Fair Trials

AUTOMATING INJUSTICE:

THE USE OF ARTIFICIAL INTELLIGENCE & AUTOMATED DECISION-MAKING SYSTEMS IN CRIMINAL JUSTICE IN EUROPE

ABOUT FAIR TRIALS

Fair Trials is an international human rights NGO that campaigns for fair and equal criminal justice systems. Fair Trials' team of experts expose threats to justice and identify practical changes to fix them. The organisation produces original research, campaigns to change laws, supports strategic litigation, reforms policy and develops international standards and best practice.

Fair Trials supports movements for reform and builds partnerships with lawyers, activists, academics and other NGOs. It is the only international NGO that campaigns exclusively on the right to a fair trial, providing a comparative perspective on how to tackle failings within criminal justice systems globally.

CONTACTS

Griff Ferris
Legal and Policy Officer
Griff.ferris@fairtrials.net

Bruno Min
Legal Director (UK & International)
Bruno.min@fairtrials.net

Misha Nayak-Oliver
Assistant Legal and Policy Officer
misha.nayak-oliver@fairtrials.net

CONTENTS

EXECUTIVE SUMMARY

Artificial intelligence (AI) and automated decision-making (ADM) systems are increasingly used by European law enforcement and criminal justice authorities to profile people, predict their supposed future behaviour, and assess their alleged 'risk' of criminality or re-offending in the future.

These predictions, profiles, and risk assessments can influence, inform, or result in policing and criminal justice outcomes, including constant surveillance, stop and search, fines, questioning, arrest, detention, prosecution, sentencing, and probation. They can also lead to non-criminal justice punishments, such as the denial of welfare or other essential services, and even the removal of children from their families.

Policing and criminal justice authorities across Europe are using these AI and ADM systems to influence, inform, or assist in criminal justice decisions and outcomes. This report considers case studies of AI and ADM systems (Section 1) including:

(i) to predict, profile and assess the 'risk' of criminality of specific individuals (Section 1.1);

(ii) to profile and predict crime in certain areas or geographic locations (Section 1.2);

(iii) in prosecution decisions (Section 1.3); and

(iv) in sentencing and probation decisions (Section 1.4).

These AI and ADM systems reproduce and reinforce discrimination on grounds including but not limited to race, socio-economic status, and nationality, as well as engage and infringe fundamental rights, including the right to a fair trial and the presumption of innocence, the right to private and family life, and data protection rights (Section 2).

The law enforcement and criminal justice data used to create, train and operate AI and ADM systems is reflective of systemic, institutional and societal biases which result in Black people, Roma, and other minoritised ethnic people being overpoliced and disproportionately detained and imprisoned across Europe. These biases are so fundamental and ingrained that it is questionable whether any such system would not produce such outcomes (Section 2.1).

Predictive, profiling and risk assessment AI and ADM systems target individuals and profile them as criminals, resulting in serious criminal justice and civil outcomes and punishments, before they have carried out the alleged action for which they are being profiled. In essence, the very purpose of these systems is to undermine the fundamental right to be presumed innocent (Section 2.2).

These systems also often have technological barriers that prevent effective and meaningful scrutiny, transparency, and accountability (Section 2.3). There are also concerns around the lack of meaningful human input into these automated decisions (Section 3.1), as well as their direct use and impact on children and young people (Section 3.2).

The use of predictive, profiling and risk assessment AI and ADM systems in law enforcement and criminal justice must be banned. No number of safeguards,

short of a full statutory prohibition, will protect against the fundamental harms outlined in this report. In addition, there must be strict safeguards around discrimination, transparency and accountability for other types of AI and ADM systems used (Section 4).

Based on these findings, we call for:

(i) A prohibition on the use of AI and ADM by law enforcement and judicial, and other criminal justice authorities to predict, profile or assess people's risk or likelihood of 'criminal' behaviour; and

(ii) Stringent legal safeguards for the use of all other forms of AI and ADM used by law enforcement and criminal justice authorities (which do not carry out predictive, profiling or risk assessment functions), including:

the implementation of mandatory, independent testing for biases in the design and pre-deployment phase, as well as continuously post-deployment;

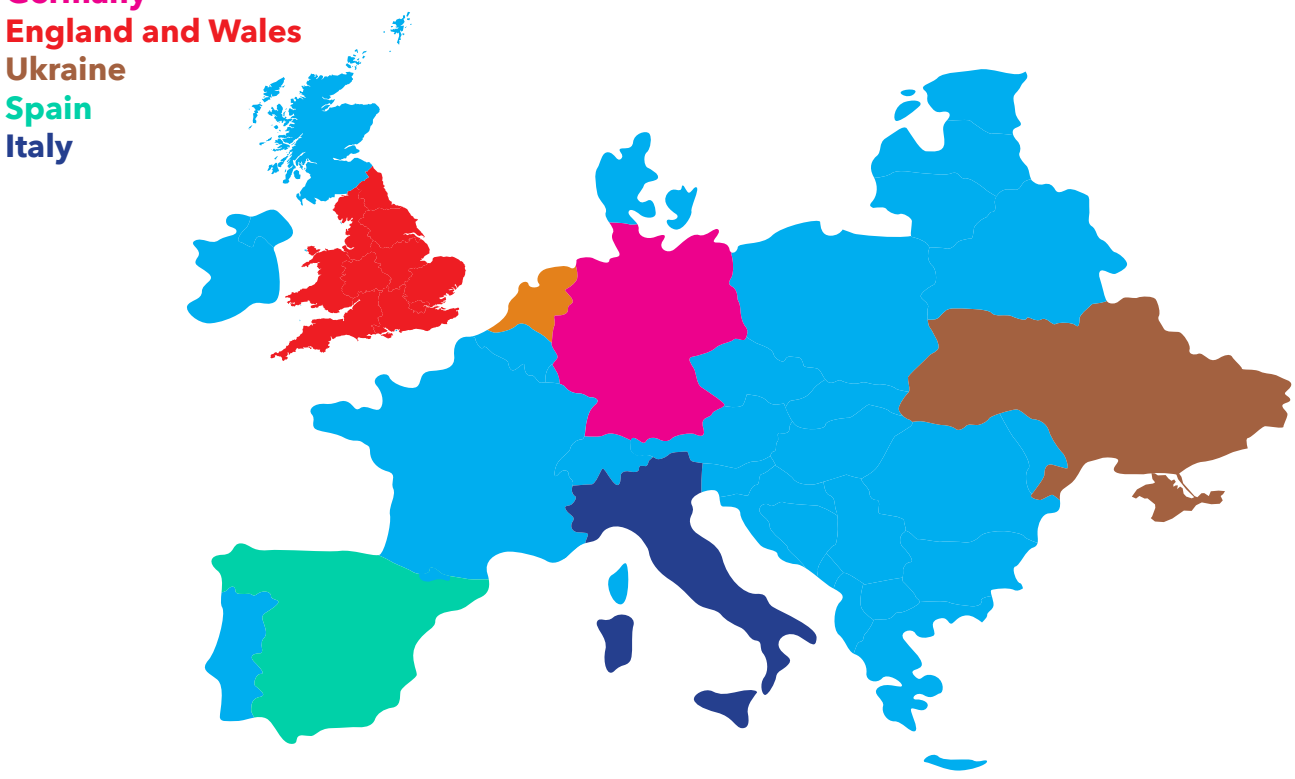
the collection of data on criminal justice that would make such testing possible, including data separated by race, ethnicity and nationality;

a requirement for AI and ADM systems used in criminal justice to be transparent, including details on system processes and input data, and for them not be subject to trade secrecy or intellectual property legal protections. Their outputs must be able to be understood and scrutinised by their controllers, subjects of decisions (such as suspects and accused persons), as well as the general public; mandatory notification to individuals, whenever there has been an AI or ADM system involved, assistive or otherwise, that has or may have impacted a criminal justice decision;

requirements for human decision-makers to evidence how and in what way decisions were influenced, through fully reasoned, case-specific, written decisions, including what factors influenced a decision, and whether this involved AI or ADM system outputs; and

clear routes for challenge or redress for individuals attempting to contest or challenge AI and ADM decisions, or the systems themselves.

Map showing the countries using the AI and ADM systems analysed in this report



INTRODUCTION

This report examines the use of AI and ADM systems in criminal justice in Europe, how they are created and operated, and their impact on people and their rights.

As with many areas of our lives, technological advancements are changing how law enforcement authorities and criminal justice decision-makers operate. These authorities are increasingly using the vast amounts of data they hold about people who come into contact with the criminal justice system, as well as other information held by public authorities and private companies, including information from health services, welfare and benefits authorities, financial and credit information and more, to assist with strategic and individual decision-making using AI and ADM systems. This trend is often driven by financial pressure, pressures for greater efficiency, and misguided perceptions about the efficiency, reliability and impartiality of these technological solutions.

