UNIVERSITY OF GOTHENBURG

Faculty of Arts

Department of Cultural Sciences

Algorithmic Injustice: Predictive Policing and Profiling in the Netherlands – the Top600 and Top400 approaches

by: Carloruggero Anastasio

Master's thesis in Gendering Practices, 30 hp
Supervisor: Lisa Åkervall
May 2024

Abstract:

This thesis is located within the emerging intersections between feminist critique and algorithmic injustice, with a specific focus on predictive policing. By employing feminist theoretical perspectives on governmentality, the non-neutrality of science and statistics, phenomenology and structural injustice, it problematizes Top600 and Top400 (AKA Veilig Alternatief), two algorithmic tools that Amsterdam police and institutions use for the identification and risk assessment of potential perpetrators and repeat offenders of high impact crimes. These tools are marked by a known over-representation of Dutch-Moroccan men and minors, with a lower socio-economic status, mild-to-borderline intellectual disabilities, and from segregated neighbourhoods. Through a reading, inspired by critical law studies, of the official regulations of personal data usage by Top600 and Top400, and through a review of literature concerning them, the thesis problematizes the gap between their promised and factual outcomes, questioning the fairness of their targeting. It asks what practices and structures specific to algorithmic usage and policing are involved and potentially aggravating, how the relations between the tools and human actors are articulated and what biases and distortions they entail. The thesis concludes with considerations about the legal frameworks in which they function, especially in light of the new AI Act in the EU.

Keywords: algorithmic injustice, artificial intelligence, predictive policing, feminist critique, actuarial statistics

Acknowledgements

I would like to firstly thank my supervisor Lisa Åkervall, whose guidance, openness and stimulating feedback have been fundamental for the writing of this thesis and have represented one of the most exciting aspects of the research process.

The professors of my courses, the counsellors, the other supervisors and examiners have rendered my years at Göteborgs universitet and this final period of writing incredibly inspiring.

My former work supervisor and my former colleague at ALDA Europe, Camilla Vedovato and Sara Afendi, as well as many of the other co-workers, may not know it, but have made sure that my confidence in my abilities of investigating and explaining social issues would not falter during a pivotal time.

I would also like to thank my family and especially my parents and my sister for always and constantly supporting and encouraging me. I think fondly of how my cousin Giulia and her friend Elisabetta Potalivo have believed in the new direction of my studies from the start, as well as so many of my long-time friends.

Last but not least, I warmly thank my classmates, and especially my friends Pia, Elise, Nicole, Max, Shaofen, Emilia and Kaja for their wonderful academic and personal support.

Table of contents

List of abbreviations

AcVZ – Actiecentrum Veiligheid en Zorg *[Security and Care Action Centre]*

ADM – automated decision-making

AI – artificial intelligence

EU – European Union

DG – drug related crimes

HIC – high impact crimes

WODc – Wetenschappelijk Onderzoek- en Documentatiecentrum *[Scientific Research and Documentation Centre]*

1. Introduction

1.1. Background of the problem, background of the researcher. Aims and research questions

The current thesis is concerned with the topic of algorithmic injustice from a critical feminist perspective, which highlights its discrimination of already disadvantaged social groups and its potential for additional harm. Algorithms have long evolved from relatively simple sets of automated instructions to entities whose mathematical and structural intricacies can make them unreadable or unpredictable to their users and designers. As such, they have also been deemed capable of surpassing human capacities in terms of cost-efficient understanding and management of complex and multifaceted problems. For example, they have been more and more "employed to inform high-stakes decision making"; yet their "treatment of particular social groups" with already "socially sensitive characteristics, like race and gender" (Kariszadeh, 2022, p.1) has been problematized, and risks for wide negative impacts have grown. And since big data – from internet-based data to digitalised administrative data – has increasingly been used for research and understanding of populations (see e.g. Connelly et al., 2016) and for computation, while more recent developments of generative and intuitive artificial intelligence have accelerated the implementation of powerful algorithms, it would seem that the dangers of unfair future uses of algorithmic tools cannot be overlooked.

This research will focus in particular on instances of predictive policing, i.e. the series of practices aimed at the prevention of crimes by an assessment of the risks of criminality linked to entire neighbourhoods and/or single persons, in this case through algorithmic and statistical computation. Such technologies can become part of a broader system where – for example – being flagged as an at-risk subject or neighbourhood by police, even in absence of formal charges or proved crimes, can have consequences on other aspects of life (loan assignment, hiring prospects, welfare benefits, custody of children, etc.). I will deal mostly with tools that have been employed in the past 10+ years in the Netherlands, namely Top600 and Veilig Alternatief (translatable as *Safe Alternative*, and formerly known as Top400), for the profiling of the criminal potential of people who have come in contact with Dutch police. As will be detailed in the next chapter, such assessments are used for means of swifter repression, as well as further discipline in the attempt to rehabilitate or reroute its subjects' lives outside of actual or predicted crime – yet it has been argued that such assessments have often been racist, xenophobic, neuroableist, classist, ageist and sexist, and the authorities' employment of them to be merely repressive and harassing. These algorithmic approaches are supposed to reach their outcomes based on a variety of factors, which are not fully disclosed to the public, but that represent attributes of someone's life and situations that computations and police intelligence estimate to be criminogenic, i.e. to lead to an increased probability of committing crime. Even in the relative opacity of such factors, authorities claim that the criteria for calculation, at least for the ProKid approach, another algorithm that was used among the building blocks of Veilig Alternatief, is, of course, race-blind: ethnicity and race are not categories that are added to the data (Jansen, 2022, p.12). How could it be, then, that the tools target for the most part Dutch citizens of Moroccan or Surinamese descent (Jansen; Fair Trials, 2021, p.12-13; de Koning, 2017, p.542-543, 550)? Is it the result of a most advanced scientific calculation, that through abundant, diverse and accurate enough data has finally determined that such individuals are more prone to criminality, regardless of their ethnicity (or that, worse, such *groups* are essentially prone to it)? Or is it perhaps that, aside from what could already be a policing bias against such minority groups, these computations are producing new forms of discrimination with new modalities? These are broad questions which cannot be answered easily, and represent perhaps tendentious or partial takes, meant more to offer alternative perspectives to problematize results and expand investigation, rather than being aimed at immediate responses and full confirmation.

The event that sparked my interest on this branch of research, in many ways also gesturing towards similar upsetting questions, was the 2023 conference *Per un femminismo delle nuove tecnologie [For a Feminism of New Technologies]* (see Internazionale, 2023) – although it would be more accurate to say that this lecture was rather a catalyst, an eye-opening moment to the connections that could be made between several other interests and experiences that I already held, many of which derived from feminist scholarship. The main speaker, Ivana Bartoletti, offered a few examples of algorithmic usage that, especially in public governance, ended up having terribly harmful consequences on some parts of the population. Notably, one of the most prominent incidents was the Dutch Childcare Benefit Scandal, which saw the Dutch Tax and Customs Administration employ an algorithm of tax fraud detection that considered attributes like dual nationality and foreign-sounding names as factors linked to such a crime. This determined the exclusion from welfare benefits and the accusations of fraud of thousands of racialized low-income families, resulting in their sudden poverty, despair, losses of children custody (see European Parliament, 2022) and, as reported by Bartoletti, several suicides. She stressed how, in many such cases, an actual racist or discriminatory intent may not even exist – rather, there are ways in which data is created, collected and used that are unfair and riddled with prejudices, that are then precipitated and enacted by the algorithms. As she spoke, I felt a few usually separate threads converging in my mind – a fascination for coding, interests in anti-discrimination and structural injustice, the criticism of the neutrality of science and the concealed partiality of data and quantitative approaches, including the statistical data I had surveyed in my previous studies in economics and marketing. It seemed like my diversified academic education could bring an interdisciplinarity necessary for bridging these fields. I was also reminded of personal experiences, in academia and on the job, where I was assessed negatively by 'objective' systems that, in hindsight, were setting me at a disadvantage from the start for some of my own personal attributes (I return to my own positionality in a longer discussion on ethics, see section 1.5).

In this thesis I choose Dutch predictive policing as a particularly interesting case study for a few reasons. As for predictive policing itself, it is linked so directly with criminalization and violence, that it may at first seem to loose some of the subtlety of the type of discrimination concealed within data that Bartoletti depicted – yet, as it will become clearer in the rest of this work, it is also the consequence of seemingly non-violent ways of assessing reality and of disciplining individual and collective behaviours. I turn to criminalization as a fundamental reminder of some of the direst consequences of potentially unfair assessments and social disadvantages – as a way to upset the norms by looking at their extreme applications. Predictive policing is also envisioned by the Artificial Intelligence Act approved by the European Parliament, which as of May 2024 is still under final corrections and will be implemented after several months. As will be discussed in section 2.9, the Act does not seem to offer increased protections or accountability regarding the Dutch systems described in this thesis. As I willingly locate my area of interest in the European Union *and* in an EU that discusses algorithms, I choose the Netherlands as an ideal example: an assessment of algorithmic predictive policing in Europe (see Fair Trials, 2021) depicts their case as the most complex and multi-layered; their digitalisation of the public sphere is particularly advanced, and previous cases of algorithmic injustice have been recorded; and their privacy laws allow such aggregation and uses of personal data. Notably, other EU countries like Sweden, where a similar path to digitalisation may be seen, could also offer an ideal techno-social framework for similar algorithmic implementation, but have not taken this direction (yet). The following months, with further delegations and implementation to the Act and related discussion, will show potential paths of development and perhaps interesting divergences, compared to countries where AI usage in governance already abounds, like the USA, which also offer interesting literature on the subject (Eubanks, 2018; O'Neil, 2016). Certainly, the Dutch cases remain an important example of a type of practical application. Although the exact inner workings of Top600 and Top400 are not disclosed, and are indeed usually obscure in most cases of algorithmic governance (see section 1.4), it is reasonable to believe they are *not* informed by the latest advancements of generative AI. However, a common statistics inference tool that draws on heaps of data to devise how to divide it efficiently into clusters or to identify the main variables that lead to a certain outcome (in this case, crime perpetration) is not much different, at least in principle, from a more advanced machine that learns how to read or respond to data without much human input. Indeed, data clustering is a “prime machine learning example of unsupervised learning” (Aradau & Blanke, 2017, p.13). And while the capacity and complexity of most recent AI ranks much higher, it is also still often at least partly based on more old-fashioned algorithms and statistics (Amoore, 2020). In general, it shares with common statistics tools characteristics of obscurity regarding the readability of its internal calculation by a human user, and of mathematical intuition (Bathaee, 2018). This work, therefore, although oriented towards slightly older algorithms, is also located in a symbolic entryway to the problematization of more advanced technology and AI that is also being discussed in the EU.

To return to Bartoletti’s lecture, I try to recapture through my work the connections she made me envision: that algorithms and AI are a new frontier of discrimination and a worthy and necessary locus of interest for contemporary feminism. I also return to what we could consider a feminist tradition of interdisciplinarity, as a way of constructing new perspectives and branches by crossing different disciplines and fields of knowledge (Lykke, 2011). My aim with this thesis is therefore to contribute to the understanding of algorithmic injustice and its relatively novel entanglements with feminist critique, especially in relation to claims of objectivity and neutrality of systems characterized by criteria for classification and assessment, which display discriminatory roots and outcomes. My work will also serve the purpose of clarifying some of the basic causes and manifestation of bias and discrimination in every algorithmic system, regardless of its level of automation, opacity and generativity, including more complex AI. In practice, my research questions are: in what ways do the tools of algorithmic profiling of potential criminal behaviour and predictive policing, Top600 and Top400, contribute to or produce social marginalisation in the Netherlands, and are there subjects and groups which are more impacted? What are the practices and structures involved, how do they work and what are their relations with human actors and with other tools themselves? What biases and distortions can be identified which specifically inform algorithmic systems and/or are amplified by them?

In the next section, I discuss the main theoretical feminist frameworks that have informed these research questions and guided my analysis. I explain my theoretically-based methodology, which crosses disciplines as well, in this case to critical legal studies, in a subsequent section. I then offer an overview of literature more specifically concerned with algorithmic justice and predictive policing, and conclude the chapter with a detailed discussion of positionality and ethical reflections. The second chapter describes and analyses Top600 and Top400 in relation to these theories, and the final one summarises the main answers to the research questions, offering further ethical and research-related considerations.

1.2. Main theoretical frameworks

My main framework consists of connections between and problematization of neoliberalism, governmentality, structural power, the alleged neutrality of science, phenomenology of misfitting and paternalistic views of children. Neoliberalism is intended as a condition where there has been a spill-over and extension of free market and capitalistic values and logic to most institutions and aspects of society: rationality has come to correspond to competitive economic sustainability, and profitability has come to dictate more than political and democratic morals and values (Brown, 2003). Brown's concept of neoliberalism builds upon Foucauldian notions of governmentality, which expresses the substitution of ancient sovereignty with a regime of power aimed less at an absolute, arbitrary control over subjects, and more towards an optimization of the wealth derived from the population. This is obtained through policing, surveillance and other forms of discipline which can be violent, but mostly try to understand and redirect society towards desirable conditions (Foucault, 1991/1978), and to favour certain groups of the population among others, or to discipline bodies and individual lives in specific directions with biopolitics and anatomopolitics (Foucault, 1976). In this endeavour, statistics and economics rose to serve the purpose of understanding phenomena which are seen as essentially *intrinsic* to the dynamics of populations (like epidemics, hunger, health, financial crises, etc.): phenomena which are, in a sense, not to be forcefully controlled but rather directed as much as possible towards preferable outcomes (Foucault, 1991/1978). Statistics and economics, it could be added, are constituted as fields of knowledge and power, both showing ideal courses of action and also constraining them within criteria of feasibility, profitability and probability. Modern algorithms in governance which deal with huge heaps of data remain founded on similar notions of statistics to detect both patterns and anomalies which are considered intrinsic in recorded phenomena (Aradau & Blanke, 2017), derived from an early modern understanding of normality as most likely statistical occurrence and as morally desirable normativity and classification among a distribution curve (p.8-10). In true neoliberal fashion, modern algorithms also seem to offer cost-effective solutions to the large-scale management of populations – an attractive factor in terms of profitability (O'Neil, 2016).

Such an opening could suggest scenarios where regimes of control have perfected the techno-instrumentalization of heaps of data and computational power towards absolute, inescapable governance of multitudes, with perfectly calculated designs. While my work has no intention of denying the pervasiveness and subtlety of many forms of discipline and governance, the existence of a clear predetermined direction or design, or at least its legibility through the analysis of governance via algorithms, and the clear correspondences of these supposed designs with instances of racism and other forms of discrimination, are much more arguable. This is not to undermine the systemic harm imposed to certain social groups, nor to deny that some truly powerful social players act out of hate or intent to damage; but the linearity of such intentions and outcomes is questioned. Instead, something closer to a politics of "muddling through", constituted of tentative political steps and attempts suggesting a certain direction, which are then re-thought or even reversed by subsequent political formations and contingencies, against an idea of "synoptic mastery" of policies (De Zwart, 2012), has informed this work. Algorithms can be understood as some among the many techno-political and administrative formations that take part into these uneven processes. As De Zwart observes, not all administrative frameworks, whatever their nature, are reversible (p.305-307): systems of classification, categorization and policy control of populations can be easily turned to rather harmful employments once political winds change, even when generated with good intent. Similarly, a reflection on structural injustice seems fitting: on how power imbalances get entrenched in institutions and result from the accumulation of micro-actions which are usually not blameworthy, nor the full responsibility or fault of single individuals (McKeown, 2021, p.5; Kasirzadeh, 2022, p.6). In a Foucauldian fashion, power is not held by isolated sovereign individuals, but rather dispersed, contested and reproduced in administrative, knowledge-related and several other social structures. Algorithms present themselves as instruments of categorization and decision-making that are similar to other administrative formations: they direct micro-actions or partial data towards final outputs that can perpetuate injustice, and their different touchpoints with human actors or data sources are difficult to frame under the lens of guilt and blame, despite their negative outcomes. Further, most algorithms are *literally* structures with blurred delimitations where the decision-making and the final outcomes happen diffusely, i.e. through a process where single deciding units and responsibilities are usually impossible to locate in a single node or connection (Bathaee, 2018; Amoore, 2020). Such structures can end up perpetuating discriminatory outcomes far beyond the intentions, awareness or fault of its several multi-located actors or even its designers; due to multiple sources of codes and data, they also escape notions of single, unequivocally identifiable authorship (Amoore, p.102-105).

A fundamental source of inspiration for this study has also been the relentless contribution of feminist studies towards a criticism of the alleged neutrality of science, starting from Haraway's contestation of an objective knowledge that is actually dependent on more or less self-aware interests and desires of control derived from specific positions of observation (Haraway, 1988). Others, like Schiebinger (1993), have deconstructed the taxonomies and categorisations of life that emerged since around the 1700s, tracing how inarticulate notions of gender, race, morality, etc. informed even what was supposed to be the mere observation and classification of nature, and still survive in many ways in contemporary science. Works like those of Fausto-Sterling (2008) offer an example of how even well-meaning knowledge production – in her case, medical studies which suggested special attention to black and brown people in terms of vulnerability to specific diseases and deficiencies – can hardly be disentangled from theoretically unsound, confusing or arbitrary notions of race that get perpetuated within the architecture of knowledge production, and can highly yet covertly mislead the understanding of data and results. Similarly, Clarke et al. (2003) discuss biomedicalization as a phenomenon where group and individual identities are re-constituted by a shift in the medical treatment and understanding of life, health and sickness due to new technological practices of data sharing and reorganization of data banks in diagnosis.

