three years of research, I was able to collect some data on the attitude, awareness, competence and general sentiment of public administrations with respect to the automation of decision-making processes and public services. Also thanks to this project, I participated and intervened in a number of public events, having the opportunity to get to know several people related to digital governance in Italy at a central level. The most relevant factor related to my positioning described here is the fact that, starting in 2019, I have entered the field and made contact with a number of privileged witnesses from the Italian context, whom I was then able to interview for this research. These include Heads of Cabinets for different ministries, technical staff working for different governments, people who have served as public decision makers in Parliament, people at the forefront of major companies offering IT solutions for government, journalists, and various other relevant actors. All of this prior knowledge favorably guided the choice of context and relevant actors to be mapped around a particular technology.

2.5. Data Collection

The collection of qualitative data through technical documents, ethnographic observation and interviews took place from March 2022 to March 2024 between Italy and the Netherlands.

In particular, between March 2022 and December 2022, I conducted a thematic analysis on the Italian context through newspaper articles and information available on the websites of different public bodies to gather more data about this topic and relevant actors to build on. Specifically, I manually searched for relevant information (see Denzin & Lincoln 2000) using keywords such as “algorithms”; “artificial intelligence” and “automation” on government and media sites. I started from (i) information gathered in this way from the public debate, especially the names of relevant actors, (ii) information previously obtained through my work with Privacy Network, and (iii) the analysis of various documents, to identify the first people I would approach to invite them for an interview.

Three smaller sections will follow in order to detail each method that has been used. To analyze the qualitative data collected through interviews, participant observation, and document analysis, I employed thematic analysis (Braun & Clarke 2006). This approach allowed for the systematic identification of recurring patterns of meaning within the data. First, I conducted open coding, identifying meaningful segments relevant to the research questions. These initial codes were then grouped into broader thematic categories, reflecting both general analytical dimensions drawn from the literature - such as actors, design choices, and embedded values - and case-specific categories, including national contexts (Netherlands, Italy) and key events (e.g. Toeslagenaffaire, SyRI, Buona Scuola). This dual-level categorization facilitated a structured yet flexible analysis, enabling a comparative understanding of how automated decision-making systems are shaped by distinct cultural and institutional determinants. In the processes of coding, selecting and transcribing the elements most relevant to this thesis that emerged from the

interviews and ethnographic observations, I opted for consistent anonymization of key informants. This decision was guided by the need to maintain uniformity among the different actors. Some of the people I interviewed, both in the Netherlands and in Italy, have public profiles and have already spoken publicly about the cases discussed in this research. While deletion of anonymity might have strengthened their positions, other people interviewed needed to maintain anonymity about their positions and would not have participated without it being guaranteed. To ensure the inclusion of their valuable testimonies, I thus chose to apply anonymization uniformly to all sources.

2.5.1. Interviews

In an overall - but not continuous - period between December 2022 and March 2023, I conducted a total of 31 interviews. Of these, 14 were conducted in Italy and 17 in the Netherlands. The decision to conduct interviews at different times over almost three years was motivated by the constantly changing nature of the object of study. In fact, discourses on the automation of decision-making processes and the use of algorithms or AI systems in public services have largely changed and evolved over the course of this research, both in Italy and the Netherlands. The distanced data collection between the two countries allowed me to follow these changes and to get a more reliable picture of the values and culture behind the management of these projects.

In order to decide which and what kind of actors I was interested in interviewing for this research, I first identified, based on the constructivist literature, which actors are the most relevant to shape a technology (Bijker and Pinch 1984; Bijker, Hughes, and Pinch 1987; MacKenzie and Wajcman 1999) or, alternatively, to understand it and have more information about it. Based on this work, I have identified five macro-categories: (i) Public agency or institution; (ii) Academia; (iii) Civil society; (iv) Independent; and (v) Media. I used these labels to

target my search for key informants, thus both starting from them and categorizing my informants ex-post in some cases in these categories. All of these categories were useful in giving me an idea as comprehensive as possible of the same cases from several different angles, providing me with complementary perspectives from different points of view (Haraway 1991; Wajcman 2004). For example, among these surely there were informants who were more privileged than others in the first group, where people also with leading positions in some public agencies or relevant agencies on these issues fell. With their consent, for this purpose, I also decided to keep the name of the organization they work or worked for when deemed useful. In the total number of key informants interviewed (N 31) between Italy and the Netherlands, N 12 informants belonged to category (i); N 6 to category (ii); N 5 to category (iii); N 6 to category (iv); and N 2 to category (v).

	DESCRIPTION	CATEGORY
1	Person working at the Data Protection Authority (DPA)	Independent
2	Lawyer	Independent; Expert
3	Person who worked in a central public agency in charge of digital innovation	Public agency or institution
4	In-house company operating in the ICT sector	Public agency or institution
5	National Social Security Agency	Public agency or institution
6	Office of the Prime Minister	Public agency or institution
7	Teacher who was affected by an algorithm	Independent
8	Teacher who supported many affected colleagues and coordinated a response as a director for a trade union	Independent; Civil society
9	Journalist	Media
10	Secretary for a teachers' trade union	Civil society
11	Staff member working on the digital agenda of a political group in the Senate of the Republic	Public agency or institution
12	Research staff member at the European Commission	Public agency or institution
13	Member of a governmental commission of experts on AI	Independent
14	Person working at the Data Protection Authority (DPA)	Public agency or institution
15	Professor at a University	Academia

Table 1: A detailed list of the people interviewed for this research in Italy

Between December 2021 and January 2022, I started trying to obtain access to organizations to carry out interviews in Italy. I started from a base of initial privileged informants (N 3) affiliated with some of the main Italian public bodies that I knew had experimented or supervised process automation in different sectors and going through different political moments. In order to follow the constructivist principle of following the actors, I then asked them both in the invitation and during the interview to point out to me other people who could play an important role in terms of development, governance, technical and political management, understanding and interpretation of automated decision-making

processes in the Italian context (cfr. Section 2.3) (Noy 2008). This chain sampling gave me in some cases a more direct and trust-based access to a number of organizations that were difficult to access from the outside. My prior knowledge of certain internal logics of organizations, acquired in previous years through my work with Privacy Network and especially through my participation as a stakeholder in a number of meetings with these people, further benefited my access to the interviewing process: in addition to being a researcher, in their eyes, I also understood the complexity of the networks in which they moved, even at the legislative level.

During January and February 2022, I carried out fieldwork and conducted an initial set of interviews (N 8) in Italy, with university professors, employees of governmental digital departments, lawyers who are experts in the digitisation of public administration, and people working in public companies that develop technologies for administrations. I attended several conferences, workshops and delved into technical documents that could drive the data collection and analysis. For the rest of the year, I thoroughly analyzed this first part of the data and manually analyzed it by looking both at the literature and the documentation in my possession so that I could organize the work in its subsequent stages.

DESCRIPTION	CATEGORY
16 Researcher in a University	Academia
17 Researcher in University	Academia
18 Staff member at the Ministry of the Interior and Kingdom Relations	Public agency or institution
19 Technical external advisor for the government	Independent
20 Person working for a governmental internal agency focused on ICT	Public agency or institution
21 Former data scientist at the <i>Belastingdienst</i> (Tax Authority)	Public agency or institution
22 Professor at a University	Academia
23 Professor at a University and technical expert	Academia
24 Person working at an NGO focused on digital human rights	Civil society
25 Person working at an international NGO on human rights, heading the digital rights team	Civil society
26 Investigative journalist	Media
27 Lawyer working for an NGO focused on strategic litigation to defend human rights	Civil society
28 Opposition Member of Parliament	Public agency or institution
29 Person working at the <i>Auditdienst Rijk</i> (National Audit Office)	Public agency or institution
30 Professor at a University	Academia
31 Person working at the Data Protection Authority (DPA)	Independent

Table 2: A detailed list of the people interviewed for this research in The Netherlands

From June 2022, I worked on the preparation of the interviews and ethnography in the Netherlands. By reconstructing the literature and public discourse around the *Toeslagenaffaire*, I gained access to some of the key relevant actors in the field of digitisation and technology in the Dutch central public administration. Between March 2023 and June 2023, I was based and affiliated at the University of Utrecht in order to carry out fieldwork in the Netherlands and conduct interviews (N 17).

Finally, after these two phases, I conducted a final set of interviews in Italy between October and December 2023 (N 6).

2.5.2. Participant Observation

The analysis aimed at collecting data on the interaction, negotiation and compromises discussed between the different actors involved (Callon 1986; Pinch and Bijker 1984; Latour 1987) was based on participant observation in seven different contexts over the overall data collection period.

	TITLE	DESCRIPTION & FORMAT	LOCATION
1	Workshop Algorithms & Human Rights in Public Services	Closed meeting addressed to public agencies; open discussion	Amsterdam
2	Workshop Technology & Human Rights	Event open to the public and different communities; keynotes and open discussions	Amsterdam
3	The Future of AI, Chamber of Deputies	Event organized by the parliamentary intergroup on AI with invited policy-makers, legislators and relevant stakeholders, open to the public; keynotes and panels	Rome
4	Artificial intelligence and the future of cultural diversity, Senate of the Republic	Closed event organized by a Senator in collaboration with a private company with invited policy-makers, legislators and relevant stakeholders; panels	Rome
5	State of Privacy 2023	Closed annual event and workshop organized by the Data Protection Authority (DPA); working groups and keynotes	Rome
6	Governance challenges for AI	Event organized by ForumPA and Microsoft Italy for public administrations and major service providers; panels	Rome
7	ForumPA 2024	Annual event and multi-day trade show dedicated to public administration and new challenges	Rome

Table 3: A detailed list of events where participant observation was carried out

The observation component in the qualitative analysis is based on the principle of technology-in-practice as drawn by Orlikowski (2000). For example, in her paper she wrote that “Over time, through repeated reinforcement by the community of users, such technologies-in-practice may become reified and institutionalized, at which point they become treated as predetermined and firm prescriptions for social action, and as such, may impede change” (2000, 411). In this sense, observation provides insights into real-time interactions, relevant to specific institutional and bureaucratic settings discussed in all three empirical chapters.

These lenses have been applied in particular on these occasions (N 7) to see how the algorithmization (Meijer et al. 2021) of public services evolves and what characteristics it shows on a socio-cultural level. The value of directly observing the interactions of some of the relevant actors in these contexts is to be able to witness their exchanges of information and behavior with respect to these technologies by seeing them within the actor-network, and not just isolating them from it. What emerges is a picture of practices that informs attitudes towards the technologies, including the language used to refer to them, the negotiations and above all the power attributed to them, e.g. degree of trust and reliance.

I attempted to attend all useful events in which I was invited, in some cases as an observer, in others also as a participant and stakeholder (as a researcher or as a member of a civil society organization) and engaged in various formal and informal conversations. These were events, workshops, conferences, roundtables, and meetings that were always followed by more informal networking sessions. All of these events share a common focus on AI, new technologies, their governance, and their use by government or public agencies. In some cases, actors representing public administrations explained what they were doing with automation in public services or internally. Both when I was part of the stakeholders (and thus, for example, participated in panels or gave keynotes) and when I participated “only” as a spectator, these were all valuable opportunities to undertake my own participant observation and get closer to a community that I knew only partially before starting this research. Since this is a very closed world, the institutional one

especially, in many cases it turned out to be the best way to weave relationships and get access to a range of actors, projects and documents. In addition to this, field notes served to document my feelings and observations, which were then layered on top of the rest of the data I collected and my theoretical background to follow a grounded theory approach (Glaser and Strauss 1967).

2.5.3. Document Analysis

	TITLE	DESCRIPTION	Owner
1	DESI Index, 2022		
2	Strategic Program on Artificial Intelligence 2022-2024		Italian Government; Agenzia per l'Italia Digitale (AgID)
3	Italia Digitale 2026	Goals and initiatives for digitalization in the National Recovery and Resilience Plan (PNRR)	Italian Government; Agenzia per l'Italia Digitale (AgID)
4	Unprecedented Injustice		Dutch Parliament, House of Representatives
5	Processing the nationality of childcare benefit applicants		Data Protection Authority
6	Dutch Digital Strategy		Dutch Government
7	An Audit of 9 Algorithms used by the Dutch Government		Algemene Rekenkamer (Netherlands Court of Audit)
8	SyRI legislation in breach of European Convention on Human Rights	The Hague District Court judgment in a case about the Systeem Risico Indicatie (SyRI)	Rechtbank Den Haag
9	Recovery operation childcare benefits	Parliamentary document	Dutch Parliament; Standing Committee on Finance
10	Xenophobic Machines	Report that investigates the <i>Toeslagenaffaire</i> case from the perspective of human rights	Amnesty International
11	Response from the Ministry of Education	FOI request	Ministry of Education

Table 4: A detailed list of the documentation analyzed

Finally, but in no chronological order, completing the data sources I analyzed for this thesis were a number of documents (N 11) of different kinds. These include government reports, relevant publicly available policy documents and legal frameworks, and also material obtained through generalized civic access requests, which often helped me to contextualize and better understand what participants were referring to in both interviews and ethnographic occasions.

2.6. Limitations

A disadvantage of traditional ethnographic methods is that they can sacrifice breadth for depth, choosing as in this case two specific contexts for a long time rather than comparing as many cases as possible. This potentially makes the research more prone to selection bias (Flyvbjerg 2016). To address this shortcoming, the ethnography carried out aimed at a selection of as many different possible contexts and actors within the two case studies as possible, with the aim of rooting the analysis in two specific and different cultures without any claim to generalization. Nevertheless, by choosing organizations with different values and magnitudes, actors with different roles and responsibilities (consistent with the constructivist ideas of following the actors and interpretative flexibility) and involved in different situations and at different historical periods in the two different countries, this work succeeds in gathering useful and original information about the differences and similarities with regard to the object of study.

But ethnography only guarantees the study of what is accessible. For this reason, the risk of bias has been minimized by adopting approaches to algorithms as socio-technical systems and collecting a range of data around them, not only by interviewing the relevant actors but also by obtaining the relevant technical documentation in some cases.

Another limitation of this type of study relates precisely to interviews and the difficulty of accessing even qualitative data about organizations, their practices and choices, overcoming their opacity with regard to the technologies used, not just the technologies themselves. Christin (2020) points to corporate secrecy, technical illiteracy, incomprehensibility and size as factors that make it inherently difficult for ethnographers to focus their analysis on algorithms. These factors also depend on culture and some contingent factors concerning the two main case studies discussed in this chapter. In particular, a difficulty in accessing certain organizations (starting with those defined to begin the analysis) was noted in Italy, as will be extensively illustrated in Chapters 4 and 5. This, however, although to be regarded as an objective limitation of the research due to limited access to information, will be culturally contextualized in this thesis and also treated as a main finding.

3. “Unprecedented Injustice”: Network Determinism in the Dutch Digital Welfare State

3.1. Introduction

Among the various tasks entrusted to automation is that of predicting a future state of things. A range of different techniques are, in this sense, increasingly employed in order to automate decision-making processes and performed through the analysis of large amounts of data by means of algorithms and mathematical models. These data analysis techniques are usually based on machine learning (ML), from simple linear regression - and many other classic statistical operations - and supervised predictive algorithms to techniques considered closer to artificial intelligence (AI) (such as unsupervised ML). In any case, this class of algorithms is specifically designed to make statistical predictions that create probabilistic models of future outcomes and events based on historical data.

Some scholars wrote about algorithmization to describe the codification of procedures in an organizational context to automate or support a process, e.g. decision-making (Gillespie 2014; Meijer et al. 2021; Veale & Brass 2019). If the automation of decisions implies a process - a trend that more and more entities and organizations are embracing due to technological readiness and the drive to innovate, algorithmization signifies the procedure through which these processes are made possible. Among the many public services, both local and national, to be automated are justice, healthcare, tax collection, education, law enforcement and social services (Dencik and Kaun 2020). In recent years, these practices have become so widespread that the term “digital” has been added to classical analysis of the welfare state to describe the use of technological tools such as AI, automation, big data, and predictive analytics to identify risks and optimize the delivery of public and social services (van Zoonen 2020). Philip Alston, U.N.

Special Rapporteur on extreme poverty and human rights, noted the initial spread of these practices in a 2019 U.N. General Assembly call and pleaded for them to «change course significantly and quickly to avoid stumbling into a digital welfare dystopia» (Alston 2019).

Although algorithmization in this context is implemented to increase efficiency and optimize processes, an extensive literature has shown how it can perpetuate bias, inequity and amplify social discrimination and exclusion (Barocas, S. & Selbst 2016; O'Neil 2016; Eubanks 2018), reduce transparency (Ananny & Crawford 2018) and accountability (Kemper & Kolkman 2018) of governmental agencies. To investigate what happens when welfare is automated and what factors contribute to the negative impacts of this process, this article draws on in-depth qualitative research conducted in the Netherlands. The study of the Netherlands, which in recent years has been considered the European “laboratory” of algorithmization and automation of public services, allows to closely analyze its impacts and some of its main characteristics. Firstly, how historical and ideological conceptions of welfare and certain social groups have shaped these processes. These, like other human and cultural values, were incorporated into the technical design and led to discriminatory and dramatic results. Secondly, the absence of human supervision and maintenance of algorithmic processes, which denotes a widespread automation bias. Thirdly, the fact that the algorithms used in this context are capable not only of automating pre-existing injustice, discrimination and social exclusion, but also of amplifying them, as the literature has already shown (Eubanks 2018). Most importantly, this article aims to contribute to the literature by answering the question: how does this happen? What are the cultural and social determinants that are involved and imprinted in the process of design and construction of these technological practices shaping them in turn (MacKenzie and Wajcman 1999) and causing the amplification of these discriminations, and in which way is this occurring? In particular, I will argue that we should interpret these projects in light of network determinism, the scenario I propose from observing the Dutch case study on automated welfare administration based on resemblance. Resemblance

through connectedness, imposed through the division into rigid categories and “clusters”, implies that people who resemble others or are similar to others in some way are condemned to the same fate. I refer to this as *network determinism*, recalling the more classic concept of *technological determinism*, the theory that interprets technology as the main force driving change in society, its institutions and structures, culture and human habits. Network determinism, as a contemporary evolution and extension of technological determinism, is how most automated decision-making practices are designed and used today by the public agencies that decide to implement them, and is based on ancient structures of computation and surveillance of social groups, rooted in statistics, old demography and, even before, eugenics (Birhane & Guest 2020). This empirical and critical analysis draws on constructivist “opening the box” methodologies (Bijker, Hughes, and Pinch 1987) to discuss how these designs are implemented today and can help prevent the implementation of potential problematic and discriminatory socio-technical systems.

The first section of the chapter is dedicated to the theoretical background, where a more detailed account of what I mean by these concepts and their historical origin will be provided, including a brief discussion of the theories and methodologies that have dealt with determinism previously and to outline the theoretical framework around automated decision-making in public services today in order to motivate the aims of this chapter. This will be followed by a brief methodological section, to illustrate the data collection methods relevant to the empirical context of the Dutch case study. The third section will get into the core of the case study, tracing the history of welfare automation in the Netherlands and the most important cases that have characterized it in recent years with problematic impacts on people’s lives. Section four, in light of the empirical context, will specifically reflect on the social and cultural determinants that have directly caused these negative impacts being codified in the design phases. This will also lead to a discussion of the different responsibilities of the actors involved and the importance of the different stages of socio-technical construction. Mainly, this

part will use empirical data to present the concept of network determinism, pointing out its harmfulness and suggesting to interpret some negative consequences of the automation of public processes - such as the amplification of discriminations - as a necessary consequence of dangerous resemblance-based correlations. In order to overcome these, we cannot only look at technical and mathematical corrections, but must necessarily identify the socio-cultural roots of each design and application context in order to prevent the fate of many in an increasingly automated public landscape from becoming more and more predetermined and oppressed.

3.2. Theoretical Background

3.2.1. From Technological Determinism to Network Determinism

«The hand mill leads to a society with the feudal lord; the steam mill to a society with the industrial capitalist» (Marx 1847). This excerpt by Karl Marx, which would deserve a long elaboration and contextualisation that unfortunately will find insufficient space here, perfectly sums up some aspects of the idea of technological determinism. Marx, though not a technological determinist in the strict sense, suggested that technological advancements in the means or forces of production would inevitably lead to changes in relations of production, i.e. social relations, and economic structures. Technological determinism, more broadly, is a theory that suggests that technology is the primary force driving changes in society, shaping social structures, cultural values, and human behavior. According to this perspective, technological innovation and development are seen as autonomous and inevitable (Smith & Marx 1994), determining the direction of societal progress and historical change, often minimizing the role of human agency, social context, and cultural factors (Huyskes 2024). Technological determinism has been convincingly supported over time (e.g. by Thorstein Veblen, Clarence Ayres,

Jacques Ellul, William Fielding Ogburn). In modern treatises on technology and its relationship to society, however, it is not necessarily embraced in a direct and forthright manner. Rather, it is the taken-for-granted classical paradigm by which the relationship between technology and society is interpreted as self-evident. While much of the discourse on technology has always been pervaded with a determinist rhetoric, at a certain point in history - coinciding with a growing critique of technological and scientific means that were being used in a way that was for some unjustified and uncontrolled for malevolent purposes, such as nuclear power, or in an extractive way to the detriment of the Global South (MacKenzie and Wajcman 1999), or on women's bodies (Huyskes 2024) - several scholars began to openly critique this paradigm and question it. Social constructivism, between the 1970s and the 2000s, developed and practiced empirical methodologies to measure and show the role of society and culture in shaping the development and adoption of technology. The relationship between technology and society cannot be interpreted as a mere cause-effect relationship, but rather technology must first be interpreted as an effect of society. Many technological constructivists, both "classical" and in its feminist articulation (Wajcman 2004), have insisted on investigating the social and cultural causes that led to technological "effects", contrary to what determinism implied (Bijker and Pinch 1984; MacKenzie and Wajcman 1999).

According to the French sociologist and philosopher Ellul, technology had become an autonomous force shaping every aspect of human life. In 1954, Ellul wrote that "in our technological society, technique is the totality of methods rationally arrived at and having absolute efficiency (for a given stage of development) in every field of human activity" (1954, 25). Ellul speaks of "technique" as meaning not only machines or tools, but the totality of methods and practices aimed at efficiency and optimisation. Technological society, according to Ellul, privileges efficiency and rationality above all else. This leads to the standardization and homogenisation of human life, where decisions are made on the basis of technical criteria rather than human values that drive towards

predetermination. For Ellul, efficiency and optimisation are what make technology and automation potentially more problematic.

In recent years, the adoption and spread of big data has stimulated and encouraged the idea that through large and unprecedented amounts of information we can understand complex social processes (Barocas & Selbst, 2016; Leurs & Shepherd, 2017). The concerns for optimisation and efficiency reported by Ellul and the tendency to explain highly complex human and social phenomena through data analysis have one root in common, that of statistics. Correlation and probability calculus were founded in the years in which Ellul wrote, with generally different aims than the discipline we know today. Some social constructivists, such as Donald MacKenzie, have sought to deepen and problematise some of the epistemological - and above all practical - foundations on which this discipline developed (1981). MacKenzie traces the evolution of statistics in Britain from 1865 to 1930, focusing on scientific knowledge, including statistics, as not purely objective or detached from societal influences, in line with constructivist thought. He discusses how statistical methods were applied to various social issues, such as eugenics, public health, and social reform. He shows how these were used to justify certain social policies and how these applications, in turn, influenced the development of statistical theory. For example, the eugenics movement heavily relied on statistical methods to promote its agenda, which had a lasting impact on the field. The founding fathers of these methods and some of the most important principles of the discipline that we still employ today were Francis Galton, Karl Pearson and Ronald Fisher (see MacKenzie 1981; Louçã 2009; Clayton 2020). Among others, they established quantitative genetics with an interest to prove hereditary traits through correlation, not necessarily only following blood ties but even within the same “race”. The principle on which much of Ronald Fisher’s work was based, for instance, was that a single observation can be considered an individual, and its repetition a population (Fisher 1948). Normally, any social or environmental analysis, investigating cultural and historical context, was left out of their analyses in order to focus on anagraphic and “fixed” data (O’Neil 2016;

Apprich et al. 2018; Eubanks 2018). Not many other authors analyzed the social implications of fundamental decisions about people based on statistical methods and correlation. For instance, the feminist literature on technology, which has contributed enormously to revealing the injustices and social exclusions tied to many technological practices, never directly addressed this. Yet, as Ellul (1954) wrote, optimisation and efficiency underpin these practices, and in order to study their implications, it is crucial to trace these values and “goals” in the design of technologies. In 1999, Geoffrey C. Bowker and Susan Leigh Star wrote a book investigating the practices of classifying, measuring and grouping information and their correlation in important contexts. They wrote that historically, the classification of individuals for the purpose of decision-making is composed and structured by rigid and externally imposed connections or correlations, full of uncertainties and contradictions (Bowker and Star 1999). While modern statistics is accompanied by social analysis, where the calculation of probability does not remain decontextualized from social life and the risks of generalization and correlation are taken seriously, early statistical practices that used to make inferences about populations and identify “normal” and “deviant” behaviors laid the groundwork for modern techniques (Apprich et al. 2018).

In algorithmic practices, correlation is often mistaken for causation (Apprich et al. 2018). This increasingly fuels a new empiricism and epistemology based not on knowledge but on data, not on causal inferences but on the collection and correlation of anything (Mayer-Schönberger & Cukier, 2013). This has profound implications for our understanding of complex human, social and cultural processes (Kitchin 2014) and according to van Dijck (2014) it is rooted in problematic ontological and epistemological claims. The risk, in these times, lies in the very possibilities offered by the ability to process and analyze huge datasets, sift through complex - especially historical - data, identify patterns and extract insights too subtle for human analysts to detect in order to make predictions about future trends, behaviors and outcomes. Their use might compromise in practice - that is,

when these techniques are developed and used to make decisions about people outside the experimental sphere - the vigilance required to prevent optimisation, efficiency and generalization resulting in the categorisation, classification and thus discrimination of certain social groups. The reason why they are increasingly used for decision-making purposes, especially in complex and stratified contexts, is that machine learning models create patterns from noise (Chun 2018). This can undoubtedly help to navigate in a sea of data, but can easily lead to loops of injustice and the perpetration of discrimination (O'Neil 2016; Eubanks 2018). In order to create patterns, these models look for homophily (Chun 2018), i.e. resemblance, with the idea that it creates connection. In big data-based predictive analytics, according to Chun, the focus is not on individuals but on network analyses that “segregate users into neighborhoods” (Chun 2018, 61). “In network science, differences and similarities – differences as a way to shape similarities – are actively sought, shaped, and instrumentalized in order to apprehend network structures. Networks are neither unstructured masses nor endless rhizomes that cannot be cut or traced.” But also “Networks, because of their complexities, noisiness, and persistent inequalities, foster techniques to manage, prune, and predict. This new method – this pattern discrimination – makes older, deterministic, or classically analytic methods of control seem innocuous.” (Chun 2018, 61-62).

3.2.2. The Algorithmization of Social Services

Datafication originates from the human need to measure. In the empirical context of essential services, datafication refers to the categorization of people and citizens by public agencies with the main purpose of classifying them in order to extract information while optimizing processes, without necessarily using additional technological tools or automated means. However, the introduction of algorithms to process these large information flows in core public services could add a layer of complexity. Algorithmization, described by Meijer as an organizational

rearrangement revolving around the introduction of algorithms (2020; 2021), follows processes of datafication in order to transform the hugely collected and registered information into insights, looking for patterns. “Algorithms” is a seemingly vague term that often provokes confusion because of the lack of shared definitions. In this case, the adopted definition of an algorithm is a computational procedure (although this is not always the case) consisting of a set of rules and introduced to solve a problem. These algorithms may be more or less “smart” (Hildebrandt 2020), e.g. they may or may not be based on artificial intelligence techniques, be rule-based or imply multi-agent systems. They may simply divide citizens into two categories or they may predict future behavior (as in the case of predictive analytics based on machine learning). Regardless, this chapter is interested in analyzing what are the main implications for decisions taken in the welfare state when such computational procedures are introduced.

When a practice is even partially automated, in fact, a certain degree of delegation of human decision-making is introduced regardless of the sophistication of the process. In the literature, automation bias describes a phenomenon that has been vastly explored and shows how people tend to perceive technology as neutral and therefore rely on it without questioning or verifying its impact and/or performance (Skitka et al. 1999; Kolkman 2022). What Alon-Barkat & Busuioc describe as the over-reliance on algorithmic advice (2023) has its ideological roots in technological determinism, regarded as an antidote to the influence that emotions, culture or values have on human judgments. According to this principle, then, algorithms and automated processes should theoretically reduce human error, be free of bias and not incorporate the same values and cultural traits as humans. Automated decision-making through algorithmization is also taking place in social policy and benefit administration, where the increasing use of data analysis aims to determine eligibility for a benefit based on specific criteria, automating their distribution (Meijer et al. 2021), or verifying that claims are not fraudulent. This is happening in contexts ranging from the United States (Singer 2015; Eubanks 2018), Denmark (Andersen 2019), Brazil (Mari 2019), Jordan, Sweden (Kaun and

Liminga 2023), Belgium (Ballon 2016), France, Australia (Lecher 2020), India and, of specific interest to this analysis, the Netherlands.