Far from promoting fair and equal justice, predictive, profiling and risk assessment AI and ADM systems currently being used within criminal justice systems in Europe are further fuelling and legitimising racial and ethnic profiling and discrimination, they are normalising pre-emptive law enforcement and criminal justice action through predictions, and they are infringing fundamental rights, including the right to a fair trial, the right to liberty, the right to a private and family life, and data protection rights. The increasing use of data-based analysis and risk assessment tools in criminal justice decisions is producing similarly dangerous results.

These systems, introduced with little or no safeguards, no consideration of their impact on the individuals and groups targeted and subject to them, and with negligible consideration for human rights and the wider societal impact, are exacerbating and deepening existing biases and inequalities, effectively automating injustice.

This report will consider how AI and ADM are created, trained and operated and how the data used can result in biased results. It also examines the issue of discriminatory policing and criminal justice practices across Europe, and how these are reflected in criminal justice data. It analyses the fundamental rights and principles that these systems engage and infringe, and the gaps and loopholes in current frameworks and legislation that these systems, and their developers and deployers, are exploiting or evading, including discussion of the ethical and legal conceptions of the right to a fair trial and the presumption of innocence, discrimination, transparency and accountability.

These issues are not unique to Europe, nor are they new. AI and ADM systems used in the same contexts in the US have been shown to exhibit the same fundamental flaws and produce and create the same harms. However, there is an opportunity for European legislators to act before these systems and the consequences become even more widespread and entrenched.

This report makes clear recommendations on how to address these fundamental issues, harms and outcomes, including a prohibition on the use of AI and ADM in criminal justice to predict, profile and carry out 'risk' assessments against individuals, alongside strict safeguards in relation to all other uses of AI and ADM in criminal justice.

Definitions of AI and ADM systems

There are differences of opinion as to the definition of artificial intelligence (AI) and its true meaning against how it is used widely in practice, and there is no widely accepted definition. Indeed, the term has been used to encapsulate the most basic of computer-based analytical systems. Automated decision-making (ADM) system is a broader term that encapsulates all systems which process data and other inputs and produce outputs, which influence or assist with human decisions, to different degrees. Below are some of the different conceptions of these terms.

The Council of Europe has described AI in the following terms:

> "AI is used as an umbrella term to refer generally to a set of sciences, theories and techniques dedicated to improving the ability of machines to do things requiring intelligence. An AI system is a machine-based system that makes recommendations, predictions or decisions for a given set of objectives. It does so by: (i) utilising machine and/or human-based inputs to perceive real and/or virtual environments; (ii) abstracting such perceptions into models manually or automatically; and (iii) deriving outcomes from these models, whether by human or automated means, in the form of recommendations, predictions or decisions."[1]

The EU High-Level Expert Group on AI (AI HLEG) uses the following definition of AI:

> "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their

> environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.
>
> As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)."[2]

The EU Commission's AI Act describes AI as:

> "software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with"
>
> (…)
>
> **ANNEX I: ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES**
>
> (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
>
> (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
>
> (c) Statistical approaches, Bayesian estimation, search and optimization methods.[3]

In practice, AI is often used as a catch-all term for systems that use or analyse data with some element of autonomy, even if this is in the form of basic algorithms or data analysis, rather than more advanced machine-learning systems. As the Council of Europe's Ad Hoc Committee on Artificial Intelligence (CAHAI) has said:

> "it can be concluded that the term "AI" is used as a "blanket term" for various computer applications based on different techniques, which exhibit capabilities commonly and currently associated with human intelligence. These techniques can consist of formal models (or symbolic systems) as well as data-driven models (learning-based systems) typically relying on statistical approaches, including for instance supervised learning, unsupervised learning and reinforcement learning."[4]

It is necessary to consider all the systems which fall under this umbrella, as well as those which do not. Many ADM systems do not meet the standard of 'true' AI, yet they are often included within common conceptions of AI. ADM systems, often comprising some form of data-based analysis and statistical techniques, are widely used by public and private authorities, including by law enforcement and criminal justice authorities, and can have a significant impact on people's lives. ADM systems have been described by AlgorithmWatch as:

> "a socio-technological framework that encompasses a decision-making model, an algorithm that translates this model into computable code, the data this code uses as an input—either to 'learn' from it or to analyse it by applying the model—and the entire political and economic environment surrounding its use."[5]

This report will consider and analyse both AI and ADM systems used in criminal justice due to the similar and overlapping issues raised in their design, creation, purpose and operation, and in order to include and address the similar harms that result from them. Its focus is on AI and ADM systems which are used to inform, influence or 'make' law enforcement and criminal justice decisions and outcomes, in place of assisting, or influencing human assessment or decision-making.

1. AI & ADM SYSTEMS INCRIMINAL JUSTICE IN EUROPE: CASE STUDIES

There are various types of AI and ADM used in criminal justice systems in Europe.[6] The use of AI and ADM systems to assist, influence or inform law enforcement and criminal justice decisions has been growing over the last decade due to the increasing availability of data and more advanced data analytics tools. Many of the systems analysed as case studies below began during that period.

Public awareness of AI and ADM tools in policing and criminal justice, including the information, processes and technologies they are based on, how they are created and deployed, where they are used, who by and against whom, is relatively low. This is understandable, due to the lack of transparency and sometimes deliberate secrecy that surrounds the creation and operation of these systems, and the poor or non-existent legal frameworks which do not require meaningful transparency, accountability and other protections in relation to these systems and their use. Despite this, AI and ADM systems used by law enforcement and in criminal justice can have very serious and real impacts on people's lives. For a detailed analysis of the harmful impact of these systems, as well as fundamental flaws with their creation, purpose and operational use, see Section 2.

The AI and ADM systems considered in this section include systems used:

(i) in policing to predict, profile and assess the 'risk' of future criminality of specific individuals, leading to surveillance, questioning, fines, stop and searches, and arrests;

(ii) in policing to profile or predict the future occurrence of crime in certain areas or geographic locations, also leading to surveillance, questioning, stop and searches, and potentially arrests;

(iii) to influence and assist prosecution decisions; and

(iv) in sentencing and probation, to decide between custodial and community sentences and release from prison.

This is not an exhaustive list, but an attempt to investigate and analyse several of the most common types of systems, many of which follow the same patterns in their design, implementation, operation and impact, and those systems which demonstrate the breadth of discrimination and most harmful impact on individuals and fundamental rights in countries across Europe.

1.1 Predictive policing: Individuals

Predictive, profiling and risk assessment AI and ADM systems are used by law enforcement and criminal justice authorities to target and assess people on an individual level, attempting to make individualised predictions, profiles and risk assessments about them. These predictions, profiles and risk assessments then inform law enforcement action, criminal justice decisions and punishments, as well as non-criminal justice and civil action and punishment.

1.1.1 ProKid - Dutch police (Netherlands)

Used for: Risk assessment
Created: 2011

Since 2011, the Dutch police have used an automated risk assessment tool, ProKid, which purports to assess the risk of (re)offending – future criminality – of children and young people.[7]

Several iterations of the ProKid system have existed over the years, with the original program, ProKid 12-SI attempting to assess the risk of criminality of children under 12 years old.[8] Later versions have focused on 12–18-year-olds, with the latest version in development focusing on under 23-year-olds (ProKid 23).[9]

ProKid is an 'actuarial' risk assessment tool, used by police to assess the risk of young children and adolescents being involved in "future violent and property offending."[10]

Actuarial (or statistical) risk assessment tools estimate the risk of specific future behaviour – in this case, violent or property offending – through the assignment of weights to certain information, or data variables, that have shown to be associated with such behaviour.[11]

It is a method of statistically estimating the likelihood of a future event occurring. In reality, ProKid does not actually predict the likelihood of criminality, rather it predicts the likelihood of a child being registered on a police system in relation to a crime.