I also derive from these contributions a reminder of how pathologization, criminalization and other forms of othering have often gone hand in hand. Traits and characteristics of bodies and behaviours have been produced as moral faults, or as dangerous tendencies even outside the intentions of the individuals, and have been naturalized through supposedly objective categorizations of human nature. For my analysis of the impactful othering effected by some of the criteria of evaluation that algorithms employ, I further use Garland-Thomson's (2011) proposal of misfitting as a shifting, but noticeable, lack of alignment with the normativity that society requires from individuals. She links it to physical abilities, while I also refer it to broader notions of race, whiteness and behaviour that "sticks out" when showing up, in the terms of Ahmed's phenomenology (2007). Indeed, privilege sometimes manifests in the form of not being noticed for one's own difference (Ahmed, 2016). Expanding these concepts, I advance the understanding of Top600 and Top400 in relation with supposedly softer nuisances and gentler correction: the targeting not only of high impact crimes, but of what in the design stage was considered anti-social behaviour in young persons (e.g. disrupting certain settings, through absenteeism or loitering). The paternalistic sides of discipline (Young, 2003), specifically when it comes to the minors and young men affected by Top600 and Top400, that critical age studies offer in depicting the shift of children between threats, learners and vulnerable subjects (Thorne, 1987; Moosa-Mitha, 2005) also inform such reflections. I will argue that an ambiguous adult-centred view of young perpetrators and of their vulnerable young acquaintances fully emerges, requiring a submission to standards much wider than those merely prescribing non-criminal behaviour.

1.3. Methodology and materials

I have constructed my interdisciplinary methodology by starting from the acknowledgement that algorithms at their core operate with sets of instructions and commands that dispose and delimit what actions they will perform, what type of data they will use and output, and even how they will classify said data. Rather logically, this leads to interrogate, even outside the inner workings of the algorithms (which are, as better discussed in 1.4, often inaccessible), the terms through which data and people are pre-disposed, interpreted and generated according to official statements and definitions of said tools and processes. Attempting to assess the grievances of such tools and policies at a systemic level motivates an exploration of sociology of law. In particular, of the tradition of critical legal studies, developed primarily in the USA in the 70s and 80s, and connected in many ways to feminist legal critique, critical race studies and other struggles for resistance of the marginalised. This tradition is credited, among other things, with having offered several arguments to rethink external influences, privileges and powers that affect law and justice, with tracing a history of how laws reflect the prevailing interests and interpretations of privileged groups, and with deconstructing centuries of long-held principles and practices or claims of scientific objectivity (see Rahman, 2018). Engaging with the heritage of critical legal studies also means that, despite a few recognized methods of work, arguably no strict guide or step-by-step methodology is presented, an attribute which has often been moved as a critique to the whole sub-field (p.23). However, and perhaps in connection with the characteristic pointed out by Rahman, Sokhi-Bulley (2013) mentions how in critical legal studies methodology can be fairly seen "as a way of thinking, an *attitude* – it is therefore not so much about theory or method but about an approach, a perspective, or a lens through which to see a project" (p.12). According to her, a methodology could be developed from centring, for example, the concept of Foucauldian governmentality as a "'creative' concept and a creative methodology – meaning it is a flexible, open-ended and above all *useful* tool", which "helps us to challenge typical conceptions of government as implying state power, sovereignty or hierarchy and examining instead the various technologies and tactics, the often mundane processes, through which power circulates in a heterarchical fashion", for example by questioning technologies of governmentality (p.15-16).

Because of insights much similar to the ones presented by Sokhi-Bulley on Foucault, my work is also considerably informed by Dean Spade's inspiring research on the symbolic advancements and actual harms of equality laws and criminalization of hate crimes in transgender activism in the USA (Spade, 2015). Spade is concerned with the effects that laws and dispositions have on trans people: the ways that individuals and populations are identified, segregated and disciplined along a gendered-racial axis that contributes to and is maintained by purportedly neutral and normalised aspects of bureaucracy and administration that impacts them negatively, but are in no way addressed by anti-discrimination laws (Chapter 2-4). Whereas hate crime and anti-discrimination laws narrate transphobia as an individualised and circumscribed *oppression* operated by irrationally hateful and easily identifiable culprits, trans lives are actually made almost impossible by a constant *subjection*, i.e. by being produced as specific subjects of disadvantaged liveability. This production of disadvantage happens through different, subtle structural disciplinary moments across different fields of their lives, from apparently neutral laws, to economic production, to allocation of resources and security, etc. dispersed in regimes of practice, recording, notation, bureaucracy, administration (p.4-6). Subjection also helps locate power and phobia not only in the hands of a few privileged or a few laws explicitly regulating the marginalised groups, but rather at seeing several moments of discipline as places of negative impact and addresses of more necessary resistance. For example, in the current era of increased surveillance justified by the War on Terror in the USA, trans people who cannot rectify their names and gender on identity documents due to all the different requirements and barriers enforced in different institutional agencies (there are different regulations in place on the matter in the agencies dealing with driving licenses, employment, health-care, etc., and additional differences across the 50 States), are more at risk of being outed as trans on the job, of being flagged as ineligible for several services, of losing their driving license, etc. (Chapter 4). Yet, considerations about gender, ethnicity, etc. as identifying protected categories are more and more avoided in laws that do not deal with individualised hate crime or discrimination, as if the status quo on which these laws are grounded should be considered fair to all social groups (p.42-43).

I espouse Spade's invitation to generalise the struggle of trans people to those of other groups whose lives are also produced as practically impossible or unfairly harsh despite egalitarian promises, along similar or intersecting axes of vulnerability, and often in connection with criminalization (p.12). Although my work is neither centred on the USA, nor on trans resistance, Spade's insights are pivotal in the construction of an interdisciplinary methodology that builds on critical approaches to the law. If his work can be seen as composed of two main movements – mostly, denouncing the ineffectiveness of equality laws, and analysing other regulations that have even wider impacts on trans lives, even when they do not address gender – my own work can be located in the aspect of a precise focuses on the deconstruction and assessment of those measures and structures that produce those very same unfair impacts and marginalisation. Spade's identifies this as a "distributive analysis" (p.12), which questions the declarations of laws and policies in regard to their fairness, to rather assess their effect instead of their purported intention (p.8-10). In my work, I therefore engage with statements, dispositions and descriptions of Top600 and Top400 by the official channels of the agencies employing them, which are not laws, but can be seen as binding in terms of the promised behaviour of the involved agents and systems.

In more practical terms, I have retrieved the main material in a quite linear way, by treating the website of the Amsterdam municipality (www.amsterdam.nl) as the main source of official information on the tools for citizens. A simple query for the names of the approaches points to a page listing the Regeling Gegevensverwerking *[Regulation of Data Processing]* documents for each approach (see Actiecentrum Veiligheid en Zorg *[Security and Care Action Centre]* – from now on, AcVZ – 2020 and 2021). The same query offers an article celebrating the 10 years' anniversary of Top600, which I have occasionally included in my arguments as it constructs another official report of events and remarks about the creation and evolution of the tool (Gemeente Amsterdam, 2021), as well as the only official review on the outcomes of Top600 (AcVZ, 2019). I deemed these documents and article satisfactory as by explaining the regulation of the data processing of individuals they identify the ways in which the lists are compiled, the inflow and outflow from the list, the official treatment of targeted individuals, as well as the most authoritative narration at least on Top600 (similar official articles and reports on the Top400, which is after all considered a derivation of the first, could not be found). In a way, this choice of material also serves the purpose of developing, through this work, an eye for being alerted to the potentially discriminatory outcomes of regulations *already* at the level of their declarations, even when they may include benign statements of fairness or, in this case, promises of rehabilitation. I develop an understanding of their impact both through the premises of their statements and the existing literature that already researches their negative bias and outcomes. Since my understanding of Dutch is limited, I aid myself with translation software (I return to my research limits in 1.5). Occasionally, I also take into account documents which are not accessible to me, but that other scholars have recorded: a letter received by targets of Top400 explaining their likelihood of committing future crimes (Fair Trials, 2021, p.13-14), as well as internal communication concerning the design of Top400, which has only been released after formal enquiries with the respective agencies (see Jansen, 2022). These documents are already presented in English by said scholars. While my work is not exactly a new empirical addition (although the policy documents I take into account are more updated), I contribute to the existing literature by offering a more overseeable account of the tools and of the relations between them and practices themselves, as well as a more focused understanding of the structural marginalisation and related responsibilities.

1.4. Research field overview

In this section I offer a quick review of critical feminist research on algorithms/AI and on predictive policing: both are a relatively recent branch of study. More and more voices, even outside the field, have started denouncing the potentially harmful impacts of the employment of algorithmic tools. These include: the temporary exclusion of female candidates by an algorithm that Amazon employed for its hiring processes (Browne, 2023, p.332-334); language-learning tools perpetuating biases or mirroring racist slurs derived from the discourse fed to them as training data; privacy violation, excessive surveillance and flagging in the anomaly detection of security programs (Aradau & Blanke, 2017); automated systems excluding teachers from hiring due to obscure decisions, likely manipulated by falsified information, which could not be appealed against (O'Neil, 2016); face recognition software reinstating social stereotypes and ancient physiognomic essentialization (Agüera y Arcas et al., 2023), etc. Others also point to the dark material pre-conditions of algorithms usage: for example, the exploitation of Third World Countries, where underpaid workers are hired to classify the content of heaps of data and images (including gruesome ones) in order for them to be readable as training data for machine learning (Bartoletti, 2020; Amoore, 2020).

Some are extremely critical of algorithmic tools, denouncing how their usage can lead to damages of mass proportions due to their scalability (O'Neil, 2016) while others focus in particular on the aggravating effects they have on the poor, observing how they are used to discipline and repress those of a lower socio-economic status, whose difficulties are essentially blamed on themselves (Eubanks, 2018). Critical reviews tend to discourage the usage of basically any predictive policing tool in Europe due to their biased outcomes (Fair Trials, 2021).

There are common biases of algorithmic usage, or negative peculiarities linked to their interactions with human users, which are highlighted. Users tend to rely on the suggestions of technical tools, even when they are supposed to merely be *assisted* in decision-making, and end up actually training their human intuition on that of the machine (Eubanks, p.141-142). The insertion of humans in the process of algorithmic decision-making as a way to counteract algorithmic unfairness thus may seem a moot solution. Further, algorithms tend to over-represent certain profiles or overestimate probabilities simply because of the way data is read (and usually, misunderstood at first glance). They will, for example, consider racialized people from poorer neighbourhoods more likely to commit crimes because they already display several contacts with the police (police contacts that are higher than other groups of the population *due* to them being racialized and from poorer neighbourhoods in the first place; see Fair Trials, 2021). Reading recorded contact with police as crime representation is already problematic (police contact does *not* correspond to culpability, nor does police focus represent the overall of crime in a society, considering how much of criminality remain undetected and unrecorded). Additionally, the reliance on past performances to allegedly predict the future ends up reinforcing the status quo in several ways, from narrations to relocation of resources, and it also persuades algorithm users that their predictions are correct, since such predictions *produce* a reality that is skewed in such a direction (O'Neil, 2016, p. 8; p.27). The systems keep receiving confirming data about their predictions, and thus keep considering the factors they used as relevant and correct.

Because of their appearance of scientific rigor, even if they supposedly just express a likelihood, a probability, algorithms are almost impossible to appeal against (O'Neil, 2016). Additionally, the sources of their decisions are difficult to narrow down to specific data or nodes in their reasoning network (Bathaee, 2018; Amoore, 2020). Browne suggests the concept of "untraceability" to identify this characteristic, including the fact that sometimes, seemingly sexist, racist, or whatever discriminatory outcomes, are obtained *despite* the intentional elimination of data about gender, race, and whatever explicit sensitive characteristic was tentatively neutralised for computation. On the surface, the source of such bias cannot be identified. She also refers to "thoughtlessness" in evaluating different, unquestioned parameters of categorization that algorithms simply treat as data, which then reinstate structural injustices (Browne, 2023, p.333-337). Amoore problematizes the actual overall reasoning of complex algorithms: they may be presented as ultimately rational, but the relentless way in which they process data without leniency or understanding of the individuals connected to them, but only of the related measurable factors and probabilities, would arguably be categorised as madness by human standards (Amoore, Chapter 4). This may also be the right place to offer some reflections on the ambiguity that the words 'bias' and 'discrimination' can entail in algorithmic fairness, at least in the measures in which such words are attributed to a 'thinking' subject. While it *is* useful to contextualise algorithms in comparison with madness to judge their human impacts, it may be misleading to speak of 'thoughts' and 'reasoning' for a machine, because of the specific human coordinates and characteristics we may attach to thinking and reason. In other words, imagining a machine like a thinking subject in these terms may darken rather than clarify. Hayles (2023) borrows from biosemanticians the concept of ecological reciprocity, according to which the response of an organism (or, in this case, of a machine) to its environments, and the information coming from it, produces a form of "meaning making", i.e. a reaction that is relevant to the organisms. We may add the example of how, however different their faculties of perception (for example, even in absence of eyes and perhaps of sight at all), and however inscrutable to us some of their inner workings (especially in comparison with thought, consciousness or feeling), plants know how to react to light, air, temperature, etc. Similarly, although programmed with human-given basic instructions, machines engage with and react to data, including human language, in ways that make sense to *them*, which do not correspond to human ones (p.14), but that humans try to harness and translate in usable ways. To return to Browne's argument about untraceability and structural injustice, when it comes to bias and discrimination in these cases, we are not talking of an entity that is enacting a prejudice or is purposely hateful (unless programmed this way). Most often, we instead talk of the consequences of partial and non-neutral data which produces harmful effects when cumulated, with no perception of a race-, ethnicity-or other- related issues until the outcome is effected. It is only following the outcome that it becomes evident that the results seem skewed against certain groups.

Such untraceability derives in no small part from the black box problem as well: the fact that decisional neural networks are mathematically so advanced that they visualise and process information in ways that cannot be fully understood by human minds (Batahee, 2018), as well as the fact that algorithms are protected by corporate interests or obscured to protect their functionality (knowing how the systems tick could allow conscious manipulation by users or targets; see O'Neil, 2016). It could be argued whether crime detection, which is aimed at people who should be even *more* prone to cheating the system, but also deals with actions which are supposed to be clearly and strictly defined by law, should be more transparent than other algorithmic applications, and what it means both for its effectiveness and for the democratic tenure of a society when it is not. But the general problem of opacity and untraceability is even more cutting – one that we can derive from what Amoore (2020) describes as the illusion of fixing complex social or political issues which would require nuanced results with single, definitive outputs. Whether advertently or not, algorithms play to the dream of finding an exact, quantitative, unequivocal solution to complex problems, when in reality they effectively "reconfigure the site of manipulation" in political conflicts towards data processing (boyd, 2023, p.238), they turn politics into management and efficiency (Eubanks, p. 197-198) and into detecting the 'natural' anomalies, aberrations and patterns that supposedly hold a key to human behaviour (Aradau & Blanke, 2017). Additionally, computations are based on data that, despite being deemed objective, *always* derives from a form of selection or simplification of reality and society (Browne, 2023). In truth, they are programs set to find a solution that will be deemed desirable through tweaking and adjusting of results, and through the arbitrary decision of drawing a line and requesting an output in what can be considered unfinished processes overall. They do not represent an entirely objective process lead to equally objective consequences, and their designers, even when they are not fully aware of it, are all but uninterested in obtaining specific outcomes (O'Neil, 2016; Amoore, 2020). When the interests *are* conscious and known, they may not be clear or transparent to users or targets of the algorithms: for example, algorithmic management of work schedules aimed at improving work may not serve the well-being of the staff members, but rather the profitability of companies through a more efficient (but less human-concerned) allocation of hours and resources, while eliminating the margin of negotiation in the conflict between profitability and workers' well-being that was inherent in previous forms of scheduling (O'Neil, p.128-130).