According to Hobsbawm (1994) technology played a key role in the development of bureaucracy and the welfare state. Database building and citizen monitoring, as much older and constantly evolving practices, have always had the function of monitoring citizens but also of tracking and assessing the needs of the population and the allocation of resources (Dencik and Kaun 2020). Given this history, welfare distribution is seen as a prominent site for the implementation of data analysis, automated processes and algorithms in public governance (Dencik and Kaun 2020; Jørgensen 2021; Dencik 2022). In exploring the contradictions of welfare states, particularly in capitalistic societies, Claus Offe argued that the attempt to balance the needs of a capitalist economy with the social demands for welfare and equality creates a tension between the economic logic of capital accumulation and the political logic of social welfare, which are often at odds (1984). One of the core contradictions is between economic rationality (efficiency, growth, and profitability) and social rationality (equity, social justice, and welfare). The automation of welfare decisions, based on efficiency, no longer establishes “normal” or “deviant” categories (Chun 2018) but crystallizes and codifies the differentiation between “deserving” and “undeserving” citizens based on their perceived moral worthiness to receive state support (Offe 1984). These practices have their roots in what Christopher Hood has called “New Public Management”, to describe an approximation by the public sector to efficiency-building logics typical of the private sector (Hood 1991; Yeung 2018) in order to make public administration “innovative”.

3.2.3. The Cultural and Social Determinants of Algorithmic Practises

A growing number of scholars in recent years have shown how algorithms learn from us to perpetuate bias and inequality to the point of amplifying social discrimination (e.g., Barocas, S. & Selbst 2016; O'Neil 2016; Eubanks 2018,

Buolamwini & Gebru 2018). This can occur due to several intertwining factors. In some cases, problematic results of artificial intelligence models depend on representational bias, the quality and heterogeneity of the data on which they were trained (boyd & Crawford 2012; D'Ignazio & Klein 2020). This may depend on the under-representation of certain groups in the datasets (Buolamwini & Gebru 2018), or conversely on their over-representation, whereby model weights favor or disfavor these categories. The latter case is well illustrated by Virginia Eubanks, who addressed the issue of the algorithmization of some public and/or social services in the United States by showing how it can automate inequalities, for example by further marginalizing certain social groups. Eubanks shows how these groups face higher levels of data collection when they access public benefits, pass through highly controlled neighborhoods, enter the health care system, or cross national borders (2018), and when more data is collected on certain groups, the algorithms charged with processing it will reflect biases due to this imbalance.

In addition to this, data can negatively bias the model due to their historical dimension. Historical bias happens when data collection perpetrates what happened in the past which, when processed by models, becomes “codified” as a strict rule. A classic example is that of gender inequality: since women have been systematically excluded or denied access to certain services for decades, such as the provision of credit, mortgages, or jobs, an algorithm employed in these fields - when trained on historical data - tends to statistically replicate that situation as more likely and suggest that a woman is less trustworthy than a man, since there is a lack of data to suggest otherwise (Hamilton 2018; Leavy 2018). The cases that will be discussed in more detail in the empirical section also describe this pattern. This will also illustrate that the connection between datafication and algorithmization is inseparable, but insufficient.

In the literature on the responsible and critical use of these technologies, the role of data in producing and provoking bias has been widely explored. In other studies, such as law and technology, some authors have noted how focusing on data is limiting for several reasons, especially for the protection from harm

(Mantelero 2022; Renieris 2022). This research wants to go beyond data as the only possible source of problems, describing algorithmization as a cultural and socio-technical process (Saver 2017; Airoidi 2021). The analysis of algorithmic practices employed in welfare will be limited neither to the datafication nor to the technical-statistical components of the process, but framed in the political and social context that required and employed them in order to trace values and goals, i.e. determinants, in the design and management that can also influence its final outcome. As social constructivists theorized, the relationships and negotiations among the actors who conceive, design, govern, market, and use it shape technological construction (Cockburn & Ormrod, 1993; MacKenzie & Wajcman, 1999). Therefore, in this thesis, algorithmic practises in the public sector will be framed as complex socio-technical systems that, as Karen Yeung put it, “include computational networks, people (data scientists, software engineers, interface designers, operators, reviewers, users), organizations, norms and practices, all connected to a broader social endeavor oriented towards knowledge production” (2018, p. 506). For this reason, the next section will aim to reconstruct in a descriptive way the events surrounding some of the main cases of automation in the Dutch welfare system. In order to reconstruct the socio-technical scenario, the different choices, the political ideas, the administrative culture and the actor networks will be pointed out as, as it will be argued, have significantly influenced and determined the negative impacts. This perspective is at the heart of the methodological choice for this research, as detailed in the following section.

3.3. The Algorithmization of Social Benefits in the Netherlands

3.3.1. Constructing the Dutch Digital Welfare State

The Netherlands has been described as the first European country to carry out a data-driven mass discrimination among its citizens (Hadwick & Lan 2021). This designation refers to a specific case, which unfolded in early 2021 and became

known as the *Toeslagenaffaire*, or childcare benefits scandal, that shook the country and led to the resignation of the cabinet in charge. The case, which will be described in detail, is only the most extensive and well-known of welfare automation in the Netherlands. Yet, it is far from isolated. Instead, it is a trend that started with projects of datafication - before algorithmization - dating back to 2003.

The Netherlands has a great history of collecting data from citizens and the general environment with the goal of improving policies and targeting resources. In 2003, the government established a multi-level working group consisting of various public agencies, from the tax authority and the Ministry of Justice but also the police, the prosecutor's office and some municipalities. The purpose was to create risk profiles by collecting citizens' personal data as a measure to counter undeclared work, welfare fraud and tax fraud. The Dutch Data Protection Authority (DPA) had to give an opinion on the project, and in 2006 it issued important recommendations focusing on the principle of proportionality and the transparency rights of individuals (Gantchev 2019) - considered crucial despite the fact that the project was established in the pre-GDPR years³.

In 2011, then, the DPA issued a very severe decision with respect to the government's use of citizens' personal data to counter fraud⁴. Most importantly, concerns centered on the fact that the government conducted investigations without a prior suspicion of fraud, which therefore did not justify such massive data collection, as there was no legal basis to regulate the processing. In addition, the principle of purpose limitation was not respected, and citizens were not assured of any right to transparency (Gantchev 2019).

On the basis of these events, SyRI, a project that will be detailed in the next paragraph, was invented. Interestingly, historically, the Dutch population has had a perception of the government as very transparent to its citizens. Equally, the

³The General Data Protection Regulation (GDPR) is a regulation of the European Union on the processing of personal data and privacy, adopted on April 27, 2016, published on May 4, 2016 and entered into force on May 24 of the same year.

⁴ CBP (2011) Besluit tot oplegging last onder dwangsom, Den Haag: College Bescherming Persoonsgegevens https://autoriteitpersoonsgegevens.nl/sites/default/files/downloads/pb/pb_20110317-dwangsom-siod.pdf

Netherlands is considered as a country where the relationship between authorities and citizens is based on trust (OECD 2023). On this basis, it is worth looking at the welfare-related digital infrastructure introduced in recent years.

The benefit (*toeslagen*) system in the Netherlands plays an important role in their welfare state and is considered to be one of the most developed in Europe: there are numerous aids that the government has been providing for years to the population in need (e.g., housing, unemployment, child raising and care, and health insurance). An interesting factor, however, of the Dutch welfare model is that it is available - and popular - even among people who are affluent or not in need. This is an important cultural factor, as the costs of some services - such as childcare, for example - are very high, so it is open to all citizens to apply for aid and to share the costs with the State. For this reason, Dutch welfare is seen as a “hybrid” system, somewhere between a social-democratic and a corporatist model (van Oorschot 2006). This model developed in the postwar period in a context of solidarity and collective support and was reinforced in the 1960s with a minister who thought that “every citizen has a right to self-realization and to equality of chances” (van Oorschot 2006, 59).

Beginning in the 1970s, this pattern began to shrink to privatizations and greater individual empowerment. The goal was to reduce the so-called “culture of dependency”, popular in countries such as the United States, the United Kingdom, and the Netherlands, and detect fraudulent behavior in welfare (Van Schie et al. forthcoming). This has led to a further shift in recent times to a different attitude towards citizens dominated by suspicion. Benefits continued to be distributed, but whereas before the idea was to create a support network, now it is to look at each applicant as being in bad faith a priori and ask them to prove themselves truly in need. This attitude has reportedly driven the introduction of algorithms into welfare. With an increasing difficulty in handling all applications and checking them, public agencies began to rely on algorithmization (made possible by the already existing datafication and profiling of citizens): key, however, is the fact that algorithms were not designed with the idea of “helping” more people, but with the

goal from the outset of “catching” criminals. This recalls what Eubanks describes as the tendency to “blame and punish the poor” (2018). On the other hand, Esping-Andersen described the Dutch welfare state as “corporatist” (van Oorschot 2006). By empirically analyzing algorithmization in the Netherlands, one can find evidence of this. For example, the algorithms that various public agencies have chosen over the years to create fraud risk profiles and measure citizens are the same used by banks to conduct credit scoring in order to discriminate between “good” and “bad” clients. This invokes a more business-like image of the welfare state, which seeks efficiency by blurring the boundaries between public and market values. Connected to this is another important issue, that of outsourcing welfare. What for years was personally guided and managed by civil servants, with their professional experience and knowledge, is now delegated to non-expert third parties. In fact, local governments in particular rely on vendors to develop these technologies, which, as we shall see, makes it very difficult for the public to maintain control of the processes and to have regular insight of what is going on. While this increases cost efficiency for the government, which would otherwise have to employ technology experts in every municipality, it also creates challenges in respecting public values and the interests of citizens.

In order to circumscribe the scope and clarify the scenario, I will outline three intertwined specific cases at different levels in chronological order: the SyRI, which involved the central government and a network of local governments; the *Toeslagenaffaire*, which involved the national government and the tax authority; and the municipality of Rotterdam, at the local level.

3.3.2. System for Risk Indication (SyRI)

In an effort to institutionalize data-matching projects and provide a legal basis, the System of Risk Indication (SyRI) was born in 2015 to detect fraudulent behavior in people applying for and receiving welfare measures in municipalities. However, it had two precursors: *Waterproof*, from 2004 to 2007, and *Black Box* from 2008 to

2014 (Wieringa 2023). It was supposed to create a shortlist of people in whom suspicious behavior was found and then put them under deeper investigation. In fact, the idea was to detect “discrepancies” by cross-referencing various data such as residency, integration, employment status, education, property, and criminal records.⁵ At a technical level, SyRI was based on a predefined risk model, represented by a decision tree, which used indicators that should inform with respect to the risk of committing welfare fraud. During the analysis phase, personal data were encrypted, and the moment they were indicated as a risk profile they were decrypted for investigation and kept for a period of two years also to be accessed by other administrations (Gantchev 2019).

SyRI was a sort of umbrella-technology for several projects that analyzed personal data to identify a risk of fraud. But after a number of access requests (FOI) made by lawyers and civil society organizations, it became very clear that the only investigations within SyRI were aimed specifically at what were called “problematic neighborhoods” with high rates of poverty, crime, and welfare beneficiaries (Wieringa 2023), particularly in Eindhoven and Rotterdam. Part of a global trend of introducing digital tools in welfare states, there was no trace in the project of an attention to “the potentially devastating consequences they may have on a range of internationally protected human rights” (Alston 2019). A civil society coalition in defense of human rights decided to sue the government for its use of SyRI in 2018 (van Bekkum and Zuiderveen Borgesius 2021). At this stage, the government was uncooperative with parliamentarians submitting questions and those from the outside calling for more transparency, refusing any informal investigation, and most of the people interviewed found this troubling and inexplicable (Ananny & Crawford 2018; Gantchev 2019).

As Maranke Wieringa (2023) described in her research on accountability around SyRI, reports on municipalities’ use of the system came out between 2018 and 2019 showing problematic data on its use, as it seemed not to bring added value to the extra scrutinies, but rather it frequently made administrative processes

⁵ for the complete list see <https://zoek.officielebekendmakingen.nl/stb-2014-320.html>.

more complicated. People and unions began to complain, and the House of Representatives started discussing more and more about it - though always with great closure from the government. In 2020 the verdict was handed down by the Court of The Hague, which was investigating after the lawsuit. The Court, while recognizing the need for such technologies in the welfare system, blocked the Dutch government's use of SyRI, and accused it of being in violation of the right to privacy, Article 8 of the European Convention on Human Rights (ECHR) (van Bekkum and Zuiderveen Borgesius 2021).

First of all, according to Dutch judges, “new technologies-including digital options for linking files and analyzing data with the help of algorithms-offer greater possibilities for the government to exchange data between its authorities in the context of their legal duty to prevent and combat fraud” (Rechtbank Den Haag 2020). The rules governing SyRI, in particular, appear to serve the “interest in the economic well-being of the state” and thus have a “legitimate purpose”, represented by adequate verification regarding the accuracy and completeness of the data on the basis of which citizens are assigned rights. But they would require an important and “fair” balance. Indeed, the reason why the Court censured SyRI is its lack of transparency, opacity and absence of necessary safeguards to protect those whose data are collected and processed. The risk model, indicators and data that were concretely processed were not, in fact, public nor were made known to the data subjects. SyRI's operation is so “secretive” that it prevented the court from even properly framing it as an “automated decision-making system”, beyond demonstrating the use of “predictive analytics, deep learning and data mining”. Those under scrutiny are prevented from knowing how the decision tree works and thus how a decision on their behalf is made. The fact that the model was developed by a private company, then, further impeded the ability to capture any conflicts between public and private interest. Moreover, the opacity of the model inevitably reverberated, according to the Court, into a lack of accountability and responsibility on the part of the public decision maker.

Specifically, the legal basis identified by the SUWI wet legislation did not offer enough transparency to welfare recipients, using an algorithm to make key decisions about them but without explaining them how. Among the testimonies of SyRI victims and their lawyers, this narrative recurs whereby there was no way to get information from the government:

«Documents were obscured and the workings of the algorithm hid, by claiming that revealing it would “favor criminals”»

(Informant 16, a Dutch researcher that worked extensively on the case)

According to Philip Alston, UN Special Rapporteur on extreme poverty and human rights, the system could have a hugely negative impact on the rights of poor individuals without according them due process (2019). The problem is made much worse by the fact that when the program was eventually given a legal basis, there was no debate in the Dutch Parliament, no media attention, and virtually no in how the system works.

The fact that the investigation found that algorithmization actually occurred in only a few specific neighborhoods (what, in this research, can be referred to as networks) validates once again what Eubanks argued, i.e. that “marginalized groups face higher levels of data collections when they access public benefits” and that “data acts to reinforce their marginality when it is used to target them for suspicion and extra scrutiny. Those groups seen as undeserving are singled out for punitive public policy and more intense surveillance, and the cycle begins again. It is a kind of collective red-flagging, a feedback loop of injustice” (2018, 11). One question many people have asked about the effectiveness of SyRI - and other similar projects - concerns precisely its overall partiality, since there was an *ex-ante* arbitrary - and discriminatory - selection of the geographic area to be monitored. These areas were typically those neighborhoods where public housing is condensed, the population has a migratory background and is pushed at the “edge” of city centers.

For the first time, the SyRI investigation showed to European policymakers the dangers of automating social services, where “new technology is important in making welfare systems not just more efficient, but also in ensuring fairness and justice. To do this requires a degree of transparency, and assurances that particular groups are not being unfairly singled out” (Alston 2019). To date, it is unclear how much harm the welfare investigations initiated by SyRI have had on people. What is certain is that after this ruling people expected a different management of algorithms in the welfare state in the Netherlands.

3.3.3. The Childcare Benefits Scandal

At the turn of 2020-2021, some members of the Dutch Parliament were engaged for months in inquiries, investigations, and research on social services and childcare allowances that soon led to an outbreking situation that they called an “unprecedented injustice” (Dutch Parliament 2020). This was the child benefits scandal, which at the time brought the Netherlands to the attention of Europe and around the world for having implemented and continued for years an automated decision-making process in public welfare that structurally discriminated against thousands of citizens.

Once again, to understand the underlying rationale and the determinants for algorithmization, we need to look back in time. The social and political construction of this process began between 2010 and 2011, when Dutch media started reporting on a number of cases of fraud that became known as the *Bulgarenfraude* (Bulgarian frauds), as they involved an organized group of people receiving Dutch welfare measures while living in Bulgaria and other EU Member States. One of the respondents interviewed for this article recalls that story through the very strong and evocative images that were shown on television at the time:

«There was actually a documentary, I remember it. The message that came through was that there was this round of Bulgarian people taking welfare money, and they were showing these

fantastic images of donkeys and the countryside where these people were with the statement 'this is where our money is going.' The narrative proposed was one of outrage, a "we have to do something." So many Dutch people saw this documentary and, indeed, were outraged.»

(Informant 23, Dutch professor and technical expert who accessed the algorithm)

Informant 23 also described the political sentiment at the time, which was to “get these people”, because they are not Dutch and they “take things that belong to us, take them away from us”. This situation, which was happening alongside the development of SyRI projects for municipalities, provided an opportunity for the government to revise the welfare system in order to modernize it. At the same time, the “suspicion shift” mentioned earlier took over. The government spoke as if potentially all people applying for welfare were doing so fraudulently: these people had to be tracked down with stronger measures. In 2010, the first Rutte government was just established, and during one of the first speeches to Parliament, the State Secretary for Finance, France Weekers, said that “fraud is not just toxic for what we as a society provide, but also erodes support for the welfare state”. He also said that “a few months after taking office, I published a fiscal agenda, based on three principles: simplicity, solidity and fraud-proofing” (Lower House of the States General 2020). In the aftermath of the Bulgarian fraud, an anti-fraud management team was established in May 2013 within the Ministry of Finance. In a 2020 document summarizing the parliamentary reconstruction effort called “Unprecedented Injustice”, the political sentiment behind the operation is apparent:

«The first report by this new management team stated that the “innocent will suffer with the guilty”, but also that this should be prevented as much as possible. No mention was made of how this was to be guaranteed.»

(Document 2)

In August 2013, the anti-fraud management team decided to set up the Combiteam Approach Facilitators (CAF), that consisted of members of different

teams at the Tax and Customs Administration in order to exchange information and suspicions more efficiently (de Witt Wijnen 2020). In fact, this administration combined two complex and normally separate functions: the tax authority and the welfare authority. To combat fraud, people from these two different functions were put to work together in the CAF. The CAF worked behind the “80/20 approach”, the assumption that 80% of the applicants were “bad” (or fraudulent) and 20% “good” (or innocent) (Kleinnijenhuis 2020).

Interestingly, all of this happened in the years when the potential of big data was emerging across industries:

«Back in the early 2010s, before GDPR, we were optimistic about big data. The Head of the Belastingdienst, the Tax Administration Office, was someone who wanted to innovate and be at the forefront and he talked very publicly about all the opportunities that big data could offer. He also gave some interesting interviews where he talked very enthusiastically about all kinds of random correlations that you can find if you link all these data streams together. Up to the point where, for example, if you drive a specific car, maybe you might not pay off and maybe there might be interesting relationships that we should look into, we should see what that means. Because the idea is we can measure everything, and there is also the political pressure to do things.»

(Informant 23 Dutch professor and technical expert who accessed the algorithm)

The Tax Administration Office (*Belastingdienst*) was the leading public agency in the childcare benefits case. At the time, the government and the tax authority decided to outsource to an external company the development of a software that could analyze the data entered by citizens in applications for specific childcare benefits. The underlying algorithm was supposed to analyze as much data as possible (similar to those already used for SyRI) to find correlations that would inform the risk of fraud, following the same principle: analysis, alert, deep investigation. According to the experts, the risk classification model operated under a “zero-tolerance” regime, flagging as potentially at risk any application form that contained filling errors. This approach was referred to as an “all-or-nothing approach”: even in the presence of minor errors during the process, such as incomplete delivery of documents or failure to pay the exact amount, the entire

child care allowance had to be fully refunded without a payment scheme (as it normally would be). One of the key informants for this thesis is a person working in a Dutch government organization that deals with ICT matters for different central administrations, who observed:

«We could have said, “Okay, then we will intervene by contacting this person and ask why she or he did not send all the forms or why she or he did not put a signature on her application or why she or he is cooperating with this office that has a bad reputation”. This did not happen. The intervention was that people were immediately flagged as fraudulent and their benefits were blocked and they received fines and, even worse, they were blacklisted and shared with other organizations. So this prediction was passed on to other organizations, who used it to make decisions and didn't know they were blacklisted.»

(Informant 20, working in a Dutch government organization for ICT)

His/her role allowed to notice how networked the decisions or “flags” on specific individuals were from one administration to another. After many years, the situation reached a peak when some MPs began to hear from lawyers that several people were turning to legal aid for help reporting that their benefits had been blocked, and with a request to return them within a short time, seemingly for no reason. A group of parliamentarians worked on a report (Lower House of the States General 2020) that reconstructed the events together with lawyers, associations, civil society organizations and victims, showing the error rates of the tax authority, the utterly chaotic situation - particularly in terms of governance - and the role of the algorithm. One of the MPs who has worked most on this report told me that at some point, together with other colleagues, they put on the reporters’ lenses and began to reconstruct the case, realizing that they had to keep going back in time to understand the full picture:

«I said well, I’m going to investigate myself. I opened an email account and people started contacting me there, saying that in different periods and years their benefits were stopped. And I was asking them what the reason was. They said, “we were just told that we were fraudsters, so they couldn’t help us. Without any trial, any explanation, any possibility to prove that you were not wrong.»

(Informant 28, opposition Member of Parliament)

Informant 28, as one of the parliamentarians who worked on the Childcare Allowance Parliamentary Inquiry Committee, had close contact with other key stakeholders in those months. What came out of their report, *Unprecedented Injustice*, was quite striking and led to a political crisis. Between 2011 and 2019, about 30,000 families were wrongly accused of fraud by the government. They were asked to repay the benefits they had previously obtained. Families, over the years, struggled to understand what was going on, why they were being charged, and sought information. As with SyRI, they encountered no transparency. After some tried to talk to lawyers about it, a widespread network of similar cases emerged. One of the first questions that came out was the role of the different actors involved (the government, the Ministry of Finance, the Ministry for Social Affairs and the *Belastingdienst*): it turned out, for example, that within the public agency different voices tried to bring up bias in the analyses. Since 2016, four different government advisory councils warned that the way the benefits system was designed could lead to serious financial difficulties and that these difficulties would lead to stress, illness and other problems (Court of Audit 2016; National Ombudsman 2016), but no one took this risk seriously until 2019, when the government commissioned some tests on the model.

Almost all key informants which knew the case and worked on it to some degree reported that this was the political and governmental decision-making framework, it was the ideological climate at the time and the human set of choices that had been made with respect to how to handle fraud, even before any model or software had been implemented. The new anti-fraud policies that followed the 80/20 and the all-or-nothing approaches regarded any error as negligence. Unintentional errors were no longer an option: the assumption was that anyone who made a mistake was at fault. Interestingly, the role of an algorithm or semi-automated decision-making system has been largely overlooked in the Dutch political-institutional narrative of this scandal. Outside the country, both in the media and academia, the presence of an algorithm has been emphasized and

highlighted, while inside the country there has been almost no mention of it. The parliamentary report “Unprecedented Injustice”, for example, never mentions “algorithms” and there is no public source to date that clearly explains the role of the technical component and the risk classification model in the *Toeslagenaffaire*. Yet, we know from subsequent assessments, such as the DPA’s, that the model contained self-learning elements that allow for the creation of risk profiles of child welfare benefit claimants who may be more likely to submit inaccurate and false information and potentially commit fraud. Tracking inaccuracies in applications was used as an early warning system for potential fraud. Dozens of risk factors were considered in the design of the risk classification model (Data Protection Authority 2020). The Tax and Customs Administration’s machine learning algorithm was trained on a combination of different data sources, but the specific nature of them is still unclear. One of the main sources was the Fraud Signaling Provision (FSV), considered as a “blacklist” where the personal data of suspected applicants were stored and recorded. In total, almost 180,000 citizens were placed on the FSV list (Kleinnijenhuis 2020). The FSV was active between 2013 and 2020, when after the first investigations that highlighted the huge GDPR violations the *Belastingdienst* stopped it. But the model also learned from historical data, that is, positive and negative cases of fraud (or “correct” and “incorrect”) from previous years. In practice, this meant that the more similar an incoming application was to one that had previously been labeled incorrect, the higher the risk score attached to this application was granted (Amnesty International 2021). No detailed information was available on the nature and quality of the training data. What some key actors reported is that, at least initially, the *toeslagen* unit had access to all the personal data of individuals and families on benefits. Presumably, among the training data was the case history of SyRI’s decision-making outputs (which, as detailed, were already problematic). Concretely, this means that based on instances deemed to be fraudulent in the past, when the same features occurred, the system would redeclare a risk of fraud.

The work of the Dutch Data Protection Authority (2020) was specifically aimed at understanding what personal data were included in the analyses and which factors led the huge number of affected families to be wrongly accused in the first place and to establish the flow of errors. What were the risk factors that the model considered as informative and weighed most with respect to the risk of fraud? The technical investigations showed that the algorithm increased the risk of fraud for individuals with dual nationality. The Dutch DPA conducted tests by creating potential “profiles” of benefits applicants and calculating their risk: all things being equal, dual nationality turned out to be a specific risk indicator. In the end, indeed, it was proved that the majority of the victims actually had a double nationality with a migratory background. In the aftermath of the parliamentary interrogation committee’s report, the government admitted that families with a non-Dutch nationality and/or dual nationality were more likely to be investigated by the Tax Administration. The MP recalls a specific episode, according to him/her of particular relevance to the broader scandal:

«There was an entire child welfare institution that was considered high-risk as a fraudulent organization in which families applied for welfare measures and then poured them back into the institution. In that case, there were about a hundred parents and all of Ghanaian origin. As a result, all parents of Ghanaian origin in the Netherlands who were on childcare benefits were flagged as at risk by the system.»

(Informant 28, opposition Member of Parliament)

This example perfectly illustrates how a machine learning algorithm works in such environments. Using statistics and machine learning techniques, they relied on problematic correlations. In this case, the variables to be positively correlated were nationality and the risk of fraud. The tax authority and the government never proved this correlation through empirical data or scientific literature that could justify such an association. For this reason, the *Toeslagenaffaire* has become known as a case of “institutionalized discrimination”, pointing out that this reflects a historical trend of racism in the Dutch welfare system. According to one of the

experts that was asked to look into the algorithm after the scandal came out and inside of the processes that led to it:

«There was a focus group, a time for brainstorming. They asked, “If we want to find these problems, what kind of characteristics can you think of?” And it seems that they came up with a list of about 100 indicators, as they call them. Many of them are censored in the public record, so we don't know exactly what they were, but we do know that the second nationality was there.»

(Informant 23, professor and technical expert)

The “feedback loop” of injustice (Eubanks 2018) depends on the fact that what belongs to historical and past cases, i.e. positive and negative cases of fraud, is codified as a rule and perpetrated in the present and future based on one or multiple traits. In this case, it is very clear how the initial fuse, the Bulgarian frauds, created a feeling of suspicion toward people with dual citizenship who were committing fraud from abroad. These cases became the government’s only target: it was like the software was not designed for fraud detection in benefits, but for fraud detection by people with dual nationality. According to the MP and the reconstruction she/he worked on,

«Another perverse factor that contributed to the result is that there was a strong push from the organization to “bring the money back”. But at that point, it all led to going to the groups of single mothers with lots of children who were getting a lot of support, rather than the wealthy people who were getting little help, so they could take back more money faster.»

(Informant 28, opposition Member of Parliament)

A widely shared feeling suggests that the *Toeslagenaffaire* was not a scandal about the use of algorithms in public services. Or that at least the algorithmic component in it was not the one to blame. It was a scandal about the welfare state and the fragmentation of administrative bodies in the management of social services that was made worse by the use of automated analyses carried out in an arbitrary and discriminatory way. Errors and faults were spread at different levels that reinforced

one another: the people (governance), the policies (values), and the model (technical). One of the actors which observed these dynamics from within the administration, noted that:

«We got this policy of all or nothing, years ago, that possible fraudulent receivers of benefits should be targeted. And that was the policy that the government chose. And then they created this algorithm that used indicators of whether you had a double nationality. And also income, I think, which I don't understand, because if you're applying for a benefit of course you have financial difficulties. Things went wrong on so many different levels at the same time. It's all wrong.»

(Informant 20, working in a Dutch government organization for ICT)

However, if we want to look at just the level of technical functioning, she/he stresses that:

«The model was also not okay, it was a bad model. But still, I think that the model wasn't the worst problem. The algorithm was a problem, but there were so many layers that went wrong, and that worries me because we are focusing on the algorithm, and to be honest, that was the easy part.»

(Informant 20)

Reinforcing the idea of these inseparable components in socio-technical terms, and thus the fundamental role of all the institutional and human roles in addition to the model itself, was another person who observed them closely:

«From what I saw, for me it was a major failure of a system that included all these components. A lot of things went wrong collectively, and they all went wrong a little bit and then together became a huge problem that nobody seemed to be responsible for.»

(Informant 23, professor and technical expert)

Algorithmization, whether based on artificial intelligence systems or not, when embedded in layered and complex contexts is always a socio-technical process that

is shaped along with the relevant culture and the decisions of the people working on it (Hughes 2004). According to constructivist thinking, the actors who construct it infuse values into the technologies and direct their operation (Bijker, Hughes & Pinch 1987; MacKenzie & Wajcman 1999). Toeslagenaffaire shows us not only how true this is even today with immaterial - and more procedural - software and technologies, but also how extensive and chaotic this network of actors can be. In fact, the networks influencing the construction and the decision-making of contemporary algorithms may be much broader. In this particular case, the responsibilities have been so spread out and so diluted over time that it is very difficult even today to distinguish them clearly. The networks of actors, for example in Belastingdienst, are part of an organizational chart that is itself very complex, so thinking about the management of an algorithmic process within this setting further complicates the case.