For the development of the ProKid algorithms, data was used on a sample of 31,769 children (20,141 boys, 11,628 girls) between 12 and 18 years of age, who were registered in official Dutch police records in 2007 because they were involved in an offence as a suspect, victim, or witness.[12]

In order to formulate and generate its risk assessments, ProKid uses police data from two police databases. The

first is a 'criminal fact' database where criminal records and evidence are registered, called the Basic Facility Law Enforcement (BFLE). The second database is a 'criminal opinion' database where "circumstantial information", such as "observations of officers" are registered, called the Basic Facility for Forensic Investigation (BFFI).[13]

Data used from these databases includes:

- reports of where children have come into contact with the police, regardless of whether that contact was as a suspect, victim, or witness;
- their addresses;
- information about their 'living environment', which includes information about parents or co-habitants, including whether they were suspects, victims or witnesses; and
- age, gender, offending frequency and whether there was a range of different criminal offences committed.[14]

ProKid assesses this information to identify children as being in one of four categories of 'risk' of committing crime in the future.[15]

The four 'risk' categories are: 'red' (indicates critical danger), 'orange' (indicates a problem child or an address where there are problems), 'yellow' (indicates that a potential risk is developing) and 'white' (no indication of any risk).[16]

These codes can apply to both children and their addresses.[17]

The system establishes correlations between the history of parents, guardians and other adults living under the same home address of children – as criminal suspects, victims, or witnesses – to assign risk profiles to those children.[18] The system assesses children

Category	Behavioural indicators	Living address of the child
White: No risk indication, probably a safe zone	Maximum of two registrations as a victim or as a witness	A minimum of two registrations of the living address as the location of an incident, and a maximum of 5 registrations of the co-habitants of the child
Yellow: Indication of a rising risk potential	3-9 registrations as a victim or as a witness, or a minimum of one registration as a suspect of a light incident, or one-time registration as a missing person	A minimum of 3 registrations of the living address as the location of an incident, or more than 5 registrations of the co-habitants of the child in the system, or the sum of the two above-mentioned registrations is 6-14
Orange: Indication of 'problem youth' / 'problem address'	A minimum of 10 registrations as a victim or as a suspect, or a minimum of 5 registrations as a suspect of a light incident (for instance, shoplifting) or a minimum of 5 registrations as a victim or as a witness related to different types of incidents, or more than one registration as a missing person, or one registration as a suspect of a serious crime	The sum of the incident registrations at the living address of the child and the registrations of co-habitants of the child is 15-29, or 1-2 registrations of domestic violence at the living address, or one registration of child abuse at the address of the child
Red: Alarm phase (indication of 'criminal youth' / 'alarm address')	A minimum of two registrations as a suspect of a serious crime	The sum of the incident registrations at the living address of the child and the registrations of co-habitants of the child is 30 or higher, or more than two registrations of domestic violence at the living address of the child, or more than one registration of child abuse at the address

Figure 1: ProKid 12-SI risk colour codes and corresponding security risks

based on their relationships with other people and their supposed risk levels, meaning that individuals can be deemed higher risk by being linked to another individual assessed as 'high risk', such as a sibling or a friend.[19] ProKid assesses risks of future criminal action that children have not carried out and judges them even if they are a victim of or witness to a crime, or on the basis of the actions of other people.[20] These children, whether victims, witnesses or those who have been falsely accused of a crime, have no responsibility for the crimes with which they are associated and no control over or responsibility for this association.[21]

The impact of ProKid

If a child has been subject to a ProKid risk assessment, it results in police registering them on their systems and monitoring them, then referring them and their families to youth care services and child abuse protection services.[22] As a result, the impact of a child being assessed as 'at risk' by ProKid can be extremely significant for both the child and their family, with children potentially being taken away from their parents and family. A family whose child was wrongly assessed as being at risk of neglect and domestic violence described it as "a huge threat hanging over your head and the worst punishment you can receive is losing your children."[23]

ProKid risk assessments have also been shown to impact family members' criminal records. In one example, a father whose criminal record on the police BFLE database should have been erased under Dutch data retention limits of five years had his record reactivated by a correlation to his child's ProKid risk profile, which referenced the father's otherwise non-active criminal record. Not only was the father's criminal record re-activated, but this in turn also amplified the perception of risk to the child, as the father's historic offence aggravated the child's risk profile.[24]

A significant number of incorrect or erroneous ProKid assessments have been recorded. An evaluation of the system commissioned by the Netherlands Ministry of Security and Justice found that one third of 2,444 children risk assessed by ProKid as red, orange and yellow, i.e. those at the varying levels of risk, involved, "system or registration errors or reports based on irrelevant incidents", and that only 1,542 of the 2,444 assessments were deemed correct.[25] In other words, over a third of children had their risk levels mis-assigned.

A ProKid profile is also the first step in a pipeline of automated risk assessments, which can lead to serious criminal justice penalties and other related outcomes. Despite its name, ProKid is anything but.

1.1.2 Top 600 – Amsterdam Municipality, police & social services (Netherlands)

Used for: Risk modelling and profiling
Created: 2012

In 2012, the Amsterdam Municipality started the 'Top600', an automated risk modelling and profiling system, in partnership with police and social services.[26]

It attempts to profile the 'top 600' young people, over the age of 16, who are most at risk of committing 'High Impact Crime' in the future.

The Top600 is part of what the Dutch police called the 'Person Oriented Approach' ("Persoonsgerichte Aanpak" or PGA) to policing.[27]

The Dutch police state the aim of this approach as follows:

> "The PGA aims to break through persistent patterns of crime and nuisance with repressive and preventive interventions. Prioritized persons are led to a care and/or judicial process. Attention is paid to the person himself and his (family) system."[28]

The Dutch police describe this approach as:

> "The police periodically draw up a police name list. The list includes all persons who may qualify for a PGA on the basis of the prioritized safety themes. The ranking of persons on the list is based on: the number of times that a person is listed as a suspect for the selected offenses and the person's score on a Risk Assessment Instrument (RTI). The system knowledge is enriched with street knowledge."[29]

The criteria used by Top600 to assess the risk of someone committing 'High Impact Crime' is chosen by the police and the Public Prosecution Service. It is stated by the Amsterdam Municipality as follows:

- "In the past 5 years, you have been arrested as a suspect for a high-impact crime: robbery, burglary, aggravated assault, murder/manslaughter or open violence against people.
- In the past 5 years, you have been presented to a bankruptcy judge.
- In the last 5 years, you have come into contact with the Public Prosecution Service at least 3 times and have been sentenced to a punishment.

  Juvenile offenders (under 21 years of age) are included on the list if they have come

into contact with the Public Prosecution Service at least twice in the past 5 years and have been sentenced. In the case of a conviction for a high-impact offense, it is also possible after the first conviction."[30]

The punitive consequences of a Top600 designation

Those on the Top600 are subject to a number of criminal justice and non-criminal justice punishments. The Dutch Public Prosecution Service describe the outcomes of the Top600 system as "punishments" and "interventions".[31] It states that the outline approach is to "punish quickly, severely and consistently", stating that "clever use is made of the combination of punishment/care".[32] The Amsterdam Municipality also ominously states that "The families of the frequent offenders on the list are visited"[33] without specifying by whom. From the experiences of those people profiled and their families, considered below, this often involves police raids and arrests.

The Public Prosecution Service has stated that they will seek higher and more severe penalties for those on the Top600 list, asserting that they will make "higher penalty demands for high-impact crimes such as robberies" in addition to ensuring "active and rapid prosecution of minor offences as well".[34] Dutch lawyers have also alleged that public prosecutors also seek "the longest possible pre-trial detention" for those listed on the Top600.[35]

Other punitive action taken against those on the Top600 includes the "active and swift deprivation and confiscation of valuables",[36] as stated by the Public Prosecution Service. The Amsterdam Municipality state that "a large part of the people on the Top1000[37] list have outstanding fines at the CJIB [Central Judicial Collection Agency]"[38] and that they are looking into "whether coercive measures can be used to keep a person fulfilling their responsibilities in this way".[39] The Amsterdam Municipality also states that Top600 partner organisations are exploring "extra efforts… to bring about behavioural change" by targeting "the three Ws (living, work and partners)".[40]

The impact of the Top600

The impact of being profiled by the Top600 stretches beyond the official punishments stated by the Amsterdam Municipality and Public Prosecution Service, with those on the Top600 also subjected to a whole host of other sanctions and interventions by state authorities.

A lawyer who has worked with individuals profiled on the Top600, Eline Groenendaal, has described a number of cases where individuals have been subjected to arbitrary police and other enforcement action, including arrest, being "constantly followed" and harassed by police, and having regular home checks, as a result of being on the Top600.[41] She describes the myriad ways in which people are harassed, oppressed and have their daily lives impacted:

"Jesse has never been convicted by a judge of a street robbery, but because he is on the list, he is arrested whenever someone is wanted for a street robbery. Then it usually turns out after one day that it was not Jesse. The bad thing is that arrests also count to get or stay on the list. Jesse never gets off the Top600 in this way".

"Mohammed was first 'bullied' with all kinds of group bans, preventing him from even going to the swimming pool, and then his residence permit was revoked. He wants nothing to do with the Top600."

"Souf gets a ticket every three meters when he rides on his moped, simply because he is on that list. When he finally got himself a job, it was ruined by the Top600 people because they called the new employer to see if he knew who he was dealing with. Then the job was cancelled. He also doesn't want to know anything about the Top600 anymore."