In alternative to usually unsatisfactory suggestions about redistributing better outcomes or alleviating bias (Kasirzadeh, 2022), solutions or pre-emptive measures to algorithmic harms have been devised within a fault-based system, where designers, users or companies employing algorithms are to be made accountable for unfair usage according to the usual proceedings of law for illicit behaviours. Others contend that, when human culprits are not identifiable, the source of algorithmic harms can still be pin-pointed to the way data is produced. However, it is important to reaffirm how this thesis employs a much more structural and non-linear understanding of the unfair outcomes of algorithms. In other words, on one hand I agree that there is a general, systemic issue in which normative ways of assessing reality and human behaviours and performances inform over-simplified and already-biased data that turns out to be damaging, as pointed by Browne (2023). I also agree with the speculation that in some cases actual, intentional, individualised hateful intents exist in some of the tools (Jansen, 2022; Fair Trials, 2021), or that the problematic legacy of Dutch xenophobia and colonialism (see e.g. Weiner, 2014; Roggeband & Van der Haar, 2018) may be lodged within them. Yet, narrowing down chunks of data of code as responsible for discrimination usually proves impossible (Browne), and so does the identification of an ultimate author of an algorithm, or of someone who is exactly able to pilot their outcomes (Amoore, 2020, Chapter 3). Blame-based punishments seem therefore ineffectual or inappropriate (Browne). Similarly, suggestions like those of "Hippocratic oaths" of sorts for algorithm designers, making them swear by a moral principle of making sure as best as they can that their machines produce no harm (O'Neil, p.205), may be insufficient. Governments and powerful social players or corporations surely must problematize the usage they make of such tools at a systemic level, but finding single culprits among technicians or in data sets may be an impossible task. In my work, I consider much more inspiring those propositions where plenty of feedback (as data, and not only) from different actors and from interested parties is given (O'Neil, p.209; Browne). In alignment with Amoore (2020, p.25), I suggest that in a metaphorical trial of law, the ensemble of those who should be consulted about algorithmic justice would be "a crowded court", a diverse group referring to the multiple responsibilities and opportunities of intervention that are possible and necessary to better judge and adjust algorithms.

As for the cases concerning predictive policing specifically, there are concrete examples of such 'untraceable' biased outcomes: of how, especially in the Dutch context, algorithms may avoid the usage of data about race and ethnicity, yet still end up targeting those social groups. Sometimes such outcomes seem to be impossibly developed from seemingly race-blind data. It could be possible in such cases to take into account the existence of proxies for race and immigration background – a term which refers to information that is used to estimate the distribution or identify the individuals which belong to a certain group (for example, when combined, information about lower economic status, education level, neighbourhood, can all be used as proxies for immigration background as well as other-than-White ethnicity). The usage of proxy as a concept suggests a scenario where subtle identification of ethnicity *is* willingly sought out with a potentially racist intent – this may be hinted at by some of the private documents that Jansen (2022) has access to when reconstructing the design of the Top400 tool. But this is more of an exceptional case, whereas the inner workings of algorithms, the actual code and data, and of course the actual intent of the designers and users, are essentially unknown, as in a black box. Work on judging such tools has therefore a detective-like, speculative aspect which has to be accepted as one of its delimitations (and from a point of acumen, also a merit), but which nevertheless strives to detect and denounce their unfairness.

Aside from (ending up with) targeting groups without *explicitly* targeting them, such tools tend to eschew a few other legal protections. Almost all algorithmic tools used in scalable ways tend to actually target, among all disadvantaged groups, the poor: except that, unlike other vulnerable categories, these are not actually univocally identified and protected by law, which makes their discrimination much harder to challenge from a legal point of view (Eubanks, 2018, p.156). Furthermore, algorithms that still discriminate, without having an explicit set of instructions for targeting certain groups, actually function by aggregating and clustering data across lines which are not know to human users, or classifying data without needing a priori categories in their inputs. This makes even a formal recognition of unfairly discriminated targets which correspond to protected categories impossible (Aradau & Blanke, 2017, p.19-20; boyd, 2023, p.237). Lastly, predictive policing in particular, for example when it results in the worsening of life conditions of people, usually operates "beyond formal legal ideas of reasonable suspicion […] in a way that undermines these legal structures, formulations and the protections attached" (Fair Trials, 2021, p.32). This is exemplified by de Koning's (2017) concept of diffuse policing, which depicts a disciplinary power that does not always act through police forces, but blurs the lines between them and other institutions, and which I will adapt to a practical understanding of structural injustice. As the related disciplinary actions are often taken, in terms of informal punishment, *before* criminal acts have been committed and proven, because the algorithms suggest it, then presumed innocence is marred (Fair Trials, p.4), as well as the principles of chances at redemption and rehabilitation (Eubanks, p.187).

Overall, Dutch predictive policing tools caused a solid literature that criticises them through much of the reasoning summarised here, as well as more specific reconstructions (e.g. de Koning, 2017; Fair Trials, 2021; Jansen, 2022), which I expand in combination with the theories seen in 1.2.

1.5. Reflections on delimitations, positionality and ethics

In this section I reflect on the delimitations of this work, as well as my positionality from an ethical and epistemic point of view, in the attempt to outline the opportunities and limitations involved. I also bring to the front some of my own personal characteristics as a source of further reflexivity.

One limitation that comes as an obvious consequence of the presentation of the material and the research field overview of the previous paragraph comes from the lack of contribution of the people targeted and marginalised by these tools. Some of the investigations benefitted from interviews and field observations from samples of affected individuals (Fair Trials, 2021; de Koning, 2017), however this approach is not particularly present in the overall literature. My work still contributes to the understanding of such tools, but any further investigation and attempts at improving their situations should have a much bigger input from the interested parties.

This thesis has also been written under the knowledge that probably none of the people affected by these tools will read it; the knowledge that most likely, the readers it will engage with are quite far removed from the socio-economically disadvantaged and criminalised individuals that it covers; and the awareness of how much these asymmetries between readers, writers and objects of study are normalised. It is concerning to reflect on how much academic careers on marginalised groups could be founded on and distanced from such groups they analyse; and how there can still be problematic distances where academics and intellectuals essentially profit from them with limited mutual positive impacts. In this sense, and in connection with violent criminality, I am once again reminded of critical trans studies and of how voices like Riley Snorton and Haritaworn (2013) have argued for the "value extracted from trans of color death" (p.67). They discuss "how trans women of color act as resources – both literally and metaphorically – for the articulation and visibility of a more privileged transgender subject" (p.71), in the way that the experiences, abuse and deaths of some of the most othered subjects are utilised and co-opted narratively and politically for the social benefit of more privileged members of the LGTBQ+ community. They offer a compelling and specific example of something that should constitute a concern of every social science with ties to ethnography and its colonial past: of who benefits from the research and the extraction of information from people's lives, and to what purpose it is produced and employed (see e.g. Alonso Bejarano et al., 2019, Introduction), and of how there are different positions of privilege and advantage even in similar or identical marginalised groups. An insight that makes promoting the well-being and emancipation of different groups problematic and often conflictual, as evidenced for example in the seminal work of Crenshaw (1991). And while the scope of this work has been limited to literary sources rather than field work, and tries to describe and understand a problem, without a bigger focus on improvement of the situation, any reading of it must keep in mind the underlying question: would any revision of algorithms derived from this work be actually able to positively impact marginalised lives?

Another important delimitation comes from my limited understanding of the Dutch language when engaging with the official documents and texts that concern Dutch municipalities and related tools (whose selection I detailed in 1.3). I have aided my work with translation tools (mostly DeepL and Google Translate – https://www.deepl.com/translator and https://translate.google.com/), and by referring to a body of research in English by authors located in the Dutch context, who at times also provide their own informed translations of official or internal documents. Similarly, my acquaintance with Dutch culture and even the physical locations of the phenomena I treat is limited. Nevertheless, as mentioned in 1.1, despite the obvious barriers, I argue that the Netherlands are an extremely interesting focus for this topic of research, as it represents one if not the most advanced example of digitalisation of public security and policing in Europe. As other countries in the EU may set out to increase their digitalisation, and as the finally approved EU AI Act undergoes a slow implementation, the Dutch experience could be used as a source of discussion, if not as a pilot and point of reference for further developments elsewhere – hypothetically, even in my current and already very digitalised country of residence, Sweden. Indeed, and of relevance to the creation of these tools, the idea of experimental procedure, i.e. trial and error at the expense of populations, appropriately signifies the willingness of institutional bureaucracy to experiment tools of control on some social groups (Jansen, 2022, p.30-32), especially when such social groups are already vulnerable and marginalised, their oppression is harder to defend and the scalability of their management is seen as particularly attractive (O'Neil, 2016, p.8). It may very well be then, that in order to understand further developments in contexts which are more familiar to them, researchers should try to engage with contexts unfamiliar.

I have been trying to find the advantages and insights that may be possible relatively to my specific position and background, despite my distance from the researched context. I have reflected on my inbetween-ness compared to a more fixed understanding of an outsider vs insider status within the researched matter, trying to embrace it rather as shifting positions in what Nash named a queering of research. In other words, there are fluid proximities of my experience with the object of research, and focusing on fixed understandings of identity, closeness and distance may foreclose the insight that could arise from it (Nash, 2010, p.141-142). For example, while it will be evident in the analytic chapter that most subjects racialized and othered by the algorithms are usually at least second generation immigrants, I have also experienced several forms of othering and racialisation as a first immigrant in Nordic countries in Europe – despite being considered white in my own country, Italy. While my movements and transfers across different countries have been marked by an essentially good socio-economic status, I have also been misfitting and sticking out in different situations because of behaviours and performances that not always matched the local expectations. This was not only due to clashes with a new culture – but it was also linked to chronic health conditions, and symptoms of what I thought was a cognitive disorder (another marker of many of the discriminated subjects by these tools), and other aspects which were essentially outside my control, but that others, as well as me, blamed on myself due to the main narratives I was immersed in. As I navigated an extremely impacting disease in my early twenties, and other chronic health conditions, I was particularly vulnerable to (I should say, thrown into despair by) the noisiness of my neighbourhood, as it deprived me of sleep I was already hardly getting due to my sickness. In such contexts, I often wished for more restrictive rules regarding noise and loitering – the same desires that, as we will see, apparently pushed for the creation of the Top600 approach in Amsterdam, which was also used to enforce actions against such smaller nuisances, viewed as not effectively tackled by the police. On paper, the rich neighbourhoods that requests stricter measures can be easily antagonised – but in my research I have tried to be mindful of the potential unrecorded fragilities of their inhabitants, too.

On a similar note, aside from my own experiences with mental health, I have reflected on my proximity with people who, either with a specific diagnosis or not, have also sought or expressed a need for mental care in one form or another. And while they have not experienced police persecution, I am also aware of their (and my) troubles fitting in several disciplined contexts of life, and of their tendency to being entangled in potentially harmful situations despite their desires of happiness. I am often reminded of the difficulties in maintaining closeness with some people struggling with mental health, and of how some discourses and narratives we are all immersed in can blame their conditions on them, or of how their choices and situations may push away others who may help them; I am aware of the general tendency towards blaming people who diverge and misfit due to how their mind works. I am concerned about the ease with which this could result in criminalization in cases of below-average IQ (a measurement which is not unproblematic in itself) and other material hardships in life, and of how the lines between responsibility, choice, fault and lack thereof become blurred and unreliable when it comes to subjects who do not fit the idealized rational and self-governing adult, even if they are impacted by outer conditions.

Across all these lines of vicinity and distance, I have also been guided by an empathetic notion of curiosity. In other words, while distance can be a challenge and must be problematized, I believe it to be an appropriate use of the privileges of research one that strives to improve the connections with something unfamiliar. Curiosity towards something deeply unknown becomes, in this case, an opportunity to nurture a more mindful understanding. It allows some progress into turning something alien into something closer, and to progress from estrangement to chances for solidarity. In this sense, I reclaim fully the distance from the individuals covered by my research precisely as I configure my research as an attempt to shorten that distance by developing a more understanding attitude.

Lastly, as for my credibility in approaching something also distant from me like a technical expertise on machine learning, while I cannot claim proficiency in it, I benefit from studies and work experiences in surveying, generation and usage of data and statistics both in a research and in a managerial perspective (especially between 2016-2018 and more recently in 2023). Relative insights – that come from my previous studies and experiences in economics and marketing – reflect an understanding of how data can be non-representative or requires considerable additional interpretation. They also inform how statistical software, through operations of clustering and factor analysis (what seems a purely mathematical aggregation and grouping of data) may be misunderstood as presenting their results as pure scientific certainties – the same ones that would blame the underperformances at work or in social or school life, both in my personal experiences and in the ones described in the literature on the Dutch algorithms.

In the following chapter I will present in a more thorough way Top600 and Top400, and I will assess, analyse and discuss their impact. A third and conclusive chapter will provide a summary of this thesis, as well as further reflections on limits and possibilities of improvement.

2. Analysis of the Top600 and Top400 / Veilig Alternatief approaches

2.1. Outline of the chapter

In this chapter I will discuss some of the algorithmic tools employed by Dutch police forces and institutions, mainly in the municipality of Amsterdam and Amstelland. I will focus in particular on the Top600 and Top400, the latter now known as Veilig Alternatief (which can be translated as *Safe Alternative*), two tools and approaches aimed at the reduction and prevention of crime, especially when it comes to young perpetrators or young potential perpetrators. They enact what has been called predictive policing, i.e. measures aimed at identifying individuals and areas that could be involved in crime *before* crime happens.

In the cases examined here, predictive policing takes the form of the screening of individuals that come into contact with, or are objects of interest of, police forces. The targets of these systems are profiled and their likelihood of committing (or re-committing) highly impacting crimes or public nuisances is assessed by looking at similarities and commonalities with actual perpetrators' profiles, and by detecting characteristics that are seen as correlated with crime and recidivism. It is also worth mentioning that the Netherlands employs several other tools for policing aside from Top600 and Top400, including ProKid+, another tool targeting young persons, which is undergoing a general re-design and that has contributed to the implementation of the Top400. Other noteworthy examples are the Sensing Project, piloted by Roermond police since 2018 to identify likely pickpockets and shoplifters who drive towards a local shopping area, and the Crime Anticipation System, a nation-wide tool in use since 2017 to predict crime rates in entire areas (Fair Trials, 2021).

Whenever possible, the materials I use for the analysis consist of official documents produced by responsible authorities on such tools. More specifically, I have mainly referred to the Regeling Gegevensverwerking *[Regulation of Data Processing]* documents for each approach (AcVZ, 2020 and 2021), as well as occasionally to the only official review on the outcomes of Top600 (AcVZ, 2019) and to an article celebrating the 10 years' anniversary of Top600, which represents another official record of events related to the creation and evolution of the tool (Gemeente Amsterdam, 2021). As mentioned in sections 1.3 and 1.5, my understanding of the Dutch language and culture is limited, and aided by translation software and the English production of scholars from the Netherlands. Still, the Dutch context has been considered particularly intriguing for a predictive policing analysis, as its social infrastructures are particularly advanced in terms of digitalisation, and as the tools themselves seem to function in a sort of ecosystem of explicit mutual interactions. A further barrier in analysing and collecting material about these tools is constituted by the obscurity which characterizes them: seldom, if ever, police corps divulge the details of their usage of such technology and its potential shortcomings; and the inner workings of the algorithms themselves are not fully readable by humans because of the mathematical complexity (as further discussed in section 2.6). Scholars and activists involved in such research take therefore on a rather investigative and speculative role.

To further clarify terms, while I have already argued for the affinity with Artificial Intelligence that many simpler or older algorithms and statistical tools display, specifically when it comes to said characteristics of mathematical unreadability and other forms of untraceability, it is assumed that the tools in focus here do not fully correspond to the most recent developments of AI. In lieu of this definition I will therefore use that of Automated Decision Making (ADM) tool (as suggested by Fair Trials, 2021). This does not mean that, either in their original design process, or during further reviews over time, these systems have not employed machine learning to advance their ability to compute data; however, ADM seems better fitting while also sufficiently open for identifying such tools.

I will firstly try to present extensively the two tools, by expanding in section 2.2 on their alleged main goals, on the criteria which determine who is included in their focus and what additional data is collected about them. In subsequent sections, I discuss some other critical themes connected to them, most of which have been outlined in the introductory chapter. Section 2.3 overviews some of the more evident criticism against Top600 and Top400, mainly their potential over-representation of specific social groups and identities and the problematization of their effectiveness. Section 2.4 briefly contextualises the seemingly racist police bias through the legacy of colonialism and xenophobia in Dutch culture and the politics of specific 'focus' and 'care', empowered by statistical collection, that have often been enacted on ethnic minorities. Section 2.5 discusses the neoliberal, governmental uses of such statistics for prediction and control and problematizes the several mistakes that can ensue in the interpretation of data and of the related results as neutral and objective. Section 2.6 further bridges the problematization of statistical thinking with the opacity and untraceability of algorithmic computation, and the specific relations that algorithms have with the different human actors involved. Section 2.7 explores the concept of diffuse policing, that represents a mix of surveillance and repression that takes place beyond encounters limited to those with police forces. It also expands the concept and combines it with phenomenology and subjection to account for all the different touch-points of data production, monitoring and discipline within the subjects' private and public lives, with the outcome of repressing and preventing 'misfitting' and 'anti-social' behaviours. Section 2.8 presents the implicit production of an ideal, well-behaving subject in contrast with the undesirable one outlined in the previous section, and further locates it in the moulding and discipline of children and minors and in their contradictory shifts between subjects of paternalistic protection and threats to security. Lastly, section 2.9 further expands the concept of diffusion, in this case to a confusion of legal responsibilities, and offers an overview of some of the ways in which, from a legal point of view, it is hard to seek recourse against such algorithmic tools, or of how they escape some legally recognised actions or malpractices.