Focusing on people and their choices rather than on the technology itself helps to hold them accountable (Orlikowski 2000) and to see more clearly how much the technical choices, which then contribute to clear effects and impacts, actually depend on them from the start. The situated culture that gave rise to these particular automation projects can be found in the approaches used in welfare, in policy, in the light-heartedness with which the methods to be used in meetings were decided, in the idea that big data along with numbers, if correlated, can find solutions to everything. But also in considering citizens receiving benefits as good or bad customers, a human choice that shaped the mathematical model from the very beginning. According to the MP, this was far from an automated and objective choice:

«These two lines that they used for profiling in the model, they searched for the people with the lowest income and they searched in the areas where they knew there were living a lot of migrants, so they were always looking to the specific side of society where people have low income, live in rental housing, have a migrant background, women. And these came in the profiling, yeah, well that really brought out the worst of, well, humankind and system thinking.»

(Informant 28)

Still in the view of the MP, all these choices describe a particular institutional, administrative, bureaucratic culture and idea of citizens and tell a particular way of doing politics:

«The model itself was a problem, because this idea of good and bad eventually used a technical model that resembled a credit risk model, and the problem with credit risk models is that they make risk assessments with the goal of giving money to a customer and verifying that the customer is not defaulting. And I'm afraid that this kind of mentality and framing and modeling punishes the poorest people. But it's a different problem. Public welfare versus private credit: it's a semantic conflict. A good or bad customer who gets a large credit from a bank is basically someone who is financially more or less stable. In this case, however, a good or bad file is represented by whether or not mistakes are made. These models keep adding up risk indicators that can be correlated if you are young and not on a permanent contract and you are studying and living with your parents. So the model that comes out is very transparent and explainable, but it deliberately discriminates against these classes and adds indicators that add up to one another.»

(Informant 28)

As a consequence, he/she adds, a mathematical system can only anchor and codify this:

«Then, I find it quite hard to only blame the algorithm. Because there've been a lot of design issues there that I would definitely have done differently, yet with this bigger setting behind it I just see it happening.»

(Informant 28)

So, there was little technicality in the choice of whom to discriminate against. It was almost entirely a human choice. At the same time, however, it is important to distinguish the different human responsibilities in the decision-making chain. If top management knew the processes and actively influenced them, the civil servants in the offices who were charged with validating the outputs had little room for action, or at least were greatly influenced by the outcome of the model. They had the power to label one application as fraudulent, but crucially officials were not given

any information as to why the system had assigned the application a high risk score for incorrectness (Data Protection Authority 2020):

«The list of people at risk was drawn up and then it was passed to a human. So they said that this was not a stand-alone system that did something on its own. There had to be a civil officer monitoring it. But that civil servant was not very educated. It was just an official who was passing papers. And for reasons of privacy or impartiality, that person only got the risk scores, but no explanation. So she/he/they could never have seen that only young mothers who had many children, who were poor and black passed the check consistently. So I can imagine that these people, from a power balance perspective, did not tend to be in a position to oppose it.»

(Informant 28)

In contrast, the people who worked in the Belastingdienst teams were at least initially very free:

«From a structural point of view, the problem is that these agency teams were free to do anything. Obviously, if you're a bit racist or have misconceptions or uninformed ideas of what leads people to act in a certain way, as an organization you have to make sure that you can't act in a way that has such a negative impact on people.»

(Informant 21, former data scientist at the Belastingdienst)

This is also demonstrated by the relationship and attitude they had with audits, monitoring and human control, as noted by the person that knows the culture, attitude and work of the teams from within:

«We are focusing a lot on the design, construction and deployment of the algorithm. Not so much on the use of the system. I can guarantee you that many organizations don't have a monitoring plan, they don't do any monitoring, and that is really bad. To see what performance is not decreasing and the bias metrics, for example, and that is also something that is very, very important. There are people inside of the tax administration who really think they did nothing wrong. These people are not asked to do another job, they are in high positions. So getting this out is hard work, it's long work.»

(Informant 20, working in a Dutch government organization for ICT)

This reveals interesting information about the administrative culture that prevailed within this organization, and especially about accountability and

responsibility, which will be addressed in more detail in the third chapter of this thesis.

3.3.4. The Case of Rotterdam

The Rotterdam municipality's case shares all the assumptions with the previous two but at its smallest scale, the local one. The other key difference is that having emerged after both SyRI and the *Toeslagenaffaire*, its impact is very informative with respect to a broader national strategy to remedy the harms caused by the algorithmization of public services and welfare.

In 2022, the Dutch Court of Audit (Algemene Rekenkamer) conducted an audit of 9 algorithms used by the government to make decisions. In addition to concluding that six out of nine of the algorithms analyzed did not meet basic principles based on ethics, justice, transparency and technical robustness, they found that the algorithms used by the Ministry of Finance, its *toeslagen* division, the *Belastingdienst* and those used by the Ministry of Social Affairs repeated a similar architecture of simple algorithms based on decision trees (Netherlands Court of Audit 2022). Thanks to journalistic investigative work conducted by WIRED and Lighthouse Reports (Constantaris et al. 2023), detailed research was published in March 2023 on the use of a machine learning algorithm, based on hundreds of decision trees and programmed to perform risk classification for welfare fraud detection, used by the city of Rotterdam. As the precedent of SyRI showed, in the Netherlands some categories of public benefits are distributed and managed at the local level. Because of this, municipalities have a certain level of autonomy, and many cities decide to experiment with different approaches to allocate welfare.

One of the journalists that for some years tried to access this data highlighted the starting point of this work:

«In 2017, a presentation by the consulting firm Accenture cited the work done for Rotterdam, explaining that the city's desire was to identify the illegitimate welfare recipients through a truly data-driven approach. After viewing these slides, we decided to submit access requests

(FOIs) to the local administration to request documentation about the algorithm and the process around it, specifically demanding access to the variables used, the dataset, the code, the model, the performance evaluations, the data protection impact assessment (DPIA, which is mandatory for these projects under GDPR), the caseworker handbook, and other technical records.»

(Informant 26, investigative journalist on the case)

Journalists and researchers also had contact with victims in Rotterdam, analyzing groups on social media where they shared their experiences. Reviewing all of this documentation made it possible to build “tests” - such as those conducted by the DPA on the *Toeslagenaffaire* algorithm - to assess the dataset and the model when prompted in a specific context and using specific indicators and analyze how the technology would perform in specific scenarios. Lighthouse Reports created two hypothetical profiles of people on welfare benefits living in Rotterdam: Sara, a single mother with two children who had to quit her job to take care of one of the two, and Yusef, an immigrant who was a teacher in Iraq, his home country, living with other co-nationals in a predominantly non-native neighborhood and who failed the Dutch language test. Out of a total of 30,000 people receiving welfare measures in Rotterdam, the two test profiles of Sara and Yusef scored in the top 10% as high risk based on their characteristics (Constantaris et al. 2023; Casimir-Braun et al. 2023). From the tests, it was clear how the algorithm discriminated against some specific social groups or certain characteristics. All things being equal, for some nationalities (deduced as a proxy from the language spoken, as after the GDPR nationality cannot be used directly as an indicator) and for the female gender the risk was higher, but also indicators such as age and relationship status increased the risk. The list of indicators used, however, was far more extensive: more than 300 very specific and “absolutely arbitrary” variables lacking scientific evidence (e.g. “a well-groomed appearance”) were critical to determine risks. Even more interesting is the layering of these risk indicators with each other: the work shows that a woman who is financially supporting her children is more at risk than a man in the same situation, and when you add that

the woman has a migratory background and is in her 30s, these are two other indicators of risk.

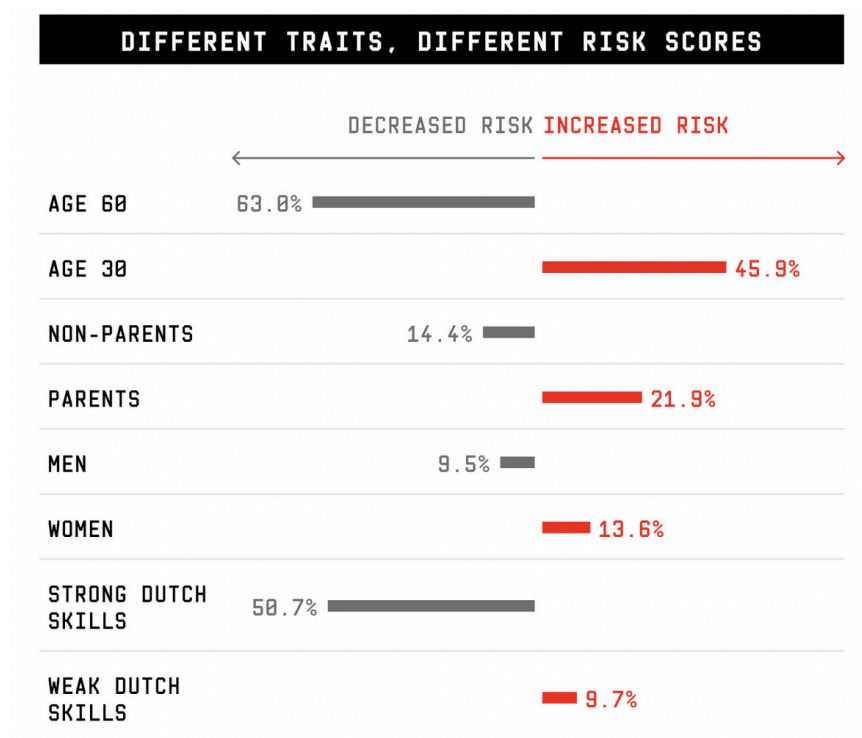


Figure 2: A visualization of the data and the rules that resulted from the testing of the Rotterdam model (Constantaris et al. 2023)

In the visualization created by the journalists, it is evident how much each individual attribute weights in calculating the final risk score. The most informative indicators are gender (female), age (younger than 30), parenthood (having children), and Dutch language proficiency (low). When all these attributes are added together, the risk becomes higher, regardless of the specific situations. For at least four years, the local government has used this model to detect fraud in benefits applications. According to the informant working for the tech organization within the central government, it is significant to reflect on the benefits brought by this system, and to weight them together with the risks and impacts:

«The case of Rotterdam, with 300 characteristics and indicators, was targeting especially young mothers who were raising their children alone. The municipality discovered 2.5

million euros in welfare fraud per year. 2.5 million. And they thought that only a fraction of the 2.5 million was discovered because of the algorithm, and I thought, two and a half million is not a lot of money, and then if only a fraction of that is discovered because of the algorithm, why on earth would you use it? Why would you take the risk, why would you create this model with 300 functions that needs to be updated, that needs a team, that needs an infrastructure, that needs audits. And this is also something that I think we should look at better: what are the gains of the model, is it proportionate? Paradoxically, we should look at the numbers of the numbers.»

(Informant 20, working in a Dutch government organization for ICT)

It was easier to perform tests controlling the model for conditional and statistical parity (Casimir-Braun et al. 2023; Carey & Wu 2022) because of the choice of the model, as in the decision tree it is possible to display the algorithm's decision path for each newly added variable. Beyond the discrimination test, this research brought out the extremely negative consequences for people who were classified as "high risk" and thus selected for in-depth investigations. This meant home visits, tracking and surveillance, and penalties in all other interactions with public services (from registration to education) with no room to challenge the decisions.

In this context, it is crucial to note what one of the journalists reported, telling how this work was possible:

«Despite the difficulties - this research was only possible because of the cooperation, availability and transparency of the Rotterdam officers. This is the first such comprehensive reconstruction of a public algorithm used for welfare (but also in other public services) in Europe. It is clear that the drive was given by the scandal related to childcare benefits, and that the goal is now to be more transparent.»

(Informant 26, investigative journalist on the case)

It is important to stress that the city stopped using the algorithm in 2021, so before the investigation even started. That is also why they cooperated, considering it important that other administrations were aware of what can go wrong in these projects. In recent years, the city of Rotterdam is one of those that has made the

most effort to manage these problems. It now has in-house ethics and algorithms experts, as well as data scientists doing research on how to avoid these risks. For example, many entities in the Netherlands have been considering whether to delete labels on nationality from their datasets and models to avoid explicit discrimination. After realizing that this is not a viable solution, Rotterdam has halted its new fraud detection project until they obtain more findings. It is also very important to point out what has changed on these issues over time, which can be identified by people who have been working on these projects for some time:

«By reporting and investigating these issues for several years, we have noticed a change. In the beginning everything was more secretive, no one wanted to give us anything. Since Toeslagenaffaire, on the other hand, it's clear that there's an extra push to show that we're not doing anything wrong. Before the child benefits scandal we had things hidden everywhere: blacklists, personal data catalogs, secret algorithms. Things are better now, but I think it's because they were so bad before. The reputational factor, which in a public agency translates into public trust, is high, but the power is still higher. It was still a turning point, because without that scandal we would never have received all this documentation about Rotterdam. Still, we had to negotiate for two years to get it, we even said we would get a court involved if they didn't share the information with us under the access request.»

(Informant 26, investigative journalist on the case)

In both the Rotterdam case and the national scandal, the technical choice is rarely clear. From the documents obtained, Lighthouse Reports reporters talk about machine learning while other informants talk about logistic regression. But by analyzing the data, especially by following the people involved in these cases, it seems that the technique used fades into the background. Regardless of how sophisticated the model was, the example of Rotterdam taken with the others demonstrates that human behavior is at least as important. After all, in both cases, one of the most relevant data for the final impact was the choice of variables to be used. A choice that, regardless of the technique, was made by human actors and combined with the use of historical data, which tend to inherit historical human biases (Barocas & Selbst 2016):

«Even in the case of Rotterdam, the choice is not random. There was a clear initial human decision of what to go for. The model was trained on historical data, so if these said single mothers were more likely to commit fraud, then a query was run to look for them.»

(Informant 26)

Among the approximately 300 indicators used by the Rotterdam model to assess the risk of fraud in people there were attributes such as “low self-esteem”, “duration of last romantic relationship” and “ability to persuade and influence others”. But, beyond these specific factors:

«It's always the same target groups. There is always a migration aspect, being a parent is always a risk indicator, typically being a woman is a risk indicator. I was very struck by that. How children were attached only to their mother. And then of all the indicators used, which are so many, many are proxies for something else. Like “having a partner living abroad” is obviously the proxy for living alone, while “Dutch language proficiency” is the proxy for nationality. So it's definitely the same groups, whether it's more complex systems or simple systems, and I think that's quite illustrative.»

(Informant 26)

The researchers and journalists concluded that a non-Dutch mother in financial distress was considered among the highest risk profiles. If the person is young, the risk was even higher. Quantitative investigative work on the system was very important in order to conduct an external audit and try to retrace the steps that led to the final results. Through this work, we can identify with particular clarity how human choices about the model influenced it, leaving little room for autonomous and unavoidable operation of the model. We can glimpse, above all, how these choices were justified by the objectivity of a mathematical system and its correlations portrayed as neutral. But putting the quantitative analysis together with the broader picture of underlying administration, policy and culture, it becomes clear how inseparable this component is from the others. As I will argue in the next section, these data show that the human intention may still be to use technological determinism as an alibi for being able to make very human choices that risk making the situation worse for some social groups. In particular, these choices

describe the intention to exclude specific social groups and monitor them more easily through their connectedness, or similarity, based on predetermined networks.

3.4. Networks, Determinants and the Amplification of Discrimination

The Dutch algorithms analyzed, their design and use contexts, rather than being isolated cases resulting from technical errors actually reproduced structural social disadvantages that are historically rooted. Moreover, far from finding “hidden patterns” (Engin and Treleaven 2019), these models and their outputs rather shed light on quite evident sexist and racist ideas in welfare and public services. Certain social groups were excluded from a public good they were entitled to at a huge scale. Even worse, in many cases the automated decision related to one context was then expanded and used by other administrations in other services, as one journalist that had access to all the documentation behind the Rotterdam case noted:

«The Rotterdam algorithm has also been used for other systems by other agencies, for example, everyone in a certain zip code is screened and, if they match one of these risk indicators, flagged for investigation.»

(Informant 26)

These situations, therefore, challenge the subjectivity and identity of certain social categories, e.g., migrants, women and parents (Van Schie et al. forthcoming), even more so when layered on each other in an intersectional way. These considerations cannot be made in a vacuum, but must necessarily be contextualized in a given socio-cultural setting. As an interviewee with technical knowledge who has also observed the situation from a political perspective points out:

«The Netherlands generally also has a rather complicated issue with diversity, at least people of color. Gender is also a bit complicated actually, but especially ethnicity with colonial history and racism issues. The Netherlands is not racist. And if you would say it is, you're killed in public basically. This is just how things are. So things that are culture are implied to be immutable.»

(Informant 23)

In this sense, culture influences technology as a *determinant*. In the Dutch case, as Chelsea Shields' book on Dutch colonialism in the Caribbean (2023) suggests, the correlations that identified certain categories of citizens with a higher risk of fraud stem from a long-standing institutional legacy. Shields traces this link back to research in the 1970s studying the relationship between being a single, black mother and welfare dependency in order to manage welfare benefits in the colonies of Suriname, Curaçao and Aruba. The historian, in analyzing how these dynamics influenced colonial economies, argued for the high normativity attributed to the family, and how it influenced the idea that some families were “deviant” at their core because of their structure and the individuals in them, rather than because of systemic inequalities (Van Schie et al. forthcoming). In fact, when injustices occur on a large scale, the structural component becomes very visible. It is related to the normativity by which each culture can judge “deviant”, “bad” or “undeserving” different behaviors, which instead of being supported are policed, excluded according to punitive logics. Related to this, an interesting and fundamental remark was made by the MP who worked on the *Toeslagenaffaire*:

«What is interesting is that you have these groups and you discriminate against groups, not against cases or single cases.»

(Informant 28)

Cultural prejudices against migrants and the identification of the “good” and deserving citizen with the white man are not something that arose with algorithmization. Rather, they are practices rooted in a much older culture and society - in this case, the Dutch one - in which gender, ethnicity and religion have

grown as discriminatory traits. The “hardworking” white Dutch citizen is thus contrasted with the inactive and unemancipated migrant (see Van Schie et al. forthcoming). On this basis, algorithmization and (semi-)automated decision-making find fertile ground for the amplification of inequality (Eubanks 2018). A dense network of preconceived biases, part of a specific culture and society and thus in a specific technical style (Hughes 1987) can provide easy grounds for connections and correlations: if the belief is that migrant people are untrustworthy, have higher rates of criminality, and single mothers are incapable, it will only take this as a starting point to search for these profiles within a database and increase the levels of surveillance and policing towards these categories. Even worse, through the use of historical data and predictive models - ranging from logistic regression and up to machine learning - these forms of social exclusion, far from being resolved, are coded to be repeated in the present and future. These analyses, based on probabilities, are normally considered non-deterministic because they involve a degree of unpredictability. In reality, these processes seem to be entirely determined by the initial conditions and the rules that govern them, and it is difficult to see large elements of randomness in these examples, especially when we consider the human role and the presence of historical data. Starting with a pre-existing social discrimination, networks are created where each data point represents an individual with those characteristics. Network determinism is how this happens today: correlations occur through *resemblance*, namely sharing at least one characteristic considered “problematic” with another person. According to a person who worked in a Dutch NGO at the time of the Toeslagenaffaire and who was specifically interested in the discriminatory aspect of the case:

«The Belastingdienst used various data including nationality, family ties and/or living conditions - to identify larger groups that were classified as one homogeneous population. Thus, the entire population was suspected of organized fraud. Moreover, officials spoke in derogatory terms about families with Caribbean ties by calling them an "Antillean nest". Officials flagged these applicants through manual selection. This indicates that the Tax Administration assumed a link between potential fraud and certain groups of people

identified by nationality and ethnicity. People of non-Dutch origin were considered more likely to commit fraud.»

(Informant 25, working in a Dutch NGO focused on victims)

These algorithms can often only perpetuate the discrimination they “find” (Chun 2018). Their operation is based on identification and repetition, and this implies their prescriptive character of modeling the future by moving from “analyzing captured data to building an epistemological model of the captured activity” (Agre 1994, 364) that originates from human normativity. These practices, based on statistical correlations, basically reflect what Fisher, among the founding fathers of statistics, wrote in 1948: that if an individual is considered as an observation, its repetition is a population.

The profiling of citizens and the creation of networks, instead of being used for the purpose of identifying those most in need of help, is used deterministically to reveal something about the identities of these people, their past and future behaviors. In this way, women share with each other this characteristic that makes them at risk, young women share two, young women and mothers three, and so on, until a dense network of filiations (Bowker & Star 1999) is created. In this way, the behavior of a person with similar characteristics in the past reverberates in the present and future in a person belonging to the same network. Network determinism grows out of technological determinism, i.e., the idea that it is technology that directly influences society, prevailing and reflected today in the reliance on automated decision making:

«One thing with the Netherlands is we tend to be very comfortable with numbers. If there is a number or if there is data evidence or we claim scientific evidence, we very quickly put it in a decision rule and say this is just the way it is. So in the Netherlands, I do have the feeling that we are very too trustful of numbers and data and too much see it as if it were natural science laws.»

(Informant 23, professor and technical expert)

The many empirical contexts of automated discrimination studied in recent years (e.g. O’Neil 2016; Eubanks 2018) show this: from methods of automated data analysis to the most sophisticated AI techniques, the most common way they are used is to track patterns, until they become the epistemological starting point for decision-making. Far from being just a problem related to the algorithmic output that “can only be as good as the data on which it was trained” (Miller 2022)⁶, the responsibilities are broad and usually it cannot be the model to be blamed. Rather, the chain of human choices that are made during the design and lifecycle - which, as the childcare benefit scandal shows, can be very long - must be pointed out to highlight the cultural-technical interaction which allows humans to choose target variables to “control”, and then “leave the algorithm free to group, create clusters, and unearth correlations at a rate no human could sustain” (Huyskes 2024). What network determinism reflects is specifically a set of human choices, from the idea that grouping people into rigid categories based on similarity leads to predetermined outcomes for those within the same cluster. As older paradigmatic ideas on technological determinism, these projects and the way they are carried out show how societal change and relations should be driven by networks. Network determinism extends this by relying on historical methods of group classification, including colonial and discriminatory assumptions on social actors. These practices influence how decisions are made, often reinforcing societal biases through computational methods. Networks on which most automated discrimination depends are created on the interaction between culture, policy and automation.

3.5. Conclusion

The case of welfare automation in the Netherlands serves as a critical examination of how algorithmic decision-making, underpinned by historical,

⁶ <https://hai.stanford.edu/news/data-centric-ai-ai-models-are-only-good-their-data-pipeline>

cultural, and ideological values, can exacerbate social injustices. The case studies of SyRI, the Toeslagenaffaire, and Rotterdam's welfare fraud detection algorithm exemplify how deeply entrenched cultural and social biases can become codified into technological systems, leading to significant real-world harm. The Dutch examples show how individuals are not only categorized but also preemptively judged and condemned based on their resemblance to others that belong to the same social groups within rigid, algorithmically defined clusters and in a deterministic way. This phenomenon can be conceptualized as a contemporary extension of technological determinism, where it is believed that the technology itself, guided by connections and correlations, becomes a driving force in shaping social outcomes. In the Dutch welfare system, a pioneer in the European landscape of public service automation, this has led to the amplification of discrimination, particularly against vulnerable groups such as single mothers, immigrants, and those struggling financially. What I referred to as network determinism - the interconnectedness of individuals based on shared characteristics - shows that the reliance on algorithmic models to manage and distribute welfare not only reflects but also amplifies pre-existing societal inequalities. This is possible through the correlation of individuals - treated as data points - and the identification of patterns that in this case were addressed by human choices. The introduction and implementation of these algorithms, specifically, were driven by a desire for efficiency, cost reduction, and the elimination of fraud. Empirical analysis of the Dutch context actually revealed that in addition to these "neutral" aspirations, there were many hidden cultural drivers guiding these projects. For example, the historical evolution of institutional attitudes toward welfare, a strong narrative about government transparency that failed to deliver on promises, the relationship with other nationalities and the culture of suspicion, judgments about trustworthy and untrustworthy citizens, and reliance on math and numbers. As proponents of social constructivism have long suggested, it is essential to study the interpretive flexibility of any technology, namely the idea that different stakeholders, or relevant social groups, can have varying interpretations of a technological artifact, including

its design, purpose, and use in order to understand it deeply in its interaction with society (Bijker and Pinch 1989). This concept firstly challenged the notion of technological determinism, which suggests that technology develops according to its own logic, independent of social influences. In contrast, tracing the stages of these modern technologies and following the key actors who influence them helps us to see how human choices still guide their design and use today. Instead of looking only at the impacts that these projects had on citizens, the analysis of these three case studies focused on wanting to highlight the cultural determinants that then cause those impacts, to show that different outcomes could be achieved with different values and intentions and that, as many of the key informants for this first analysis highlighted, have more to do with humans than with machines.

What emerges is the understanding that algorithmic tools in public administration often reinforce and amplify institutionalized cultural biases, leading to a potential worsening of systemic discrimination and injustice. The concept of network determinism aptly captures this phenomenon, where individuals are not judged solely on their own circumstances but are categorized and treated based on their resemblance to others within certain groups. This approach inevitably leads to a form of determinism, where the fate of individuals is increasingly predetermined by their algorithmically assigned category rather than their actual needs or circumstances. The Dutch experience underscores the need for a critical reassessment of how technology is integrated into public welfare systems and highly bureaucratic human regimes. Moreover, it calls for a deeper examination of the socio-cultural factors that shape these technological practices and a commitment to addressing the root and contingent causes of bias and discrimination embedded within them. The projects discussed here demonstrate how technological innovations, if well intentioned, can lead to unprecedented injustices when not carefully managed. However, these are only a few examples: behind each automated project used in public services to take important decisions, if studied carefully, we can unveil the cultural assumptions that influenced them. Doing this is especially urgent at this moment in history and can be helpful in

understanding deeply how to increasingly prevent network determinism from driving these projects. In this, the lessons learned from the Netherlands have great potential for us to understand how culture influences the design of digital technologies. Realizing that different technologies may be designated in different contexts, we can pay attention to their contingent aspects by studying them in their context.

4. The Machinery of Government: Bureaucracy, Automation and Institutional Black-Boxing

4.1. Introduction

In recent years, the automation of public services through algorithms and mathematical models has emerged as a significant and controversial issue in governance globally and has become a defining feature of contemporary bureaucratic institutions. While it promises efficiency, cost-effectiveness, and enhanced decision-making, its increasing opacity raises critical questions about transparency, accountability, and democratic oversight (Pasquale 2015; Burrell 2016; Meijer et al. 2021). A whole research area, often referred to as digital governance, e-government, or GovTech (government technology), is focused on exploring how technology, particularly automation, artificial intelligence (AI), and digital systems is transforming the operation, management, and delivery of public services at the organizational level (Tangi et al. 2022; Misuraca & van Noordt 2022). Analysis on the implications of these changes for governance, democracy, public administration and society at large are urgent and particularly needed. However, most of these explore the general implications of automation and the use of technologies in these settings, rarely linking these reflections to the specific cultural context in which they are situated to highlight different approaches. Looking at the European context, the scenario appears very varied: some countries are far ahead on the automation of public services while others seem to have not yet chased this trend (Sun & Medaglia 2019; Tangi et al. 2022). Moreover, how this digital transformation is governed also varies from country to country.

This article takes Italy as a case study and seeks to explore the deeper institutional and cultural reasons behind its approach to the automation of public administration. The country provides a striking case to study these dynamics as it

was designated as one of the top adopters of automated systems in public services (AI Watch 2022) while its approach in integrating these technologies into bureaucratic workflows remains shrouded in opacity. This is perfectly represented in the controversial Buona Scuola case, where an algorithm designed to allocate teachers to schools resulted in widespread errors, causing significant disruptions in the education system. Yet, this event has not triggered a significant precedent for future public debate or transparency on the wider use of algorithms in Italian governance. Through a constructivist methodological lens, which aims to follow actors and institutions rather than focusing only on technology (Bijker and Pinch 1987; Latour 1987; 1999) this article draws on a range of different qualitative sources and personal experience to navigate the challenges of accessing information from Italian institutions. Far from being a purely technological issue, this chapter argued that cultural resistance and other relevant factors within these organizations play a significant role in the black-boxing of these processes and posits that the true locus of black-boxing in the Italian case study lies in the opaque, human-driven processes that manage and obscure the use of technology within the public sector.

The chapter analyzes the idea of bureaucratic automation in the first section, starting from Weber's theories on bureaucracy and contemporary research on algorithmic governance and discussing the concept of black-boxing in different domains. This provides the foundation for understanding how bureaucratic processes have long been fertile ground for automation due to their focus on standardization and efficiency. The findings are discussed delving into the different institutional and legal challenges about the automation of Italian public services. The section also examines the Buona Scuola case in order to demonstrate these challenges empirically, particularly how algorithmic errors can have far-reaching consequences when embedded in opaque systems. Finally, the chapter emphasizes the cultural and institutional factors characterizing the Italian opacity around

algorithms, contributing to research on algorithmic governance and the uses of digital technologies by public administrations.

4.2. Theoretical Background

4.2.1. The Machinery of Government

*«I am not a client, a customer, nor a service user.
I am not a shirker, a scrounger, a beggar nor a thief.
I am not a national insurance number, nor a blip on a screen.
I paid my dues, never a penny short, and was proud to do so.
I don't tug the forelock but look my neighbor in the eye.
I don't accept or seek charity.
My name is Daniel Blake, I am a man, not a dog.
As such I demand my rights. I demand you treat me with respect.
I, Daniel Blake, am a citizen, nothing more, nothing less. Thank you.»*

The excerpt above belongs to the letter that Daniel Blake, protagonist of Ken Loach's eponymous movie, leaves after his death and represents one of the most expressive and moving denunciations of the bureaucratic machinery. Blake is a citizen who relies on public services to obtain various benefits to which he is entitled, and who in the course of his life continually encounters the rigidity and - at times - absurdity of administrative procedures. His denunciation touches on several interesting and characterizing aspects of the modern bureaucratic apparatus: its relationship with the private sector, which often infuses capitalist logics into administrations by turning citizens into clients (Yeung 2022); acting out a logic of suspicion toward citizens and considering them more or less "deserving" of support (see Chapter 3); and treating them as numbers. These characteristics suggest that bureaucracy has always been a particularly favorable terrain for technological automation, since even before the most recent processes of algorithmization and digitization, the founding goals aimed at standardization of procedures and greater efficiency, impersonality and objectivity. In order to

investigate the evolution of this, it is necessary to put into perspective the relationship between the two basic components of these processes: the bureaucratic one, related to administrative and governmental processes, and the technological one, related to the computerization and mechanization of procedures.