"Rachid was arrested for the slightest thing and has now been detained for half a year for a deal of nothing. But when he recently became a father and wanted to visit his wife and baby on a short leave from detention, the Top600 did nothing to help him. He too has lost all faith in the Top600. He's still stuck. When he is released later, there will be no escorts, there will be no home and he will simply take a garbage bag out into the street."[42]

Groenendaal concluded that the system often "works counterproductively" and that while it "looks good on paper… in [her] experience it is nothing in practice".[43] In another case, a 22-year-old man from Diamantbuurt on the Top600 was arrested and detained twice in one day, and held overnight, just for being in a group outside in his local area.[44] He said that he is "constantly approached" by police. His lawyer, Groenendaal stated that "he is not even allowed to go out with his family".[45]

Another resident from Diamantbuurt, a 24-year-old man, was profiled as part of the Top600 due to crimes committed before 2009 – even though the Top600 risk assessment system only began to be used in 2012. He was arrested three times in 2011 and 2012 for street robberies, and despite being released each time before a judge had assessed the arrest, this was enough to keep him on the Top600.[46] His lawyers alleged the arrests were all unlawful and said that such profiling could only legitimately use information on convictions, not just arrest(s) or 'contact' with the police.[47]

A mother of a son profiled by the Top600 has described the significant impact it had on both her and her son's life. Diana Sardjoe has said that police detectives came "all the time" and her son was "constantly accused of doing things with the guys he hung out with".[48] On one occasion, stolen property was tracked to their apartment block, and the police came for her son again, as well as taking away her other 15-year-old son; both were released.[49] When there was an assault and robbery in their neighbourhood, police again came for her son.[50] She described being called to pick up her son from the police detention centre late at night after he had been taken there by police. She said that as a result, her children "withdrew further and further".[51] She has also said that the mothers of children on the Top600 and Top400 are threatened with having their other, younger children taken away from them. Sardjoe has since founded an organisation, De Moeder is de Sleutel (The Mother is the Key), which supports mothers and young people impacted by being profiled by the Top600 and Top400.

The Top600 approach dangerously blurs the lines between care and punitive coercion. Those profiled by the Top600 are subject to the "threat of repression", formally referred to as "Very Irritating Policing". Here, Top600 authorities encourage police to "use their discretionary power to act on 'small infringements'" against those on the Top600 who are deemed not to 'comply' with requirements made of them.[52] A police officer has described that:

> "the Top600 [list] gets undivided attention…. Everyone knows who's on [the list], everything is registered. We have a complete picture, almost hour by hour."[53]

One young Dutch-Moroccan man from Diamantbuurt profiled on the Top600 has spoken about being "followed and monitored wherever he goes" and that "police officers regularly called out to him by name in public, outing him as a person of interest for the police".[54] He said that "When the police drive by and call you by name, you really feel put on the spot."[55] A police officer assigned to the Top600 has said that such action "can be a form of intervention", and described other action the police can take against young people on the Top600:

> "They ask for their IDs, call into the precinct and hear that four of them are in the Top600. The uniformed police will then probably say, 'Hey guys, you're in theTop600. We don't have anything on you now, but mind you, we're keeping an eye on you'. So, yes, they do get that stamp among police officers."[56]

A lawyer described how the system become a "self-fulfilling prophecy":

> "Because you're on the list and you're from that neighbourhood, it means you'll be arrested quickly. And you'll be sent home after two days because they actually have nothing on you, but you do have a citation next to your name. When that happens a few times, they'll say, see, you have a lot of police contacts... Even though you are actively sought out by the police because you are on that list. Because they have a description of the subject: guy on a black scooter with a black jacket, and, well, that can be anyone in Amsterdam. And then they start stopping guys from the list, that's just how it works."[57]

The criminal justice and non-criminal justice interventions and punishments detailed above can have significant impacts on an individual, resulting in both criminal justice action, and other, non-criminal justice punishments, without any judicial oversight or remedy, such as a trial. These criminal justice consequences occur without any formal trial or assessment of the relevant evidence by a judge or judicial process. The Top600 assessment is only carried out and updated every six months, and people can only be removed after a year without arrests for a list of violent crimes, so individuals profiled as part of it have to endure a year of this intensified monitoring and action from police.[58]

Among those people profiled by the Top600, there are clear themes, both in terms of their ethnicity and their neighbourhoods, which suggests a clear element of discrimination on ethnic and socio-economic grounds. The Dutch national public broadcaster NPO reported in May 2020 that

"more than one third of the Top 600 boys are of Moroccan descent".[59]

In January 2020, the Top 600 risk assessment list contained only 469 people, as opposed to 600.

The majority of those on the list lived in one over-represented district, Amsterdam Oost, while 61 lived in Amsterdam Nieuw-West and 56 lived in Amsterdam Zuidoost, the second and third highest respectively.[60] Many of those individuals who have shared their experiences are from the same neighbourhoods, such as Diamantbuurt,[61] which has a notable Dutch-Moroccan population,[62] and many whose experiences of the Top600 are detailed above have Dutch-Moroccan heritage.

1.1.3 Top400 – Amsterdam Municipality, police & social services (Netherlands)

Used for: Risk modelling
Created: 2015

The Top400 is another risk modelling and profiling system run by the Amsterdam Municipality alongside police, prosecutors and social services. It was first used in 2015.[63] The Top400 focuses on children under 16 years old, in contrast to the Top600. The Top400 is run by the Amsterdam Municipality, which describes the system as "the same for the Top600: combining punishment and care", and that it "place[s] more emphasis on care for some people, and more on punishment for others."[64] Its objective is stated as not just targeting 'High Impact Crime' but all forms of crime, and it aims to prevent "brothers and sisters" from being involved with crime.[65]

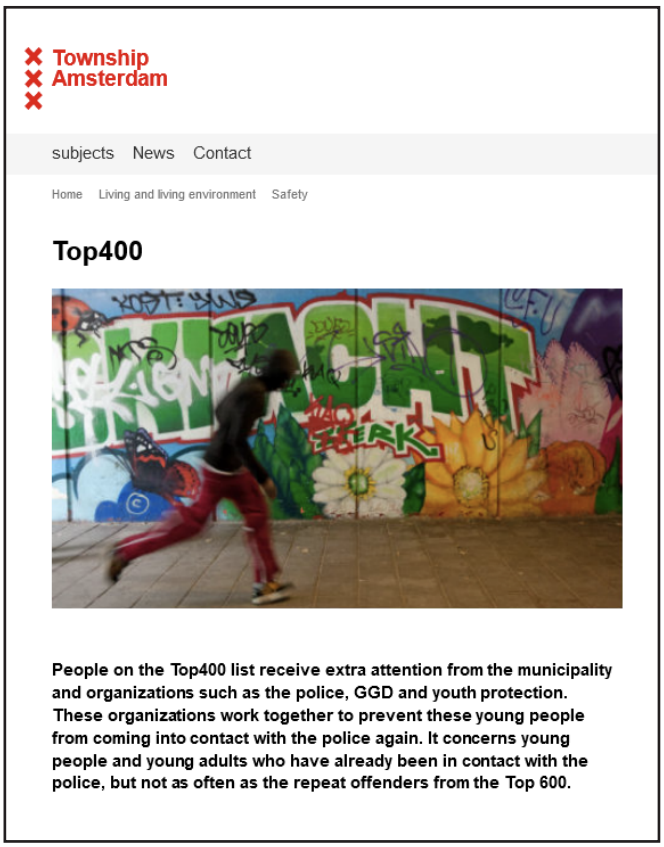


Figure 2: A description of the Top400 on the Amsterdam Municipality's website

The Top400 was developed out of and using the ProKid risk assessment model and risk assessment outcomes as a basis, with some people put on the Top400 because of previous risk assessments by ProKid. In July 2016, 125 people were put on the Top400 on the basis of assessments made by ProKid, and the Amsterdam Municipality currently states that ProKid is used to assess people as part of the Top400.[66]

The Netherlands based Public Interest Litigation Project (PILP)[67] has discovered further information on the Top400 via freedom of information requests that suggests the genesis of the Top400 was even more problematic and unjust.

The Amsterdam Municipality wrote to sub-municipalities and neighbouring municipalities asking for the 'top ten' most 'troublesome' young people in those areas.

Once this information was received, these lists were then cross-checked against the Top600, and those that were already on the Top600 were excluded. The remaining children were assessed and correlating factors between them were used to create the original criteria for the Top400. This exercise shows a clear confirmation bias, the retrospective allocation of criteria based on pre-conceived notions of 'troublesome', embedded with discriminatory perceptions, as well as fundamental errors in placing importance on correlation rather than causation.