2.2. Overview of the Top600 and Top400 approaches

The Top600 approach has been used by the Amsterdam municipality, police and social services since 2012, to identify the top 600 younger people (older than 16 and younger than 30, at least at the moment of inclusion in the list) with high chances of committing or re-committing major crimes in the future. As of the latest revision of the regulation of data processing regarding the tool, it is specified that list members tend to be young men (AcVZ – 2021). Top600 was introduced in connection with media sensations about the lack of swift policing in certain areas of Amsterdam (de Koning, 2017, p.540) and was strongly promoted by then-mayor Van der Laan and related Social Democrats policies (p.543) as well as, later, by mayor Femke Halsema, with her push towards an expansion of criteria towards drug-related crimes in 2019 (Gemeente Amsterdam *[Amsterdam Local Authority]*, 2021). As of AcVZ (2021), its main purpose is to serve public interest by reducing high impact crimes and drug-related crimes. The goals are to punish and repress more effectively such crimes; to offer perpetrators better life alternatives in lieu of recidivism; and to ensure that their siblings, which are estimated to have a higher chance of partaking in criminal activities just because of their relation to them, do not follow in their footsteps. The tool employs a so-called person-centered approach, by processing a plethora of information permeating what is recorded about the identity of the target of the approach in several aspects of their lives.

More precisely, there are two main groups of criteria for inclusion in the list: one related to so-called high impact crimes (crimes which are supposed to have bigger consequences, including a considerable unrest and sense of insecurity in society), which include robberies, street robberies, residential burglary, aggravated assault, murder and violence against persons, and one related to drug- and violence-related crimes, including drug possession and possession of weapons (AcVZ, 2021, p.3). Each main group has two sub-groups of criteria, one determined by the police and one by the Openbaar Ministerie (essentially the public prosecution). It is sufficient to meet *both* the police and public prosecution criteria for *just one* of the main groups (so either high impact or drug-violence related crimes) to be added to the list. In their entirety, the criteria are (see AcVZ, 2021, p.9-10):

For high impact crimes (HIC) (both the police and the public prosecution criteria must be met):

- Police criteria:
    - all persons arrested at least 3 times in the past 5 years on HIC
        - who have also been suspected for HIC at least 1 time in the past 2 years
        - who have also been arraigned to a prosecuting officer (OvJ, *Officer van justitie*) / examining magistrate (RC, *Rechter-commissaris*) for HIC at least 1 time in the past 5 years

    OR, alternatively:
    - all persons arrested at least 1 time in the past 5 years on HIC
        - who have also been suspected for HIC at least 1 time in the past 2 years
        - who have also been arraigned to a prosecuting officer (OvJ, *Officer van justitie*) / examining magistrate (RC, *Rechter-commissaris*) for HIC at least once in the past 5 years
        - who have been identified as suspects for an incident in the police registration system at least 33 times in the past 5 years

- Public Prosecution criteria:
    - all persons younger than 21 who have been convicted at least 1 time in the past 5 years with sentencing for HIC

    OR, alternatively:
    - all persons aged 21 or older who have been convicted at least 3 times in the past 5 years with sentencing, of which at least 1 time for HIC

For drug-violence related crimes (DG) (both the police and the public prosecution criteria must be met):

- Police criteria:
    - all persons younger than 30 the date the list is compiled
        - who have been arrested at least 1 time in the past 5 years for trafficking hard drugs
        - who have been arrested at least 1 time in the past 5 years for a violent crime, threatening or weapon possession

- Public prosecution criteria:
    - all persons sentenced (be it to custodial or community service) for a crime at least 2 times in the past 5 years
        - who also reflect even just 1 of the following criteria
            - having a sentence (for a specific section of the Opium Act, *Opiumwet*) without having indicated a dealer
            - having a custodial sentence for one of the articles in the Weapons and Ammunition Act (*Wet Wapens en Munitie*)
            - having a custodial sentence for a serious violent crime (even if an only attempted one)

A case director is assigned to each member of the list, to overlook the resulting profile and to devise an action plan that can put in collaboration more than 40 institutions and organisations which can serve as touchpoints and sources of discipline and change in the targeted person's behaviour. It is also up to the case director's discretion whether to utilise further data from the targeted person's life for other assessments or not. Members of the list spend on average 2.5 years on it (AcVZ, 2019) and the related policing can last up until 3 months after their removal (AcVZ, 2021, p.16). More precisely, the list is updated twice a year to make sure that profiles that do not meet the entry criteria anymore can flow out. However, even when the criteria of entry are not met anymore, list members will be kept in in case they have been arrested or detained under certain conditions (arrested for HIC or DG offenses in the last year; or, if they have been generally detained, in custody or imprisonment, for at least 100 cumulative days in the previous 2 years if they do not meet either the police or the public prosecutor criteria, or for at least 180 days if they do not meet both the police and public prosecutor criteria anymore).

People in contact with the list members are also potentially monitored, especially with the purpose of assessing the right approach to their household and minors linked to them (AcVZ, 2021, p.15). More exactly, linked people whose data can also be processed are family members and partners, co-perpetrators or other links which are already know to the AcVZ, people who intervene to their aid (including trainers, social workers, partners), and other acquaintances. That data gathered for such profiles, as well as the data that is processed for the list members too, varies, and includes ethnicity, religion, sexual and health data, housing, income, debts, criminal data, and other administrative identification coordinates. According to AcVZ (2019), in the retrieval of such data, alleged "criminogenic factors" which are taken into account also include the absence or low level of education of parents, and loitering.

We could summarise the Top600 like this: it compiles a list of members with a *top-down* approach, with an attempt to gather all profiles meeting the criminal and legal data reported above. It then moves from the collective to the individual and engages in a *horizontal* approach, where data regarding the list members and their contacts is used to further assess their case and determine courses of action for their specific situation.

The Top400 Approach was instead developed in cooperation with the Amsterdam municipality, police and social services since 2014, in light of the alleged success of Top600. Its function was to identify the top 400 younger people (younger than 16, as opposed to the original Top600, however the age of entry ranges from persons aged between 12 and 23), which a high likelihood of future involvement with any form of crime (not only major ones), and had the goal of preventing or reducing such involvement for them and for their siblings. The starting data was composed of top lists of the 'most troublesome children' suggested by neighbouring municipalities (Fair Trials, 2021): their names were drafted from the different districts, asking the local security forces what individuals they wanted to *make sure* that would be included in a focused programme, and analyses of similarities at a statistical level produced the criteria of access to the list for a more generalised version of the software (Jansen, 2022, p.17). Such criteria have changed often over time. At some point, they were also combined with ProKid+, another tool of assessment of future criminal behaviour by children. Similar to Top400, ProKid+ was created by looking for commonalities and relevant factors in a sample of children that were registered by police officers as suspected perpetrators, victims or witnesses, by extracting their public records in the years previous to their registration. The purpose at the creation of ProKid+ was to track, by subjecting this data to a regression analysis (a procedure which is discussed in section 2.6), factors that would predict their contact with the police, with a main difference being that the sample was selected at random from police archives (Wijentes et al., 2017, p.252-253). ProKid+ used to provide 4-tiers of risk classification, some of which were considered alarming and would require the child and/or their address to be registered in the police systems for monitoring and intervention (Fair Trials, 2021, p.8-10). In 2016, results from ProKid+ were used to draft an additional 125 minors to the Top400 list (which had never actually reached the amount of 400 and thus had a budget for further additions). The usage of ProKid+ within the Top400 approach was discontinued soon after, but the 125 minors remained in the Top400 list (Jansen, 2022, p.31-32).

Criteria for entry in this case are divided into 3 main groups: one determined by the police, one derived from signals from care-related institutions, and additional criteria. List entry is determined by being flagged cumulatively for all the criteria groups. According to AcVZ (2020, p.7-8) they are as follows:

- Police criteria:
    - persons apprehended as suspects at least 1 time in the past 5 years for one offense (including just attempted) among: robbery, street robbery, home burglary, aggravated assault, murder/manslaughter, violence against persons
    - who have been arrested at least 2 times in the past 2 years as suspect for a crime (including those in the previous point)

- Criteria derived from care-related signals:
    - persons who meet at least 3 of these requirements:
        - having a judicial order (this is worth 2 requirements out of 3)
        - having a civil action in progress or in the past
        - having a record under the Compulsory Education Act (*Leerplichtwet* – this means having had an official record drawn by a compulsory education officer due to school absenteeism. See Inspectie Veiligheid en Justitie – Security and Justice Inspectorate, 2016, p.11), being a student that stays at home unjustifiably or an absolute truant, or being a drop-out
        - having changed school at least 3 times in primary education or at least 2 times in secondary education
        - having been involved in a domestic violence incident at least 1 time in the past 5 years
        - having been apprehended while aged between 12 and 14 in the past 5 years
        - having been arrested in the past 2 years for dealing fake drugs
        - being part of a prioritized youth group

- Other criteria:
    - the persons must not be or have been in the Top600 list
    - and must not have been in the Top400 already

Members of the list receive a statement that explains they have been placed on it for a number of reasons, among which arrest records, having been witnesses or victims of domestic violence; school changes and absenteeism are also mentioned (Gemeente Amsterdam, n.d.). Fair Trials (2021, p.13-14) also records a version of this statement where it is explicitly said that facts and figures suggest that members of the list may go down a criminal path and that officers seek to collaborate with them to avert that. To this purpose, the municipality employs a personalised project of care, just like in the Top600: it results in collaboration and exchange of information between the national health system, police, social services and youth protection to devise plans of actions under the supervision of a case director.

The minimum stay in the list is at least 2 years: only at the end of that period the validity of their continued inclusion is tested and data entry outflows if the entry criteria are not met anymore. Members stay on the list if they have been apprehended as suspects for a crime at least 2 times in the past year, or if they have been cumulatively detained for 100 days in the past 2 years, but will necessarily flow out when they turn 26; list members may also be transferred instead in the Top600 list if they end up meeting the related criteria (AcVZ, 2020, p.13).

As for Top600, close links and acquaintances of list members will also be screened out in regard to their personal information, especially with the aim of identifying connected minors to prevent further damage to them (p.14).

To summarise then, similarly to the Top600 approach, the Top400 screens its members with a *top-down* movement, compiling all profiles corresponding to certain criteria, to then move *horizontally*, i.e. retrieving other relevant data in the environment and records pertaining the specific persons it targets.

2.3. First sources of concern

At first glance, it may seem like the two approaches should not deserve the criticism ascribed to them at the beginning of this work, i.e. that they target certain individuals unfairly. Both the tools clearly state that among the criteria for inclusion in the lists, it is necessary to have had multiple contacts with the police as suspects of, or formally accused for, some serious criminal offences. While the following sections will also try to upset and problematize what leads the police or the public prosecution to apprehend or target as presumed offenders certain individuals, this work does not intend to undermine the general idea that, in some cases, communities may be genuinely safer with the policing of some known perpetrators.

However, upon closer reading the mere overview of the approaches leads to some troubling questions. It is not so much that known criminals or criminally involved persons are in contact with the police – but what does it mean that, outside of the usual forms of detention or persecution, with which they have already had their contacts, they are also subjected to a tailored project of repression and rehabilitation? Why do this otherwise novel approaches, which lead to a joining of institutions and forces which usually do not act in such close connection with each other, target these individuals specifically? While the Top600 has entry criteria which are all clearly linked to criminal persecution, more glaringly concerning is the fact that the Top400, among known suspects and perpetrators with a certain rate of arrests and police contacts, will *single out* those whose only added fault may be that of having witnessed domestic violence, having skipped or changed schools too often, or being part of a group of special interests: how have these non-criminal entry criteria been determined? What does the retrieval of all the other personal data – among which information that is *not* necessary to verify the criteria for targeting – of the list members *and* of their acquaintances lead to? And can the approaches at least claim effectiveness towards its goals?

Activists and scholars who have researched these specific cases point out the remarkable lack of independent reviews of the two approaches (Wetenschappelijk Onderzoek- en Documentatiecentrum *[Scientific Research and Documentation Centre]* – from now on, WODc – 2023). It is mostly official channels that speak of their results, and they usually do so favourably, stating for example that the Top600 contributed to reducing the recidivism of the targeted individuals and crime rates in general (AcVZ, 2019). WODc however have argued that, despite determining considerable public expenses, in the case of the Top600, results in terms of recidivism and prevention cannot be unambiguously attributed to the approach (WODc, p.45). The literature regarding these and similar tools seem also to indicate that they lead to over-representation and over-criminalization of people with a migratory background (in particular, young men of Moroccan descent), as well as families and individuals in a poorer socio-economic bracket, or even homeless (p.17) and with mild-to-borderline intellectual disabilities or IQ scores sensibly below average (see Verheijen et al., 2022), with rippling effects for their entire neighbourhoods and districts. In essence, these tools are accused of reinforcing already-existing negative assumptions regarding these social groups, compounding their disadvantages and marginalization (Fair Trials, 2021; de Koning, 2017).

The consequences in terms of institutional behaviour towards the list members are also particularly concerning. Since it is known that list members have a history with crime, and in line with the explicit aim to repress more swiftly when necessary, they often become the target of what looks like mere police harassment. Indeed, institutional voices reclaim this form of "Very Irritating Policing", which is supposed to keep list members in check by disrupting and pestering their routines (de Koning, 2017, p.543). The term "irritation" is however perhaps too euphemistic and far removed from the actual violence, which does not take the form of physical repression but remains abusive and impactful. List members may end up enduring all sorts of pestering treatments with serious consequences on their activities, livelihoods and families. Police officers learn to recognise them and will, for example, call their names out loud in public, exposing them as subjects of interests for the law – which is intended as a way to prevent supposed illicit actions they may commit (Fair Trials, 2021, p.12). Targets will be often constantly followed and monitored, receiving frequent unannounced home visits as well. Some will also be involved in a plethora of other sanctions and informal but grave consequences, like having their new employers informed that they are included in the list, leading them to loose a job, or being banned by groups and membership, including for example the loss of the possibility to practice certain sports and social activities (p.11). Police action on even the smallest infringements by list members is encouraged. In addition, when there is a robbery or another HIC near a member of the list, they tend to be the first ones to be suspected by the police, leading to more arrests based on such suspicions. This will, in turn, feed more arrest records back into the systems, making some of them extend their stay in the lists, in a vicious circle of self-confirming predictions. When it comes to the Top400 approach, refusal of collaboration and to abide by the sanctions and "care" side of the approach will result in threats "to remove your children from home" (p.14).

In a way, these considerations do not stray far from an already existing concern, within gender studies and antiracist and anti-discriminatory scholarship and activism, towards structural violence and the racist and xenophobic bias of police forces. In the following section I delve on the possible connections of Top600 and Top400 with police racism and briefly contextualise other forms of racist bias and discourse in the Netherlands.

2.4. The vanishing acts and reappearances of race

Historical reconstructions and contextualisation of the Top600 and Top400 tools reveal what could be an at least racist or xenophobic matrix at their roots. A negative stereotype of specific groups of Dutch-Moroccan young men pre-existed the creation of the tools: police would write notes and records in their digital systems concerning some of them even in absence of any clear wrongdoing, which would also lead other officers to target them or to write more records once they had entries into the system (de Koning, 2017, p.537). Groups of Dutch-Moroccans were also the targets of xenophobic political discourse and subjects of public notoriety during the years of design of the tools – in particular of the xenophobic inputs of Van der Laan's party, the Mayor of Amsterdam at the time (p.542-543). A scrutiny of internal documents concerning the design of Top400 in particular, released only years after its initial deployment, also suggests that the over-representation of Dutch-Moroccan and Surinamese people was detected, but ignored as a problem. Rather, the thresholds for inclusion were tweaked in the first years to make sure to include more problematic young persons, even those related to merely "lighter" nuisances. The number of 400 list members was seen as a goal to reach (in its first year, Top400 included less than 400 people), and a wiser way to spend a budget that had now been allocated anyway (Jansen, 2022, p.12, p.18). Even a new leftist government proceeded to utilise these new infrastructures and pushed for their expansion. Street coaches which were part of the reporting and observing bodies named to identify new potential records for the tool data were concerned with reporting even just noise, littering, football games, night-time disturbances, loitering (p.24-25). Here, I would suggest that the target was what gave a *perception* of degradation of the neighbourhood, rather than more profound material problems.

Generally speaking, several scholars highlight how racism and xenophobia take frequent forms in the Netherlands. On one hand, just like in many European countries who strayed away from explicit collection of data about race after the problematic memories of World War II (Simon, 2017), in the Netherlands, where some sort of ethnic statistics is nevertheless collected (De Zwart, 2012), racial discourses and reflections have largely disappeared (Weiner, 2014). More precisely, together with a discomfort about race in the aftermath of Jewish persecution and targeting of other ethnicities, the historical heritage of colonialism and its embeddedness with racism have also performed a sort of vanishing act from collective memory (p.736-738). In this way, however, the exploitation of people and resources which determined the cumulated privileges of some social groups, and disadvantages of others are also made to conveniently disappear, dissociating them from a responsibility for contemporary inequality and depoliticizing them (p.737). Far after the times of slavery, such erasure also conceals the convenient bouts of immigration for cheap labour, and subsequent family reunification, that happened in the Netherlands since the 60s. Racism and xenophobia then emerge more subtly in other forms: while "Europeans" transfer their perception of superiority to the alleged moral and ethical markers of their work culture and of their institutions as "the zenith of civilization" (p.738), minorities and other racialized groups which are marginalised due to the intricacies of colonial heritage or difficult integration are blamed for their socioeconomic status (p.737). Class difference is presented as an individual responsibility against and allegedly neutral and favourable status quo, a characteristic of most colour-blind perspectives (as in Spade, 2016, p.58-62). Poverty or disadvantage are presented as a result of group failings rather than structural and societal ones, and are attributed to an essentialized "backward cultural traditionalism" (Roggeband & Van der Haar, 2018, p.80) of certain groups.