The mechanization of bureaucratic organizations has evolved significantly over time, reflecting broader changes in technology, governance, and organizational theory. Since the early adoption of machines in public offices, mechanization facilitated the standardization of administrative processes, enabled more efficient record-keeping, data processing and management of information (Fountain 2001). Max Weber noticed the relationship between the rapid changes in bureaucracy in the nineteenth century and the industrial revolution and famously defined bureaucracy as a “technology of control” that inevitably moves toward rationalization (Fountain 2001, p. 8). According to Weber, decisions and actions in a bureaucracy are made based on objective criteria and rules, rather than personal preferences or relationships. This impersonality is meant to ensure fairness and equality (Bonazzi 2002). Bureaucracies are designed to be highly efficient, and one of their key advantages is precisely the standardization of procedures that allow for predictable and reliable outcomes. Rationalization can imply that individuals are trapped in this system of rigid procedures, iron cages leading to a loss of creativity, autonomy, and humanity (Weber 1904). In 1991, Christopher Hood argued that the rise of New Public Management (NPM) marked a shift away from traditional, rule-based bureaucracy towards a system that emphasizes efficiency, performance measurement, and competition - often borrowing heavily from private-sector management practices. Hood identifies several key characteristics of NPM, including a strong focus on professional managerialism, explicit performance targets, cost-cutting, and the use of market mechanisms like competition and outsourcing to improve public service delivery. This model contrasts sharply with earlier approaches to governance, particularly Weberian bureaucracy, which relies on hierarchy, formal rules, and procedural accountability, and is largely driven by

economic pressures, growing skepticism about government efficiency, and the broader ideological shift towards neoliberalism. One of the key takeaways from Hood's analysis is that while NPM presents itself as a generalizable approach to public management, its effectiveness and consequences vary depending on institutional traditions, political cultures, and historical administrative structures. While he does not provide a comparative analysis of different national contexts in this paper, his discussion implies that NPM does not automatically translate well across different governance systems.

From the 1970s, with the development of more powerful computers and the rise of management information systems, computers became integrated into the day-to-day operations of bureaucracies. In 1990, Stuart Bretschneider was referring to computers in public agencies as an "ubiquitous tool for managing information, a tool which is technologically sophisticated and one which continues to experience rapid technological change" (1990, 1). The two main tools were decision support systems (DSS) and management information systems (MIS) (Ibid.). The ability to process large amounts of data allowed for more sophisticated planning, forecasting, and decision-making in government and corporate bureaucracies. That is why, looking at the evolution of information systems in bureaucratic organizations is a useful lens on the evolution of technological tools for standardization and decision-making support. Although it does not directly have to do with the relationship between bureaucratic organization and internal automation, e-government has developed as a paradigm since the 2000s due to the widespread adoption of the Internet and the development of Web-based technologies that have enabled the shift of many public services and administrative functions online. Citizens began interacting with services such as tax filing, license renewals, and inquiries to various departments online, with the promise of greater proximity to government and increased transparency. In 2004, Jane Fountain was among the others analyzing the implications of information technologies and the Internet for governance, noting that from 1993, the US government was increasingly "virtualized", or transferred online in a "web portal model" (2001).

She argued that it would carry both extraordinary possibilities for coordination, accessibility and communication and an increase in surveillance and control (Ibid.). In recent years, advances in AI, machine learning, and big data analytics have opened new possibilities for automating complex bureaucratic processes and decision-making, leading a number of scholars from various disciplines to study this new complex evolution. Geologist Rob Kitchin deals extensively with the implications of datafication in the field of critical data studies including the implications of algorithmic governance for privacy, data security, and democratic participation (Kitchin 2014; Kitchin & Dodge 2014). In *The Age of Surveillance Capitalism* (2019), Shoshana Zuboff focused on how digital technologies, including those used by governments, are reshaping power dynamics and citizen autonomy. Zuboff famously inaugurated a new season of studies on technological surveillance and control in the age of AI and big data (2015; 2019), after many other thinkers addressed it at other historical moments looking at other information technologies (see Deleuze 1990; Hardt and Negri 2017). In 1990, Gilles Deleuze wrote that societies of control - which were beginning to emerge and of which he anticipated the full realization in the future - were characterized by numbers and the tracking of patterns. With several references to Franz Kafka's *Trial*, where the bureaucratic machine results in more of an illusion of rationality, Deleuze remarkably describes the individual of societies of control as a "dividual" (1990, 7) that results from all the information contained in the different databases, in other words the different indicators and categories with which each of us is identified in analytics. Deleuze's idea about individuals as sets of indicators and numbers echoes one of the original characteristics of modern bureaucracies, as already traced by Weber, which were anticipated at the beginning of this section and extracted from Daniel Blake's letter. Far from being exhaustive, those three characteristics are in my opinion the most relevant ones for studying and observing administrative automation and the mixing of bureaucratic practices with an increasing use of information technology. Another of these characteristics is the intermingling of public and private, as Hood theorized with NPM to describe an approximation by the public sector to

efficiency-building logics typical of the private sector (Yeung 2018; Hood 1991) that make public administration “innovative”. Although it is not the purpose of this research to trace the reasons for the rise of these processes, one of the most important drivers for the algorithmization of public services is certainly the optimization of resources, also considering the reduction of skilled personnel (Yeung 2018). Karen Yeung traces the evolution of these practices in contemporary settings by arguing for a paradigm shift from management to analytics. Yeung discusses how governments are increasingly relying on large-scale data collection and analysis to inform decision-making processes. In the *New Public Analytics* (NPA) (2018), she describes a trend towards what she calls algorithmic governance, where human judgment is partially or entirely replaced by automated systems. For Yeung, this is a manifestation of the neoliberal shift in public management. The use of automation and algorithmic data analysis can also occur as an end in itself as a form of techno-solutionism and subjection to market rules (Purtrova & Huyskes 2024).

A whole area of research focused on algorithmic governance or digital governance has emerged in recent years. The difference between these two topics is subtle: in the former, as Yeung (2018) does, the implications of automation of governance and public decision-making processes are critically investigated, while in the latter, how to govern and regulate them is studied. In both cases, they explore how technology, particularly automation, artificial intelligence (AI), and digital systems are transforming the operation, management, and delivery of public services at the organizational level (Ziewitz 2016; Veale & Brass 2019; Danaher 2016; Sætra 2020). In recent years, with the burst of artificial intelligence, its interplay with bureaucratic organizations has been investigated more specifically (Mehr 2017; Meijer et al. 2021; Veale & Brass 2019). This new shift is the basis of this chapter’s interest.

In recent years, the European Commission has invested in research on the use of AI in the public sector of Member States. Looking at the geographic distribution of these cases of automation in both the national and local public

sector reveals a surprising result. The country with the most cases mapped by the European AI Watch is the Netherlands (according to the authors because the government itself maps them, so they are easier to find) and the second is Italy:

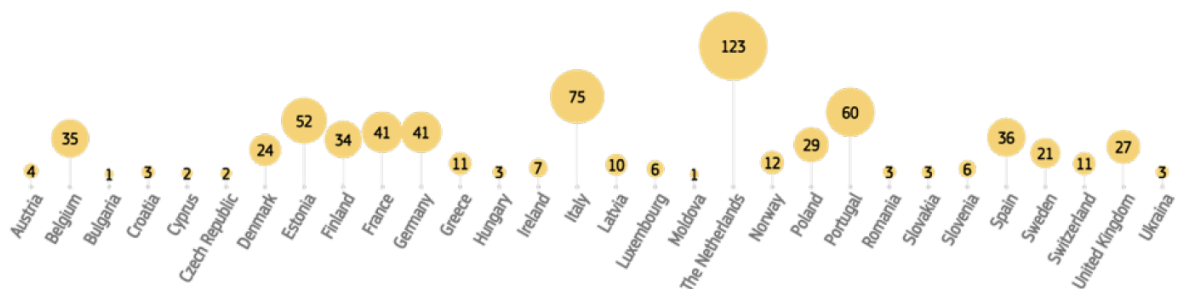


Figure 3: Mapping the use of AI in public sector in Europe (Tangi, L. et al. 2022).

It should be noted that the report had also included “logic/knowledge-based approaches, inference and deductive engines, symbolic reasoning, and expert systems, etc.)” among AI systems (Tangi et al. 2022). Nevertheless, the number of cases detected in Italy is still surprising, for reasons that will be detailed in the discussion of the data. Van Noordt & Misuraca analyzed several cases in order to map what are the major implications and changes in public administrations as a result of machine learning (2022). Them, among others (e.g. Sun & Medaglia 2019), pointed out that one of the first things that one notices is that in reality the adoption of these technologies by public administrations is very complex and often inconclusive. Michael Milakovich analyzed the premises of the alleged e-government revolution in various U.S. contexts and explored the impact of new digital technologies on public service delivery and citizen participation (2011). His book discusses how the adoption of information and communication technologies (ICTs) has transformed different public administrations, shifting from traditional bureaucratic models to more flexible, networked forms of governance, where digital tools play a central role in enhancing service delivery and administrative processes. The transition includes a change from traditional bureaucracy to electronic government (or e-government) leading to digital governance, “the advanced use of ICTs as enabling strategies for improving organizational performance” (Milakovich 2011, xv). According to the author, change in the public

sector was already “limited by internal and external factors, including, among others: bureaucratic protectiveness, existing laws and regulations, political interference, privacy concerns, obsolete technologies and infrastructures unsuited for the adoption of high-level technologies. In addition, accountability, transparency and lack of trust are major problems facing nearly all governments” (Ibid.). But, if looking at the characteristics discussed in this paragraph looks like traditional bureaucracy was already so mechanized before the arrival of early forms of advanced digitization, is there anything that more recent automation trends would change? And how do this relate to the external and internal factors - identified, for example, by Milakovich (2011) - that make technological advancement in the public sector difficult?

4.2.2. Blackboxing as a Practice

The availability of big data and increased computing power over the years have caused the models underlying AI and the analysis that rely on them to become increasingly complex (Mayer-Schönberger & Cukier 2013). An exponentially greater amount of information, sophistication of analysis and thus in the identification of patterns and correlations have led to these systems being defined as progressively inscrutable and difficult to interpret even for those who create them (Faraj et al. 2018; Stone et al. 2016). For this reason, the black-box metaphor has returned to use in recent years to refer to the AI patterns underlying many operations (Pasquale 2015; Burrell 2016). From this backdrop, an entire field of research on explainability and interpretability has emerged to seek mostly technical solutions that would make these models more interpretable and explainable (e.g., Rosenfeld & Richardson 2016). Famously, in the years of the promised revolution in big data analytics, Frank Pasquale wrote a book that critically examined the increasing role that opaque algorithms and data-driven processes were playing in societies (2015). Almost ten years ago, Pasquale brought the metaphor of the “black box” back to describe systems whose inner workings are not visible or

understandable to the public, leading to different concerns related to fundamental rights and surveillance. In various industries, algorithms were already central to decision-making processes despite the fact that they were inaccessible to public scrutiny and their incorporation into many procedures, especially in the case of government and public administrations, was not politically and publicly debated (Pasquale 2015; Burrell 2016). One of Pasquale's and many other scholars' main concern was towards the lack of transparency and accountability of these processes, that prevent the inspection for bias, discrimination and social exclusion.

Among the first scholars to use the black box concept, back in 1987, was Bruno Latour, to describe how scientific and technological processes become accepted and stabilized in society, to the point where their internal complexities and the controversies surrounding their development are no longer visible or questioned. According to Latour, "when a machine runs efficiently, when a matter of fact is settled, one needs focus only on its inputs and outputs and not on its internal complexity. Thus, paradoxically, the more science and technology succeed, the more opaque and obscure they become" (1999, 304). According to him and many other sociologists working within the STS, the key to the success of technological and scientific progress lay there, and they very often made explicit their goal of "opening the black box" (Pinch & Bijker 1987). For social constructivism, the phenomenon of blackboxing - which more than a phenomenon is the way in which almost all technological development has always been conceived and treated - has been a central element of both theoretical and empirical inquiry. In fact, the black box represents not so much the technical workings, the hidden "gears" of any technology, but more the social, political, and economic factors that influenced their development and that go unseen (Ibid.). These include the networks of actors who work in different ways for that technology to be feasible, whose actions, relationships, goals, ideas, negotiations and intentions normally remain obscured (Ibid.; Latour 1987; 1999). Winner commented on this conceptualization by social constructivism criticizing it (1993). According to the political scientist - who also distanced himself from technological

determinism - constructivism risked “opening the black box and finding it empty” because it often focused on some specific and contextual factors and failed to see other more structural and political elements that influence technology, such as power and its political consequences, marginalization of particular groups, or other social shapings that are not always evident during the development phase (Ibid.).

Less frequently, bureaucracy itself has been called a black box, although the metaphor is also perfect for referring to processes or decision-making systems that operate opaquely, that is, without the people involved (citizens, officials or even other bureaucrats) having a clear understanding of how decisions are made. This happens when the procedures, rules and internal mechanisms are so complex that they are inaccessible or incomprehensible to those outside the system. According to Graeber, in general, there is much less talk and study of bureaucracy in recent years than in the 1960s and 1970s, when, for example, there was a great focus on its analysis. In those years, social movements were also, according to Graeber, movements of opposition to bureaucracy, its lack of creativity, progress and flexibility. These instances were widely interpreted, in the old welfare states, in the manner of the private market (Graeber 2015). The presence of bureaucracy is felt less and Kafkaesque narratives are very far away because, according to the author, “we’ve just become accustomed to it. Bureaucracy has become the water in which we swim” (Graeber 2015, 6). Kafka described it as a huge and mysterious machine against which all attempts to penetrate are doomed to failure. The consequence is, precisely, forced inaction. As Latour (1999) wrote, the black box emerges when a technological process stabilizes, we begin to take it for granted, for inevitable, and we stop questioning it. Today, that instance of progress is identified with the digital transition of the bureaucratic apparatus. But what we see in many cases is that rather than progressing with technology, public administrations that become digitized risk becoming more opaque and less accountable. In the automation of government, public services, and bureaucratic practices, blackboxing is sublimated. Various opaque practices are layered until they create a dense and complex network that mixes human and nonhuman actors (Latour 1987). Algorithmic practices in

the public sector are complex socio-technical systems that, in the words of Karen Yeung, “comprise computational networks, people (data scientists, software engineers, interface designers, operators, reviewers, users), organizations, norms, and practices, all connected to a broader social effort oriented toward knowledge production” (2018, 506). This structure is reminiscent of what Karen Yeung calls New Public Analytics (NPA), structures that are so complex and so oriented towards optimizing their internal processes instead of citizens’ experiences, that they worsen the public good. As public analytics often involve complex algorithms and vast datasets, the decision-making process can become opaque, leading to challenges in ensuring accountability. At the same time, originally, the digitization of government always promised greater transparency as online platforms, social media, and other digital tools can facilitate greater public involvement in policy-making, leading to more informed and responsive governance (Milakovich 2011). Bovens and Zouridis documented the transition in large executive public agencies from street-level bureaucracies, to screen-level bureaucracies, to system-level bureaucracies (2002). The computer has played a symbolically important and more profound role than it may appear as it has significantly shifted the discretion that previously occurred at the street level, by one person and directed at other people, to the computer. This image is perfectly captured by the British humorous TV series *Little Britain*, which in 2004 popularized the saying “the computer says no”: the main character was a perfect screen-level bureaucrat who, before answering any customer-citizen question, would approach the computer monitor to “hear” what the computer had to say and reply “the computer says no” (Huyskes 2024).

As different authors argued, computers can be interpreted as actors in the network (Latour 1999; Yeung 2018) or as something that rather influence the network and reframe the relationship among its members, responsibilities and competences (Wihlborg et al. 2016). As Christin (2020) notes, ethnographic methods in the study of algorithms are useful in overcoming algorithmic opacity by decentralizing the analysis: instead of focusing on the “black boxes” of algorithms,

ethnographers can study how collectives of human and nonhuman actors emerge, solidify and evolve over time. However, this proposal, which qualifies the active and more or less intentional role of human actors, can be complex to apply, as the next section discussing the data analyzed will illustrate.

4.3. Transparency Barriers in Italian Algorithmic Governance

The primary objective of this chapter is to comprehensively understand how Italy is approaching the automation of public services, with a particular focus on how social and cultural values influence the adoption and governance of these technologies. However, despite a structured methodological approach that included interviews and ethnographic participant observation with a representative sample of informants, I encountered significant obstacles. A considerable barrier emerged when attempting to engage with certain key figures, particularly those within ministerial teams and public agencies known to be involved in algorithmic governance. Despite prior knowledge of their involvement in such practices, gained through my work within the Automated Administration Observatory project, many potential informants were unresponsive. Out of the 35 individuals initially identified for semi-structured interviews, in fact, only 14 agreed to engage. Hence, the research presented here can be viewed as a partial failure, albeit one that is revealing in its own right. It is not uncommon to encounter a situation like this when investigating algorithms, platforms, or other technological components. For example, Gandini and Bonini experienced these difficulties when they wanted to study music curators working in major online music streaming platforms (2020). The authors pointed out that much of the knowledge we have about the workings of the cultural industry comes from the work carried out by social scientists who were able to access and investigate it with ethnographic methods in the past (Ibid.). This is what ethnographic methods are intended for: to analyze certain social and cultural practices “from within”, and it is also what social constructivism - particularly in its feminist expression of gender and technology studies -

recommended doing with artifacts (Bijker & Pinch 1989; MacKenzie & Wajcman 1999). For example, one study that guides this research is Cockburn and Ormrod's famous work on the microwave oven (1990). The authors, interested in investigating the social construction and in particular the influence of gender relations behind this domestic technology, compiled their ethnography through a study that "followed the actors" with a constant inside look that started from the factories and went all the way to the stores. They talked to the designers, engineers, and workers who worked to build those technologies, discovering what their intentions and values were; the legislators who regulated their production and sale; and the salespeople who had to convince customers to buy them. Today it seems that this practice, so fundamental to unveiling the values with which technologies are constructed, is increasingly difficult. Similar to Bonini and Gandini (2020), the hypothesis of this research was that by extracting information directly from the (bureaucratic) organizations that decide to employ algorithms (and then, therefore, at least theoretically, govern them) - in this case in Italy - it would be possible to motivate the use of these technologies and its methods as related to the culture they belong to. Although the kind of access to this information went differently than I expected, I cannot say that I completely failed to test my initial hypothesis. As Seaver already documented, "A great deal of information about algorithmic systems is available to the critic who does not define her object of interest as that which is off limits or intentionally hidden" and that "challenges to access hidden meetings, reluctant interlocutors, non-disclosure agreements are part of the field, not simply barriers around it" (2017, 7).

When looking closely enough - even if "circling around" the topic from different vantage points and not from within - and accessing various sources of information regarding the governance of the algorithmization of public services in Italy, one can see that the reasons for the lack of transparency are many and varied. Based on what I gathered in the field and from the analysis of various other sources, I propose some of these reasons, namely (i) legal barriers; (ii) lack of awareness and competences; (iii) private interests; (iv) lack of governance and

strategy; (v) accountability; and (vi) bureaucratic resistance, to which I will devote individual sections in the next pages.

4.3.1. Legal barriers

First, it should be noted that public agencies and governments, compared to private companies, are more unlikely to be able to entrench themselves behind secrecy, if only industrial secrecy. Moreover, as one key informant working at an independent supervisory authority reported, public organizations should have different aims than private companies:

«When we talk about this issue (the automation of public services or bureaucratic activities) we should never forget that the public good should always be a fundamental variable of the project. Because the public is the public, it should not guarantee profits or protect products: it should act for the benefit of citizens.»

(Informant 1, person working at the Data Protection Authority (DPA))

Acting in the interest of citizens also means protecting them. This is the idea that justifies some of the exceptions of public transparency. One of these is the article contained in the Legislative Decree of 14/03/2013 No. 33, which regulates the right of civic access and the obligations of publicity, transparency and dissemination of information by public administrations. Within it, it is stipulated that requests for information may be refused if sharing them would jeopardize any of the public interests inherent to: a) public security and public order; b) national security; c) defense and military matters; d) international relations; e) state policy and financial and economic stability; f) the conduct of investigations into crimes and their prosecution; and g) the orderly conduct of inspection activities. After reading this list, it is quite immediate to recognize that given its breadth and comprehensiveness many requests for information can be rejected by appealing to these interests, especially considering that the person submitting the request generally does not know the contents of the documentation and, therefore, the administration can call this article into question without any possible verification of

whether there is actually a risk to the public interest. Another possible response by government offices to avoid answering this question concerns citizens' personal data. Several of the responses the no-profit organization Privacy Network received in recent years when submitting requests for the Automated Administration Observatory's mapping cited the inability to share, in particular, the Data Protection Impact Assessment (DPIA). The DPIA is required under the General Data Protection Regulation (GDPR) for every new project that is likely to involve a "high risk" to other people's personal information. Given the amount of personal data that a public administration normally holds and uses for automation projects, many public digitization projects require these assessments.

Name of the project	Owner	Documentation received
Redditometro	Agenzia delle Entrate	No
SARI	Ministero dell'Interno	No
Graduatorie GPS	Ministero dell'Istruzione	Yes
VeRA	Agenzia delle Entrate; Ministero dell'Economia	No
Frozen	INPS	No

Table 5: main active algorithmic processes in Italian public administrations and central agencies mapped by the Automated Administration Observatory up to September 3, 2024⁷

Apparently, however, these exceptions are used by administrations even when unnecessary:

«Many times, when administrations are unwilling to share the DPIA, it is because they do not have it. They therefore hide behind the excuse of not sharing personal data - when they could just obscure it from the document - because they haven't done it. It should be noted that the impact assessment under the GDPR is not always mandatory, quite the contrary. Certainly, however, having it done and shared would be an important step toward transparency and accountability.»

(Informant 1, person working at the Data Protection Authority (DPA))

⁷ The Observatory is accessible at: <https://privacy-network.it/osservatorio/sistemi-automatizzati/>

4.3.2. Awareness and competence

Thus, there exists an active intention and deliberate choice to avoid scrutiny regarding the presence of algorithms and/or automated processes. However, a second relevant aspect that emerged from the analysis of qualitative data is an unintentional, or unconscious, component related to non disclosure:

«There is a gap in awareness within public administrations about the algorithms being used, especially in the context of purchasing software solutions from vendors. Many administrations may not even realize that an algorithm is part of the solution they are using.»

(Informant 1)

This is one of the most striking pieces of evidence that emerged from the interviews and the participant observation sessions. Several people, in fact, on these occasions reiterated how widespread unawareness of the presence of these technologies is for a large number of public personnel at different levels. Many of the interviewees who were willing to speak expressed either a limited understanding of the algorithms in use within their institutions or a hesitance to discuss the specifics of these systems, often citing bureaucratic or legal constraints.

«Only a small portion of administrations are consciously using algorithms, with many relying heavily on vendors without fully understanding the technological implications.»

(Informant 1)

One of the key issues highlighted in relation to unconscious non-transparency is the lack of digital literacy within Italian public administrations:

«We are a country in which digital competence is not there...if we do not have the basic digital competencies, how can I expect that these are all systems they can govern?»

(Informant 3, person who worked in a central public agency in charge of digital innovation)

The lack of specialized technical expertise is a significant issue within the public administration, another informant stressed:

«There is an a priori adverse selection: the administration does not hire computer scientists but lawyers and economists in the best case. People like me are the exception, and we are here because in fact we came through agreements with private companies.»

(Informant 5, person who worked for the National Social Security Agency)

Without the necessary skills, administrations are unable to critically assess the solutions they are implementing, leading to a high probability of misuse of technology. Analyzing the collected data, lack of expertise emerges as the first reason for distrust and skepticism about the implementation of algorithms and other IT infrastructure useful for policy-making and important decision-making processes in public services. In all the opportunities for open discussion and participant observation, conferences and meetings with various officials from public agencies and ministries and external experts that I have been able to attend, the issue of competencies has always emerged. Even in the National Strategy it is recognized as a problem, and in fact it is proposed “a training system that aims at excellence, on the growth of talent possessing skills consistent with the emerging scenario and the efficiency of Public Administration services through the introduction of AI solutions” as cross-cutting priorities (AgiD 2024, 5). Despite this, speaking with some of the stakeholders, no structural solution appears to be planned to modernize the professional figures involved in the digitization of public agencies and significantly increase their skills, except with one-off courses that delve into the topic of AI.

4.3.3. *Private interests*

To analyze the adoption and use in Italy's central public administrations of artificial intelligence (AI) algorithms and systems, I realized almost immediately that it was not enough to simply engage with public actors. It was crucial to also talk to private actors, the real executors and controllers of public algorithms in Italy. This, in terms of personal data, requires special attention:

«It's hugely important keeping data within the control of public institutions rather than outsourcing to private companies, which may pose risks related to data privacy and sovereignty. All our data is stored by Google, Microsoft etc.»

(Informant 5, person who worked for the National Social Security Agency)

Moreover, in Italy, there are several large service provider companies that are juridically private but are controlled by the State, in some cases even completely. This is also the case with technology service providers. The fact that these are always external companies, even if they are state-controlled, leads to a risk of over-reliance:

«I would say that, first of all, my suspicion is that the public administration does not always know how to use an algorithm because most of the public administrations are in the hands of service providers from this point of view. So they buy a solution, and the dynamic of operation, the technology behind that solution, they know nothing about.»

(Informant 1)

Public administrations rely and depend extensively on these providers. That of external providers is a significant role. They often drive the process more than the public administrations themselves, as one person working for one of these companies told me:

«Dependence on outside vendors highlights a potential weakness in the public administration's approach to technology. Without a strong internal capacity to manage and understand these technologies, the government becomes dependent on external contractors,

who do not always prioritize the public interest or fully meet the administration's needs.»

(Informant 4, person working for a state-controlled company operating in the ICT sector)

Furthermore, agencies normally adopt technologies because companies propose it to them, convincing them commercially of its usefulness. It is rarer for the need to arise within the administration. This leads to a lack of decision-making sovereignty on the part of the administrations themselves and further reinforces over-reliance on external partners:

«The relationship between public administrations and private companies is characterized by an imbalance of power, with private companies often taking the initiative to provide solutions without adequate oversight or understanding on the part of the public sector.»

(Informant 6, person working at the technical office of the Prime Minister (Presidenza del Consiglio dei Ministri))

So a vicious cycle is set in motion: by outsourcing, public agencies do not invest in in-house expertise, all of which results in not only the development, but also the monitoring and governance of these tools remaining outside the public administration:

«A loop starts because the automated or algorithmic process is often driven by immediate needs, such as simplifying bureaucratic tasks or adhering to new regulations. However, these solutions are sometimes adopted without a deep understanding of their impact or even an awareness that an algorithm is involved, because everything happens really fast and they rely on companies' promises.»

(Informant 2, lawyer)

And so, the vicious cycle ends exactly as it began: with general unawareness.

«The administration, in good faith, does not know that it is using an algorithm...the user, the administration, may not even perceive it, in many cases, they do not perceive it.»

(Informant 1)

4.3.4. Governance and strategy

The implementation of algorithmic processes in Italian public administration is often reactive rather than proactive. In particular, from the data, and especially from the analysis of official documents released by the Agency for Digital Italy (AGiD) in recent years, such as the last National Strategy on Artificial Intelligence, two aspects emerge that are particularly relevant for the purposes of this article: (a) the absence of a strategic plan, because there is much mention of investments and a national industrial plan but no detail of how - beyond development and economics - the adoption and governance of these processes are to be managed; (b) a thorough understanding of its implications, because there is no explicit mention in the documents of strategies for managing or observing the societal implications of introducing AI into public services (AGiD 2024). Moreover, there is a notable gap between the policies and their implementation on the ground:

«Execution often falls short due to outdated systems and resistance from those responsible for implementing these policies. This is not a country where you can innovate. At some point, everything comes to a standstill. As with the creation of 3-i, which is also promised in a government act, which is still not there and has become a mere bureaucratic exercise of funds allocated and people moved around.»

(Informant 5, National Social Security Agency)

3-i S.p.a is the first Company with wholly public capital and participation that brings together the IT skills and services of Public Administrations and in particular three national institutes: for social security (INPS), for accidents at work (INAIL) and for statistics (ISTAT) to “develop technological innovation, interoperability, optimizing investments and ensuring increasingly high levels of security and cybersecurity in the delivery of digital public services”. Promised in

2021, three years later it is still not operational, although there is a specific website announcing it.⁸ Automation is perceived as a tool for efficiency, but there is a mismatch between its adoption and the understanding of its broader impact. The consequence for the absence of sufficient foresight or planning is a substantial lack of transparency and accountability, which increases the risk of inefficiencies and potential problems related precisely to social impacts.

4.3.5. Accountability

The reluctance to disclose the use of algorithms reflects a deeper issue with transparency in public administration. The Italian Data Protection Authority (DPA) noted that algorithms are not always managed or disclosed in ways that meet privacy regulations. In case these processes are found to violate citizens' rights or if they result in discriminatory outcomes, this could lead to significant legal repercussions. All of this suggests that these matters are not talked about within administrations, and that technological innovation is a goal in its own right and not something integrated or to be integrated into pre-existing processes:

«There is a certain reluctance to declare it...what might seem like a human error, obviously, it's hidden behind the system.»