The Top400 uses even more dangerously broad criteria in its risk modelling than the Top600, including non-criminal justice data as indicators of criminality, as well as mere suspicion of involvement with crime, without actual evidence. The Amsterdam Municipality describes the Top400 risk modelling as including not just criminal justice data, but also "serious care signals", stating that:

> "After four years of Top600 approach, we know that characteristics other than arrests and convictions for HIC [High Impact Crime] offenses are also indicative of a (developing) criminal career"[68]

As a result, the Top400 aims to "allow these other characteristics such as 'care signals' to play a part in determining the composition of the target group."[69] According to the Amsterdam Municipality, the criteria for being assessed by the Top400 risk model is as follows:

> "You are on the Top400 list because you meet a set of (criminal and care) criteria (variant 1), or because you have been arrested as a suspect and appear in the police records (variant 2).

In variant 1, the (criminal and care) criteria consist of that you have been or have been suspected of one or more crimes in the past 5 years. You have also been suspected of a high-impact crime at least once: a robbery, street robbery, burglary, serious assault, murder/manslaughter or open violence against people. We also have concerns about you because we see that at least 3 of the points below apply to you:you have (had) a juvenile probation measure (this point counts double);

- you have been or have been placed under surveillance;
- for example, you have been absent from school a lot or did not finish school;
- you have changed primary school at least 3 times;
- you have been involved in a domestic violence incident (as a victim, witness or suspect);
- you have been arrested as a suspect between the ages of 12 and 14;
- you've been arrested for dealing fake dope for the last 2 years.

In variant 2 you have been placed on the Top400 list because you have been arrested by the police at least once in the past 5 years, and because we are concerned about you because of the contacts you have had with the police and/or the contacts that others in your immediate vicinity have had with the police. Figures show that there is a good chance that you will also go down the wrong path and end up in crime. We want to avoid this together with you."[70]

The use of arrest and 'suspicion' data without any objective evidence of involvement in crime, as well as the use of 'care' criteria and a child's school history as part of this risk model which has serious criminal justice implications, is extremely concerning.

The impact of the Top400

Children and the families of children on the Top400 receive letters from the Amsterdam Municipality to tell them that they have been identified as one of the Top400.[71] The Amsterdam Municipality states that young people on the Top400 list "receive extra attention from the municipality and organizations such as the police, GGD [public health department] and youth protection" and that "information, including personal data, is exchanged between these organisations."[72]

One mother of a child on the Top400 described that if people refuse to participate in the "care" side of the Top400 approach, "they will report you" and "threaten to remove your children from home".[73]

The consequences of a Top400 risk assessment, therefore, show the same dangerous blurring of boundaries between 'care' and punishment' as the consequences of a Top600 risk assessment.

There are clear possibilities for structural discrimination to find its way into this system. Young people, under the age of 16, are labelled as criminals by this system purely on the basis of previous 'contact' with the police. There is no need for any objective evidence of criminality other than this 'contact', or being 'suspected' or, sometimes, actually arrested, in relation to a crime – but no conviction is needed to label someone in this way. Young people who are victims or witnesses are also targeted by the Top400 and subsequently police, even if they have not been convicted of any crime.

Unlike the Top600, the Top400 assessment list is not updated every six months, and once someone has been assessed as at risk, they will remain on the system for "at least two years" according to the Amsterdam Municipality, so individuals profiled as part of the Top400 – including victims and witnesses – and their families have to endure at least two years of these policing interventions.[74]

1.1.4 National Data Analytics Solution – West Midlands Police, UK Home Office & other police forces (England, United Kingdom)

Used for: Risk assessment and profiling
Created: 2016

The National Data Analytics Solution (NDAS) is a data analytics, risk assessment and crime prediction tool, created by West Midlands Police in England, in partnership with eight other police forces, including Greater Manchester Police and the Metropolitan Police Service,[75] as well as the UK Home Office.[76] Accenture, the multinational consultancy firm, has also assisted in the creation and development of NDAS as a contractor.[77] The NDAS has been given millions of pounds in funding every year by the UK Home Office

since its inception in 2016.[78]

West Midlands Police state that NDAS uses "advanced analytics and statistical techniques"[79] and describe it as:

> "[A] new, scalable and flexible analytics capability for UK law enforcement using advanced analytics to deliver insights to partners on agreed high priority operational and organisational issues."[80]

The NDAS uses machine-learning and predictive analytics to conduct "behavioural analysis" and "predictive modelling"[81] in order to create and provide individual predictions and profiles about people and their likely future actions. These are intended to inform and influence pre-emptive policing interventions.[82]

West Midlands Police has stated that the NDAS will "create meaningful insight and identify value driving patterns which should ultimately lead to crime prediction and prevention", enabling police "to action the insights generated", "make early interventions" and "evidence-based local interventions", in order to "prevent criminality… by proactively addressing threats".[83]

NDAS uses data from each participating police force "to create a rich picture of the law enforcement landscape", including police intelligence reports on individuals and 'events', stop and search data, drug use data and custody information.[84] West Midlands Police has said that it intends future partners providing data for the NDAS to include the National Health Service, the Department for Education, the Department for Work and Pensions, and the Department for Communities and Local Government,[85] giving it the ability to:

> "pull in data from other local public service providers (such as social care services, local authorities, education providers and other emergency services), private sector organisations, or open source data to deepen understanding of local services and the social context."[86]

This raises the prospect of policing and criminal justice decisions and outcomes being based on automated predictions that have been informed by data from essential public services including health, education, social welfare, and local authorities, among others.

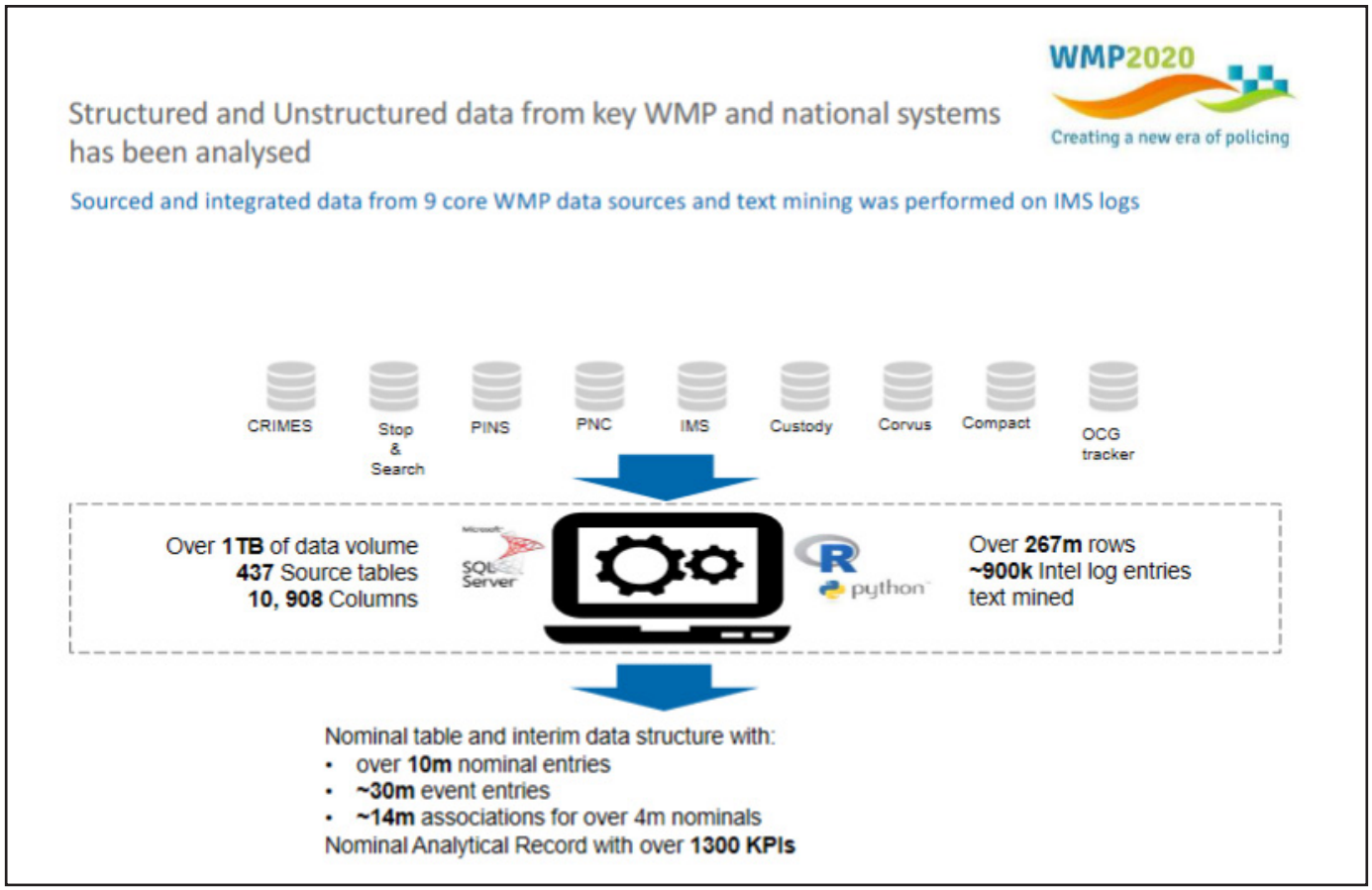


Figure 3: A slide from West Midland police's NDAS development programme 'Data-Driven Insight & Data Science Capability for UK Law Enforcement" detailing the data used.