In the complex relation with and attempts of multiculturalism over the decades, as summarised by De Zwart (2012), Dutch governments and public discourse show a tendency to deal with minorities with strategic redistribution of resources for quick pacification rather than deeper address of rooted issues. The first historical precedent for such an approach would be the responses to the emergency determined by the violent tensions between Moluccan minorities and Dutch government in the 70s, which ended with a mix of initiatives to preserve their culture (e.g. via schooling in their language) and aim for assimilation (p.304-305). At the time, the Dutch top-down approach to determining ethnicities became particularly useful for this sort of policy-making. Within the margins of Dutch constitutional law, which prohibits the collection of data about race on grounds of discrimination, the collection of everyone's age, place of birth and parents' place of birth was used to determine some ethno-statistics. In other words, individuals cannot choose how to identify in a census, but data is automatically collected on their immigration background to identify so-called 'allochthones'. Their relative numbers have been used to guide reallocation strategies similar to the one employed for the Moluccans, with integration essentially becoming not a matter of cultural relativism but a "benchmark of statistical facts" and of efficient design of policies aimed at solving the problems revealed by differences in ethnic statistics (p.309). I use ethnicity and race in strict relation here because I agree with de Koning about how, despite the problematic nature of the concept of race and the preference of Dutch public discourse for the term ethnicity (due to that discomfort with race-related discourse), the generalisations and ascriptions of identities from above remain a phenomenon of *racism* (de Koning, 2017, p.539). Roth's extensive discussion of the multiple dimensions of race enlightens the inconsistencies and arbitrary nature of the different racial identifications (Roth, 2016, p.1327). I do not consider this a lack of scientific nature in the concept of race since it is evident that scientific discourse and instruments are central in the production of new, partial forms of racial identification with a veneer of objectivity (e.g. racial ancestry, p.1319-1320, or genetic investigation, p.1328). Dutch ethno-statistics do precisely the same: through apparently objective tracking of lineages and ancestry, which may be function as a rough proxy for some cultural background and socio-economic status in a segregated society, they also barely take into account the much more complex process of integration, distancing or proximities from owns origins, admixtures and self-identification in everyday life and in family histories. Nevertheless, these arguably arbitrary ethnic subdivisions from above become the roots of stereotyping: the perception of 'quantitative' compactness of the identified groups through the statistical data leads to a problematic essentialization of them, of their culture, uses and conditions (de Zwart, p.312). Although employed with, at least on paper, positive intents, infrastructures and uses of ethnicity retrieval become institutionalised, and are maintained over the decades and the changing political winds (p.307), paving the way for far more negative stereotyping when xenophobic and racist parties surface (Roggeband & Van der Haar, 2018, p.82-84). Possibly building on what Schinkel considers a moral panic about criminal "senseless violence" lodged into Dutch public discourse (Schinkel, 2008), in the years preceding the design of the Top600 and Top400 approaches a new political discourse targeted Dutch-Moroccan citizens as a source of violence and insecurity, othering and essentialising them as an entire ethnic group who has struggled with integration due to rooted and unchangeable cultural incompatibilities. Very little awareness of structural injustices in their integration is evident, while they are instead treated as a group that requires additional discipline and parenting, turning previous targeted policy-making into a repressive one with the purpose to appease the "threatened" Dutch society (Roggeband & Van der Haar, p.88-93).

Correspondences and reflections of the over-representation of Dutch-Moroccans in the predictive policing approaches with the racist and repressive political discourse of the same period appear quite evident. Perhaps one of the consequences attributed to the Top- approaches, that people who simply look Dutch-Moroccans also get stopped by police on the streets (Fair Trials, 2021), would hold true even in absence of the additional tensions and attention raised by the tools. Reasoning on the internal documents behind the design of Top400 reported by Jansen (2022) and mentioned at the beginning of this section, I suggest that if the potential over-representation was noted, but not addressed, then there may even have been a precise intent by some actors to enact racist outcomes, either for political convenience or in the belief that the targeted groups *do* represent a critical source of criminality, and that personal repression is better than structural rethinking. This speaks of the possibility of concealing discriminatory intentions within techno-scientific and administrative approaches that purport objectivity and neutrality. In the following sections, I will discuss more of the non-neutrality of such data and of its relations with governmentality. I will also reflect on how such structures and inequalities maintain themselves, possibly at the hands of other social actors that are *not* intentionally discriminatory and possibly not personally responsible.

2.5. Actuarial statistics, governmentality and non-neutrality of data

The development of similar systems of classification and identification of targets, starting from ethno-statistics to more nuanced algorithmic tools like Top600 and Top400, is, in a way, no surprise in a neoliberal state. As argued in the Introduction, it reflects the neoliberal tendencies to hold the efficient allocation of resources among its utmost practical principles. It is not only that political dilemmas are replaced with economic ones (as suggested by Eubanks, 2018, p.197-8), but it is precisely the economic and the managerial mind-set that spills over all dimensions of public and private life (Brown, 2003), as the purpose of governance becomes not only sovereign control but the optimization of national wealth (Foucault, 1991/1978). To this aim, statistics, archives and taxonomies are deployed both as an instrument of control of populations, as well as a tool for learning what are deemed its intrinsic phenomena: rather than *forcing* them they are now preferably understood, disciplined and policed in desired directions, specifically in the realm of economics, and even at the unawareness of the involved subjects (again, Foucault, 1991/1978). And just as in economics, where future earnings and costs are modelled based on data and trends, for improved allocation of resources and expenses in the present (they are *actualised* to the present moment), so are the lives of different groups of populations, or individuals that come into contact with certain institutions, subjected to actuarial statistics, i.e. the risks and expenses represented by their likely future life is assessed (Schinkel, 2011). Modern data infrastructures (i.e. the way data collection, storage, retrieval and consumption have been designed with increasing power, interconnectedness and versatility), as well as the willingness or complacency of many users of digital tools to the disclosure of their own personal information and behaviour, have certainly changed the relationship of the surveillance and of the aggregation of macro-data. However, in many regards such concepts of governmentality can still be relevant, as well as it is the understanding of anatomo-political and bio-political Foucauldian projects that are involved with the ranking, promotion and repression of desirable and undesirable behaviours and attributes of both macro-populations and single individuals (Foucault, 1976). In this perspective, we better understand the pressure (as mentioned in 2.4) for the Top400 workers to hit the number of 400 list members *as a goal*: it is not so much about finding individuals or pressing social issues that *objectively* need to be tackled by the approach for everyone's benefit, and much more about ensuring that resources already allocated to the operation are not wasted (Jansen, 2022, p.18). In terms of population governance, instruments targeting disadvantaged social groups with a mix of repression and care – promises of benefits and improvements to their conditions that require their abidance to rules and certain behaviours – also bear a striking resemblance to the scientific charity movement in the USA from the late 1800s. In that context, the indiscriminate allocation of benefits towards the poor, invalid and destitute had come to be considered parasitic for society, and instead case-by-case approaches with personalised caseworkers were deemed necessary to investigate the targets' worthiness of receiving aid, as well as to discipline them or dispose of them to lower the numbers of less desirable subjects (Eubanks, p.17-26).

Data represents one of the most attractive resources of the present and of the actualised future: unlike other costly resources, personal information can be re-used over time with little degradation of its value (Bartoletti, 2020), and current infrastructures of data may ensure a stable inflow of relevant updated information. In a way, it could be argued that the increased capacity of nowadays technology seems to bring us closer to a positivistic vision of complete recording, understanding and optimization of the material conditions of populations. Whether this may turn in a triumph of science and detection and resolution of injustice for all, or in more inescapable systems of control and management of the expanding numbers of the poor and destitute of our countries (as in the many examples by O'Neil, 2016 and Eubanks, 2018), is still up for debate. What should however always be kept into consideration in the employment of data is the lack of objectivity of its recording, even when they are often misunderstood as undiscussable and verified aspects of reality. It is not only the establishment of racial identifiers that is murky and non-neutral, as mentioned in 2.4. Rather, scientific research and discourse can be notably complicit in biased truth-making in every field, from the either intentional or inadvertent partiality in the selection of information and the inarticulate biases that circulate through taxonomies, archives and recording systems (Schiebinger, 1993, offers several examples of biased notions that still survive since Early Modern science), to the effect that the goals of the recording and its imagined futures and prospective uses have on the recording process itself (Schinkel, 2011, p.376). Similarly, there are power relations inscribed in the categories of classification and in the truth-making discourse, and biases that lead a specific gaze to consider relevant only certain aspects of it observation; discourse and categories are, in the end, received and inhabited by subjects, rather than controlled, and determine the possibilities of their expression (Foucault, 1971).

This is more an observation of statistical flaws and failures than a criticism to any and all quantitative approaches. Within such a quantitative-bound discourse, statistical insights can, indeed, be used for friction – to highlight precisely the different power and (dis)advantage positions in said algorithmic instruments. As highlighted in 2.3, it is through statistical analysis that scholars and activists that criticize algorithmic policing have highlighted the over-representation (at around 90%) of persons with mild-to-borderline intellectual disabilities in the Top600 list (WOCz, 2023, p.45; Verheijen et al., 2022), the over-targeting of persons with Dutch-Moroccan descent and related neighbourhoods (Fair Trials, 2021; de Koning, 2017), as well as the low-to-destitute income levels of many of the list members (WOCz, p.18). It has also been verified that Top600 cannot in full certainty be considered a determining factor in the reduction of general crime rates and of the recidivism of its targets (p.14-15).

Further, it is statistical thinking and knowledge about the validation and verification of surveyed data that offers room for deconstruction of the alleged positive outcomes of the tools. For example, it is observed that crime records only offer an *attempt* at a picture of total crime, but only represent the crimes that authorities are *able to detect* (Fair Trials, 2021). Among the many obvious reasons why crime and illegality tend to be more concealed than manifest, to escape any attempt at giving a full picture of themselves, perhaps we can also agree with Eubanks' remark that richer neighbourhoods, families and citizens will have a higher tendency to avoid the involvement of the police or of more repressive, institutionalised, or even just public, forces, and to use the wealth at their disposal to address their issues in more private ways (Eubanks, 2018). That is at least, when dealing with problems that could signify the criminalization of themselves or of their peers – wealthy neighbourhoods may be quicker in calling upon disciplinary interventions of othered individuals or group that they perceive as outsiders, as in the examples of the nuisances caused by groups of Dutch-Moroccans in Amsterdam's Diamantbuurt (de Koning, 2017, p.540). On the other hand, poorer families will be more reliant on public services and, sometimes, police involvement. Increases of police intervention and disciplining measures that depart from previous police contacts tend therefore to also result in over-policing of the less wealthy. This also sparks another feedback loop: targeting specific neighbourhoods and profiles with further surveillance inevitably skews the system towards more police contacts, compounding their crime records and reinforcing the narration of them as a locus of illegality and danger (similar to how, in section 2.4, I have mentioned that increased contacts of the list members with the police results in more arrests and apprehensions, since police is invited to be repressive with them even for the smallest infringements, and may result in keeping them for longer in the approaches).

The following section further locates these types of bias and misinterpretations within the functioning of algorithms and the specificity of their relations with human users.

2.6. Black boxes and automation bias

So far, the analysis presented here has focused mostly on how statistics and related statistics-based approaches can reflect existing policing bias and societal racism and discrimination. It has also been suggested that techno-scientific approaches can become a concealing mask for potentially conscious discriminatory intents (see section 2.4). It should already be evident that, other than clearly identifiable oppressors and conscious discriminating actors, statistics and technical approaches also contribute to maintain such understanding and social position of certain groups, and frame and influence the actions and judgements of other social actors and institutions who may not be actively trying to unfairly discriminate, or even be aware of doing it because of how persuasively their targets' identities are produced by the assessments.

Here, I further discuss the ethico-political positionality of algorithms and their complex and specific dynamics: in particular, their effects and relations with the humans involved (whether in the roles of general users, officers, targets, etc.). I suggest that it is worth analysing the practical functioning of such tools as the meso-practices that involve them can in turn shape the human reactions and material outcomes in ways that are quite peculiar. For example, Eubanks (2018, p. 141-142) highlights how human workers who screen target individuals through algorithmic tools, which are supposed to merely *assist* their decision-making, can actually be *trained* by said tools, as they may doubt their own personal assessments and try to align them to the technological ones; this can also be called an automation bias (Fair Trials, 2021, p.34).

It could be added that to this phenomenon contributes something I would define like an aura of objectivity of algorithms – a notion and common narrative that they are objective, unbiased, infallible – which is probably in no little part due to the opacity regarding their functioning. On one hand, general users of such tools do not have an accurate understanding of the statistics and technicalities involved in the production of its outcomes, and there is also a push to keep their inner workings secret to prevent outsiders from gaming the system (O'Neil, 2016, p.28-29), or for other competitive needs. However, on the other hand, as mentioned in 1.4, certain algorithms and AI systems *do* function like a black box, as they take decisions intuitively, from massive heaps of data in which they track patterns and visualize dimensions that humans "may not even be able to perceive" (Bathaee, 2018, p.891-892). They also use their neural network in a diffuse and de-centered way, meaning that it is usually impossible to locate what node, factor or piece of data has been more or less relevant in the decision-making or by which unit it has been processed (p.902-904). This is a component of the untraceability highlighted by Browne (2023). And even if Dutch ADM tools probably do not employ algorithms with the same level of sophistication of more recent generative AI, as I have already argued, many of the characteristics that make the inner reasoning of AI illegible to humans are in common with any statistical inference tool (Bathaee), which have certainly been used at least for the logistic regressions at the base of ProKid+, the tool which was partly repurposed for Top400. For further clarity, a regression is an equation that models an expected outcome (in this case, criminal involvement with police) as being the result of different variables with different weights (the different personal attributes regarding families, schools, previous criminal records, etc.) at a certain degree of accuracy; for ProKid+, and with a certain amount of human input, the related calculations were also meant for finding the most significant variables within personal records (Wientjes et al., 2017, p.253).

It is important to keep in mind Eubanks' insight about how technical tools can be used to generate an "ethical distance" needed by human officers "to make inhuman choices", and help concealing and substituting moral and political dilemmas with issues of efficiency and engineering (Eubanks, 2018, p.13; p.197-198). Similarly, I also agree with boyd's consideration on how algorithms can be used to merely "reconfigure the site of manipulation" of social outcomes to "data infrastructure" (boyd, 2023, p.238). However, while it *is* necessary to keep these reflections in mind, and it *is* even possible to track political pressures in the set-up of the tools in object (Jansen, 2022, p.20; AcVz, 2019; de Koning, 2017, p.539), we should not only consider scenarios where algorithms are designed with clear discriminatory intent. The nature of the algorithms and the way they take their decisions, their relationships with human users and the importance of the influx of external data, raise issues about the quality and non-objectivity of data as well, which can have unfair consequences, largely undetected or justified by its own systems, even in the hands of well-meaning authorities.

In particular, whereas the Top600 approach can be seen as more linear when it comes to its criteria for addition to the list – which mostly consist of serious criminal offences – the Top400 emphasises a focus on its targets not only for their brushes with the law, but because of added factors that show no direct relation with serious crime – i.e. persistent truancy, school changes, involvement in domestic violence even just as witnesses or victims. How have these non-criminal criteria been outlined? First, we need to note that, in its inception, compared to the Top600 which may have started as a grouping of some particularly problematic targets of the police, the Top400 derived from more than arbitrary lists of the 'top 10 most troublesome children' of the city districts. Indeed, the personal records of the members of these top-10 lists were subjected to statistical analyses to find common factors (i.e. personal data, performances, recorded behaviours, etc.) deemed relevant predictors for a criminal evolution, and the list of such factors was later generalised to be employed as predictors within the records of other young profiles (Fair Trials, 2021, p.13). In light of this, it can be speculated that in the tweaking of Top400 to amplify its scope, driven by the budgetary and political reasons mentioned in 2.4, inference tools and processes (akin to those used for ProKid+) have identified progressively less severe factors that were estimated to be predictors of future criminal behaviour. The strength or approximation of this relation is not disclosed, but the aura of objectivity of the mathematical calculation, and the fact it leads to focusing on young persons that must have already been perceived as unruly, certainly solidified the belief that the computation had suggested an appropriate target, and may help convince users and officers of the necessity and goodness of their intervention. Naturally, there is not much neutrality in the constitution and usage of such a tool – from the bias of the evaluation and creation of data itself to the problematic inclusions of more or less criteria – and so, when such an algorithm "runs, it manifests a thousand invisible choices […] under a cloak of evidence-based objectivity and infallibility" (Eubanks, p.168). And while the official documents of Top400 do not explicitly state that corresponding to the profiles outlined by the criteria *determines* one's path towards criminality, the ensuing practices go very close to implying that statistical relation coincides with causation, which is not something that such calculations have proven. In other words, it may have been established (and it is not clear with how much statistical significance and accuracy) that there is a connection between criminality and non-criminal markers identified by the criteria, but not that such issues effectively cause criminality. The approach, however, revolves also around addressing these markers as well – school absenteeism and other non-criminal circumstances – as if tackling them means intercepting the causes of such behaviours.