(Informant 2, lawyer)

The informant speaks with this certainty that a human error is attributed to the system because indeed in recent years, in Italy, in cases where a public algorithm jumps to the headlines, it is made responsible and, in a sense, humanized by attributing blame or credit to the software or calculation itself. The tendency to “blame the algorithm” is certainly not a typically Italian attitude. As shown, for example, by Maasland and Weißmüller, in various contexts algorithms are given the most “unpleasant” tasks so that they can then be blamed for errors, unethical or discriminatory choices (2022). In this way, rational calculation justifies unethical conduct that no longer passes for human beings. Doing so, however, contravenes

⁸ <https://www.3-i.gov.it/>

one of the most important principles of modern algorithmic governance, namely that of leaving the ultimate decision - and thus responsibility - to the human, also mentioned in the latest National Strategy (AGiD 2024) and in an important court ruling, as will be detailed below.

4.3.6. Bureaucratic resistance

Automation is not necessarily seen as a solution by Italian public administrations. Although we can certainly say that there is a general interest with respect to its use and that it is growing (the Automated Administration Observatory, for example, has intensified its work over the years due to a growing attention to the topic but also in the number of technologies adopted), the rate of interest and adoption found in Italy is not comparable to that of other European countries in public administrations, such as the Netherlands. Going back to the mapping done by Tangi et al. for AI Watch (2022), there appears to be a large gap between the cases detected and the awareness of the Italian public sector with respect to automated processes:

«In general, automation seems more of a challenge than a solution within public administration.»

(Informant 5)

While it offers an opportunity to streamline processes and reduce manual labor, there is skepticism about its impact on jobs and what it might cause on the overall functioning of the administration. This is probably related to the lack of expertise within organizations to “lead” these projects from a strategic perspective as well. Another piece of evidence relates to the type of automation most prevalent in the processes of Italian public bodies, which has so far been limited to very simple techniques. Mostly, in fact, these are basic algorithms, simple mathematical formulas and rules that are coded. The last AI Watch report decided to include

these in a broader definition of AI, while today they would not fall into the scope of the definition provided by the European Artificial Intelligence Act (AI Act). Despite this, transparency regarding their existence and use remains unsecured. Quite remarkably, last year we received a response to a FOI request that adds an interesting element to this scenario. As is our practice, we were asking one of Italy's most important government agencies for information with respect to the active or inactive status of an algorithm that had been reported to us, and the administration replied that they "did not use algorithms, only computer systems" and therefore would not share the information. Despite the terminological confusion, a computer system is something even more complex than an algorithm, in that it encompasses software and hardware elements and, as an infrastructure, could also accommodate multiple algorithms.

All these shortcomings contribute to the difficulty of implementing and managing automated processes effectively. As a result, public administrations are afraid to fail. A key actor who witnessed these sentiment from the inside highlights the role of bureaucracy as a significant barrier to the effective implementation of technology:

«We currently experience a public administration that is basically schizophrenic, hovering between maintaining traditional processes and trying to incorporate new technologies.»

(Informant 4, person working for a state-controlled company operating in the ICT sector)

This creates a situation where technology is introduced but not fully integrated into decision-making and operational processes, leading to increased complexity rather than simplification, as he/she stressed:

«Bureaucracy is a stopping force that puts the brakes on any kind of activity.»

(Informant 4)

The persistence of traditional bureaucratic structures and mindsets seems to be one of the main notable obstacles to the adoption of new technology. This is particularly evident when, for example, automation occurs on already problematic bureaucratic bases, or on procedures on which critical issues have already been found:

«Even when technology is adopted, it often adds layers of complexity rather than streamlining processes, as seen in the example of SPID (Public Digital Identity System), which is overly complex and not easy to use.»

(Informant 4)

Thus, on the one hand, there is a resistance to changing pre-existing procedures, and on the other hand, there is a belief that forms of digitization can be introduced without adapting processes:

«There is strong resistance to digitization within the public administration, often rooted in cultural and bureaucratic inertia. Many decision makers prefer to maintain the status quo, in part because of a lack of technical understanding and fear of losing control over processes.»

(Informant 6, person working at the technical office of the Prime Minister (Presidenza del Consiglio dei Ministri))

Importantly, this was also emphasized by a key actor working for the European Commission who contributed to the mapping of European administrations and their relationship with automation for the AI Watch:

«The Italian public administration tends to be more resistant to the adoption of new technologies compared to other European countries.»

(Informant 12, research staff member at the European Commission)

This resistance is deeply rooted in bureaucratic culture and reluctance to innovate, which contrasts with the more forward-looking and innovative

administrations of countries such as the Netherlands. Looking at a specific case of algorithmization in the Italian context can help to better understand the scenario that has been built so far around these technologies. In particular, in the next section, I will describe and summarize the setting related to the facts that marked the first case of algorithmization of public procedures and led to several problems. Among them, several of those discussed in this section emerge.

4.3.7. A Lesson from the Algorithm: the Buona Scuola Case

In 2015, the Renzi government introduced a school reform that among other objectives had the goal of making the working conditions of teachers more stable. Among the various measures included in the “Buona Scuola” one of the most relevant was the permanent hiring of about 100,000 precarious teachers through an extraordinary recruitment plan. In order to carry out this measure, the Ministry of Education made use of an algorithm. Specifically, the algorithm was supposed to support the allocation of tenure and distribution of newly hired teachers to various schools and regions in Italy. The factors it relied on to make these distributions included preferences expressed by teachers, seniority score, and curriculum. However, the procedure created numerous problems, causing errors in the assignment of workplaces. Many teachers found themselves assigned to schools very far from their preferences or even in regions they had not listed. This created great discomfort, especially for those who lived in one area and suddenly found themselves transferred hundreds of miles away. This led to a phenomenon that has been described as a real “exodus” of teachers from the South to Northern Italy (Zunino 2019). Many teachers complained about the lack of transparency in the process, as it was unclear how scores were actually calculated and assignments made. For example, one of the concrete impacts of this procedure was to favor teachers with less experience and lower scores (but perhaps younger) rather than those with higher scores and experience. This has created a strong sense of injustice and disaffection. Following numerous protests and appeals, the Ministry

of Education had to intervene, trying to remedy the errors with corrective reallocations and legal appeals, but the damage was already done. The chaos created by this process has left a deep mark on the careers and lives of many teachers.

In 2019, a number of court rulings intervened to judge what happened and the role of the algorithmic procedure in the allocation of positions. Both rulings, though very different, provided for the first time a set of principles to be respected in the use of automatic calculations in decision-making processes involving Italian citizens, such as the centrality of human judgment. According to the Regional Administrative Tribunal (TAR) of Lazio, in the case of Buona Scuola, it was a “serious administrative flaw” to leave decision-making power to an algorithm that was not checked and validated by humans in its outputs and that a technical report from the Court described as “confusing, lacunose, ampullary, redundant, elaborated in two different programming languages, one of which dates back to the prehistory of computer science, built on input data managed in the wrong way” (Zunino 2019). According to the judges, this was a method that conflicted with constitutional principles and fundamental rights: “computer procedures, even where they reach their highest degree of precision and even perfection, can never supplant, truly fully replace, the cognitive, acquisitive and judgmental activity that only an inquiry entrusted to a person officer is capable of performing” (Ibid.).

The Buona Scuola case echoed in Europe as an example of a practice not to be repeated when implementing automated procedures for decision-making purposes. Precisely because this algorithm dates back almost ten years, when these practices were not yet widespread, and despite the fact that it was a technically very simple procedure and that quite some time has passed, the case still serves as part of the Italian culture on public algorithms and the identification of the case with the broader topic is strong. According to one interviewee from the public sector:

«In my opinion, the Buona Scuola and everything that happened afterwards - the rulings, the courts, the media attention and resonance even outside of Italy - has set a strong precedent in

Italy. It showed how translating administrative and bureaucratic practices into mathematical language is nothing obvious and simple, and it created fear in administrations.»

(Informant 4, person working for a state-controlled company operating in the ICT sector)

Some interviewees shared that the Buona Scuola was probably a case that nipped in the bud the Italian desire or attempt to automate some administrative procedures. Returns the idea that automation of a procedure can solve everything, and that there is no need to review the procedure itself first, before it can be coded:

«There is a certain easiness to make things simple so much technology intervenes that then solves but it is not true.»

(Informant 4)

This observation in particular underscores a critical issue in technological adoption: the tendency to underestimate the complexity of integrating new systems into existing processes. The expectation that technology will simplify processes often leads to disappointment when the reality is that these technologies introduce new challenges that must be carefully managed. While there is a significant potential for using data to improve public services, there is also a pervasive fear of violating laws, which can inhibit the adoption of technologies. This appears contradictory to the numbers found by AI Watch, which in 2022 noted a high number of algorithmic processes employed locally and centrally. In fact, the Ministry of Education has never stopped using algorithms. In recent years, from the Buona Scuola onwards, a number of algorithms have been employed to automate the assignment of teacher positions, and in particular those in the rankings for qualified teachers (GAE) and those for substitutions (GPS).

Bizzini (2023) reconstructed for AlgorithmWatch the functioning of the most recent version of the teachers' algorithm, dedicated to intersecting these two rankings and moving from an in-person to a digitized recruitment system. The

factors used to make the decision are very similar to those of 2015, and the procedure is complex. Errors continue from year to year: “thousands of teachers all over Italy have been subjected to discriminatory choices by the algorithm. Some teachers have been overtaken in the rankings by colleagues with lower scores or been assigned to support roles for those applying to cover a specific subject.” (Ibid.). According to a teacher that I interviewed and that was among the many whose job was affected by the automation of the procedure:

«Despite my qualifications, I remain stuck in a temporary employment cycle, a situation many educators face. My situation exemplifies the widespread issue of job precarity in the Italian education sector, where even qualified teachers face instability due to the bureaucratic and algorithmic nature of recruitment processes.»

(Informant 7, teacher who was affected by an algorithm)

With a privileged perspective as an affected person, he/she adds that:

«The algorithmic system that allocates teachers is often perceived as unfair and arbitrary. Despite having higher scores and ranking in the first tier, I was assigned a part-time role, while lower-scoring teachers received full-time positions. This points to a lack of transparency in how the algorithm functions, frustrating those who rely on it for fair treatment.»

(Informant 7)

On May 28, 2022 and September 29, 2022, the Privacy Network team that works on the Observatory sent two requests to the Ministry of Education to obtain information about this algorithm and its role in the overall procedure, the functioning of the process, and the relationship between algorithmic recommendation and human decision. Having not received a response to the first request, we sent another request, receiving a response on October 27 of the same year. At that point something completely unexpected happened: the Ministry, responding to our requests, sent us a lot of documentation, the entire

documentation we requested, including the algorithm code. Something very similar happened in Rotterdam, the Netherlands, in 2023, when the municipality sent Lighthouse Reports - which sent out the request - exhaustive documentation about a risk profiling machine learning algorithm that assigned welfare benefits to citizens (Casimir-Braun et al. 2023). In the Dutch case, this was likely because of the welfare scandal that hit the country in early 2021, which subsequently prompted the opening of the black box especially with respect to algorithms used in bureaucratic administrations. There would be nothing unusual about this, except that in our case it was unprecedented that an administration openly and transparently shared the required information about an algorithm. It is particularly interesting to put this data together and to note that the first and only one was the Ministry of Education, which was the only central public administration in Italy to be previously involved as a protagonist in a case of negative automation that made it to the public debate.

4.4. Behind Closed Doors: Measuring Technological and Institutional Black-Boxing

As Wieringa (2023) points out, public accountability is severely challenged by algorithmization, especially when these systems are opaque and difficult to scrutinize. Having closely observed the contexts in which public decision makers work around major central agencies, their intentions, and having followed some of the cases involved, it emerges that generally these processes are opaque and difficult to access. This is so regardless of the level of technology, algorithmic and technical sophistication. As reported in the theoretical background, Max Weber and many others already recognized bureaucracy as a technology of control. Processes are optimized, the relationship with the public becomes more client-based, and the citizen is rarely seen in its specificities but more as a number. The bureaucratic apparatus, the machinery of government originates to serve for

efficiency and rationalization, and because of this, algorithmization resides in its very purpose, and automation is nothing more than its sublimation. In this genealogy, as many previously identified (Milakovich 2011; Kolkman 2020; Vydra and Klievink 2019), there are several limitations and barriers to an enabling use of digital technologies.

Bureaucratic processes are much similar to statistical ones as they operate with structured objectives but are inherently opaque, often resulting in a system that neither the public nor administrators fully understand. One interviewee captured this perfectly:

«(Technological) processes are 90% undocumented, 100 percent unshared, and least of all explained.»

(Informant 4)

All of this, focused on the Italian case, demonstrates the various phases of a single loop that shows back the inextricable nature of technological and institutional black-boxing, which together amplify the complexity of these systems and reinforce one another. If bureaucratic structures can cause harm and negative impacts in subtle and intangible ways, the automation of these procedures - when they already start as problematic - through mathematical models could result in unprecedented levels of injustice (Eubanks 2018).

Probably the most problematic consequence is what another informant, thinking more generally about the automation of bureaucratic processes, reported to me:

«If you don't know what's wrong, you can never put it right.»

(Informant 28)

There are concerns about the legal and political responsibility for errors or issues arising from the use of automated systems. Since Frank Pasquale's (2015) *The*

Black Box Society, where he warned about the growing opacity of automated systems, and the discourse around black-box algorithms, we have become very accustomed to taking almost for granted the impenetrable nature of the technical features that underlie them. My sense is that by doing so we are supporting those who should be in control and doing the most to ensure transparency, understanding and governance of the processes and instead have the excuse to hide behind the complex nature of these systems. The Italian case, in its entirety, shows this well: despite the technical simplicity, damage can be done and continues to be done if the skills and governance structures are not adequate. However, one cannot simply blame an algorithm for the problems that have arisen, for which instead several actors bear responsibility. The reluctance of ministerial teams and public agencies to engage in this research underscores an institutional opacity that mirrors the technological black-boxing of automated systems. The difficulty in accessing these informants, therefore, represents a crucial finding in itself. It suggests that the processes governing the use of algorithms in Italian public services are not only technologically opaque but are also institutionally, and at least partially intentionally, black-boxed. Furthermore, it suggests a broader cultural resistance within Italian institutions to scrutinize and critically assess the role of automation in governance. The question which arises is: what if it is the lack of institutional transparency that makes these technological processes opaque, and not the other way around? At the expense of the more recent and mainstream narrative of algorithms as black boxes, this evidence is more akin to what Latour originally noted, namely that the black box does not so much represent technical functioning but rather the social, political and economic factors that surround it, the actors that govern it and the actions, goals and ideas that drive it (1987; 1999). Along these lines, this research shows that even in the face of a partial failure related to accessing direct data on technologies and their governance, it is still possible to obtain important information - by adapting the methodology, inspired by constructivist ethnography, to the practice of “scavenging” information (Wieringa 2023) - and reconstruct the human opacity that intersects on different levels with the algorithmic one. In most

cases, instead of studying institutions as black boxes, there is a tendency to shift the focus to technology and blame it for its lack of transparency, justifying the inability to access insights and knowledge about the underlying processes. Notably, this narrative deflects responsibility away from institutions, reducing public accountability. Focusing on the institutional barriers to technological transparency also allows us to unveil the cultural factors that perpetuate this lack of openness in the face of increasing automation.

4.5. Conclusion

In the first section, I analyzed the meaning of the idea of bureaucracy as a machinery. Far before technological automation intervenes, the bureaucratic apparatus has been interpreted as an impersonal machine, aimed at rationalization (Weber 1904) and efficiency often at the expense of other important traits (Fountain 2001). When early technologies such as decision support systems (DSS) and management information systems (MIS) were introduced in public administrations, this prompted the first reflections and studies on the relationship between the machinery of government and technological means (Ibid.; Kitchin 2014). The individual, for bureaucracies, was already seen as a number, a component of a whole to be treated mechanically in the same way, but this reaches more critical levels with algorithmization (Deleuze 1990; Kitchin & Dodge 2014) until a regime of public analytics and algorithmic governance, which describes the tendency of governments to rely increasingly on data analysis and algorithms to make important decisions (Yeung 2018). Among the most troubling implications of this shift is that of the opacity of processes, which reached major attention with the work of several scholars who used the famous black box metaphor to describe them (Pasquale 2015; Burrell 2016). In fact, the same metaphor was commonly used by constructivist scholars investigating the relationship between technology and society, and in particular by Latour (1987; 1999), who considered black boxes

those machines (but not necessarily technological ones) that are stabilized and no longer questioned. Therefore, their call was to “open the black box” and look for the social influences that could enable the analysis of processes deemed as inevitable. In line with this literature and proposal, section two outlines the methodology used for this chapter, specifically the technique of “scavenging” to access difficult and opaque knowledge (Wieringa 2023) and my own experience in accessing the field. The next section analyzes the evidence that emerged from data collection in the Italian context, describing from different perspectives the relationship between Italian policy and culture and the algorithmization of public services. The picture that emerged describes a profound lack of transparency on multiple levels, which, however, importantly turns out to be much more multifaceted and complex than what emerges from the literature. With the example of the Italian context, in fact, we can observe that the transparency problems related to bureaucratic automation are both intentional and unintentional, and the barriers are not imputable to technology alone but rather to several human, organizational and institutional components, such as policies, lack of skills and awareness, the influence of private interests, aspects of governance and strategic choices and a more general bureaucratic resistance to the full adoption and governance of algorithms. In this chapter, I argue that the black-boxing of algorithmic systems and the processes that govern them are not just a technological phenomenon, but are deeply rooted in institutional practices that resist scrutiny. The Buona Scuola case, where teachers were wrongly assigned to schools far from their preferences due to a flawed algorithm, is a stark reminder of the real-world consequences of such institutional opacity. Wieringa (2023) points out that public accountability is being challenged by the rise of algorithmization, and this article extends this thesis by showing how the institutional opacity of Italian governance is primarily responsible for the problem. The reluctance of public officials to participate in this study, coupled with the difficulties in accessing data, highlights a broader cultural resistance to transparency. This resistance is not merely a result of technological complexity but is deeply embedded in the

governance structures that oversee the implementation of automation. Rather than viewing the algorithms themselves as inherently inscrutable, the lack of transparency is equally, if not more, a product of institutional practices that obscure decision-making processes related to the use of such technology and thus need more specific attention for research. The findings suggest that the lack of institutional openness is a key factor in the opacity surrounding algorithmic systems, and that addressing this issue requires a cultural shift within public administration. The significance of these findings lies in the profound impact that institutional black-boxing can have on governance, public trust and future technologies. The Italian case provides a cautionary tale for other countries grappling with similar challenges, emphasizing that effective governance of automation must go beyond technical solutions and address the institutional and cultural barriers that perpetuate opacity and hinder accountability. The black-boxing of these processes allows institutions to deflect responsibility, framing failures as purely technological rather than acknowledging the institutional failures that contribute to these errors. Therefore, it is important now to delve deeper into the cultures that root accountability in technologies and to investigate more deeply the relationship between control, oversight, human judgment and the automation of processes.

5. “The Algorithm Made Me Do It”: Automation, Culture and the Erosion of Accountability

5.1. Introduction

Modern technologies, especially those involving algorithms and automation, are increasingly relied upon to manage critical public services, from healthcare and welfare distribution to legal and bureaucratic processes. As algorithms take on more roles in decision-making, they begin to serve as a new form of authority, one that is often seen as more impartial than human officials (Cabitza et al. 2023). However, this perceived objectivity and neutrality can obscure the fact that algorithms are designed by humans and reflect specific values and priorities (Van de Poel & Kroes 2014). Evidence was provided in the previous two articles on how different cultures respond differently to technological implementation (Hughes 1987) and how this results from human and deliberate choices. For this reason, we can observe that technologies - in this case, algorithms and automated systems used in administrative and bureaucratic contexts - are not predetermined and inevitable, because the same technology can be developed, used and interpreted in different ways depending on the cultural and social context in which it is embedded (Bijker, Hughes & Pinch 1987; MacKenzie 1993). However, by looking closely at these technologies, particularly through ethnographic methods that allow us to observe primarily the humans who gravitate around them, it is also possible to note more general behaviors in the social construction of technology that have to do mainly with processes of legitimization and accountability. We can thus observe that even in different cultural contexts some traits in how technologies are justified, monitored, and controlled actually reoccur.

In this chapter, I will ground my analysis in the literature on legitimacy and public trust. The increasing reliance on algorithms has led to a shift in how legitimacy is conferred in public systems (Suchman 1995), with algorithms becoming sources of legitimacy in themselves, often without the transparency and scrutiny that would traditionally accompany human-led decision-making processes. In the theoretical background, algorithmic accountability will also be discussed as a relevant conceptual framework to discuss notions of responsibility, trust and reliance on automated procedures. Although, as many authors highlighted, the history of technology revolves around errors and failures (Hughes 1987), different examples prove how we take for granted the infallibility of automation. Automation bias, a well-documented phenomenon, occurs when humans defer to technology, assuming its outputs are more reliable or accurate than human judgment. Yet, this attribution of accuracy and objectivity masks the fact that responsibility is increasingly shifted from people to data and algorithms without adequate scrutiny or accountability. Following a brief outline of the methods used for data collection, I will discuss the key findings, focusing on commonalities that emerged from interviews with individuals in Italy and The Netherlands, and argue that automation is not only a reflection of the cultural values of specific societies but also a force that amplifies an erosion of accountability across different contexts. In this last analysis, I seek to explore the cross-cultural factors that emerge when automation intervenes in public services. Building on the findings from the first two empirical chapters, where I demonstrated how distinct social, historical, political, and cultural values can determine the way automation is adopted, this final section shifts the focus to broader dynamics to shed light on more general practices related to automation, its governance and control.

5.2. Theoretical background

5.2.1. Algorithmic Legitimacy

Technologies are rarely simply “revealed” to who works on their construction. Since the 1970s, social constructivists have worked to show how technology is a collective process resulting from diverse human interests, intentions and powers over time, even at the expense of certain social groups (Hughes 1987; 2004; MacKenzie & Wajcman 1999). Historian of technology Thomas P. Hughes showed that invention is not and cannot be regarded as “a sudden flash of genius” from which a new “ready-made” device emerges, rather as developing over time (Sperry 1930), emphasizing that they are complex networks composed of not only technical components but also social, political, and economic elements (Hughes 1987). Importantly, he recognized organizational elements as components of technological systems. A central figure in Hughes’ theory is the “system builder”, an individual or group responsible for creating and expanding technological systems which include engineers, entrepreneurs, and government officials who shape the direction and growth of technology (1983). To stress this, according to Donald MacKenzie outside the human, intellectual, and material networks that give them life and strength, technologies cease to exist (1993). In his work, MacKenzie frequently addressed the issue of human and social *trust* in relation to technology. Whether in the military or in financial markets, his studies show how technologies are trusted to function accurately, and how this trust is built up through social networks, institutions, and expertise. It is this trust, in fact, that fills technologies with meaning and credibility, legitimizing them. Technology is thus legitimated when it is accepted and justified in a particular social, political and cultural context. In other words, legitimation is the perception that a technology is appropriate, desirable, or necessary, based on the values, norms, and beliefs of the people that make it available, including engineers, users, policymakers (MacKenzie 1993). As the notion of technological style offered by Hughes (1987) suggested, different

societies have different standards for what makes a technology acceptable. What is seen as legitimate in one cultural context may be contested in another. In his famous account on legitimacy, Suchman also illustrates that for Meyer and Scott (1983a; Scott 1991) legitimacy derives from the congruence between the organization and its cultural environment (1995). Legitimacy is thus culturally located and performed, and often depends on the role of experts and institutions that serve as gatekeepers of technology (MacKenzie 1993).

Different institutional theories also supported constructivism and the relevance of cultural and social norms, rather than just formal rules and structures. Powell & DiMaggio (1991) pointed out that organizational behavior is not only shaped by technological or material imperatives, but rather from “cultural norms, symbols, beliefs and rituals” (Suchman 1995, 571). According to the authors, legitimacy is gained by conforming to social norms, laws, and expectations (Powell & DiMaggio 1991). In his important synthesis on organizational legitimacy, Suchman defines it as “a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions.” (1995, 574). “Legitimacy is a perception or assumption” he adds, “in that it represents a reaction of observers to the organization as they see it; thus, legitimacy is possessed objectively, yet created subjectively. An organization may diverge dramatically from societal norms yet retain legitimacy because the divergence goes unnoticed” (Ibid.). The definition introduced by Suchman resonates with the black box concept introduced in 1987 by Bruno Latour to describe how scientific and technological processes are accepted and stabilized in society. In this sense, what Latour was describing can be interpreted as a legitimation mechanism: “when a machine works efficiently, when a question of fact is settled, one must focus only on its inputs and outputs and not on its internal complexities” (1999, 304). When stabilization occurs, according to Latour, science and technology become opaque and obscure. This is what Thomas P. Hughes already predicted with his theory of “technological momentum” when arguing that, in the life cycle of a technology, a phase where human intervention

would be possible and can give the technology various developments and different trajectories through its choices is followed by a phase of “maturity” - or stabilization - where society “adapts” to the existence of that technology and mutually shapes cultural norms and institutions (Hughes 1994). Far from being a determinist stance, this can be rather understood as a description of the mechanism of technological legitimation as seen from the perspective of society. Essentially, very often this legitimation occurs unconsciously. Affirmation and stabilization are based on legitimacy, which in the case of technology is rooted in the objectivity and neutrality with which it is automatically, unquestioningly and uncritically invested (Noble 1987; Feenberg 2010). This taken-for-grantedness is also one of the main features of what Schuman called “cognitive legitimacy” and described as the most subtle and powerful form of legitimation: “institutions not only render disorder manageable, they actually transform it into a set of intersubjective ‘givens’ that submerge the possibility of dissent” (1995, 583). What is legitimized turns out to be inevitable, so legitimization is by default. Cognitive legitimacy, more broadly, is based precisely on taken-for-grantedness and *automaticity*. It reflects an organization’s ability to align with shared cultural patterns and norms, making it appear natural or inevitable (Schuman 1995). In the case of the stabilization of a technology inside an organization and procedure, it can be understood as a moment, over time, when certain technologies become embedded and achieve cognitive legitimacy, meaning that they are no longer questioned because they are perceived as part of the organization. In this account of legitimacy, both in its original context and in its transposition to the context of technology, trust plays an important role. This view on the relationship between legitimacy, organizations and technology is reminiscent of the Weberian perspective on the instrumental and rational nature of social relations, as explored in the previous article in this thesis in reference to the machinery of bureaucracy and government. According to Jan Pakulski, in the process of modernization “relations in political institutions which may have been based on personal trust have been gradually replaced by more universalistic, impersonal and generalizable ties based

on legitimacy.” (1992, 25). Trust, intended as personal, has been progressively eroded in favor of formal and impersonal processes and legal procedures (Pakulski 1992). In traditional public services, trust was often based on personal relationships and interactions. Bureaucratic organizations - now also equipped with automated systems - lack this, hence they need to build trust through other means - such as legitimacy, transparency, accountability, and explainability of decisions:

«Because they rely on purely formal maximization of calculability of means, regardless of the particular ends they may serve, formally rational apparatus cannot be tied to any 'substantive' principle, such as equality or justice. Because they are impersonal and rule-oriented, they minimize the role of trust in the 'client-official' relations. The price of increased efficiency of bureaucratic organizations is an 'iron cage' of rationalism marked by impersonality, legal formalism, and calculative rule-orientation.»

(Pakulski 1992, 27)

As discussed in the previous chapter, algorithms are extensions of this bureaucratic logic, “rules of rationality” (Dalston 2012), where legitimacy stems from the perception that they follow established rules, are objective, and make decisions in a standardized, unbiased manner. In bureaucratic systems, following Pakulski (1992) legitimacy without trust can lead to a scenario where automated services are procedurally accepted but fail to gain public confidence.

5.2.2. Algorithmic Accountability

When automated systems prioritize efficiency, optimization, and rationalization, other important values fall aside. The issue of accountability in algorithmic and automated contexts has long been debated in the literature (see Johnson and Nissenbaum 1995; Lessing 1999). In 1994, Helen Nissenbaum was already writing about computing accountability noting that computers - especially those used in decision-making contexts - complicate traditional notions of accountability.

Moreover, in complex computing systems, responsibility is often distributed among multiple actors - designers, programmers, operators - making it difficult to pinpoint accountability when something goes wrong (Nissenbaum 1994). More recently, scholars who critically investigate harms related to automation have recommended greater accountability by responsible human actors (O'Neil 2016; Crawford 2021). In this sense, algorithmic accountability refers to the responsibility required when developing, implementing, and overseeing automated systems or algorithms. As Wieringa points out, in algorithmic accountability “the topic of explanation and/or justification is an algorithmic system” (2020, 2) and although it might sound misleading, the concept does not refer to an alleged responsibility of algorithms. According to Kroll et al. (2017), traditional oversight mechanisms, designed for human decision-makers, often fail in these contexts. Accountability here can take various forms: for example, some have argued that code disclosure is necessary to make an automated process accountable (Diakopoulos 2014), while others insist on third-party scrutiny. Kroll et al. (2017) argue that these measures are not sufficient and challenge the idea that transparency solves everything (while sometimes it is not even desirable). A traditional and generic account on accountability is outlined by Bovens, and illustrates well the centrality of transparency:

«A relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor may face consequences.»