West Midlands Police has also said that the data they use as part of NDAS may include social media data.[87] It has said it may use commercial marketing data from Experian,[88] which profiles and categorises individual postcodes into geodemographic classifications, using information including exam results, child benefits and income support, family and personal names linked to ethnicity, and 850 million data points.[89]

West Midlands Police acknowledges that it uses personal data and special category personal data and conducts 'sensitive processing' under the Data Protection Act 2018 (the UK law that implements the EU General Data Protection Regulation and Law Enforcement Directive) as part of the NDAS.[90] It also acknowledges that there is an "absence of a framework regulating analytics in law enforcement" and has said that it is "developing a proposed framework" itself.[91]

In early iterations of the NDAS, West Midlands Police developed a predictive risk model of 'co-offending – the commission of a crime by more than one person – using police records to identify the strongest 'predictors' that indicated whether someone was an 'influencer' of co-offending.[92] The police data used included the number of times an individual was stopped and searched, the number of intelligence reports about an individual in police records, the number of solo crimes committed by their associates, and mentions of the individual in drug habit or addiction records.[93]

The use of stop and search data in NDAS predictive models is deeply concerning, given that it is highly likely to result in discriminatory outcomes. Stop and search is a policing strategy that is consistently used in a discriminatory manner in the United Kingdom (as well as across Europe).[94] In 2019/20, Black people were four and a half times more likely than white people to be stopped and searched in the West Midlands Police area of England.[95] People from mixed ethnic backgrounds were six times more likely and Asian people were two and a half times more likely.[96] During this time, just 14 per cent of stop and searches led to an arrest.[97] Nationally, across England and Wales, Black people were more than nine times more likely to be stopped and searched than white people in 2019/20.[98] In this same time period, 76 per cent of stop and searches resulted in no further action.[99] This is the type of data being fed into West Midlands Police's predictive models.

Numerous other 'use-cases' for NDAS have been developed, designed and promoted by West Midlands Police over the last few years, including a model aimed at predicting 'most serious violence'.[100] The 'most serious violence (MSV)' model, trained using machine-learning tools,[101] intends:

> "to predict which individual nominals, who are already known to the police, are likely to commit their first most serious violence offence in the next 24 months."[102]

West Midlands Police state that the MSV predictions are conducted for the purpose of "enabling early interventions",[103] and this process is described as follows:

> "This use case looks to use advanced predictive analytics to identify indicators that lead to an individual committing their first serious violence offence with a gun or a knife. Through applying these indicators, the model can output a risk score for all nominals known to the police, which can be managed and utilised as supplementary intelligence."[104]

The MSV model uses police-held information about individuals, including uncorroborated police 'intelligence' reports:

> "Behavioural KPIs [key performance indicators] – these provide a summary of an individual's past behaviour derived from appearances within police data. Some examples of these behaviours are the past number of offences committed, the number of times a person has been a victim, or the number of times a person has been mentioned within an intelligence report"[105] (emphasis added).

The MSV model also draws predictive value from the relationships that individuals are alleged to have with other people, such as whether they are "closely linked" to "known individuals charged with MSV crimes".[106] In the context of the use of police data and 'intelligence' records, West Midlands Police have even admitted that:

"There is potential for bias to be present in the underlying dataset in terms of the recorded incidents of harmful/most harmful offences and within the intelligence reports."[107]

Yet, West Midlands Police continues to use this kind of deeply problematic data in their predictive models.

In its own analysis of NDAS, and specifically the MSV model, West Midlands Police admits that "there is currently no formal system in place" to analyse and demonstrate whether the MSV model will "improve the current system".[108] West Midlands Police states that it uses a "precision score" of only 50 per cent for the outputs of the model, meaning that the prediction must only be around 50 per cent certain to actually occur for it to be used to inform interventions.[109] Initial testing of the MSV model by West Midlands Police resulted in a precision score of just 54 per cent[110]

An independent review of the NDAS by the Alan Turing Institute Data Ethics Group (ATI DEG) and Independent Digital Ethics Panel for Policing (IDEPP) concluded that there were:

> "serious ethical issues… concerning surveillance and autonomy, as well as the reversal of the presumption of innocence on the basis of statistical prediction".[111]

The independent review questioned:

"whether it is ethical to use data in order to intervene for the public good against individuals before they have offended even though this approach will single out individuals who, like the public generally, may not have committed a criminal offence, or who will perhaps not go on to commit a future offence".[112]

The review also criticised the "reliability or biases in the 'evidence base'", noted that this would impact the "accuracy as well as the legitimacy of preventive action",[113] and stated that the NDAS "seeks to legitimise proactive and preventative policing."[114]

Despite these numerous and significant issues, West Midlands Police continue to develop NDAS for the purpose of operational deployment. Other predictive models in development include a geographic 'violent crime' prediction tool.[115]

1.1.5 The Sensing Project – Roermond police (Netherlands)

Used for: Risk assessment and profiling
Created: 2018

In Roermond, the Netherlands, police use an algorithmic risk assessment and crime prediction system called the 'Sensing Project', which attempts to profile and 'predict' the "likely perpetrators of pickpocketing and shoplifting" in a local shopping centre.[116] While the Sensing Project is described as a 'pilot', which was begun in 2019, it is clear that the system influences and results in active law enforcement action and operations.[117]

Roermond police created the Sensing Project algorithm following an internal study into pickpocketing and shoplifting at a local shopping centre in Roermond, using its own specific criteria in this analysis. Its 'study' found that around 60 per cent of suspects of those crimes in that location were Dutch nationals, with around 22 per cent of suspects being nationals of 'Eastern European' countries.[118] However, Roermond police selectively focused on a conception of crime it calls 'mobile banditry', which it identified as comprising economic crimes "committed by foreign groups of so-called 'bandits'",[119] which, it claims, is committed by people coming to the Netherlands from Eastern European countries.[120]

The model itself is specifically biased against non-Dutch nationals. Roermond police took the clearly biased step of excluding Dutch nationals from the definition of 'mobile banditry' and narrowing the focus of the Sensing Project.[121] Roermond police do not clarify which nationalities are defined as 'Eastern European', but associate it with people from Poland, Bulgaria, Romania, Lithuania, Bosnia and Herzegovina and Serbia. Roermond police also specifically associate mobile banditry with Roma. The latter association is not based on statistical fact as there is no national law enforcement data on suspects' ethnicity in the Netherlands, because Dutch police do not record it.[122]

Roermond police has created a profile of this conception of 'mobile banditry', and the Sensing Project algorithm assesses those travelling in Roermond against these criteria. The Sensing Project's criteria include:

- "Target travels by car. The Sensing project focuses on suspected pickpockets and shoplifters travelling by car
- Target is accompanied by passengers (other targets) in the car. The police consider multiple individuals in one car as an indication of a group of targets.
- Car takes a specific route. The police can retrace the route of the car from the ANPR cameras. The police predict suspicious routes using such data. A car travelling from Germany and headed towards the shopping centre is regarded as suspicious.
- Car may have a Romanian or German licence plate. (…) the police stated that a Romanian licence plate would generate points in the risk model.

- Car may be a (…) rental car from Germany, three to five years old.