Another point that seems to be lost in dealing with statistical results is the actual meaning of their forecasting models. As much as they can be presented as instruments for predicting the future of individuals with certain degrees of significance, they are not assessing individuals: rather, a predictive tool "merely indicates that a person shares traits with a group who did or did not carry out certain (criminal) behaviour at a certain rate. But the prediction or assessment cannot and will not provide individualised information about how that individual will behave or act" (Fair Trials, 2021, p.27).

In addition to all these consideration, once such debatable uses and understanding of data computation are validated and normalised in analysis and practice, what happens to the additional personal data that Top600 and Top400 will amass from the families and acquaintances of the list members? How will it be assessed, and how will it feed further calculations?

In the following section I will explore the ways in which the policing entailed by the Top-approaches blurs the lines between police forces and other public institutions, as well as between public life and surveillance. I will also refer to concepts of phenomenology, misfitting and subjection that help us understand the ways in which such a system determines a constant disciplining and monitoring in several touch-points of the individuals within their public and private life, warranting their repression when they adopt behaviours that cause 'disturbances'.

2.7. Diffuse policing and surveillance

When it comes to the understanding of the specific consequences of capillary data infrastructure and assessments, and of how data retrieval and subsequent action turn to most of the recorded aspects of list members' lives, I employ the concept of diffuse policing offered by de Koning. Diffuse policing suggests an understanding of police coercion that is not evidently focused and narrow or clearly repressive, but employs the wide institutional landscape of a welfare state as digitalised as the Netherlands (de Koning, 2017, p.536-538). It is used to denounce how related practices blur the lines within institutions, as they are presented both as caring advice and threat of repression in cases of non-compliance (p.543-544).

Institutions in contact with other aspects of life become a surveillance system for the profiling of the individual and risk fostering mistrust towards the different aspects of the State, ranging from neighbours' reports to schools (Jansen, 2022, p.27), with the aforementioned role of absenteeism as a determining factor the Top400 approach, as well as with teachers being asked to provide information and insight on the behaviour of Top400 members in their class. Families are also seen as a site of determination of the individual and the screening of parents and siblings has also been promoted at some point. Notably, discussed criteria for potential future screening of parents (p.29), especially when there is a suspicion of intellectual disability for the child in the Top400 approach, are impressively similar to those of the subjects which are screened for the Top600, although here I supply the alarming reminder that the individuals screened in the approaches are already suspected or established perpetrators of however severe crimes, and the involved parents are not (or not always). I argue that this is telling of the perspective scalability of such tools: as O'Neil suggests, they are seen as powerful tools which can be applied outside of their initial goal, as long as similar input data can be produced for new targets (O'Neil, 2016, p.31). In a way, they have a potential for an almost "epidemic" (Eubanks, 2018) application towards larger and larger groups of the population whose main fault is to either be in contact with othered subjects, or to be themselves on the verge of being othered for other reasons (among which, of course, a relation with a member of either Top- list constitutes itself alarming data in the eyes of the assessment systems).

Diffusion as a transfer of the disciplinary action towards several other touch-points in someone's life presents itself like a particularly active application of Spade's usage of subjection. Subjection replaces a linear and individualised understanding of oppression of marginalised subjects (in Spade's case, trans people in the USA). Rather than being seen as an oppressive act that can be singled-out and has specifically identifiable perpetrators and power-holders, subjection supplies a concept of how disciplining power is instead not located in a single actor or institution. Discipline and maldistribution of life chances are instead dispersed in many aspects of our public and private lives, as discipline permeates and encompasses them (Spade, 2015, p.6). It often conceals itself precisely in the "purportedly banal and innocuous daily administration" (as could be school performances, street behaviour, and all the other recorded aspects of life and personal networks pulled in by Top600's horizontal extractions), but uses them to distribute life possibilities and vulnerabilities (p.73), to impact and manage entire population groups through the standardization of their performances and the alleged promotion of collective health and well-being (p.57-58).

In the diffuse aspects of data retrieval and disciplinary enclosure, I also think of subjection in combination with Sara Ahmed's phenomenology, and the possibilities her work lends to better frame what could be termed a phenomenology of difference, although departing from that of whiteness and racialized subjects. Despite the main focus that the previous sections of this work have had on racism and socio-economic class, I propose this expanded phenomenology of difference as an adept way to understand the over-disciplining of those who do not fit. Ahmed discusses how people have a sort of bodily familiarity with the spaces and institutions they inhabits, that grants their bodies an orientation: a knowledge or sense of how to reach things or, when it comes to difference, a knowledge of what is or is not in reach, depending on one's own privileges and circumstances (Ahmed, 2007, p.152-154). In essence, orientation shows certain lines of flight that direct movements as well as, we could add, possible future developments and paths of life. More specifically though, bodies are drawn to familiarity among themselves as well – are brought to likeness (to similarity with each other), which has a mutual relation with proximity. The main familiarity and likeness, the implicit starting point and orientation of how the world has been shaped by colonial and cultural history, is indeed whiteness, and whiteness is also sought out, maintained and reproduced by institutions and spaces that, inadvertently or not, are only familiar with it (p.157). In essence, there are certain spaces that interpellate subjects based on whiteness: when they cannot inhabit them without an effort in that sense, they stand out, or even disturb the spaces (p.159). In the cases I propose, it is not only skin colour or cultural background that stick out (although they certainly contribute to the societal attitudes towards those targeted by Top600 and Top400 that are Dutch-Moroccans), but other recorded behaviours and characteristics that probably mark them in many cases as evidently unruly or undisciplined.

Ahmed also offers the image of a skin inflammation as a tentative analogy of the effect that certain types of diversity have in certain environments and discussions. Skin itself represents a viable metaphor for affect and phenomenology – a feeling organ that deals with the sense of exteriority; or that represents the effects and constant contact with exteriority itself. This representation of skin, however, relies too on familiarity, proximity, likeness, and accommodation of diverse needs. There are certain types of diversity and discourses that disturb it, that bring up an inflammation or an irritation, that immediately elicit tensions in a discussion or environment. Except, unfortunately, the inflammation may be felt by other subjects, but it is also mostly experienced by specific skins and bodies that stand out – to the point where they may begin to be considered, or consider themselves, as the sources of inflammation itself. Privilege in these cases becomes an absence of such disturbances, “what does not come up” when certain persons turn up (Ahmed, 2011, in particular p.27-28).

Such notions can be also intertwined with Garland-Thomson’s usage of misfitting. Here, I instead expand on a concept that is originally thought for employment in a feminist materialist account of bodily disability. Misfitting arises as a shifting idea of lack of alignment with a world that is constructed for the able body; an able body that especially refuses vulnerability and lack of independence (Garland-Thomson, 2011, p.598-600). Fitting, to be able to not stand out, like in Ahmed’s explanations, in this case rather means being able to access certain spaces, to perform all the expected tasks by a subject (Garland-Thomson, p.596). Clearly, such reflections can be easily extended to neuro-diversity, for example to the impressive aforementioned statistics about around 90% of the Top600 list members displaying a low-to-mild intellectual disability, and, in general, on any misfitting behaviour. Misfits are not necessarily or intentionally defiant – but they will act in clear misalignment with what is requested by them by social situations, spaces and institutions where privileges are so normalised that they become invisible. They will stick out in mostly white spaces, they will have anti-social behaviours, they will perform (in the sense of even missing it altogether) badly at school or in testing, they will not come from an ideal family (where domestic violence does not occur or is not reported), etc.

In short, I suggest that, while misfitting and phenomenology of whiteness can be used to interrogate diversity and criminalization in many ways, they also offer an additional key to the problematization of Top600 and Top400 and diffuse policing. I return to the skin metaphor to imagine how diffuse policing, with its assessments and disciplinary acts of almost every exterior touch-point of recording and evaluation of its subjects, locates what I would define a sort of digital epidermis that engulfs them as much as possible as entities that are surrounded by data generation. It also acts like a skin in the sense of having different junctions and layers. On one, the ideal body, behaviour and subject is produced and maintained without the expositions and discomforts of those who stick out. On another, undesirable misfits are also compacted within a group and a series of contacts that direct and condition their life possibilities through subjection and interpellation as inadequate or even unsafe individuals.

The following section further highlights the construction of an ideal subject that is implicit in the repression of undesirable behaviours through 'care', focusing in particular on the ambiguous characterization of children and minors in the interplay between their agency, their autonomy and their alleged innocence. Their complex shift between vulnerable subjects and perpetrators ultimately proves to resonate with paternalistic notions of repressive protection and security.

2.8. Paternalistic security, moulding and 'prepression'

In the previous section, I highlighted how non-criminal, but still undesirable behaviours are targeted by the tools. Of course, where there is an undesirable subject, there is a desirable and idealized one, whose production is further discussed in this section. This is particularly evident in the aspects of care that go together with that of repression in the Top600 and Top400 approaches, and even more so when the targeting moves towards younger people and children.

The desirable subject for policing is, obviously, the non-criminal (or the effectively redeemed one). But when actuarial statistics and profile assessments are held as true, then the non-criminal also has to be a rather disciplined child or young person with a similarly adequate family. In other words, for a path far from crime it is not only important to reduce one's involvement with police-persecuted situations, but also to perform well in intelligence assessments, to attend school regularly, to have properly educated parents and non-violent families, to not have potentially criminal or already flagged acquaintances, and (if it were possible, and this impossibility marks a further layer of structural inequality) to have lived in better neighbourhoods from the start. Even outside the scope of the more children-oriented Top400, the Top600 also requires discipline in terms of avoiding to be even just the source of nuisance – which is deemed as being protagonists of "anti-social behaviour" (de Koning, 2017, p.545). In essence, the subjects of the approaches endure what has been called a "prepression": they are repressed in several aspects of their lives *before* having the chance of committing other illicit acts, in the attempt to remove their tendency to commit them (de Koning, p.541; Schinkel, 2011).

The exemplification that this provides of the close ties between unruliness, misfitting, pathologization and criminalization is striking. It may also show a combination of governmentality-related statistics with what Foucault called pastoral power: a form of power and subject-making that ties directly to one's inner voice and outer perspectives, in short becoming a marker of their most intimate identity and imagined futures (Foucault, 1982). In this case, subject-making happens by the essentialization through statistical grouping and the treatment that list members have to endure (including, as mentioned for the case of Top400, receiving a communication that explicitly states that one incurs the high risk of becoming a criminal unless they work to prevent such a development (see Fair Trials, 2021, p.14, for an old version of such letters, as well as Jansen, 2022, p.27, for similar letters sent to the schools).

Children and minors are also further produced as objects of attention and vulnerable subjects even when they are *not* among the list members. Indeed, it could be said that Top600 and Top400 have targets outside of their designated targets; that they aim at increasing contacts with persons that are not on the list, as both approaches specifically focus on minors in close proximity (usually in the same household) with list members. It is also explicitly declared that case directors have the faculty to involve social services for the minors if deemed necessary (AcVZ, 2021, p.15; AcVZ, 2020, p.14). Certainly, this is done in line with the persuasion that having siblings already involved in crime can be an accurate predictor of one's own future criminality (Jansen, 2022). It is easily established that:

> [S]imilar risk factors that make children vulnerable to abuse or exploitation, render them likely to offend. Poverty, disrupted education, family dysfunction, mental health issues and experiences of abuse or trauma are known to affect both the child protection and juvenile justice cohort. (Sacher, 2022, p.9)

Unfortunately, police-heavy rehabilitating (or 'pre-habilitating', as it should be called in this case) approaches seem to forget that, at least according to scholarship from the USA, "being stopped or arrested increase[d] deviant attitudes and negatively influence[d] future interactions with the police" (p.26).

These approaches all reveal to have more or less in common a certain perception of security, avoidance of danger, childhood and adulthood. Children and minors are usually perceived in shifting ways, oscillating between being seen as victims and vulnerable subjects, as threats and as learners (Thorne, 1987, p.89-95). Notably, it can be argued that neither approach really strives to bring closer to an understanding of children and their persons from children's point of view. An idealised adulthood is constructed in relation to these representations of childhood: adults are the opposite of victims, and the aspiration point of young learners. Indeed, children are supposed to be moulded into the independent, self-regulating, well-behaving (here I add: properly fitting, conveniently anonymous) ideal adult. At the same time, societal groups which are not treated to full freedom, full contribution to society, in essence, full adulthood, could be said to be treated as children (p.96), infantilised. Seeing children as learners of adulthood unfortunately tends to overlook how adults are often all but functional, independent, well-behaved and happy (p.93-94), which I argue also mystifies the reasons for criminal behaviour.

Dangerous children, on the other hand, often walk a very blurred line towards victimization, and vice-versa. Moosa-Mitha provides a compelling example, through that of the involvement of children in sex trafficking, of how vulnerable children, depicted as innocent and incapable of providing independently for themselves, i.e. incapable of adult self-sufficiency and agency, can immediately become invested of responsibility, guilt and agency when they commit something inacceptable for victimhood status (e.g. when they became more active in their engagement with the sex trade). A consequence in such cases is that children then become also invested of the threatening part of adulthood and are criminalized and persecuted similarly to adults. These approaches are too ambiguous in denying children of their agency until they are considered agentic because they do something that only supposed adults could do (Moosa-Mitha, 2005, p.382). I additionally suggest that they also associate agency and self-governance only with positive, constructive and legal choices, and do not understand how agency can also mean being protagonist of harmful choices.

The stricter hold that actuarial statistics impose on (anti-)social behaviour and adulthood-oriented discipline also always hinges on the tendency towards securitization and the aversion of threats – in this case, the threat of highly impacting criminality. A paternalistic approach seems evident in the way this is enacted. Notions of patriarchy are invoked here not to essentialize gender – although the essential lack of female representation in the Top600 and Top400 lists is certainly telling of the differences either in police reactions to behaviours from different gender, or of the gendered socialization into different behaviours, or of both. Rather, I mention paternalism and masculinist protectionism to seek to understand "how a certain logic of gendered meanings and images helps organize the way people interpret events and circumstances, along with positions and possibilities for action" (Young, 2003, p.2). One of the images connected to masculinity is that of the protector, sometimes even in chivalrous terms. In other cases, masculinity lends a less gentle face, constituting an appointed and much less flexible leader, or security state, in responses to perceived emergencies (like the 'senseless violence' and the unruly Diamantbuurt loiterers – see sections 2.4 and 2.5, respectively). Perceived emergencies can be both present or future, both external or coming from enemies within (p.8). Indeed, the figure of the protector always implies the complete subordination of the protected subjects – women, children (p.4) and other vulnerable subjects, be it with a stricter or gentler approach in directing lives and people, that is by no means less powerful or constraining in its dimensions of purported love (p.6) – or, in our case, purported care. In the end, however, the existence of emergencies and threats is necessary to justify the added securities, surveillance and power afforded by masculinist protectionism: in other word, the identification of a "selfish aggressor", of "bad men" (p.4) is convenient to the imbalances of power – and, in our case, 'bad men' are indeed almost only male.

In the following section I provide a brief summary of the interactions of algorithmic tools of predictive policing with the legal frameworks they inhabit – or, as it will become more evident, to the laws, principles and rights they tend to upset and undermine.

2.9. Diffuse responsibilities and other legal considerations

I conclude this analytic chapter with a short review of the main ways in which these tools eschew certain protected boundaries and categories of the law, while also not acting explicitly against legality, making it extremely difficult to appeal against them. From a perspective of governmentality, laws are not the source of power anymore (Foucault, 1991/1978), nor they are sometimes seen as much more than an instrument in the assessment and management of population groups (Spade, 2015, p.73). However, it is interesting to observe the intricate relations and possibilities of existence of techno-scientific systems of policing with their legal frameworks, especially when observing the striking contrast between target groups that have intensive contact with the law and numerous legal incriminations, and systems that are instead in no way made to stand trial.

Additionally, it is of particular interest towards the field of this research to acknowledge the process of discussion and approval of the Artificial Intelligence Act in the EU, a regulation of the strongest kind within the Union, which is supposed to be binding for all members States (Future of Life Institute, n.d.-a), and has been under elaboration since 2021 (Future of Life Institute, n.d.-b). The AI Act represents the first international regulatory agreement on AI usage (European Parliament, 2023) and could certainly set an example for other provisions on algorithmic usage in governance. Mostly agreed upon since the end of 2023, it has been approved in March 2024 (although it is still under a process of final corrections – see European Parliament, 2024), and will be fully effective within approximately 2 years from the approval, with the first bans coming into effect within approximately half a year.