(Bovens 2007, 447)

As Wieringa (2020) notes, in the transition to algorithmic accountability - where at least part of a process is delegated to a more or less complex computer system - the first matter to consider is how to understand this “actor”. Designers are normally considered as having a certain degree of responsibility, as they programmed the technology and have the technical skills to interpret what may arise from it (Martin 2018). However, responsibilities are much more fragmented

among different actors. Especially in the case of public services, they are distributed on different levels and Bovens (2007) anticipated them by talking about hierarchical and also collective accountability. In this sense, social constructivism texts that insisted on the roles and importance of different relevant social groups influencing a technology are again of particular relevance (Bijker and Pinch 1987; Latour 1999). Similarly, MacKenzie (1993) argued for more critical reflection on who is accountable for technological failures and risks. He emphasizes the need to examine who benefits from technological systems and who bears the risks or consequences when they fail. Growing awareness about the risks of algorithmization, sparked by both academic and public debate, has led to a growing emphasis on the question of who is responsible and on accountability as a key principle of algorithmic processes (van Schie et al. forthcoming). Thinking of designers and developers as the first potential responsible actors, the “accountability gap” between what they can control and the unanticipated outputs that can happen with algorithms that learn in part autonomously has been highlighted (Mittelstadt et al. 2016). However, what characterizes these technologies is precisely the variety of actors who have an active role in influencing them and deciding their trajectories (Bijker and Pinch 1987), so in addition to the developers there are a number of other actors that may bear responsibility for them (Huyskes 2024). Furthermore, there are varying degrees of control that human actors may have over algorithms, which may influence the respective level of accountability: for example, there are semi-automated procedures that humans must necessarily address in order to run them, and others that are more entirely automated (Wieringa 2020). Although the level of human understanding of the procedure can certainly change according to these different levels, it is difficult to think that the accountability of a process - especially when particularly sensitive - could vary according to the human actors' understanding of it, if they were the ones who decided to delegate part or all of that process to a machine. As Wieringa concludes after their literature review on algorithmic accountability:

«Algorithmic accountability concerns a networked account for a socio-technical algorithmic system, following the various stages of the system's lifecycle. In this accountability relationship, multiple actors (e.g. decision makers, developers, users) have the obligation to explain and justify their use, design, and/or decisions of/concerning the system and the subsequent effects of that conduct. As different kinds of actors are in play during the life of the system, they may be held to account by various types of fora (e.g. internal/external to the organization, formal/informal), either for particular aspects of the system (i.e. a modular account) or for the entirety of the system (i.e. an integral account).»

(Wieringa 2020, 10)

Drawing on Schuman (1995) and Pakulski's (1992) work on legitimacy, it follows that accountability can be "placed" on, or expected from, actors. There is an expectation that the actors or the organization will be accountable for its actions, i.e., that it will justify its decisions, be transparent, and address the consequences of poor performance or unethical behavior, and this creates people's "reliance" on it. Although from the outside it may appear as if nothing changes, procedures actually change profoundly when they are automated, so it is necessary to maintain high levels of legitimacy to maintain a high level of accountability (Wihlborg et al. 2016). According to Wihlborg et al. (2016), who analyzed a case of decision automation in Swedish welfare, legitimacy is the result of different values such as accountability, impartiality, openness, and transparency, which must necessarily be balanced.

In this context, "reliance" plays an important and relevant role. Over-reliance on automated systems without appropriate oversight can erode trust, especially if the system makes errors or if the process lacks transparency (Dzindolet et al. 2003). In the literature, the human tendency to over-rely on automated systems has been largely investigated as "automation bias". The underlying assumption is that their outputs are more accurate, objective, or reliable than human judgment (Parasuraman & Riley 1997; Mosier et al. 1997; Skitka et al. 1999; Alberdi et al. 2009; Cabitza et al. 2023). Parasuraman and Riley analyzed several ways in which human-computer interaction can go wrong. Among these, they mention automation bias in terms of "misuse", i.e., over-reliance on automation, where

humans are more likely to rely on computer advice (even if it is wrong) rather than their own judgment; but also in terms of “abuse”, when technology is developed without due consideration of needs or consequences and the role of the human being is forgotten (Alberdi et al. 2009). Skitka, Mosier and other colleagues investigated automation bias in the late 1990s in hi-tech contexts of the time, such as aviation (1997; 1999). Automation bias has been initially measured in this context as users were over-relying on automated systems which should raise an alarm in critical situations. When they don’t and the users are relying exclusively on the automated system, the failure to respond - in other words, the omission - can lead to important errors and accountability issues (Skitka et al. 1999). Many cases of human error involve trusting the system more than the user’s own judgment. For example, a radiologist might miss signs of cancer because the computer support system did not flag it (Alberdi et al. 2009). Other types of errors occur when a user follows incorrect advice from an automated system, even when other cues indicate it’s wrong (Ibid.). These errors are the most prevalent as they are more procedural compared to alarming situations: the problem does not rely on omission, but rather on the absence of human judgment and/or monitoring. For example, if an algorithm incorrectly denies someone benefits, users may not question the decision due to their perception that the system is inherently fair and accurate. As in the Dutch case discussed in the third chapter, risk assessments carried out on welfare recipients both at a national and at a local level have been repeatedly carried out using machine learning algorithms without humans questioning its outputs even when they indicated structural targeting of certain social groups (see Chapter 3). In the context of automation of essential services, automation bias can heavily influence the accountability of these procedures and complicate the relationship between trust and legitimacy. The challenge with automation bias is that it creates a situation where trust is built on the assumption of infallibility in automated systems, which can undermine responsibilities on the systems when failures occur. In this sense, algorithmic accountability looks like a loop which can lead to situations where trust is misplaced, and the socio-technical

system is perceived as legitimate without actually being accountable or reliable: if there is no perception with respect to the possibility of error, there is no planning with respect to how to deal with errors. All of this can have significant consequences for accountability, as automation bias shifts responsibility from human decision-makers to algorithms. The discussion section will illustrate in detail through empirical data how pragmatic responsibility is often delegated to machines.

5.3. Responsibility, shifted: “it’s not me, it’s the data!”

In the first empirical chapter of this thesis, through the discussion of the Dutch case, the considerable decision-making responsibility that different human actors had in creating and leading to problematic situations concerning the automation of public services and the consequent harm to people was shown. In particular, in the *Toeslagenaffaire* national case, these responsibilities were widespread and diverse, given the scale and complexity of the project, and most importantly, were distributed along its lifecycle. Following the accounts given on legitimacy and constructivist theories - which identify technologies as the result of “possible” and chosen trajectories amidst a wide field of possibilities (MacKenzie and Wajcman 1999) - I argue that when humans choose one technological path over another, or even more simply make a choice related to the design and implementation of that technology, they should be responsible for it. As Kate Crawford argues, focusing just on the outcome of an algorithm “forecloses more complex readings of the political spaces in which algorithms function, are produced, and modified” (2016, 79). Responsibilities lie in the construction, the governance and the general shaping - other than their actual use - of technological practices. This is even more true when dealing with cases of automation of bureaucratic practices, involving political and institutional actors subject to democratic election and changes of government that succeed and inherit these projects over time. In the Italian case, responsibilities are more difficult to identify because many of the behaviors toward automation in

the bureaucratic sphere occur unconsciously. This certainly does not imply that responsibilities are not present, but the lack of awareness and intentionality of those responsible adds further complexity. Despite their cultural and infrastructural diversity on technology, accountability was an important missing factor in both the cases discussed in the Dutch and Italian contexts.

The data collected in the Netherlands proved to be particularly illustrative with respect to this. One of the informants, for example, told me about the reality inside the administrations during and after the outbreak of the *Toeslagenaffaire*:

«We went in and we said... Well, give us the documentation. They said there was no documentation. We said ok, can we talk to the people who made this algorithm? "They are not working here anymore", they said. The algorithm was running all by itself.»

(Informant 20; person working for a governmental internal agency focused on ICT)

The reality faced by the person I interviewed points back to automation bias and the reliance placed on algorithms to the point of letting them run on their own without anyone supervising, knowing, or maintaining them. On the substantive issue of the childcare benefits scandal, another key informant commented on automation bias and its persistence in human work, in the automated validation of algorithmic (even sensitive) recommendations:

«Our assumption was that when you are using an algorithm to predict high, medium or low risk and based on that prediction you're handing it over to a caseworker, there will be inclined to follow it. I mean, that's very well researched.»

(Informant 15, University professor)

At a later point in the interview, Informant 20 raised another interesting point, telling me that she was asking herself whether these different actors were actually talking to each other about the algorithm and the procedure in question. She told

me that it appeared quite clear from the behavior of the programmers she was talking with that there was no contact between them and the people who previously worked on the algorithm in question. This is a relevant issue with respect to the distribution of actors' responsibilities, which recalls the constructivist idea of the plurality of groups influencing a technology but also the complication with respect to the fragmented nature of accountability and the difficulties of identifying and attributing it (Wieringa 2020). In this regard, another interviewee suggested:

«So even here I'm probably skeptical honestly on what comes out of the formal investigation because probably we're all going to conclude everyone was at fault, so nobody actually is!»

(Informant 23, professor at a University and technical expert)

Here the expert refers to the *Toeslagenaffaire*, a case in which so many errors emerged on not only a technical but also a political and institutional level. The informant shares a skepticism based on the assumption that responsibilities are largely shared at least when automated processes are embedded in complex sociotechnical frameworks. On this, she added:

«The reality is nobody really was responsible or an owner to this because all these different components had an issue so you've already been seeing that the people start pointing at others like “yeah, they gave us bad data...” or “yeah, but they did not give us the right features yet”. That was not good and it's true, all of that is true.»

(Informant 23)

This behavior is related to another relevant aspect emerging from the literature and that contributes to explain the lack of algorithmic accountability, an avoidance of responsibility which she further explored:

*«You get a lot of these arguments like - “yeah, but **the data said it!**” Or “the data scientists found it”, or “**it's not me it's the data**”, so there's also a lot of this shifting of responsibility, or refusing to take own ownership»*

(Informant 23)

“They gave us bad data”, “the data said it” and “it’s not me, it’s the data” are manifestations of data determinism and blame-shifting when dealing with technologies, algorithms and automated processes. Responsibility for decisions and outcomes is deflected onto the data, rather than acknowledging the role of human choices in their selection, interpretation, and use. This perspective disregards the fact that biases can also arise from model design, governance, or implementation choices and has deeper roots that go far beyond the individual empirical case. Beginning in 2015, when research on the ethics and politics of technologies expressed a more specific interest in artificial intelligence (AI) and machine learning in particular, a saying became rapidly and widely established in the literature: garbage in; garbage out (GIGO). In the context of AI and computer science, this mantra highlights the idea that the quality of the output is directly related to the quality of the input. If a system is given poor-quality data (garbage in), it will produce poor-quality results (garbage out) no matter how advanced the model or algorithm is (Gupta et al. 2021). In AI - and in algorithmic processes in general - the integrity, accuracy, and relevance of input data are critical to generating meaningful and reliable results (Rovatsos et al. 2019), and GIGO is meant to emphasize precisely that. Nevertheless, its stabilization in public and academic debate over the years has also had other effects, for example, as Sara Hooker notes, that of anchoring the idea that “a machine learning model merely reflects existing algorithmic bias in the dataset and does not itself contribute to harm” (2021, 1). In reality, many authors have contributed to dismantle this idea, showing how much the choice of the model can affect the final outcome (Suresh & Guttag 2019; Mitchell et al. 2021; Barocas et al. 2023). Suresh and Guttag (2019) also critiqued the common belief that machine learning harms primarily result

from biased data. Instead, they emphasize how harms can occur at different stages of the system lifecycle and introduce a relevant framework for understanding at which particular stages they might creep in. What does it mean, for example, to have accurate data? These, the authors argue, can always be influenced by societal inequalities (e.g., higher crime rates in marginalized neighborhoods may reflect structural inequality rather than inherent criminality) (Suresh & Gutttag 2019). This and other design choices (such as the model architecture, loss function, and optimization techniques) may all potentially contribute to problematic outcomes (Hooker 2021). Human oversight and accountability are minimized, as individuals or organizations distance themselves from the results of automated systems by placing all the blame on the data, rather than considering their own role in its processing or application. Blaming the technical components has not always been about the data. As detailed in this dissertation, the dominant paradigm of the relationship between technology and society has been precisely that of determinism, which originally referred to technological artifacts. For example, Nissenbaum (1994) was already writing about the tendency to “blaming the computer” rather than the people behind its creation and operation. In fact, another testimony on the *Toeslagenaffaire* mentions the model as the focus of blame-shifting:

«One of the consequences of the scandal was that there was a demand and push for more 'trustworthy AI', which must be more explainable and interpretable. Here, I actually saw one of the most explainable models in my career. Nevertheless, it has produced so much harm.»

(Informant 21; former data scientist at the Belastingdienst)

Once again, the implication is that there is a tendency to shift the responsibility outside, to some technical parameter or factor. In this case, to the lack of explainability of the model and its black-boxed nature. However, the words of one of the people who witnessed the development of these models and their use explains how much the “automatic” work of people, in this case the programmers and designers, is taken for granted first and foremost by themselves:

«Who is to tell me whether one attribute is more problematic than another, who decides what is the threshold and what score should be considered problematic? We build the model yet we are not accountable. Because we think we are building a tool, something given and objective, to then hand over to those who have the real responsibility for it: the public decision makers.»

(Informant 21; former data scientist at the Belastingdienst)

Working with parameters, indicators, and correlations that are unsubstantiated by scientific data and imposed from above, this group of actors often finds itself merely executing requests following either their prior knowledge automatically (Birhane et al. 2021) or some other human actor to whom actual control of the system is then passed. It was addressed in detail how widespread the black box metaphor is to describe how these algorithms work, and yet how this assumption can be used to act and perpetrate broader blackboxing at the institutional level that encompasses - but overrides - the algorithmic one. In Italy, as argued in the fourth chapter, this has been observed in the use of algorithms in bureaucratic processes. An interesting parallel was revealed by one of the key informants in that context:

«Usually teachers, when interviewed for newspapers, always say it is “the algorithm's fault,” the problem is that even the ministry says so, the regional school office... Everyone says that; they may not believe precisely that, but they describe it that way.»

(Informant 8, teacher who supported many affected colleagues and coordinated a response as a director for a trade union)

Informant 8 mentioned the tendency to blame the algorithm for issues in the teacher assignment process. This tendency arises from a lack of transparency and the complexity of the system, making the algorithm a convenient scapegoat. He explained that when problems occur, such as incorrect job assignments or missing job positions, both school administrations and local educational offices - which, in Italy, depend directly from the Ministry of Education and are responsible for the

management of a certain number of schools at a local level - often shift responsibility onto the algorithm. In reality,

«These errors are frequently caused by poor communication between schools and offices, missing or outdated job availability data, or administrative delays. However, because the algorithm is perceived as a “third party” actor, both sides often avoid accountability by blaming the system.»

(Informant 8)

Teachers often believe that the algorithm is at fault, when in fact the underlying issues are related to human or procedural errors. They are not unaware of the bureaucratic procedures behind algorithms, but they still blame it as a layer that added complexity and opacity to an already complicated process. Moreover, the system as it currently works affect the quality of education, as:

«Inexperienced teachers may struggle to manage large or challenging classrooms, impacting students' learning experience. Hopefully, the Ministry of Education will eventually open a dialogue to fix the flaws in the system and we can have a more transparent and fair procedure that prioritizes teachers based on experience and qualification, rather than solely on points or bureaucratic categories.»

(Informant 7, teacher who was affected by an algorithms)

For this reason, most of them, represented by different trade unions, ask for a return to the manual procedure. As Mackenzie (1993) wrote, for a technology to maintain its legitimacy it must inspire trust and demonstrate accountability, especially when things go wrong. Technologies that are opaque or lack accountability may face challenges to their legitimacy, especially when they fail. Pakulski's (1992) argument that trust and legitimacy must balance each other is highly relevant here. If automation bias leads to excessive trust in flawed systems, legitimacy can falter when those systems fail to meet expectations or when their errors are exposed. If the public becomes aware of flaws in algorithms (e.g., biased decision-making or unfair outcomes), the system's legitimacy may quickly

deteriorate. In Italy, there is not yet sufficient evidence to determine how much this legitimacy has been eroded due to a lack of accountability, perhaps due to the fact that the errors that continue to occur, however problematic and highly impactful, occur in a closed sector such as education and never scale up to the extent of a national scandal. In the case of the Netherlands, on the other hand, where errors related to automated processes have been more frequent and more alarming, various accountability mechanisms have now been proposed, in the same way that several authors have been proposing for some time. For example, Vecchione, Levy, and Barocas (2021) have proposed algorithmic audits based on audits guided by the social science community as a means of addressing accountability. Several other methods have been proposed for fostering responsibility (Raji et al. 2020; Metaxa et al. 2021). Initially, these mechanisms were put in place by oversight and supervisory authorities on government and public entities. For example, one respondent working at the Dutch DPA recounts:

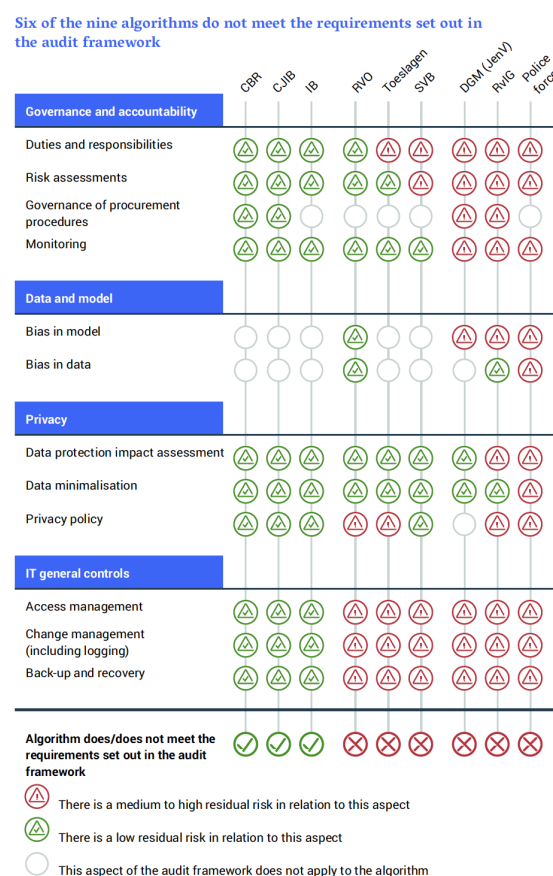
«After the child benefit scandal in the Netherlands, the Dutch Data Protection Authority set up a new department focusing on algorithm supervision. We assess risks and effects of algorithms in various sectors, not just those related to personal data. This information is shared with policymakers and other supervisory authorities. Beyond the algorithm itself, governance is crucial. Proper oversight, human involvement in decision-making, and awareness of the algorithm's limitations are essential for preventing harm.»

(Informant 31, person working at the Data Protection Authority (DPA))

Since 2023, when the Dutch DPA (*Autoriteit Persoonsgegevens*) created a new internal department focused on the Coordination of Algorithmic Oversight, every six months a report is published to monitor the possible risks and effects of the use of algorithms and AI on public values and fundamental rights. In the last report, the Authority shared that with a continuous and rapid adoption of algorithms and AI, especially generative AI, risk management practices and legal frameworks are lagging behind (Dutch DPA 2023). The document also noted that many citizens are unaware of when and how algorithms affect them, and how this

opacity could lead to more public dissatisfaction and the erosion of public trust, especially when errors occur (Ibid.).

Another example is the work of the Dutch Court of Audit (*Algemene Rekenkamer*), which in recent years - and especially after the *Toeslagenaffaire* - also started reporting on algorithmic audits conducted independently aiming to assess how responsibly the government uses algorithms in decision-making processes. The last one, published in 2022, shows that only three out of nine algorithms meet the audit framework's standards. While on the one hand this proved that responsible use of algorithms is possible, on the other it clearly showed the unreadiness of governmental organization to a responsible use of algorithms in public services (Dutch Court of Audit 2022).



An Audit of 9 Algorithms used by the Dutch Government, *Algemene Rekenkamer 2022*

This audit, based on clear and specific criteria, was very useful for the Dutch government and anyone interested in learning about it. Interestingly, among the

factors observed were the biases present in the model - as well as in the data - which was proven to be a category particularly affected by problems. Ministries were asked to review the six high-risk algorithms to ensure that immediate action is taken to mitigate the risks (Ibid.). After the outbreak of the scandal, the government and its various offices in the Netherlands also took proactive measures to avoid bias and other errors in automated projects. The *Auditdienst Rijk* is the independent internal auditor of the central government. Routine tasks include the auditing of the income and expenditure of ministries, the design of organizations and whether all data are kept safe, but lately a team dedicated to AI controls has been created:

«I work on audits of algorithmic systems used by various government ministries. While auditing can reveal issues like bias, our role is more about helping ministries prepare for external scrutiny. I would like to guide them on different ethical choices.»

(Informant 29; working for the National Audit Office)

The strategy adopted by the Netherlands has not been to stop or slow down the use of algorithms in decision-making and bureaucratic processes. Rather, a number of measures were introduced to publicly ensure greater transparency in their use and accountability. During an ethnographic session, I observed the interaction between different Dutch public officials, and this is what I pinned down:

«A person who works for the municipality of Amsterdam is talking about “applied ethics” in public innovation: he says that “we have to figure out what values we want to incorporate in our processes, because it's inevitable to harm someone's interests, you can't act in everyone's interest. But how do you decide which interests to prioritize?»

(Field note #1, Amsterdam, 13/06/2023)

The sense I got was that I was listening to very knowledgeable people with respect to risks, intent on a very open and direct multi-level brainstorming on how

to avoid them beyond “ethics washing”. I was surprised by the composition of some of the teams from the administrations described there, who were talking about experts in applied ethics, computer vision, machine learning, and other specific skills. From another note I wrote:

«The teams manage to be very self-critical of some past choices. It is very nice to observe.»

(Field note #2, Amsterdam, 13/06/2024)

Among the different activities and initiatives carried out to ensure more accountability and risk-mitigation, the most relevant is probably the Fundamental Rights and Algorithms Impact Assessment (FRAIA), a framework which guides and helps the government's use of algorithms, AI, or other decision-making systems and ensures it respects and safeguards fundamental rights, such as privacy, non-discrimination, and the right to a fair trial. By integrating FRAIA, the aim is to assess risks to fundamental rights at an early stage. Commissioned by the Ministry of the Interior and Kingdom Relations and developed by Utrecht University, FRAIA was initially to be mandatorily introduced as an evaluation tool for all public administrations that want to introduce algorithms into their processes. Two years after its introduction, it has not yet been made mandatory - and probably never will be, given the approval and imminent adoption of the European Regulation on Artificial Intelligence (AI Act) that will automatically apply to all Member States. However, despite the non-mandatory status, the sentiment in the Netherlands is such that a push for greater transparency is now present nonetheless. As a key informant working at the Ministry of the Interior and Kingdom Relations reported:

«There's a push to make sure that governments are transparent about using AI in decision-making, particularly in areas like welfare. This transparency is still developing.»

(Informant 18, staff member at the Ministry of the Interior and Kingdom Relations)

Recently, evidence has been published regarding empirical cases where FRAIA was carried out in order to assess the broader societal implications of Dutch governmental interventions involving technology and algorithms. Between 2023 and 2024, several government bodies used FRAIA for the brainstorming and discussion around 15 public algorithms. Lessons learned were compiled into a report, sharing the experience and some practical recommendations (Ministry of the Interior and Kingdom Relations 2024). Another example is the creation of a national algorithms register,⁹ with the goal of making the use of algorithms more aware and transparent. Two of the people I interviewed commented that “formal” transparency is not necessarily being matched by “substantive” transparency, where people are actually informed when automated processes are being used on them, and are helped to understand why they are considered as suspects.

The evidence discussed in this section shows how the accountability expectations “placed” on human actors is essentially legitimate but not respected, since with the introduction of automated processes there is a delegation of responsibility - as described by the automation bias - that, as Nissenbaum had already noted in 1994, complicates traditional notions of accountability. Failures of algorithmic accountability are evident in both the Dutch and Italian cases, involving a process of “othering” of the algorithm (Gandini et al. 2023), model or technology more broadly that shifts responsibility away from human actors, at least until major errors emerge and, in this “glitch” that brings out the non-perfection of the system in front of the law, those responsible are to be identified.

5.4. A Cross-Cultural Erosion of Accountability

According to the account on legitimization provided by Schuman (1992), even when it is taken for granted - as he expressed in its cognitive version - legitimization requires an alignment with societal, cultural norms and values of a given context.

⁹ Accessible here: <https://algoritmes.overheid.nl/en>

In the case of automation of administrative and essential processes, however, this alignment with cultural norms may not always be necessary for the system to be deemed legitimate. Automation involves transferring power and responsibility from human actors to technological systems, “the tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing” (Mosier & Skitka 1996, 205), which may challenge traditional ideas of legitimacy.

On the one hand, automated systems could bypass the need for cultural alignment because they are often justified on grounds like efficiency, neutrality, or objectivity rather than on cultural values. On the other end, cultural values are actually embedded in automated procedures, but more as an imposition, meaning that they are implemented by designers or institutions rather than being organically shared or agreed upon by the people affected by the system (Costanza-Chock 2020), which can lead to a gap between the technology’s implementation and its acceptance or legitimacy in the eyes of the public. The disconnect between the system’s design and its perceived legitimacy embeds culture in two ways: a situated culture, linked to the context of application and belonging, and a shared, horizontal culture, which transcends the situated culture and is common to projects of this type regardless of context. In short, some of the elements that clearly emerge from the Dutch case are over-reliance and the trust placed in the various automated systems employed not only in welfare, but also in the criminal justice system (Mutsaers & Nuenen 2023). For example, as for the *Toeslagenaffaire*, the Dutch Crime Anticipation System (CAS) was based on neural networks, regularly updated with new data, to forecast potential high-risk areas or crimes. The system often flagged locations associated with minority groups, especially young ethnic men from low-income neighborhoods, as high-risk, perpetuating systemic discrimination through algorithmic outputs. By focusing solely on statistical correlations and predictive models, the system ignored deeper social causes of crime and community concerns, instead relying on probabilistic judgments that can exacerbate existing biases (Ibid.). In this case, far from being neutral, CAS emphasized the intertwining of social norms and technological systems and

demonstrated that it was shaped by institutional practices and assumptions about crime and risk, many of which reinforce existing power dynamics and discrimination. The statistical and probabilistic calculus outsourced to the algorithmic process was deterministic in nature, resting on specific cultural underpinnings (such as historical data and discriminatory guidance on which variables to consider, which encapsulated a racist and colonial interpretation of risk) (van Schie et al. forthcoming) and endorsed a network determinism, judging all people who shared certain characteristics to be at risk. Looking at other contexts where predictive models similar to these have been adopted, one can see that the institutional assumptions are always very similar. From Eubanks (2018) which describes the feedback loops of injustice in which poor US families found themselves in, with their benefits cut off because of an automated decision without any explanation - very similar to what happened in India, where some people were categorized as deceased by government agencies and no longer received pensions, despite being alive (Majumder 2024) to Spain, where an algorithm was used for years to predict risk in incidents of gender-based violence, which made several important mistakes without anyone noticing (Algorithm Watch 2020). Interestingly, an investigation conducted on a risk assessment model identical to the Dutch one and used in France by the national social security agency found that the model systematically assigned higher scores to vulnerable people, including single parents, people struggling to pay bills, and people receiving disability benefits (Romain et al. 2023). Contextual historical and cultural values may influence, for example, which social categories are inevitably considered to be most at risk, but it seems clear that over-reliance and trust on statistical models, whether based on artificial intelligence or not, is a cross-cultural fact. This occurs regardless of the specific context and seems to have more to do with a horizontal, generalized culture toward technology and shared across geographic contexts. In the presented cases concerning public services, algorithms seem subjected to a process of “othering” in order to carry out certain actions - such as the systematic targeting of certain social groups - or outsourcing of responsibility. In this context, the erosion of accountability occurs

precisely because of this shift of attention, trust and credibility to automated systems, such as algorithms or data, instead of being clearly attributed to the human actors involved in their design, implementation and supervision (Latour 1999). In essence, an accountability gap occurs when human responsibility is abdicated or obscured in favor of misplaced trust in technology, leading to situations in which systems operate with less control, transparency or corrective measures, and errors are less likely to be recognized and addressed. Cross-cultural comparison allows us to recognize that although the specific expressions of this bias may differ, the underlying dynamics of over-reliance on algorithms are remarkably similar. This suggests that there is a horizontal and generalized culture of reliance on technology that transcends national boundaries, reinforcing the need for a critical approach to the design and governance of automated systems in public services. Although these dynamics have been proven and discussed even more recently (Cabitza et al. 2023), interesting and relevant social research has been carried out in automated contexts that is highly relevant to this discussion today. In addition to the misuse and abuse about automation already measured by Parasuraman and Riley (1997), years ago another study showed that, normally, people trust automation and consider it reliable until shown that machines can produce wrong results. At that point, the most common feeling is one of distrust and suspicion. The results of this study indicate that the understanding - or the semblance of an understanding - is a key factor for trusting decisions to rely on automation: only an explanation of why they can make mistakes brings back trust (Dzindolet et al. 2003). The tendency of human actors to shift and outsource responsibility when dealing with automation is not new. In different cultural contexts, they currently do so claiming that the data is wrong or that the model fails. But when the assumption about the inherent accuracy and objectivity of algorithmic systems is repeatedly proven wrong and the socio-technical system significantly fails, the responsibilities at stake are recognized and recklessness lapses. Crucially, Stitka et al. (2000) proved that human errors in automation cases can be reduced under conditions of increased accountability. Making participants

more accountable for their overall performance or the accuracy of their decisions reduces rates of automation bias. Some errors turned out to be the result of decreased cognitive vigilance, while others were the result of a combination of failure to consider relevant information and a belief in the superior judgment of automated tools (Ibid.; Huyskes 2024). A strong claim, consistent with the result of these studies, was made by one of the actors approached for this research:

«It's probably sad to tell, but you need a scandal in order to wake up.»