- Car might be stolen, associated with previous criminality, or displaying false licence plates."[123]

The algorithm calculates an overall risk score, and when a high-risk score is produced, the car, its driver and the passengers are designated as 'suspicious' and the Sensing Project system generates a 'hit'. Police are notified of these 'hits'. The police officers on patrol have a wide discretion in relation to these notifications. They can choose whether or not to respond. If they do respond, the car is often intercepted, and the police check the identities of the driver and passengers and can record them in connection with the 'hit'.[124]

The system has a problematic flaw which facilitates a harmful and erroneous 'feedback loop': a car without previous connection or registration on police databases can be flagged because it meets the wide non-criminal criteria of the Sensing Project algorithm, but even if its occupants are not identified as suspected criminals following a 'hit', the car's details are still entered into the police system. Next time the car is driven in Roermond, this will result in the Sensing Project designating it with a higher risk score, due to the previous registration on the police system, despite there being no evidence or even police suspicion of criminality beyond the Sensing Project risk assessment. Amnesty International alleged that when they contacted Roermond police, it was:

> "unable to demonstrate the effectiveness of the Sensing project and admitted that the design of the project does not allow them to adequately measure its effectiveness in the prevention of pickpocketing and shoplifting."[125]

Amnesty International allege that, "the project is therefore inadequate to give trustworthy intelligence about the operational methods of pickpockets and shoplifters in Roermond."[126]

1.1.6 RADAR-iTE – Federal Criminal Police (Germany)

Used for: Individual risk assessment tool
Created: 2017

In 2017, the German Federal Criminal Police Office (BKA) launched a violence risk assessment tool called RADAR-iTE, a rule-based analysis tool intended to assess the 'acute risk' of 'potentially destructive offenders' of 'Islamist terrorism'.[127] The BKA describe the purpose and utility of RADAR-iTE as follows:

> "RADAR-iTE will provide important assistance in risk assessment through improved structuring and documentation of biographical histories of already known persons of the militant-Salafist spectrum. A nationwide uniform traceability of the assessments will be possible. By means of RADAR-iTE, the resources of German security authorities will be more specifically targeted at those persons who are identified as being at high risk of committing a violent act in Germany."

The tool uses police information on a person's 'observable behaviour', with case workers "drawing on as much information as possible about events in the person's life that are necessary for a better overall understanding". The tool classifies people as either 'high', 'conspicuous' or 'moderate' risk, and these are provided to a case handler, who decides on "individually appropriate intervention measures".

RADAR-iTE was apparently developed using the procedure for already established risk assessment instruments for the assessment of violent offenders. A further risk analysis system, RISKANT, is also being developed to build on RADAR-iTE to produce a "case-by-case threat assessment for identified high-risk persons".[128]

1.2 Predictive policing: Areas or locations

Certain predictive, profiling and risk assessment AI and ADM systems target areas or locations. These systems seek to make predictions or profiles about those areas and the actions – often alleged crimes and specific types of crime – that will be committed in those areas and locations, as well as when those crimes will be committed. These predictions are then used to inform law enforcement decisions, such as increased attention on the areas in which such future crime has been 'predicted'.

1.2.1 Crime Anticipation System – Dutch police (Netherlands)

Used for: Geographic crime prediction
Created: 2017

The 'Crime Anticipation System' (CAS) is a geographic crime prediction system used across the Netherlands since 2017.[129] The system tries to predict crime rates in specific areas using data from three sources: previous crimes, locations and known criminals data from BVI (Central Crime Database); aggregated socio-economic data including ages, incomes, benefits, house prices, population density, family sizes and number of parents from CBS (Demographics from Statistics Netherlands); and street location data from GBA (Municipal Administration).[130]

Originally among the data predictors used by CAS to predict crimes in a particular area was the number of 'non-Western' individuals with at least one foreign-born parent living in that area.[131] The software not only presupposed the existence of a correlation between ethnicity and crime, but also singled out a category of ethnicities to be of particular concern, given that the presence of 'Western', 'autochtone' individuals were not used as indicators. Furthermore, given that 'Western' was defined somewhat subjectively (for example, including individuals of Japanese or Indonesian origin, and including all European nationalities, apart from Turkish), CAS incorporated

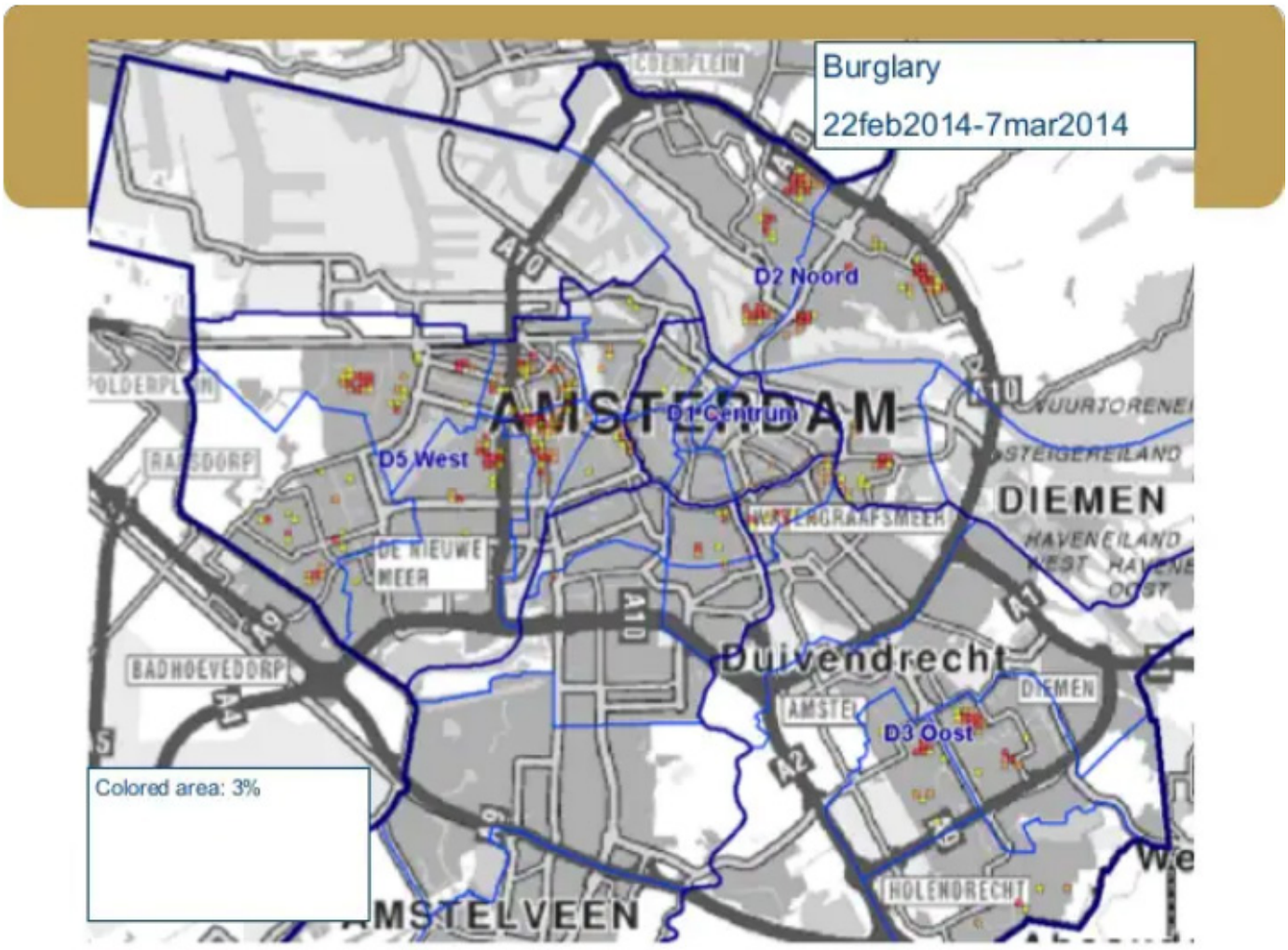


Figure 5: 'Experian's Mosaic 'Asian Heritage' profile (with added emphasis). Credit: Big Brother Watch.

highly questionable societal categorisations and biases. This indicator was removed in 2017.

However, numerous other data indicators used in CAS are also highly problematic and likely to be proxies for ethnicity data, as well as being discriminatory themselves. The way the system is designed, and the CBS data used – family sizes and parents, income and benefits, among others – shows a targeting of specific family structures and economic factors, profiling criminals and victims as part of "broken families".[132] The use of benefits data indicates that the police see poorer people as having a higher likelihood of committing crime, as there is no comparable data at the other end of the spectrum, such as the number of millionaires.[133] It is particularly concerning that an individual might be profiled for policing purposes on the basis of their access to essential services, such as welfare or benefits. These factors should not be regarded as relevant factors for determining whether someone may commit criminal offences. Areas are also profiled by the presence of previous offenders. The indicators from the BVI database refer to data about previously convicted criminals or so called "known offenders".[134] The distance to the closest known offender is calculated, as well as the number of recently active, known offenders that live within 500 and 1000 metres respectively.[135]

The CAS is ultimately a social construction of an extremely limited and biased form of reality, shaped by the socio-economic and crime data available to Dutch police, and a conception by Dutch police of "deviant physical traits, economic situations and behaviours".[136] CAS outputs are likely to reproduce and ultimately reinforce those same biases and inequalities.