While the scope of this thesis does not extend to a full critical assessment of the AI Act or EU algorithm-related legislation, it is nevertheless remarkable that the final text also concerns predictive policing practices, both in terms of biometric detection and identification for surveillance, as well as for what is called in the Act "risk assessment of natural persons" by law enforcement to predict the likelihood that someone will commit a criminal offence, as mentioned in Article 5.1.d, as well as points 3 and 42 of the preliminary section (European Parliament, P9_TA(2024)0138). Risk assessment in this case is prohibited when it comes to the categorisation of persons on the base of their profile and their recorded behaviour; however, the same points and article also explicitly allow for the permission of profiling and recording when it comes to individuals with a criminal track record, as in this case profiling would be based on what are deemed “objective and verifiable facts directly linked to a criminal activity” (Article 5.1.d). This passage seems to allow for the usage of approaches like Top600 and Top400 so long as they base their entry criteria first on criminal records, and it will be interesting to see if the ban on behaviour-based prediction will in some way lead to more legal frictions against them. In addition, the Regulation specifically allows for uses in cases of national security where the Regulation does not apply (preliminary consideration 24). Compared to stricter regulations initially envisioned, the approved version of the AI Act has been deemed to be significantly weaker in terms of protections against predictive policing (Fair Trials, 2023). Commentators suggest that despite its claims of fairness, perhaps the Act does not really provide concrete measures for such cases, especially for what we have termed automation bias and the bias deriving from data themselves (Verte, 2024). Considering that the AI Act also clearly outlines essential requirements and harmonised standards for manufactures of algorithmic tools (of any kind) to operate freely in the EU (Future of Life Institute, n.d.-a, para. 3.2.1), here I suggest to once again use Spade’s cautionary invitation to assess laws not for what they declare, but for the effects they enact (Spade, 2015, Introduction). In short, it may seem that while it declares to defend and protect its citizens from aberrations and unfair uses of AI, specifically for “high-risk” ones, like predictive policing (European Parliament, 2024), the Act also produces a series of standards that will actually allow market producers to *increase* the design and usage of algorithms in governance.

Additionally, although there are expected to be further implementing and delegating acts to the Act in the future, as of today it still does not account for actual liability for algorithm producers in cases of harmful uses (although it of course establishes non-compliance to standards) (Future of Life Institute, n.d.-a).

When it comes to liability in general, and not only in EU legislation, the analysis I have conducted in this chapter should have already suggested a few insights on the personal responsibilities of designers, users and other institutional touchpoints of Top600 and Top400 list members. Although there may have been consciously discriminatory decisions at a higher political level, these systems are immersed in an understanding of data, statistical assessment and social narrations that sustain, justify and reproduce themselves, directing the actions of the actors involved (including, we can speculate, some members of the police) and possibly even earnestly persuading them of the correctness of their conduct. We can once again use de Koning’s (2017) concept of diffusion to extend it to a diffusion of responsibilities. In other words, an injustice is in a way enacted against the targets of the approach, but in practice it is produced by a combination of actions, subjections, monitoring and changes in other situations that are institutional and societal touchpoints within their lives. This essentially mirrors a textbook definition of Youngian structural injustice, where unfairness is dealt because of the accumulation of micro-actions – which are considered legitimate, within their acceptable scopes and interests – of different actors who are not exactly individually responsible for the negative outcomes (Browne, 2023; Kasirzadeh, 2022). This type of reasoning does not deny in any way the injustice and inequality; it does, however, upset fault-based systems of responsibility. Single culprits cannot, as of now, be so easily identified from a legal point of view, and therefore discussions within the EU about the enforcement of protections against AI injustice with increased fault-based legislation risk to prove ineffective in practice (Browne). It should perhaps be demanded of more powerful social actors and collective powers to address such issues instead of narrowing them down on the actions of individuals since, however structural the injustice, the public can be argued to be more responsible than the single persons (Kasirzadeh, p.7-8). However, as just explained above, it seems that as of now, EU members are rather outlining situations in which local governments can acceptably utilise high surveillance and policing algorithmic technology for emergency and security reasons (European Parliament, P9_TA(2024)0138).

The structural untraceability discussed by Browne (p.337) is also interesting from a legal point of view. In short, it is particularly difficult to determine the sources of bias or harm within algorithmic functioning – to trace and pinpoint the type of data, the input, the decisional node that has determined it. This would also imply that, according to O'Neil (2016), it is particularly difficult to criticize the output of an algorithm (even outside of the difficulties in legal appeals) because it is hard to figure out how exactly it has made its decision, what is responsible for its distortion, and even if any bias has actually been manifested by it or not. Similarly, it can be extremely difficult to understand which human input may have caused such outcomes (if that is even the case). It may be suggested that designers of the tools are at least partly or more responsible of their consequences than others. However, algorithms can (and actually, systematically do) lead to unpredictable errors – and errors and bias themselves are *needed* by their designers to learn how to regulate them and adjust their models (Amoore, 2020, p.47; and Chapter 4). Indeed, even the attribution of an algorithm to a definitively responsible author is debatable. Authorship of algorithms is particularly blurry and distributed, in between the different versions and reuses of others' codes and tools and the employment of data from different sources (Amoore, Chapter 3). In a way, we are all engulfed in a mutual ethical relation as suppliers of data – with our countless actions in contact with algorithms – against which aberrant, criminal behaviour is modelled (p.87). Classification of certain subjects and groups requires data from the 'total it must be distinguished from, too. In the case of biased predictive policing in the Netherlands, who are the final authors of the algorithms? Those who have designed the original Top600 and Top400? The designers of ProKid+, on which some of Top400 infrastructure and inflow was based? The authors of significant later modifications of the criteria and selection of the tools? Or all the different agencies and institutions that provide data and information for computation, that has been assessed and validated in certain (perhaps already biased) ways?

Another peculiarity of predictive systems is that they allegedly act *before* crime is committed, but also *after* other criminal activities in the past have singled out certain profiles. However, in this endeavour they harm a right (or at least a moral principle) at redemption, at starting over in the eyes of the law after having already dealt with its punishment (Eubanks, 2018, p.187-188). Top600 and Top400 will indeed further target individuals that have already had their course with the law.

Acting *before* crime also means disrupting a right to fair treatment, fair trials and presumed innocence, since action is taken without a judicial sentence, or sometimes even formal charges. This translates in a lack of protection, which could be invoked only by those *formally* accused of something (Fair Trials, 2021, p.30-31). With algorithmic and ADM systems directing police efforts, "law enforcement and criminal justice authorities have moved beyond formal legal ideas and concepts of 'reasonable suspicion', 'suspect' and 'charge', operating in a way which fundamentally undermines these legal structures, formulations and the protections attached to them" (p.32). In addition, while research and activism on the over-representation of e.g. Dutch-Moroccan could perhaps lead to tackling the racist bias with anti-racist laws, other groups impacted by the tools cannot so easily be defended by law. Specifically, while the tools appear clearly also as a system of management of the socio-economically disadvantaged, "the poor are not recognized as legally protected class" (Eubanks, p.156). Yet poverty or class marginalisation seem to be one of the main ties among the different individuals targeted by the tools.

Lastly, another reason for the difficulty in appealing against the Dutch tools lies in the responsibilities of the different agencies and institutions involved. Firstly, Top600 and Top400 are already criticized in what seems to be a misplaced environment – despite dealing with criminalization and penal law, they would be judged by a court of Administrative Law, since they are considered merely administrative measures. In a 2014 appeal of a young man against his inclusion in the Top600, the court responded by pronouncing that it was only possible to trial the approach in regard to the way it exchanged information among the different agencies involved. As for the rest, however, the single actions undertaken by representatives of the different institutions were considered legitimate and under the scopes of their functions. This also meant that any criticism towards the treatment endured was to be brought to the different plaintiffs of different institutions (de Koning, p. 551), rendering an actual appeal incredibly difficult. Privacy laws concerning data may in this light result one of the best defences against algorithmic aberrations – however, I argue that law protections tend to fade away when conditions similar to that of a security or emergency state are triggered. I additionally argue that the technical potentialities of these tools may still tempt the illicit user, possibly from positions of power where even illegal behaviour can be defended from the worst consequences of any protective law.

In the chapter that follows, I will summarise the main insights of these research in relation to the research questions outlined in 1.1, and will supply further considerations on research ethics, limits and developments.

3. Conclusions

3.1. Summary

In this work, I have tried to reach a better understanding of algorithmic injustice and its entanglements with feminist critique. In particular, I have focused on predictive policing, a security practice that claims to be able to target individuals, groups and areas which are forecasted to be involved in crime through computational insight, *before* they are actually able to do harm, and that has often been denounced as discriminatory. I argue for the relevance of such a research in light of the techno-scientific dream of producing certain, unambiguous outputs to otherwise complex socio-political issues.

My research approach has been inspired broadly by critical law studies, in a sense also because of an understanding of algorithms as a set of rules and dispositions of data and resources (even though it should be noted that more advanced algorithms are less readable and less plain in determining how they allocate outcomes). It has also been considerably informed by Dean Spade's work on the marginalisation and impossibility of trans lives in the USA because of discriminatory moments of subjections and discipline dispersed and concealed across allegedly neutral administrative structures. Despite these methodological suggestions from the USA, I have striven to locate my research within a European context, in light of the recent approval of the EU Artificial Intelligence Act. In particular, I have identified the Netherlands as a particularly interesting and advanced case of digitalisation in public administration and layered adoption of predictive policing tools. Despite the language and cultural barriers, I have decided to focus on two Dutch tools in particular: Top600 and Top400 (now called Veilig Alternatief, translatable as *Safe Alternative*), which are used in Amsterdam to profile and identify known or suspected perpetrators (according to criminal records) of crimes that have a particularly destabilizing impact on society, and that statistical forecasts suggest are most likely to commit crimes again in the future. Many of the list members are young people ore even minors. I have mainly analysed some of the official documents and webpages of the Top600 and Top400 approaches, I as well as existing scholarship on predictive policing and specific literature on its employment in the Dutch context, in an attempt to understand the existence and dynamics of potentially harmful impacts.

My guiding questions have been:

Q1) In what ways do tools of algorithmic profiling of potential criminal behaviour and predictive policing contribute to or produce social marginalisation in the Netherlands?
Q2) Are there subjects and groups which are more impacted?
Q3) What are the practices and structures involved, how do they work and what are their relations with human actors and with other tools themselves?
Q4) What biases and distortions can be identified which specifically inform algorithmic systems and/or are amplified by them?

I can claim to have reached a better understanding of the questions I have posed, although, while they are outlined as separate sentences, answers to them have been overlapping and diffuse in a more complex analysis.

Regarding Q1, I have inquired on how the two algorithmic approaches work. While reading the related criteria for inclusion in the lists can be enlightening, and it has been possible to reconstruct at least some of the pressures and statistical procedures leading to some parts of the tools, I have also argued about the black box nature of the such algorithms. In other words, the details of their functioning are kept unclear to prevent external targets from gaming the systems, and the mathematical complexity of the statistical operations involved (which do not need to be advanced AI to be unreadable, but could simply be the result of statistical inference tools) makes the way the system *reasons* particularly hard to read. Statistical complexity makes it also hard to trace the specific portion of data or the internal link that may have produced a specific output (including those that are deemed unfair). Statistical forecasting determines the attributes of individuals that are predicted to be recidivists, and such attributes are translated into criteria of entry within the lists of the approaches. In the case of Top600, such criteria are mostly related to criminal arrests or formal charges and accusations or detention days within a certain time period, while for the Top400, which targets even younger persons, there is a lighter focus on previous contact with the police and other criteria are introduced, regarding aspects linked to family life, domestic violence, and school absenteeism, among others. Here the black box problem returns again, as the inclusion of such non-criminal criteria has been speculated to be the result of the tweaking of the forecasting regressions, which must have identified a series of additional factors that are supposed to predict criminal behaviour, although probably more weakly linked with the expected result. In sum, the approaches first proceed with a *top-down* movement, by identifying a list of targets from above; they then shift to a personalised approach, where it collects information *horizontally*, linked to what is all around the targets (family, acquaintances, finances, debts, etc.). The purpose is to allow for a better assessment of the person-centred courses of action which are meant for the more efficient repression and rehabilitation into a crime-less life, as well as to identify other minors in contact with the list-members, to verify if they could be under threat. The approaches rely heavily on the combination of different information coming from different institutions, and with actions deferred to them to treat each specific case.

The way the approaches contribute to potential criminalization is also discussed below, as it also needs answers to the other research questions to be better understood. In essence, however, they seem to compound disadvantages and contacts with the police for people and groups who are already marginalised or criminalised, while also producing and justifying such over-targeting because of their supposed scientific rigor.

Regarding Q2, I have observed how the population targeted by the tools has some specific markers: almost all of them are men, almost all of them are Dutch-Moroccans, almost all of them, within Top600, have what is defined a mild intellectual impairment (i.e. they perform sensibly below the average on an IQ test). In addition, they often have a lower socio-economic status, and come from specifically segregated areas and neighbourhoods. Here, I have brought briefly into question the specific discomfort about race, and yet inherent racism, that characterizes the Netherlands according to some scholars. The Top- approaches seem to reproduce the targeting of what had been seen like problematic groups by media and politicians and which were particularly identifiable in terms of race and ethnicity. However, they do not only reflect existing bias, but they also reproduce and aggravate it, further marking out individuals for a supposed criminal development, and spreading towards other subjects that do not do well in certain social contexts. In particular, they also discriminate against, or harm particularly, children and minors, who are in a way configured by their functioning and in their outcomes as subjects of additional repression, monitoring and discipline.

Indeed, as responding to Q3 and Q4 reveals, there are some specific traits in algorithmic functioning that produce discrimination rather than merely mirroring it. As for the practices and structures interrogated by Q3, it is possible to employ the concept of diffuse policing as a form of discipline and repression that does not involve only physical one, but rather manifest along many of the touch-points of the individual with other public institutions and social situations. Disciplining, here, is always tied with monitoring, recording and surveillance, certainly leading to the erosion of trust in said institutions. In particular, aside from the racist and xenophobic components, these touch-points and sources of data could be seen as the relevant junctions in which a marking of the perspective criminal is produced, and by opposition, the functional citizen is also reinstated, for example when it comes to school attendance, family status, etc. I argue that a lens of something akin to a ‘phenomenology of diversity’ and an expanded understanding of material disability is particularly apt at revealing the targeting of individuals that ‘stick out’ in disciplinary settings. I also observe how in diffuse policing, all touch-points become *also* subtle disciplinary moments that further attempt to constrain identities and life possibilities with a quite literal expression of pastoral power.

Of central relevance in my discussion, and necessary to finally respond to part of Q3 and to Q4, is an exploration of the non-neutrality of science and knowledge. I argue in length about how the classification and recording of data – for administrative or scientific purposes, and often for both – usually already contains inarticulate biases and notions that colour what are proclaimed to be objective observations of facts and nature. Fundamental to understand this employment of knowledge is also the context of neoliberalism and governmentality, as identifiers of a socio-political environment where the utmost principles are the efficient allocation of resources in a managerial perspective. In such contexts, sovereign, arbitrary power is replaced by the expectation of understanding intrinsic phenomena in populations and nature through statistics and economics, in the hope of reordering and optimizing populations rather than dominating them in less efficient ways. Statistics and taxonomies imply classifications and reordering, in more or less subtly hierarchical ways, essentially determining which population groups and which bodies will enjoy more privileges or suffer more disadvantages – a matrix of discriminatory outcomes from which contemporary scientific discourse does not seem able to disenfranchise itself.

In this context, criminalization (and neighbouring conditions) obviously becomes a hinge of population management; and it has a tendency to rely on scientific observations, which naturalize it. It also becomes specifically tied with education of minors and conceptions of childhood, as similar tools build on (and enable) statistical approaches that actualize the criminal development of younger citizens in particular. I observe how what I term an aura of objectivity and infallibility seems to permeate complex quantitative processes and machines, resulting in what, when it comes to the specific human relations with algorithms interrogated in Q3, leads to an automation bias: a tendency to defer to the calculations of the algorithms when it comes to human participation in the automated decision making, rather than to questioning them.

Other biases and specific phenomena connected to the algorithms depart from an understanding of the data and its processing itself. Profiling tools may be understood as assessing individuals, but in reality they compare their records to those of similar profiles: they assess commonality from collective references, and it is problematic to consider them actually tailored on the individuals. There is also a practical confusion of significant relation with causation: some of those that are statistically established as identifying attributes of future criminals are misunderstood as *causes* of their future behaviour, since they are treated as areas of intervention.

Additionally, criminal records may be considered a representation of the full picture of crime, and not questioning their partiality is a glaring mistake. Crime goes undetected for a large part, therefore any representation of any particular profile in criminal records is already going to be skewed and partial. The aforementioned xenophobic and racist bias also informs the actions of many law enforcement authorities, who will tend to stop or persecute certain social groups more based on the perceived emergency of Dutch-Moroccan criminality or because of their own prejudices – therefore, criminal records data will also already contain an over-representation of this group. There are structural conditions and inequalities, too, that contribute to making certain population groups more involved in criminal affairs, as well as ways in which sensitive factors like ethnicity reappear through proxies (correlated approximations) of other types of data.

Lastly, the algorithms will also *cause* more of such records – police officers memorize list members and tend to harass and persecute them more easily as some consider them the main suspects when a crime similar to their past accusations happens in their neighbourhoods. This results in more days of detentions, or more formal charges and arrests against them, in some cases effectively producing their permanence in the Top- approaches, where the cycle or targeted repression starts again. This can also be considered a self-confirmation bias, where the calculations of an algorithm actually organize and produce reality in a way that is aligned with its predictions. Worse, in future revisions based merely on the update of data, which gets progressively more skewed, the predictive operations and mechanisms will be considered valid.