(Informant 20; government technical advisor)

The need for a scandal, as the informant mentioned, or algorithmic failure to spur recognition and action on algorithmic accountability is coherent with the fact that without a visible failure or publicized error, the systemic issues created may go unnoticed as for the taken-for-grantedness that Pakulski (1992) already addressed in the context of institutional practises and that Latour (1999) discussed in relation to technologies. In the absence of such failures, there is little incentive to question existing systems, particularly when they are perceived to be working well enough in terms of numbers, efficiency or cost savings. In the absence of proactive regulation or strong ethical frameworks, a visible failure can be a push for action, leading to reforms, audits, or the creation of oversight bodies that would otherwise not be prioritized. Although this is not a strategy to be encouraged, more evidence related to the erosion of accountability in modern automated and algorithmic contexts due to human delegation is needed in order to explicitly highlight failures connected to human actors and thus to their responsibility.

5.5. Conclusion

This chapter critically examined the role of algorithmic accountability in public services, focusing particularly on how automated practices are integrated,

interpreted and legitimized within the cultural and social contexts of Italy and the Netherlands. It aimed to show, in relation to the cases discussed in the two previous articles, the importance of situated values and specific cultures in the governance of technological innovation (Bijker & Pinch 1987; Latour 1987), particularly at the bureaucratic level. In addition, however, it aimed to consider how in addition to specific cultures, a “macro” culture with respect to technology can be observed which transcends geographic boundaries and is characterized by an over-reliance on algorithms and shift in human responsibility that challenges traditional norms on accountability and legitimation (Pakulski 1992; MacKenzie 1993; Suchman 1995). In this culture, blind trust in automated outputs often exceeds critical scrutiny. One of the key findings is the tendency of public service workers and institutions to shift accountability on algorithms, data and more broadly technical features as something that stands beyond human control. Whether in the assignment of teachers in Italy or the risk assessments for welfare recipients in the Netherlands, human decision-makers frequently trust algorithmic outputs even when they lead to problematic or discriminatory outcomes. This over-reliance is not only a technical issue but a cultural one, reflecting broader societal beliefs about the authority and neutrality of technology as a cross-cultural attitude towards automation.

Despite their portrayal as objective and impartial, algorithms and automated systems are deeply influenced by the cultural and political contexts in which they are used. As feminist studies highlighted in the past, complexity and contingency are fundamental lenses to study and interpret any technology (Wajcman 2004). In both Italy and the Netherlands, historical, political, and social factors shaped the design, implementation, and reception of these technologies in public services. The widespread phenomenon of automation bias, that in this context highlights how human actors defer to the decisions of algorithms without controlling or questioning them (Cabitza et al. 2023), manifests in both settings, though it takes slightly different forms. In the Netherlands, for example, cases such as the *Toeslagenaffaire* - the child benefits scandal - exposed how reliance on algorithms can

obscure responsibility, allowing political and technical actors to avoid accountability for systemic failures. Similarly, in Italy, the opaque use of algorithms in education and public administration shows how these systems can deflect blame away from human decision-makers, who often blame the algorithm when errors occur. On the one end, despite differences in governance structures and cultural traditions, striking similarities between Italy and the Netherlands were shown in how public institutions approach automation. In both cases, the integration of automated systems into public services has been justified largely on the grounds of efficiency and objectivity, often at the expense of transparency and fairness. On the other hand, their trajectories regarding accountability and governance diverge significantly. In the Netherlands, accountability towards automation has evolved in response to a major national scandal that exposed the systemic flaws in the use of algorithms, particularly in how they disproportionately affected vulnerable populations, and triggered widespread public outrage. As a result, a number of strategies and oversight mechanisms have been introduced, such as algorithmic audits, the establishment of a national algorithms register, and the development of frameworks like the Fundamental Rights and Algorithms Impact Assessment (FRAIA). These measures have led to an increase in public sensitivity, alertness, and involvement from independent authorities like the Dutch Data Protection Authority and the Court of Audit, signaling a shift towards greater transparency and accountability in how automated systems are managed in public services. In contrast, Italy has not experienced a national scandal of the same magnitude that would have motivated similar responses. Although issues within the educational sector, particularly related to the algorithmic assignment of teachers, continue to lead to some legal rulings in favor of teachers, these cases have not yet reached the level of a national reckoning. As a result, the bureaucratic culture in Italy continues to support a more fragmented and less scrutinized implementation of algorithms, with a noticeable lack of centralized oversight or accountability frameworks. The absence of a unifying event suggests a lack of public and institutional pressure to address the ethical and governance issues surrounding automation in public

services. Thus, the key difference between the two countries lies not just in their technological practices, but in the cultural and political conditions that shape their responses to automated systems. In the Netherlands, a crisis prompted urgent responses, while in Italy, the inertia of the bureaucratic system has delayed the development of similar accountability structures.

Ultimately, this article should be considered as an initial exploration of the extent of automation bias as applied to algorithms and AI systems in public services and its relationship with algorithmic legitimacy and accountability, an area that necessarily requires more evidence and more contingent ethnographic observations. It encourages in-depth research in specific cultural contexts to observe public accountability and legitimacy, but also the horizontal culture of trust in technology that transcends national boundaries, reinforcing the need for a critical approach to the design and governance of automated systems in public services. The study calls for a reassessment of how these systems are designed and implemented, urging policymakers, designers and public institutions to remain vigilant in addressing the ethical challenges posed by the increasing use of automated decision-making in public services.

6. Conclusions and future steps for research

This thesis set out to explore how automation in public services is influenced by contextualized socio-cultural determinants. The overarching aim was to understand the conditions under which automated decision-making systems are designed and implemented in essential and administrative processes, and to analyze their effects on public governance, fairness, and accountability. By focusing on two distinct cultural contexts, the thesis provided insights into how social values, political drivers, historical and bureaucratic traditions shape the trajectory of public sector automation. In contrast with a deterministic and reductionist approach that interprets developments in digital technologies, AI models and their uses as something inevitable, these pages made use of constructivist theories on technology to reveal the willingness, motivations, negotiations and choices among various actors (Bijker, Hughes & Pinch 1987; Pinch & Bijker 1984; MacKenzie & Wajcman 1999; Callon 1980; Wajcman 2004). While featuring algorithms, AI systems and other digital technologies as the object of study, in fact, this thesis has been primarily interested in observing and highlighting the role of human actors who influence and shape them in particular contexts, with their choices and responsibilities. Especially in the case of public services, as Bovens (2007) points out, responsibilities are distributed both hierarchically and collectively. But while the literature has now widely shared how automation in public sectors like education, welfare, and administration does not happen in a vacuum but reinforce and sometimes amplify pre-existing social inequalities (Eubanks 2018; O'Neil 2018; boyd 2010), research that study these technologies “in practice”, as Orlikowski (2000) suggested, and their effects in their contexts, to analyze the specific surrounding causes that shaped them in a certain way (boyd 2010) is still undeveloped. In other words, researching the impact of the introduction of certain technologies is fundamental, but a widespread focus on it by the social sciences “leaves a prior, and perhaps more important, question unasked and therefore unanswered. What has shaped the technology that is having ‘effects’? What has

caused and is causing the technological changes whose ‘impact’ we are experiencing?” (MacKenzie & Wajcman 1999, 36).

These specific lenses were used to pursue these research objectives using a combination of ethnographic methods, qualitative interviews, document analysis based on two case studies, providing an in-depth look at the specific socio-technical conditions that lead to the implementation of automation in public services. The opportunity for a detailed analysis was made possible by the choice of two contexts, Italy and the Netherlands, described as two areas that introduce public sector automation and manage it differently. Looking at the amount of projects already implemented to automate public services, I initially observed that Italy was one of the most underdeveloped countries in both automation and transparency related to these practices, while the Netherlands was ranked as one of the most automated and transparent governments in Europe (DESI Index, 2022). Despite these starting differences, both countries have to some extent invested in decision-making automation, which has also had significant and negative impacts in some cases. In the third chapter, several examples of automation of the Dutch welfare state have been exposed. These show how the opacity of technical systems can lead to significant public harm.

Through the discussion of the *Toeslagenaffaire* (child benefits scandal), the historical tendency to classify, categorize and objectively judge individuals based on their general characteristics was highlighted. Rather than being evaluated solely on their own actions or characteristics, individuals are treated as part of a broader network of data points that may include demographic markers, behavioral patterns, or social relationships. This leads to decisions being made not just on individual contingencies and experiences, but on their perceived similarity to others within these networks. For instance, algorithms used in welfare ended up grouping individuals based on socio-economic status, race, or geographical location, leading to biased outcomes because the system “determines” behavior and eligibility based on the network an individual is perceived to belong to. This is what I have referred to as “network determinism”, to echo the older determinism (Bijker and Pinch

1984) and the way in which nowadays and with these technologies, individuals are categorized and judged on the basis of algorithmically defined networks, where their data and behaviors are compared with others within a system. What emerged is a pattern where human biases, institutional power imbalances, and bureaucratic traditions are baked into the design of the systems, leading to varied outcomes that disproportionately impact marginalized populations. As several informants reported to me, the responsibility for what happened in these cases was very fragmented, but entirely political, institutional, and human, e.g., the idea of relying so heavily on correlations - without proven causal links - discovered among big data. In chapters 4 and 5, accountability was addressed extensively. First, by discussing the Italian case and in particular the intertwining of the history of bureaucracy and that of technology, drawing connections between the bureaucratic structures and the technological processes that define public administration in the context of automation. The analysis in Chapter 4 builds on Max Weber's classic view of bureaucracy as an "iron cage", where administrative processes prioritize efficiency, rationalization, and objectivity at the expense of individual creativity and autonomy (1904). This mechanization of governance is further intensified by the introduction of algorithms and automated systems, which heighten the rational calculation, formalism and opaqueness of public services. Opacity is a common trait between bureaucratic practises and algorithms. Traditionally, the black box metaphor is used for technologies, particularly algorithms, that operate in ways that are opaque or difficult for users and even programmers to fully understand (Pasquale 2016). These systems produce decisions without revealing the internal workings and trajectories of how those decisions were made (Burrell 2016). Drawing from Latour (1999), I extended the use of the concept to describe how Italian public administrations have institutionalized opacity, not only through technological design but also through bureaucratic resistance to transparency. According to Latour, black boxes were all those technologies that "settle" inside specific procedures: when human actors stop questioning them, and they become inevitable. Social constructivism has largely pushed for social studies on technology

to “open the black box” (Pinch & Bijker 1984) revealing the connections, intentions, relationships and choices of social groups behind it more than the technical functioning of machines (Ibid.). In taking this specific approach and shifting it to algorithms and AI systems, this thesis has discussed in this chapter the black box metaphor to address the opacity of Italy’s public sector, focusing on how this lack of transparency is not simply an inherent feature of the technology but a result of administrative practices and cultural tendencies within Italian bureaucratic governance. Among these, I identified (i) legal barriers; (ii) lack of awareness and competences; (iii) private interests; (iv) lack of governance and strategy; (v) accountability; and (vi) bureaucratic resistance.

Blackboxing is identified as a practice where accountability is obscured by administrative practices that prevent scrutiny, where the use of algorithms can only exacerbate this institutionalized lack of transparency. Rather than actively promoting transparency, Italian public institutions often deflect responsibility for automated decisions by treating the technology as “neutral”, especially when errors or controversial outcomes arise, as seen in the Buona Scuola case involving teacher allocations. The refusal to subject these systems to public scrutiny, coupled with limited public discourse on their governance, makes it difficult to “open” the black box. This issue is ultimately explored in Chapter 5, where the Italian and Dutch cases are discussed together. While in the Italian case, in fact, institutional blackboxing even prevents a public discourse on algorithms and their effects, this is in contrast to the Dutch context, where public scandals like the *Toeslagenaffaire* have spurred greater transparency efforts, such as algorithmic audits and the creation of governance frameworks. The Dutch experience definitely serves as a critical case study in how national crises can drive institutional change and foster a more transparent and accountable approach to ADM systems. The final empirical discussion investigates how automation shifts responsibility away from human actors and towards technical systems, examining cross-cultural differences in how these shifts manifest. In both the Italian and Dutch cases, public administrators and policymakers often defer to algorithmic outputs, reinforcing a well-researched

bias (Parasuraman & Riley 1997; Mosier et al. 1997; Skitka et al. 1999; Alberdi et al. 2009; Cabitza et al. 2023). Over-reliance on technology, especially a blind trust in algorithmic outputs, has manifested in both contexts. The phenomenon seems so embedded in institutional algorithmic practices that I have argued it represents a cultural construct about automation that transcends cultural boundaries and, combined with the lack of adequate accountability mechanisms, creates a cross-cutting practice in which responsibility is diverted from human decision makers to the algorithms themselves. In both the Italian and Dutch cases, the research found a tendency for human actors to distance themselves from the outcomes of automated processes, reinforcing the idea that “it’s not me, it’s the data.” This shifting of responsibility not only exacerbates the opacity of decision-making processes but also undermines the democratic principles of accountability, legitimacy and transparency.

Taken together, the two empirical cases presented provide a comprehensive understanding of how automated processes function within different cultural and institutional settings, and how they can produce divergent outcomes based on the social and political contexts in which they have been constructed. The divergence between Italy and the Netherlands in terms of algorithmic governance highlights the value of cross-national comparative research, but also the need to monitor these processes as institutional practices evolve. Although the Netherlands, for example, has pioneered frameworks for algorithmic accountability, there is still much to be learned about the effectiveness of these mechanisms and their replicability in other settings that differ culturally and in the adoption of these technologies.

This research acknowledges several limitations, primarily related to the scope and depth of empirical data on the social construction of automated decision-making ADM systems. One initial aim was to closely track the development of ADM projects from conception through design and deployment, similar to the approach used in Cockburn and Ormrod’s study on the microwave

(1995). However, this level of detailed tracking proved challenging due to restricted access to certain public offices and reluctance among key actors to share comprehensive data. Additionally, such extensive tracking would require a broader research scope beyond the scope of this dissertation. Despite these challenges, this research achieved meaningful insights into the role of human values in the construction of these projects, as the intentionality behind ADM systems' design, especially concerning variable selection, which aligns with constructivist frameworks. Another limitation lies in the difficulty of "opening the black box" of ADM systems and establishing effective methodologies for investigating technology-in-practice (Orlikowski 2000). For example, the Italian case in Chapter 4 illustrates how current civic access laws do not fully facilitate transparency and are difficult to enforce (see Savage & Hyde 2012). While the barriers encountered as a researcher served as findings in their own right and lead to results about institutional blackboxing, there remains a need to develop new methods to better study these complex systems. Furthermore, difficulties such as language barriers and information asymmetry in the Italian context occasionally hindered data collection, as stakeholders used varied terminology and approaches when discussing similar concepts.

Building on these findings, future research should insist in exploring the cultural and social factors influencing design choices, advancing the constructivist agenda by examining why certain trajectories are prioritized over others. While recent studies on technology's societal impact often focus on how to overcome bias, there remains a gap in understanding the causal and shaping factors in ADM construction, such as actor interactions, design decisions, and the socio-political lifecycle of technologies. Much of the current debate centers on ADM tools themselves, rather than on how they are actually used in practice. To address this gap, researchers should investigate the actual practices and decision-making processes surrounding ADM systems in public services, aiming to reveal the socio-political influences on design and adoption. I truly believe in the power of

social sciences to actively investigate how to adapt the design of technology to the society to which it will apply (MacKenzie & Wajcman, 1999). This approach could benefit from the critical perspectives of feminism, critical studies and many more who called for applying social analysis, materialism, and empirical inquiry to technology studies. Reviving these principles could facilitate an intersectional approach that addresses not only gender but multiple forms of social exclusion, enhancing our understanding of how ADM systems affect diverse groups and amplify inequalities within society and providing more contextualized solutions to mitigate discriminatory impacts.

A focus is needed to evaluate current frameworks and explore how they can be adapted and adopted in different cultural and institutional contexts. Ultimately, although automation has the potential to improve the efficiency of public services, it never does so alone, and this research emphasizes the centrality of construction and human labor in designing and governing these systems. The different design and cultural choices that were explored over the course of the three papers highlight the presence of trajectories and different possibilities in the technological field (Bijker & Pinch 1987; Latour 1987), providing another perspective than the more common tendency to blame data and algorithms when the output of a mathematical model creates problems. More research is needed to show the weight that specific design choices attributable to one or more humans have on the final outcome.

Bibliography

- Adams, R. (2021). Can artificial intelligence be decolonized?, *Interdisciplinary Science Reviews*, 46(1–2): 176–197. doi: 10.1080/03080188.2020.1840225.
- Airoidi, M. (2018). Ethnography and the digital fields of social media. *International Journal of Social Research Methodology*, 21(6), 661–673.
<https://doi.org/10.1080/13645579.2018.1465622>
- Airoidi, M. (2021). *Machine Habitus: Toward a Sociology of Algorithms*. Cambridge, UK: Polity Press.
- Akrich, M. (1992). The De-scription of Technical Objects. *Shaping Technology- Building Society: Studies in Sociotechnical Change* in Bijker, W. & Law, J. *Shaping Technology/Building Society. Studies in Sociotechnical Change*. MIT Press, pp.205-224.
- Alberdi, E., Strigini, L., Povyakalo, A. A., Ayton, P. (2009). Why Are People's Decisions Sometimes Worse with Computer Support?. *Computer Safety, Reliability, and Security. Lecture Notes in Computer Science*. Vol. 5775. Springer Berlin Heidelberg. pp. 18–31.
- AlgorithmWatch. (2019). *Automating Society: Taking Stock of Automated Decision-Making in the EU*. Berlin. Available at:
https://algorithmwatch.org/wp-content/uploads/2019/01/Automating_Society_Report_2019.pdf (accessed 8 April 2024).
- AlgorithmWatch. (2022). In Spain, the VioGén algorithm attempts to forecast gender violence. *AlgorithmWatch*. Available at:
<https://algorithmwatch.org/en/viogen-algorithm-gender-violence/>. (Accessed 09/09/2024).
- Alston, P. (2019). *Digital Welfare States and Human Rights*. Report of the Special Rapporteur on Extreme Poverty and Human Rights. Available at
<https://undocs.org/A/74/493> (accessed 4 July 2024).

- Amnesty International (Netherlands). (2021). Xenophobic Machines. Available at: https://www.amnesty.nl/content/uploads/2021/10/NL_Xenophobe-Machines_Amnesty-International (accessed March 8 2023).
- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20, 973 - 989.
- Andersen, T. (2019). Legal experts: Gladsaxe's algorithm violates personal data law. Version 2, IT News. Available at <https://www.version2.dk/artikel/juridiske-eksperter-gladsaxes-algoritme-overtraeder-persondata-lov-1087407> (accessed 4 July 2024).
- Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals and its biased against blacks. ProPublica. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (accessed 8 April 2024).
- Apprich, C., Chun, W. H. K., Cramer, F., Steyerl, H. (2018). Pattern Discrimination. Minneapolis: University of Minnesota Press.
- Bainotti, L., Caliandro, A., & Gandini, A. (2020). From archive cultures to ephemeral content, and back: Studying Instagram Stories with digital methods. *New Media & Society*, 1461444820960071. <https://doi.org/10.1177/1461444820960071>
- Barocas, S., Hardt, M., & Narayanan, A. (2023). Fairness and Machine Learning: Limitations and Opportunities. MIT Press.
- Barocas, S., Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review* 104, no. 3: 671–732.
- Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1–13. <https://doi.org/10.1080/1369118X.2016.1216147>

- Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J. T., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., and Zhang, Y. (2018). Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. arXiv preprint arXiv:1810.01943.
- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK: Polity Press.
- Beretta, E., Santangelo, A., Lepri, B., Vetrò, A., De Martin, J.C. (2019). *The Invisible Power of Fairness. How Machine Learning Shapes Democracy*. In: Meurs, MJ., Rudzicz, F. (eds) *Advances in Artificial Intelligence. Canadian AI 2019. Lecture Notes in Computer Science()*, vol 11489. Springer, Cham. https://doi.org/10.1007/978-3-030-18305-9_19
- Berg, A.-J., Lie, M. (1993). Feminism and constructivism: Do artifacts have gender? *Science, Technology & Human Values*, 20(3), 332-351.
- Bijker, W. E.; Hughes, T. P.; Pinch, T. (1987). (eds.), *The social construction of technological systems: new directions in the sociology and history of technology*, Cambridge, Massachusetts: MIT Press, pp. 83–103.
- Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2021). The Values Encoded in Machine Learning Research (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2106.15590>
- Bogen, M. (2019). All the ways hiring algorithms can introduce bias. *Harvard Business Review*, May 6.
- Bonazzi, G. (2002). *Storia del pensiero organizzativo, la questione burocratica*. Milan: Franco Angeli.
- Bonini T., Gandini A. (2020). The field as a black box: Ethnographic research in the age of platforms. *Social Media + Society*, 6(4), 205630512098447. <https://doi.org/10.1177/2056305120984477>

- Bonini, T. and Trere, E. (2024). Algorithms of resistance - the everyday fight against platform power. Cambridge, MA: The MIT Press.
- Bovens, M. (2007b). Public accountability. In Ferlie E, Lynn LE and Pollitt C (eds), *The Oxford Handbook of Public Management*. Oxford University, Oxford, pp. 182–208.
- boyd, d., & Crawford, K. (2012). Critical Questions For Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- boyd, danah. (2015). Social Media: A Phenomenon to be Analyzed. *Social Media + Society*, 1(1). <https://doi.org/10.1177/2056305115580148>
- Braidotti, R. 1995. “La molteplicità. un’etica per la nostra epoca, oppure meglio cyborg che dea”, in D. J. Haraway, *Manifesto Cyborg. Donne, tecnologie e biopolitiche del corpo*, Milano, Feltrinelli, pp. 9-38.
- Bretschneider, S. (1990). Management information systems in public and private organizations: An empirical test. *Public Administration Review*, 536-545.
- Brooker, P. (2022). Computational ethnography: A view from sociology. *Big Data & Society*. <https://doi.org/10.1177/20539517211069892>
- Broomfield, H., & Reutter, L. (2022). In search of the citizen in the datafication of public administration. In *Big Data & Society* (Vol. 9, Issue 1, p. 205395172210893). SAGE Publications. <https://doi.org/10.1177/20539517221089302>
- Browne, J., Cave, S., Drage, E., & McInerney, K. (eds.). (2023). *Feminist AI*. Oxford University PressOxford. <https://doi.org/10.1093/oso/9780192889898.001.0001>
- Bucher, T., & Helmond, A. (2018). The Affordances of Social Media Platforms. In *The SAGE Handbook of Social Media*.

- Buçinca, Z., Pham, C., Jakesch, M., Ribeiro, M., Olteanu, A., & Amershi, S. (2023). AHA!: Facilitating AI Impact Assessment by Generating Examples of Harms. *ArXiv, abs/2306.03280*.
- Burrell, J. (2016). How The Machine ‘Thinks’: Understanding Opacity In Machine Learning Algorithms. *Big Data & Society*, 3, 1–12.
- Cabitza, F., Campagner, A., Angius, R., Natali, C., & Reverberi, C. (2023). AI Shall Have No Dominion: on How to Measure Technology Dominance in AI-supported Human decision-making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Vol. 21, pp. 1–20). ACM.
<https://doi.org/10.1145/3544548.3581095>
- Caliandro, A. (2018). Digital Methods for Ethnography: Analytical Concepts for Ethnographers Exploring Social Media Environments. *Journal of Contemporary Ethnography*, 47(5), 551-578. <https://doi.org/10.1177/0891241617702960>
- Caliandro, A., & Gandini, A. (2016). *Qualitative Research in Digital Environments: A Research Toolkit* (0 ed.). Routledge. <https://doi.org/10.4324/9781315642161>
- Caliandro, A., & Gandini, A. (2016). *Qualitative Research in Digital Environments: A Research Toolkit* (0 ed.). Routledge. <https://doi.org/10.4324/9781315642161>
- Callon, M. (1980). Struggles and Negotiations to Define What is Problematic and What is Not: The Socio-logic of Translation. pp. 197–221 in *The Social Process of Scientific Investigation*, edited by Karin D. Knorr. Dordrecht: Reidel Publishing.
- Callon, M. (1987). Society in the making: the study of technology as a tool for sociological analysis. In Bijker, Wiebe E.; Hughes, Thomas P.; Pinch, Trevor (eds.), *The social construction of technological systems: new directions in the sociology and history of technology*, Cambridge, Massachusetts: MIT Press, pp. 83–103.
- Callon, M. (1991). Techno-economic networks and irreversibility. In Law, J. (ed.), *A sociology of monsters: essays on power, technology and domination*, London: Routledge, pp. 132–165.

- Cardano, M. (2011). *La ricerca qualitativa*. Milano: Il Mulino.
- Carey, A.N., Wu, X. (2023). The statistical fairness field guide: perspectives from social and formal sciences. *AI Ethics* 3, 1–23. <https://doi.org/10.1007/s43681-022-00183-3>
- Casimir-Braun, J., Constantaras, E., Aung, H., Geiger, G., Mehrotra, D., and Howden, D. (2023). “Suspicion Machine Methodology.” *Lighthouse Reports*. Available at: <https://www.lighthousereports.com/suspicion-machines-methodology/>. Accessed on March 8 2023.
- Castells, M. (1998). *The Rise of the Network Society, The Information Age: Economy, Society and Culture, Vol. I*. Cambridge, MA; Oxford, UK. Blackwell.
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data* 5(2): 153–163.
- Choung, H., David, P., & Ross, A. (2022). Trust in AI and Its Role in the Acceptance of AI Technologies. *International Journal of Human–Computer Interaction*, 39(9), 1727–1739. <https://doi.org/10.1080/10447318.2022.2050543>.
- Christin, A. (2017). Algorithms in Practice: Comparing Web Journalism and Criminal Justice. *Big Data & Society* 4(2): 1–14.
- Christin, A. (2020). The ethnographer and the algorithm: beyond the black box. *Theory and Society* 49, 897–918. <https://doi.org/10.1007/s11186-020-09411-3>.
- Chun, W. H. K. (2017). *Updating to Remain the Same: Habitual New Media*. Cambridge, MA: MIT Press.
- Chun, W. H. K. (2021). *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Regno Unito: MIT Press.
- Citron, D. K. & Pasquale, F. A. (2014). The Scored Society: Due Process for Automated Predictions. *Washington Law Review*, Vol. 89, p. 1-, U of Maryland Legal Studies Research Paper No. 2014-8, Available at SSRN: <https://ssrn.com/abstract=2376209>.

- Clayton, A. (2020). How Eugenics Shaped Statistics. Nautilus. Available at: <https://nautil.us/how-eugenics-shaped-statistics-238014/> (accessed 4 March 2024).
- Cobbe, J., Lee, M.S.A. & Singh, J. (2021). Reviewable automated decision-making: A framework for accountable algorithmic systems. In FAccT 2021—Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. ACM, New York, pp. 598–609. <https://doi.org/10.1145/3442188.3445921>.
- Cockburn, C., & Ormrod, S. (1993). *Gender and Technology in the Making*. SAGE Publications Ltd.
- Coeckelbergh, M. (2020b). *AI Ethics*. Cambridge, MA: MIT Press.
- Constantaris, E., Geiger, G., Casimir Braun, J., Mehrotra, D., and Aung, H. (2023). Inside the Suspicion Machine. *Wired*. Available at: <https://www.wired.com/story/welfare-state-algorithms/> (accessed March 8 2023).
- Cooper, A.F., Moss, E., Laufer, B. & Nissenbaum, H. (2022). Accountability in an algorithmic society: relationality, responsibility, and robustness in machine learning, In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22). ACM, New York, pp. 864–876. <https://doi.org/10.1145/3531146.3533150>
- Corea, G. (1985). *Man-Made Women: How new reproductive technologies affect women*. London: Hutchinson.
- Court of Audit (Netherlands). (2016). Tackling problematic debt. Available at: [Rekenkamer.nl/publications/reports/2016/06/30/tackling-problematic-debt](https://rekenkamer.nl/publications/reports/2016/06/30/tackling-problematic-debt) (accessed March 8 2023).
- Cowan, R. S. (1979). From Virginia Dare to Virginia Slims: Women and Technology in American Life. *Technology and Culture*, 20(1), 51–63.
- Cowan, R. S. (1983). *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave*. New York: Basic Books.

- Crawford, K. (2016). Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics. *Science, Technology and Human Values* 41, 1 (2016), 77–92.
<https://doi.org/10.1177/0162243915589635>
- Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Crisis of Artificial Intelligence* New Haven, CT: Yale University Press.
- Crenshaw, K. W. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics, 1989 U. Chi. Legal F. 139.
- Criado-Perez, C. (2019). *Invisible women*. New York: Abrams Press.
- Data Protection Authority (Netherlands). (2020). Belastingdienst/Toeslagen: Processing the nationality of childcare benefit applicants. Available at:
<https://www.autoriteitpersoonsgegevens.nl/documenten/boete-belastingdienst-kinderopvangtoeslag> (accessed March 8 2023).
- D'Ignazio, C. & Klein, L. (2020). *Data Feminism*. MIT Press.
- Danaher, J. (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy and Technology* 29 (3):245-268.
- de Boer, N., & Raaphorst, N. (2021). Automation and discretion: explaining the effect of automation on how street-level bureaucrats enforce. *Public Management Review*, 25(1), 42–62. <https://doi.org/10.1080/14719037.2021.1937684>
- de Witt Wijnen, P. (2020). House of Representatives enforces greater compensation for parents affected. NRC Handelsblad (9 June 2020). Archived on 16 January 2021. Accessed on May 15, 2023.
- Deleuze, G. (1992). 'Postscript on the Societies of Control', *October* 59 (Winter): 3–7. Original publication: Deleuze, G. 1990. 'Post-scriptum sur les sociétés de contrôle', *L'autre journal*:1, mai.