My work is also informed by an understanding of structural inequality that seems quite fitting in the case of these tools and disciplinary actions. There are underlying structures that determine the disadvantages of the subjects involved and the privileges of others; there are certainly some responsibilities from more powerful actors, and possibly some convincedly racist or willingly discriminatory police officers and users of the Top- approaches. But if there is any intentional discrimination, it is dispersed through apparently more banal, as well as colour-blind and race-blind, acts of bureaucratic and administrative notation. There is also an aggravated injustice that is produced through the accumulation of all the other disciplinary and repressive moments and evaluations in the complex networks around the list targets – something I conceptualize as a digital epidermis, a ‘skin’ of data production and monitoring that engulfs them. All elements are interconnected between each other and often contribute to their mutual maintenance or even expansion.

In connection with this, I also try to briefly problematize the legal framework concerning AI and algorithms. The AI Act does not seem able to eliminate these kinds of discriminatory uses of predictive policing, since it allows risk assessments based firstly on past criminal behaviours, clearly not accounting for bias in pre-existing criminal records. Even if it were to condemn designers and users of the tools, it is hard to pin-point the small acts, located and diffused among different agencies and institutions, that lead to a discriminatory outcome, or even to trace specific, individual responsibilities. In this respect, advanced algorithms also upset an understanding of individualised authorship, as they are constituted of codes with different sources, and employ data from different sources and with different authors – data in whose production we are all, in many ways, entangled.

These approaches also harm the right to presumed innocence, since they act *before* a supposed crime would be committed; yet because they do so, they also eschew legal protections in place precisely for the benefit of specific categories, like official suspects, which in this way are never formally designated. Some of the social groups they harm, like the economically disadvantaged, are also not a protected category, while there could be room for appealing against more evident racist discriminations.

Lastly, the black box characteristic and the obscurity of their operations make it much harder to appeal against them, since it is often not possible to reconstruct how they have reached their conclusions (or, in the case of the Top- approaches, how exactly they have shaped some of the non-criminal criteria).

3.2. Further limits and considerations

Lastly, I also address some further considerations about this research, and its limits.

One dimension that was not within the scope of this research was that of the public and political narration of Top600 and Top400. I have mentioned in passing how they have been heralded as the proud products of certain political figures, and how there was certainly a political construction or exploitation of the criminality of Dutch-Moroccans. I suggest that it would be of additional interest to further explore how the political and media narratives react to, include, produce or even contest the approaches. In particular, as, according to my analysis, they tend to reflect and generate existing and new discrimination, a research in this sense would probably start from the assumption that the alleged success of the tools is specifically co-opted and exploited for the advancement of certain political positions, probably in a lens of the masculinist protectionism and securitization that I have also highlighted.

I have also already discussed how both epistemically and ethically, the research has obvious limits in terms of its understanding of the Dutch language and culture and of the realities of the targeted social groups. A more extensive research would also require a better understanding of the legal framework of the Dutch institutions mentioned. My focus has been on elucidating the technical and relational functioning of these tools and approaches to clarify in what specific ways they produce and perpetuate harm, rather than on possible solutions to them. Any insight derived from my work, any attempt to translate it into operative consequences and solutions, or any further investigation on the topic, should also benefit from an extensive contribution of the targeted groups and individuals, through field research and/or other forms of participation and inclusion, which have not been part of my research.

Despite these delimitations, I consider this thesis a solid addition to the growing field of research in algorithmic injustice through a feminist and anti-discriminatory lens. My broad aim, of offering a research that shows the interdisciplinary adaptability of feminist critique to recent technological advancements, as well as my implicit claim, that feminist scholarship is already quite adept at addressing the non-neutrality of techno-scientific approaches, are I believe reached and confirmed by my work. My additional aim of providing a guiding exploration of how to question administrative dispositions and choices, by departing from the intent they declare and by focusing on the consequences and further effects they conceal or anticipate, is also exemplified. To a phenomenon that has a growing related literature, and whose many components and aspects are already criticized by certain traditions and directions in feminism, I contribute with a specific investigation and overview of the dynamics and relations of relatively new tools, and a necessary reflection on their diffuse, structural, interconnected nature and functioning.

References

- Actiecentrum Veiligheid en Zorg [Security and Care Action Centre] (2019) *Top600 in beeld: Ontwikkeling, inzet en resultaten van de Aanpak Top600 [Overview of Top600: Development, deployment and results of the Top600 approach]* https://openresearch.amsterdam/image/2020/11/3/top_600_in_beeld.pdf
- Actiecentrum Veiligheid en Zorg [Security and Care Action Centre] (2020) *Regeling Gegevensverwerking Aanpak Top400 [Regulation of Data Processing for the Top400 Approach]* https://www.amsterdam.nl/bestuur-organisatie/organisatie/overige/acvz/verwerking-persoonsgegevens-acvz/#ha2022200-b142-4301-b12d-7c84688ef505
- Actiecentrum Veiligheid en Zorg [Security and Care Action Centre] (2021) *Regeling Gegevensverwerking Aanpak Top600 [Regulation of Data Processing for the Top600 Approach]* https://www.amsterdam.nl/bestuur-organisatie/organisatie/overige/acvz/verwerking-persoonsgegevens-acvz/#ha2022200-b142-4301-b12d-7c84688ef505
- Agüera y Arcas, B., Mitchell, M., Todorov, A. (2023) Physiognomy in the Age of AI. In Browne et al. (Eds.) *Feminist AI. Critical Perspectives on Data, Algorithms and Intelligent Machines*. Oxford University Press.
- Ahmed, S. (2007) A phenomenology of whiteness. *Feminist Theory, 8*(2), p.149-168 https://doi.org/10.1177/1464700107078139
- Ahmed, S. (2016). An Affinity of Hammers. *TSQ : Transgender Studies Quarterly*, *3*(1-2), p.22–34 https://doi.org/10.1215/23289252-3334151
- Alonso Bejarano, C., López Juárez, L., Mijangos García, M., Goldstein, D. (2019) *Decolonizing Ethnography – Undocumented Immigrants and New Directions in Social Science.* Duke University Press
- Amoore, L. (2020) *Cloud Ethics. Algorithms and the Attributes of Ourselves and Others*. Duke University Press https://chooser.crossref.org/?doi=10.1515%2F9781478009276
- Aradau, C., Blanke, T. (2017). Governing others: Anomaly and the algorithmic subject of security. *European Journal of International Security*, *3*(1), p.1–21

https://doi.org/10.1017/eis.2017.14
- Bartoletti, I. (2020). *An Artificial Revolution. On Power, Politics and AI*. The Indigo Press.
- Bathaee, Y. (2018) The A.I. Blackbox and the Failure of Intent and Causation. *Harvard Journal of Law & Technology*, *31*(2), p.890-938
- boyd, d. (2023) The Structuring Work of Algorithms. *Deadalus, 152*(1), p.236-40 https://doi.org/10.1162/daed_a_01983
- Brown, W. (2003) Neo-liberalism and the End of Liberal Democracy. *Theory and Event, 7*(1) https://muse.jhu.edu/article/48659
- Browne, J. (2023) AI and Structural Injustice. A Feminist Perspective. In Browne et al. (Eds.) *Feminist AI. Critical Perspectives on Data, Algorithms and Intelligent Machines*. Oxford University Press.
- Clarke, A. E, Shim, J. K., Mamo, L., Fosket, J. R., Fishman, J. R. (2003) Biomedicalization: Technoscientific Transformations of Health, Illness, and U.S. Biomedicine. *American Sociological Review, 68*(2), p.161-194 https://doi.org/10.2307/1519765
- Connelly, R., Playford, C. J., Gayle, V., Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*, p.1–12 https://doi.org/10.1016/j.ssresearch.2016.04.015
- Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, *43*(6), p.1241–1299 https://doi.org/10.2307/1229039
- De Koning, A. (2017). 'Handled with care': Diffuse policing and the production of inequality in Amsterdam. *Ethnography, 18*(4), p.535–555 https://doi.org/10.1177/1466138117696107
- De Zwart, F. (2012). Pitfalls of top-down identity designation: Ethno-statistics in the Netherlands. *Comparative European Politics (Houndmills, Basingstoke, England)*, *10*(3), p.301–318 https://doi.org/10.1057/cep.2012.8
- Eubanks, V. (2018) *Automating Injustice. How high-tech tools profile, police and punish the poor.* St Martin's Press. https://doi.org/10.5204/lthj.v1i0.1386
- European Parliament (2023, December) *EU AI Act: first regulation on artificial intelligence* https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

- European Parliament (2024, March) *Artificial Intelligence Act: MEPs adopt landmark law* https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law
- European Parliament, P9_TA(2024)0138 *European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts* https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html
- European Parliament (2022) *The Dutch childcare benefit scandal, institutional racism and algorithms* https://www.europarl.europa.eu/doceo/document/O-9-2022-000028_EN.html#def1
- Fair Trials (2021) *Automating Injustice: the Use of Artificial Intelligence & Automated Decision-Making Systems in Criminal Justice in Europe* https://www.fairtrials.org/app/uploads/2021/11/Automating_Injustice.pdf
- Fair Trials (2023, December) *Partial ban on 'predictive' policing and crime prediction systems included in final EU AI Act* https://www.fairtrials.org/articles/news/partial-ban-on-predictive-policing-included-in-final-eu-ai-act/
- Fausto-Sterling, A. (2008) The Bare Bones of Race. *Social Studies of Science, 38*(5), p.657–694 https://doi.org/10.1177/0306312708091925
- Foucault, M. (1971) Orders of Discourse. *Social Science Information, 10*(2), p.7-30 https://doi.org/10.1177/053901847101000201
- Foucault, M. (1976) *Histoire de la sexualité. 1. La volonté de savoir*. Éditions Gallimard.
- Foucault, M. (1982) The Subject and Power. *Critical Inquiry, 8*(4), p.777-795.
- Foucault, M. (1991/1978) Governmentality. In G. Burchell et.al. (eds.) *The Foucault Effect. Studies in Governmentality*. University of Chicago Press
- Future of Life Institute (n.d.-a) *Institutional Context* https://artificialintelligenceact.eu/context/#:~:text=Regulations%20are%20binding%2C%20and%20directly%20applicable%20in%20all%20member%20states
- Future of Life Institute (n.d.-b) *Timeline of Developments* https://artificialintelligenceact.eu/developments/
- Garland-Thomson, R. (2011) Misfits: A Feminist Materialist Disability Concept. *Hypatia 26*(3), p.591-609

https://doi.org/10.1111/j.1527-2001.2011.01206.x

- Gemeente Amsterdam *[Municipality of Amsterdam]* (2021, December) *10 jaar Top600: "Het gaat om afschuwelijke delicten" [10 years of Top600: "These are heinous crimes"]* https://www.amsterdam.nl/nieuws/achtergrond/10-jaar-top600/
- Gemeente Amsterdam *[Municipality of Amsterdam]* (n.d.) *Informatie over Veilig Alternatief [Information on Safe Alternative]* https://www.amsterdam.nl/bestuur-organisatie/organisatie/overige/acvz/informatie-veilig-alternatief/#h962521a8-a5a9-4638-9e21-b0cbb875a52f
- Hayles, N. K. (2023) Technosymbiosis. Figuring out our relation to AI. In Browne et al. (Eds.) *Feminist AI. Critical Perspectives on Data, Algorithms and Intelligent Machines*. Oxford University Press.
- Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575–599 https://doi.org/10.2307/3178066
- Inspectie Veiligheid en Justitie *[Inspectorate of Security and Justice]* (2016) *Schoolverzuim: de strafrechtelijke aanpak [School absenteism: the criminal approach]* https://vng.nl/files/vng/20170607-rapport-schoolverzuim-strafrechtelijke-aanpak.pdf
- Internazionale (2023) *Internazionale a Ferrara. Il programma 2023 [Internazionale in Ferrara. Programme of 2023]* https://www.internazionale.it/festival/programma/2023#internazionale
- Jansen, F. (2022) *Top400. A top-down crime prevention strategy in Amsterdam*.
- Kasirzadeh, A. (2022, August 1-3) Algorithmic Fairness and Structural Injustice: Insights from Feminist Political Philosophy. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES'22)*, p.1-13 https://doi.org/10.1145/3514094.3534188
- Lykke, N. (2011) This Discipline Which Is Not One. Feminist Studies as a Postdiscipline. In Buikema, R., Griffin, G., & Lykke, N. (Eds.) *Theories and methodologies in post-graduate feminist research: researching differently* (Vol. 5). Routledge.
- McKeown, M. (2021). Structural injustice. *Philosophy Compass*, *16*(7), p.1-14 https://doi.org/10.1111/phc3.12757
- Moosa-Mitha, M. (2005). A Difference-Centred Alternative to Theorization of Children's Citizenship Rights. *Citizenship Studies*, *9*(4), p.369–388 https://doi.org/10.1080/13621020500211354

- Nash, C. J. (2010). Queer Conversations: Old-time Lesbians, Transmen and the Politics of Queer Research. In Nash, C. J., Browne, K. (Ed.) *Queer Methods and Methodologies* (1st ed.). Taylor & Francis.
- https://doi.org/10.4324/9781315603223-9
- O'Neil, C. (2016) *Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy*. Crown.
- Rahman, M. M. (2018) Methodologies in Academic Legal Research: Lessons from Critical Legal Studies. *University of Asia Pacific. Journal of Law and Policy.* p.13-30
- Riley Snorton, C., Haritaworn, J. (2013) Trans Necropolitics: A Transnational Reflection on Violence, Death and the Trans of Color Afterlife. In Stryker, S., and Aizura, A. (eds.) *Transgender Studies Reader*, *2nd edition*. Routledge
- Roggeband, C., van der Haar, M. (2018). "Moroccan youngsters": category politics in the Netherlands. *International Migration*, *56*(4), p.79–95 https://doi.org/10.1111/imig.12419
- Roth, W. D. (2016). The multiple dimensions of race. *Ethnic and Racial Studies*, *39*(8), p.1310–1338 https://doi.org/10.1080/01419870.2016.1140793
- Sacher, S. (2022) Risking Children: The Implications of Predictive Risk Analytics Across Child Protection and Policing for Vulnerable and Marginalized Children, *Human Rights Law Review 22* (1), p.1-31 https://doi.org/10.1093/hrlr/ngab028
- Schiebinger, L. (1993) *Nature's Body. Gender in the Making of Modern Science.* Beacon Press.
- Schinkel, W. (2008). Contexts of Anxiety: The Moral Panic over `Senseless Violence' in the Netherlands. *Current Sociology*, *56*(5), p.735–756 https://doi.org/10.1177/0011392108093833
- Schinkel, W. (2011) Prepression: The Actuarial Archive and New Technologies of Security. *Theoretical Criminology 15* (4), p.365–80 https://doi.org/10.1177/1362480610395366
- Simon, P. (2017). The failure of the importation of ethno-racial statistics in Europe: debates and controversies. *Ethnic and Racial Studies*, *40*(13), p. 2326–2332 https://doi.org/10.1080/01419870.2017.1344278
- Sokhi-Bulley, B. (2013) Alternative Methodologies: Learning Citique as a Skill. *Recht en Methode in onderzoek en onderwijs 2013 (3)*2

https://doi.org/10.5553/ReM/221225082013003002002
- Spade, D. (2015) *Normal Life. Administrative Violence, Critical Trans Politics, & the Limits of the Law*. Duke University Press.
- Thorne, B. (1987). Re-visioning women and social change : where are the children?, *Gender & society (1)*, p.85-109 https://doi.org/10.1177/089124387001001005
- Verte, V. (2024, April) *Person-based predictive policing under the AI ACT* https://blog.ai-laws.org/person-based-predictive-policing-under-the-ai-act/?cn-reloaded=1#:~:text=The%20AI%20Act%20Article%205,their%20personality%20traits%20and%20characteristics.
- Weiner, M. F. (2014). The Ideologically Colonized Metropole: Dutch Racism and Racist Denial. *Sociology Compass*, *8*(6), p.731–744 https://doi.org/10.1111/soc4.12163
- Wetenschappelijk Onderzoek- en Documentatiecentrum [Scientific Research and Documentation Centre] (2023) *De effectiviteit van de Top600-aanpak. Een vergelijkend recidiveonderzoek onder veelplegers [The effectiveness of the Top600 approach. A comparative recidivism study among repeat offenders]* https://open.overheid.nl/documenten/dpc-0314177a161db83ac35855991dd45a6f25f56dd0/pdf
- Wientjes, J., Delsing, M., Cillessen, A., Janssens, J. and Scholte, R. (2017) Identifying potential offenders on the basis of police records: development and validation of the ProKid risk assessment tool. *Journal of Criminology, Research, Policy and Practice, 3*(4), p.249-60 https://doi.org/10.1108/JCRPP-01-2017-0008
- Young, I. (2003). The Logic of Masculinist Protection: Reflections on the Current Security State. *Signs: Journal of Women in Culture and Society*, *29*(1), p.1–25 https://doi.org/10.1086/375708