- Dencik, L. and Leistert, O. (2015). eds. Critical Perspectives on Social Media and Protest: Between Control and Emancipation. Rowman & Littlefield International, pp. 163-180.
- Dencik, L., Kaun, A. (2020). Datafication and the Welfare State. *Global Perspectives* 1 (1): 12912. doi:10.1525/gp.2020.1291.
- Denzin, N., and Y. Lincoln. (2000). Handbook of Qualitative Research. Thousand Oaks, CA: Sage.
- Diakopoulos, N. (2014). Algorithmic Accountability Reporting: On the Investigation of Black Boxes. Columbia University. <https://doi.org/10.7916/D8ZK5TW2>.
- Dignum, V. (2018). Ethics In Artificial Intelligence: Introduction To The Special Issue. *Ethics and Information Technology* 20, pp. 1–3.
- Dijck, J. van, Poell, T., & Waal, M. de. (2018). The Platform Society: Public Values in a Connective World. Oxford University Press.
- Ditchfield, H. (2019). Behind the screen of Facebook: Identity construction in the rehearsal stage of online interaction. *New Media & Society*, 1461444819873644. <https://doi.org/10.1177/1461444819873644>
- Drucker, J. (2011). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly* 5, no. 1.
- Dutch Data Protection Authority. (2023). AI & Algorithmic Risks Report Netherlands: Winter 2023/2024. Department for the Coordination of Algorithmic Oversight. Available at: <https://www.autoriteitpersoonsgegevens.nl/en/current/ai-and-algorithm-risks-on-the-rise-amidst-increased-use-master-plan-necessary-to-prepare-the-netherlands-for-a-future-with-ai> (Accessed 09/09/2024).

- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd Conference on Innovations in Theoretical Computer Science*, pp. 214–226. ACM.
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. In *International Journal of Human-Computer Studies* (Vol. 58, Issue 6, pp. 697–718).
- Easlea, B. (1983). *Fathering the Unthinkable: Masculinity, Scientists and the Nuclear Arms Race*. London: Pluto Press.
- Ebers, M. (2021). Regulating Explainable AI in the European Union. An Overview of the Current Legal Framework(s). Liane Colonna/Stamley Greenstein (eds.), *Nordic Yearbook of Law and Informatics 2020: Law in the Era of Artificial Intelligence*, Available at SSRN: <https://ssrn.com/abstract=3901732> or <http://dx.doi.org/10.2139/ssrn.3901732>
- Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, 14(4), 532–550. <https://doi.org/10.2307/258557>
- Ellul, J. (1964). *The technological society*. New York: Vintage Books.
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St Martin's Press.
- European Commission (2019) Ethics guidelines for trustworthy AI. Available at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (accessed 8 April 2024).
- European Commission (2022). The Digital Economy and Society Index—Countries' Performance in Digitalization. Available at: <https://digital-strategy.ec.europa.eu/en/policies/countries-digitisation-performance> (accessed 3 September 2023).

- European Commission. (2024). Proposal for a Regulation laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act). Digital Strategy. Available at: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (accessed 24 September 2024).
- Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62-70.
- Faulkner, W. (2001). The technology question in feminism: A view from feminist technology studies. *Women's Studies International Forum*, 24 (1). 79-95.
- Feenberg, A. (1999). *Questioning Technology*, London and New York: Routledge.
- Feenberg, A. (2010). *Between Reason and Experience: Essays in Technology and Modernity*. Cambridge: The MIT Press.
- Firestone, S. (1970). *The Dialectic of Sex: The Case for Feminist Revolution*. New York: Bantam Books.
- Fisher, R. (1948). *Statistical Methods for Research Workers*. Oliver and Boyd.
- Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy and Technology* 32, pp. 185–193.
- Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford, UK: Oxford University Press.
- Foucault, M. (1977). *Discipline and punish: the birth of the prison*. New York: Pantheon Books.
- Fountain, J. E. (2004). *Building the Virtual State: Information Technology and Institutional Change*. Brookings Institution Press.
- Fountain, J. E. (2004). *Building the Virtual State: Information Technology and Institutional Change*. United States: Brookings Institution Press.

- Friedman, B. & Nissenbaum, H. (1996). Bias in computer systems. *ACM SIGCHI Bulletin* 14(3), 330–347. <https://doi.org/10.1145/249170.249184>
- Gajane, P., & Pechenizkiy, M. (2017). On formalizing fairness in prediction with machine learning. arXiv preprint arXiv:1710.03184.
- Gandini, A., Gerosa, A., Giuffrè, L. *et al.* (2023). Subjectivity and algorithmic imaginaries: the algorithmic other. *Subjectivity* 30, 417–434. <https://doi.org/10.1057/s41286-023-00171-w>
- Garfinkel, H. (1974). The origins of the term ethnomethodology, in R. Turner (ed.), *Ethnomethodology*, Penguin, Harmondsworth, pp 15–18.
- Gill, R. & Grint, K. (1995). *The Gender-Technology Relation*. London; Bristol, PA: Taylor & Francis.
- Gillespie, T. (2010). The politics of ‘platforms’. *New Media & Society*, 12(3), 347–364. <https://doi.org/10.1177/1461444809342738>
- Gillespie, T. (2014). The Relevance of Algorithms. In Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot (eds), *Media Technologies: Essays on Communication, Materiality, and Society*. MIT Press Scholarship Online.
- Glaser, B., & Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Mill Valley, CA: Sociology Press.
- Godin, B. (2006). The Linear Model of Innovation: The Historical Construction of an Analytical Framework. *Science, Technology, & Human Values*, 31(6), 639–667. <https://doi.org/10.1177/0162243906291865>
- Guilherme, A. (2019). AI and education: The importance of teacher and student relations. *AI & Society*, 34(1), 47–54.

- Hamilton, I. A. 2018. Why It's Totally Unsurprising That Amazon's Recruitment AI Was Biased Against Women. *Business Insider*. Oct 13. Available at: <https://www.businessinsider.com/amazon-ai-biased-against-women-no-surprise-sandra-wachter-2018-10?IR=T>. [Accessed 21 June 2020].
- Haraway, D. (1991). A Cyborg Manifesto: Science, Technology, and Socialist- Feminism in the Late Twentieth Century in *Simians, Cyborgs and Women: The Reinvention of Nature*. New York: Routledge, pp.149-181.
- Harding, S. G. (1986). *The science question in feminism*. Ithaca: Cornell University Press.
- Hardt, M. and A. Negri. (2017). *Assembly*. Oxford: Oxford University Press.
- Hardt, M., Price, E., Srebro, N., et al. (2016). Equality of opportunity in supervised learning. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS'16)*, pp. 3315–3323.
- Helmond, A. (2015). The Platformization of the Web: Making Web Data Platform Ready. *Social Media + Society*, 1(2).
- Hobsbawm, E. (1994). *Age of extremes: the short twentieth century, 1914-1991*. London: Michael Joseph.
- Hood, C. (1991). A public management for all seasons? *Public administration* 69.1:3–19
- Hooker, S. (2021). Moving beyond “algorithmic bias is a data problem.” In *Patterns* (Vol. 2, Issue 4, p. 100241). Elsevier BV. <https://doi.org/10.1016/j.patter.2021.100241>
- Hughes, T. P. (1987). The Evolution of Large Technological Systems. In *The Social Construction of Technological Systems*, eds. Wiebe Bijker, Thomas P. Hughes, and Trevor Pinch. Cambridge, MA: MIT Press. 45-76.
- Hughes, T. P. (1994). Technological momentum. In Merritt Roe Smith and Leo Marx, ed., *Does Technology Drive History?: The Dilemma of Technological Determinism*. Massachusetts Institute of Technology, pp. 101–113.

- Hughes, T. P. (2004). *Human-Built World: How to Think about Technology and Culture*. United Kingdom: University of Chicago Press.
- Hull, D. A. (1996), Stemming algorithms: A case study for detailed evaluation. *Journal of the American Society for Information Science*, 47: 70-84.
- Jasanoff, S., Markle, G., Peterson, J., & Pinch, T. (1994). *Handbook of Science and Technology Studies* (eds). Thousand Oaks, California: Sage.
- Johnson, D. G., & Nissenbaum, H. (1995). *Computers, ethics & social values*. Prentice-Hall, Upper Saddle River.
- Kaun, A., & Liminga, A. (2023). Welfare service centers: Maintenance, repair, and care at the analog interfaces of the digital welfare state. *New Media & Society*, 0(0).
<https://doi.org/10.1177/14614448231220362>
- Kellogg, K. C., Valentine, M. A., & Christin, A. (2018). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14 (1), 366–410.
- Kempeneer, S., Ranchordas, S., van de Wetering, S. (2024). AI Failure, AI Success, and AI Power Dynamics in The Public Sector. Available at SSRN: <https://ssrn.com/abstract=>
- Kirsten, M. (2018). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics* 0, 0, 1–16. <https://doi.org/10.1007/s10551-0183921-3>
- Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1–12.
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & their consequences*. London: Sage.
- Kitchin, R., & Dodge, M. (2011). *Code/space: Software and everyday life*. Cambridge, MA: Mit Press.

- Kleinnijenhuis, J. (2020). Reconstruction: how parents were destroyed 'efficient and effective'. *Trouw* (14 November 2020). Available at: <https://www.trouw.nl/politiek/wie-wist-wat-in-de-toeslagenaffaire-de-kluwen-van-hoofdrolspelers-ontward~b721c834/> (accessed 15 May 2023).
- Kolkman, D. (2020). “F**k the algorithm”?: What the world can learn from the UK’s A-level grading fiasco. Available at: <https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/> (accessed 8 April 2024).
- Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165, 633.
- Lange, A.-C., Lenglet, M. and Seyfert, R. (2019). On Studying Algorithms Ethnographically: Making Sense of Objects of Ignorance. *Organization* 26(4): 598–617.
- Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, Massachusetts: Harvard University Press.
- Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In Bijker, W. E. and Law, J. (eds) *Shaping Technology/Building Society: Studies in Sociotechnical Change*, Cambridge, MA, MIT Press, pp. 225-58.
- Latour, B. (1999). *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge: Harvard University Press.
- Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-network-theory*. Oxford: Oxford University Press.
- Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.
- Law, J., & Hassard, J. (1999). *Actor Network Theory and After* (eds.). United Kingdom: Wiley.

- Leavy, S. (2018). Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning. In: 1st International Workshop on Gender Equality in Software Engineering (GE). New York NY, US: ACM, pp.14-16.
- Lecher, C. (2020). What happens when an algorithm cuts your healthcare. The Verge. Available at <https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy> (Accessed 8 April 2024).
- Lepri, B., Oliver, N., Letouzé, E. (2018). Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy and Technology* 31(4): 611–627.
- Lessig, L. (1999). Code and other laws of cyberspace. New York: Basic Books.
- Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanit Soc Sci Commun* 7, 9. <https://doi.org/10.1057/s41599-020-0501-9>
- Lohan, M., & Faulkner, W. (2004). Masculinities and Technologies: Some Introductory Remarks. *Men and Masculinities*, 6(4), 319–329.
- Louçã, F. (2009). Emancipation Through Interaction — How Eugenics and Statistics Converged and Diverged. *Journal of the History of Biology*, 42(4), 649–684
- Lower House of the States General. (2020). Report - Parliamentary Interrogation Committee Childcare Allowance: Unprecedented Injustice. Available at: tweedekamer.nl/sites/default/files/atoms/files/20201217_final_report_parliamentary_interrogation_committee_childcaresurcharge.pdf (accessed 8 May 2023).
- Maasland, C., & Weißmüller, K. S. (2022). Blame the Machine? Insights From an Experiment on Algorithm Aversion and Blame Avoidance in Computer-Aided Human Resource Management. *Frontiers in psychology*, 13, 779028. <https://doi.org/10.3389/fpsyg.2022.779028>

- Mackenzie, D. (1981). *Statistics in Britain 1865-1930, The Social Construction of Scientific Knowledge*. Edinburgh: Edinburgh University Press.
- MacKenzie, D. (1993). *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance*. Cambridge, Massachusetts: MIT Press.
- MacKenzie, D. and Wajcman, J. (1999). *The Social Shaping of Technology (Second Edition)* (eds.). Buckingham, Philadelphia: Open University Press.
- Majumder, S. (2024). In India, an algorithm declares them dead. They have to prove they're alive. *Al Jazeera*. Available at: <https://www.aljazeera.com/economy/2024/1/25/in-india-an-algorithm-declares-them-dead-they-have-to-prove-theyre>. (Accessed 09/09/2024).
- Mantelero, A. (2022). *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI. Information Technology and Law Series*. Asser Press.
- Mari, A. (2019). Brazil gears up to enhance data analytics to tackle benefit fraud. *ZD Net*. July 12. Available at <https://www.zdnet.com/article/brazil-gears-up-to-enhance-data-analytics-to-tackle-benefit-fraud> (accessed 4 July 2024).
- Marx, K. (1847). *Miseria della filosofia*, Roma: Editori Riuniti.
- Marx, K. (1867). *Capital: A critique of political economy*. Volume 1, Part 1: The process of capitalist production. New York, NY: Cosimo.
- Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray Publishers.
- McCradden, M. D., Joshi, S., Mazwi, M., & Anderson, J. A. (2020). Ethical limitations of algorithmic fairness solutions in health care machine learning. *The Lancet. Digital health*, 2(5), e221–e223. [https://doi.org/10.1016/S2589-7500\(20\)30065-0](https://doi.org/10.1016/S2589-7500(20)30065-0)

- Meijer, A. & Grimmelikhuijsen, S. (2020). Responsible and accountable algorithmization: How to generate citizen trust in governmental usage of algorithms. In Schuilenburg M and Peeters R (eds), *The Algorithmic Society: Technology, Power, and Knowledge*. Routledge, London, pp. 53–66.
- Meijer, A. and Bovens, M. (2003). Public accountability in the information age. In Palmirani, M., Van Engers, T. and Wimmer, M.A. (eds), *E-Government. Workshop in Conjunction with JURIX 2003, International Federation for Information Processing*, Laxenburg, pp. 16–28.
- Meijer, A., Lorenz, L., Wessels, M. (2021). Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems. *Public Administration Review*, Vol. 81, Iss. 5, pp. 837–846. doi:10.1111/puar.13391.
- Merchant, C. (1980): *The Death of Nature: Women, Ecology and the Scientific Revolution*. New York: Harper and Row.
- Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors* 46(2), 196–204.
- Mies, M. (1987). Why do we Need all this? A Call Against Genetic Engineering and Reproductive Technology. In Spallone, P. and Steinberg, D. (eds). *Made To Order: The Myth of Reproductive and Genetic Progress*. Oxford: Pergamon Press.
- Milakovich, M.E. (2011). *Digital Governance: New Technologies for Improving Public Service and Participation* (1st ed.). Routledge. <https://doi.org/10.4324/9780203815991>
- Ministry of the Interior and Kingdom Relations. (2024, June 20). FRAIA in action. Government of the Netherlands. Available at: <https://www.government.nl/documents/reports/2024/06/20/fraia-in-action> (Accessed 09/09/2024).
- Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual review of statistics and its application*, 8(1), 141-163.

- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society* 3, 2 (2016), 205395171667967.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The Ethics Of Algorithms: Mapping The Debate. *Big Data & Society*, 3 (2): 1–21.
- Mökander, J., Morley, J., Taddeo, M., Floridi, L. (2021). Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations. *Science & Engineering Ethics* 6;27(4):44. doi: 10.1007/s11948-021-00319-4.
- Moore, M. (1995). Creating public value: Strategic management in government. Harvard University Press.
- Mosier, K., Skitka, L.; Heers, S., Burdick, M. (1997). Automation Bias: Decision Making and Performance in High-Tech Cockpits. *International Journal of Aviation Psychology*. 8 (1): 47–63. doi:10.1207/s15327108ijap0801_3. PMID 11540946.
- Mosier, K.L., Skitka, L.J. (1996). Human decision makers and automated decision aids: made for each other?, in: Automation and Human Performance: Theory and Applications, Erlbaum, pp.201–220.
- Mutsaers, P., & Nuenen, T. van. (2023). Predictively policed. In Policing race, ethnicity and culture. Manchester University Press. <https://doi.org/10.7765/9781526165596.00010>
- National Ombudsman. (2016). Citizen's perspective on debt relief. Available at: nationaleombudsman.co.uk/news/investigations/2016050-citizen'sperspective on debt relief (accessed March 8 2023).
- Netherlands Court of Audit. (2022). An audit of algorithms: Nine algorithms used by the Dutch government. Available at: <https://english.rekenkamer.nl/publications/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government> (Accessed 09/09/2024).

- Newell, S. and Marabelli, M. (2015) Strategic Opportunities (and Challenges) of Algorithmic Decision Making: A Call for Action on the Long-Term Societal Effects of Datification. *The Journal of Strategic Information Systems* 24(1): 3–14.
- Niederer, S. & van Dijck, J. (2010). Wisdom of the Crowd or Technicity of Content? Wikipedia as Socio-technical System. *New Media & Society*, 12, 8, 1368-1387.
- Nissenbaum, H. (1994). Computing and accountability. *Communications of the ACM* 37, 1: 72–80. <https://doi.org/10.1145/175222.175228>
- Noble, D. F. (1978). Social Choice in Machine Design: The Case of Automatically Controlled Machine Tools, and a Challenge for Labor. *Politics & Society*, 8(3-4), 313–347.
- Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
- Noy, C. (2008). Sampling Knowledge: The Hermeneutics of Snowball Sampling in Qualitative Research. *International Journal of Social Research Methodology*, 11(4), 327–344. <https://doi.org/10.1080/13645570701401305>
- O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.
- OECD. (2019). Recommendation of the Council on OECD Legal Instruments Artificial Intelligence. Paris. Available at: <https://www.oecd.ai/ai-principles> (accessed 8 April 2024).
- Offe, C. (2018). *Contradictions of the Welfare State* (J. Keane, Ed.). Routledge. <https://doi.org/10.4324/9780429464829>
- Orlikowski, W. J. (2000). Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. *Organization Science* 11(4): 404–28.

- Oudshoorn, N., Rommes, E., & Stienstra, M. (2004). Configuring the User as Everybody: Gender and Design Cultures in Information and Communication Technologies. *Science, Technology, & Human Values*, 29(1), 30–63. <https://doi.org/10.1177/0162243903259190>
- Parasuraman, R., Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 39 (2): 230–253. doi:10.1518/001872097778543886. S2CID 41149078.
- Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, London: Harvard University Press.
- Pettigrew, A. M. (1990). Longitudinal Field Research on Change: Theory and Practice. *Organization Science*, 1(3), 267–292. <http://www.jstor.org/stable/2635006>
- Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*, 14(3), 399–441.
- Plant, S. (1996). On the matrix: cyberfeminist simulations. In Rob Shields (ed.), *Cultures of the Internet: Virtual Spaces, Real Histories, Living Bodies*, pp. 181-2. London: Sage.
- Plant, S. (1998). *Zeros and Ones*. Regno Unito: Fourth Estate.
- Purtrova, N. & Huyskes, D. (2024). General Principles of Code. *Technology and Regulation*, 2024, 132-142. <https://doi.org/10.26116/techreg.2024.013>
- Rainie, L., Wellman, B. (2014). *Networked: The New Social Operating System*. United Kingdom: Penguin Random House LLC.
- Rechtbank Den Haag. (2020). SyRI legislation in breach of European Convention on Human Rights. Available at:

<https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBDHA:2020:1878> (accessed 12 June 2024).

- Renieris, E. M. (2022). *Beyond Data: Reclaiming Human Rights at the Dawn of the Metaverse*. Cambridge, Massachusetts: MIT Press.
- Richardson, R., Schultz, J.M., Southerland, V.M. (2019). *Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems*. New York. Available at: <https://ainowinstitute.org/litigatingalgorithms-2019-us.html> (accessed 8 April 2024).
- Rieder, B. (2016). Scrutinizing an Algorithmic Technique: The Bayes Classifier as Interested Reading of Reality. *Information, Communication & Society* 0.0: 1–18.
- Rogers, R. (2010). Internet Research: The Question of Method. *Journal of Information Technology and Politics*, 7, 2/3, 241-260.
- Rogers, R. (2015). *Digital methods*. The MIT Press.
- Rogers, Y. (2012). "HCI Theory: Classical, Modern, and Contemporary". *Synthesis Lectures on Human-Centered Informatics*. 5 (2): 1–129.
- Romain, M., Senecat, A., Pénicaut, S., Geiger, G., & Braun, J.-C. (2023). France's digital inquisition. *Lighthouse Reports*.
<https://www.lighthousereports.com/investigation/frances-digital-inquisition/>
- Rosenfeld, A., & Richardson, A. (2019). Explainability in human-agent systems. *Autonomous Agents and Multi-Agent Systems*, 33, 673-705.
- Rovatsos, M., Mittelstadt, B., & Koene, A. (2019). Landscape Summary: Bias in Algorithmic Decision-Making. In *What is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it?*. UK Government.

- Rowland, R. (1985). Motherhood, patriarchal power, alienation and the issue of "choice". In Corea, G. (eds) *Man-Made Women: How new reproductive technologies affect women*. London: Hutchinson.
- Rowland, R. (1988). *Woman Herself - A Transdisciplinary Perspective on Women's Identity*. Melbourne: Oxford University Press.
- Saegert, S. (1980): Masculine Cities and Feminine Suburbs: Polarized Ideas, Contradictory Realities. *Signs*, 5, 3, pp. 96-111.
- Sætra, H. S. (2020). A shallow defense of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government. *Technology in Society*, 62, 101283.
- Santoni de Sio, F., & Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them”, *Philosophy & Technology* 34(4): 1057–1084.
- Savage, A., & Hyde, R. (2012). Using freedom of information requests to facilitate research. *International Journal of Social Research Methodology*, 17(3), 303–317.
- Scantamburlo, T. & Charlesworth, A. & Cristianini, N. (2018). Machine Decisions and Human Consequences in Yeung, K. & Lodge, M. 2019. *Algorithmic Regulation*. Oxford: Oxford University Press.
- Schleiner, A. M. Does Lara Croft wear fake poly-gons? Gender and gender-role subversion in computer adventure games. *Leonardo*, 34, 4 (2001), pp. 221-6.
- Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4(2).
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency. FAT* '19: Conference on Fairness, Accountability, and Transparency. ACM.

- Shaw, W. H. (1979). "The Handmill Gives You the Feudal Lord": Marx's Technological Determinism. *History and Theory*, 18(2), 155–176.
- Schiels, C. (2023). *Offshore attachments: Oil and intimacy in the Caribbean*. University of California Press.
- Singer, N. (2015). Bringing big data to the fight against benefit fraud. *The New York Times*, February 20. Available at <https://www.nytimes.com/2015/02/22/technology/bringing-big-data-to-the-fight-against-benefits-fraud.html> (accessed 4 July 2024).
- Skitka L.J., Mosier K.L., Burdick M., (2000). Accountability and automation bias. *International Journal of Human-Computer Studies*, 2000, vol. 52.
- Skitka, L.J., Mosier, K., Burdick, M.D. (1999). Does automation bias decision making? *International Journal of Human-Computer Studies* 51(5), 991–1006.
- Smith, A. (2015). *The Wealth of Nations: A Translation Into Modern English*. Manchester, England: Industrial Systems Research.
- Smith, M. R. & Marx, L. (1994). *Does technology drive history?: the dilemma of technological determinism*. Cambridge, Massachusetts: MIT Press.
- Srnicek, N., & De Sutter, L. (2017). *Platform capitalism*. Polity.
- Stanworth, M. (ed.) 1987: *Reproductive Technologies: Gender, Motherhood and Medicine*. Cambridge: Polity Press.
- Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Julia, H., Kalayanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016). *Artificial intelligence and life in 2030: One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*. Stanford University. https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai_100_report_0831fml.pdf

- Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *The Academy of Management Review*, 20(3), 571–610. <https://doi.org/10.2307/258788>
- Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368-383.
- Suresh, H. & Gutttag, J. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '21)*, NY, USA. ACM, New York, NY, USA. <https://doi.org/10.1145/3465416.3483305>
- Suresh, H., & Gutttag, J. V. (2019). A framework for understanding unintended consequences of machine learning. arXiv preprint arXiv:1901.10002, 2(8), 73.
- Tangi, L., Van Noordt, C., Combetto, M., Gattwinkel, D. and Pignatelli, F., (2022). AI Watch. European landscape on the use of Artificial Intelligence by the Public Sector. Publications Office of the European Union, Luxembourg. doi:10.2760/39336, JRC129301.
- Turkle, S. (1995). Life on the Screen: Identity in the Age of the Internet. Simon & Schuster Trade.
- 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>
- van de Poel, I., Kroes, P. (2014). Can Technology Embody Values?. In: Kroes, P., Verbeek, PP. (eds) The Moral Status of Technical Artifacts. *Philosophy of Engineering and Technology*, vol 17. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-7914-3_7
- van der Vlist, F. N. (2016). Accounting for the social: Investigating commensuration and Big Data practices at Facebook. *Big Data & Society*, 3(1), 2053951716631365. <https://doi.org/10.1177/2053951716631365>

- van Dijk, J. (2006). *The Network Society: Social Aspects of New Media*. India: SAGE.
- van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. *Government information quarterly*, 39(3), 101714.
- van Zoonen, L. (2020). Data governance and citizen participation in the digital welfare state. *Data & Policy*, 2, e10. doi:10.1017/dap.2020.10
- Vanek, J. (1974). Time Spent on Housework. *Scientific American*, 231, 5, pp. 116-20.
- Veale, M. & Brass, I. (2019). Administration by Algorithm? Public Management Meets Public Sector Machine Learning. In: *Algorithmic Regulation* (Yeung, K. & Lodge, M. eds., Oxford University Press, 2019), Available at SSRN: <https://ssrn.com/abstract=3375391>
- Vecchione, B., Levy, K., & Barocas, S. (2021). Algorithmic auditing and social justice: Lessons from the history of audit studies. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (pp. 1-9).
- Vered, M., Livni, T., Howe, P. D. L., Miller, T., & Sonenberg, L. (2023). The effects of explanations on automation bias. In *Artificial Intelligence* (Vol. 322, p. 103952). Elsevier BV. <https://doi.org/10.1016/j.artint.2023.103952>
- Wajcman, J. (1991). *Feminism Confronts Technology*. Polity Press.
- Wajcman, J. (2004). *TechnoFeminism*. Polity Press.
- Wajcman, J. (2006). TechnoCapitalism Meets TechnoFeminism: Women and Technology in a Wireless World. *Labour & Industry: A Journal of the Social and Economic Relations of Work*, 16(3), 7–20. <https://doi.org/10.1080/10301763.2006.10669327>
- Wajcman, J. (2009). Feminist Theories Of Technology. *Cambridge Journal of Economics*, 34(1), 143–152.

- Wajcman, J. (2021). Reflections on Gender and Technology Studies: In What State Is the Art?, *Social Studies of Science*, 19.
- Weber, M. (1904). *The Protestant Ethic and the Spirit of Capitalism*. New York: Scribner's Press.
- Wieringa, M. (2023). "Hey SyRI, tell me about algorithmic accountability": Lessons from a landmark case. *Data & Policy*, 5(2). <https://doi.org/10.1017/dap.2022.39>
- Wihlborg, E., Larsson, H., & Hedström, K. (2016). "The Computer Says No!" A Case Study on Automated Decision-Making in Public Authorities. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 2903-2912). IEEE.
- Wihlborg, E., Larsson, H., & Hedström, K. (2016). "The Computer Says No!" A Case Study on Automated Decision-Making in Public Authorities. 49th Hawaii International Conference on System Sciences (HICSS). Koloa, HI, USA, pp. 2903-2912, doi: 10.1109/HICSS.2016.364.
- Winner, L. (1993). Upon Opening the Black Box and Finding It Empty: Social Constructivism and the Philosophy of Technology. *Science, Technology, & Human Values*, 18(3), 362-378. <https://doi.org/10.1177/016224399301800306>
- Woodforde, J. (1970). *The Story of the Bicycle*. London: Routledge and Kegan Paul.
- Woolgar, S. (1982). Laboratory Studies: A Comment on the State of the Art. *Social Studies of Science*, 12(4), 481-498. <https://doi.org/10.1177/030631282012004001>
- Yeung, K. (2022). The New Public Analytics as an Emerging Paradigm in Public Sector Administration. 27(2) *Tilburg Law Review* pp. 1–32. <https://doi.org/10.5334/tlr.303>
- 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. In *Perspectives on Public Management and Governance*. Oxford University Press (OUP). <https://doi.org/10.1093/ppmgov/gvz014>

- Ziewitz, M. (2016). Special Issue Introduction: Governing Algorithms: Myth, Mess, and Methods. *Science, Technology, & Human Values*, 41(1), 3–16.
- Ziosi, M., Watson, D., & Floridi, L. (2024). A Genealogical Approach to Algorithmic Bias. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4734082>
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.
- Zuiderveen Borgesius, F. Council of Europe. (2018). Discrimination, Artificial Intelligence, And Algorithmic Decision-Making. Available at: <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>. [Accessed on 19 June 2024].
- Zunino, C. (2019). Scuola, trasferimenti di 10mila docenti lontano da casa. Il Tar: “L’algoritmo impazzito fu contro la Costituzione”. *La Repubblica*. Accessed 20 June 2024. Available at: https://www.repubblica.it/cronaca/2019/09/17/news/scuola_trasferimenti_di_10mila_docenti_lontano_da_casa_il_tar_l_algoritmo_impazzito_fu_contro_la_costituzione_-236215790/