1.2.2 Dynamic Evolving Learning Integrated Algorithm (Delia) – Italian police (Milan, Italy)

Used for: Geographic and individual crime prediction
Created: 2008

The Dynamic Evolving Learning Integrated Algorithm system (Delia),[137] formerly known as KeyCrime, is a crime analysis and prediction system, described by its developer as using "artificial intelligence and machine learning techniques".[138] The algorithm analyses criminal behaviour, particularly sequences of crimes allegedly committed by specific individuals or groups,[139] in order to create profiles of offenders as well as predictions of future criminality in specific areas. It was initially created in conjunction with Italian police in Milan. The software provider, called KeyCrime after the original name of the system, states that Delia:

> "creates and compares profiles compiled from millions of combinations of variables in order to link crimes that were committed by the same criminal of criminal gang." [140]
> and "helps users interpret the time intervals, target types and geographical areas (…) so that officers can predict the next possible event".[141]

Delia analyses not just geographic crime data, but purportedly 1.5 million variables,[142] from general information such as the date, time and place of an event, to more detailed information such as details about a suspect from witness and victim interviews.[143] This can include specific identifying details such as perceived age, height, body structure, skin, hair, eye colour, clothing and accent of the individual, information on any weapons involved, and the type, make, model and license plate of the vehicle used.[144] This information is then combined with police reports and crime records – more police 'intelligence' – as well as footage from video surveillance cameras.[145]

Delia allegedly predicts the precise place where a crime will take place, as well as a specific profile of the individual who will commit the crime. The predictions and profiles are then intended to be used by police at the specific time and place to identify and apprehend the alleged suspect, with KeyCrime stating that its predictive output:

> "allows Law Enforcement Agencies to make informed decisions about where and when to position officers for targeted activities of crime prevention or repression and use police resources in the most efficient manner possible."[146]

KeyCrime's Delia software specifically uses ethnicity data. The developer has openly said that:

> "In terms of investigation, information on the ethnicity of those who committed the crimes is essential; if developers of other software have decided not to collect this data, I have serious doubts about the effectiveness of their software."[147]

KeyCrime does not provide the accuracy of its software's predictions, with the developer of the system stating that, "this is a fact that we do not provide and that I personally consider unimportant".[148] KeyCrime alleges that Delia's predictions can be used in criminal prosecutions, stating that "Prosecutors benefit as well as Delia's documentation helps them prosecute criminals for multiple crimes" and can help to "convict perpetrators of their multiple crimes instead of just the last one for which they were caught".[149] Confusingly, the developer has also said that KeyCrime software "has no scientific value in court".[150]

1.2.3 SKALA – State Office of Criminal Investigations in North Rhine Westphalia (Germany)

Used for: Geographic crime prediction
Created: 2015

The System for Crime Analysis and Anticipation (SKALA) is a geographic crime prediction tool used in North Rhine Westphalia, in Germany, by the State Office of Criminal Investigations, to predict domestic burglary and car thefts in residential areas. It was originally tested and introduced between 2015 and 2017 and has since reportedly been implemented in 16 major police departments, making it the biggest predictive policing tool in Germany.[151]

SKALA uses police data and crime reports data, in addition to geographical, meteorological, and demographic data on specific residential neighbourhoods. The police data included information about the geo-locations of crimes, times of offences, methods, and items taken.[152] North Rhine-Westphalia authorities buy residential demographic data used in SKALA from a commercial location profiling and marketing company called Nexiga,[153] which provides "location intelligence and geomarketing solutions".[154] SKALA's predictions are represented on a map, which police then use to take specific action including covert surveillance, traffic controls and increased police presence at forecasted areas.[155]

An evaluation of SKALA by North Rhine Westphalia police in 2018 found that:

> "There are no robust statistical results that indicate an effect of SKALA - in the sense of a connection between measures and the subsequent events (e.g. arrests, prevention of WED [residential burglary])."[156]

Other similar systems used in Germany include the Pre-Crime Observation System (Precobs), a geographic crime prediction system focusing on burglary used by Bavarian police,[157] which an independent study found had no significant influence on crime rates;[158] Predictive Mobile Analytics for Police (PreMap), another geographic crime prediction tool focusing on burglary used in Lower Saxony;[159] and the burglary prediction systems KLB-operativ, used by the State Office of Criminal Investigations in Hessen, and KrimPro, used by Berlin police .[160]

1.3 Prosecution

This section considers a system which is used to profile and assess the risk of whether an individual will re-offend in future, for the purpose of determining whether that individual should be prosecuted.

1.3.1 Harm Assessment Risk Tool – Durham Police (England, United Kingdom)

Used for: Risk assessment and reoffending predictor
Created: 2017

The Harm Assessment Risk Tool (HART) is a machine-learning predictive system, used by Durham Constabulary to profile suspects of crime and predict their 'risk' of re-offending in the future.[161] HART produces risk scores: high, moderate or low. This automated risk score is used to decide whether to charge the suspect or deal with them via an out of court disposal programme called Checkpoint. If an individual is chosen to be dealt with via Checkpoint, they are not prosecuted, but instead have to agree to go through a programme of structured interventions aimed at rehabilitation.[162] If individuals successfully complete the Checkpoint rehabilitation programme, they will not be prosecuted. The automated risk score produced by HART can therefore have significant criminal justice consequences – the difference between prosecution or rehabilitation.

Only those who are rated as moderate risk by HART are eligible for the Checkpoint diversion and rehabilitation programme. The HART risk categorisations assess high risk as someone being likely to commit a serious offence within two years (murder, attempted murder, aggravated violent offences such as grievous bodily harm, robbery, sexual crimes, and firearm offences). Individuals assessed as likely to commit non-serious crimes within two years are designated as moderate risk, and those who are predicted to commit no new offences are considered low risk.[163]

The aim and principles behind the overall strategy and Checkpoint programme – to prosecute less, rehabilitate rather than criminalise, and divert people away from the criminal justice system – are extremely positive. However, the use of the automated system, HART, in making such decisions, the data it uses in producing its outcomes, and even the concept behind making criminal justice decisions based on predictions of future actions, is discriminatory, unethical and unacceptable.

The HART algorithm is based on a 'random forest' model, a form of machine-learning. As explained by its developers, academics and members of Durham Constabulary:

> "The random forest is constructed from 509 separate classification and regression decision trees (CART), which are then combined into the full forecasting model. Essentially, each tree is a model in and of itself, and produces a forecast which is then used as one vote out of 509 total votes. The votes are counted, and the overall forecast for the full model becomes the outcome which receives the most votes."[164]

HART was built on a dataset using approximately 104,000 custody events over a five-year period, from 2008 to 2012. It used 34 different pieces of information ('predictor variables') to produce a risk score. 29 of these 34 pieces of data focus on an individual subject's history of criminal behaviour as recorded in police and crime records and are considered 'behavioural predictors'. The number of police intelligence reports relating to the individual is also used as a predictive variable, as well as the individual's age, gender, and two forms of residential postcode.

Following an investigation by Big Brother Watch, a UK-based rights organisation, it was discovered that one of the postcode variables used in the original design and operation of HART was a 'consumer classification' marketing product called 'Mosaic', created and sold by a global data broker, Experian.[165] This commercially available marketing product profiles and categorises areas (postcodes in the UK) into socio-geodemographic profiles to give, according to Experian, a "pin-sharp picture of today's UK consumer".[166] These 'classifications' are created using information including census data (household composition, employment, occupations, ethnicity and health), land registry data, exam results, welfare and benefit data, "family/personal names linked to ethnicity", and many more pieces of information, among more than 850 million data points.[167]

All adults in the UK are profiled under this classification system, based on their postcodes, into what are often deeply discriminatory and offensive stereotypes. The Mosaic profiles included the following categories: Asian Heritage, Disconnected Youth, Crowded Kaleidoscope, Families with Needs or Low Income Workers.[168] Further offensive 'characteristics' were attributed to each of these profiles. People classified as being Asian Heritage were described as:

> "extended families with children, in neighbourhoods with a strong South Asian tradition… living in low cost Victorian terraces… when people do have jobs, they are generally in low paid routine occupations in transport or food service".169

Another offensive profile, Crowded Kaleidoscope, was considered to be made up of "multi-cultural" families likely to live in "cramped" and "overcrowded flats". The profiles were even given stereotypical names associated with them, with Abdi and Asha linked to Crowded Kaleidoscope. Low Income Workers were described as having "few qualifications" and being "heavy TV viewers" with names like Terrence and Denise, while Families with Needs were considered to receive "a range of benefits" and have names like Stacey.[170]

Durham Constabulary paid Experian for this offensive profiling information,[171] specifically "the 28 most common socio-geo-demographic characteristics for County Durham",[172] in order to use it as one of the predictive variables in HART, where it influenced criminal prosecution decisions. HART's developers and operational users at Durham Constabulary were aware of the potential for the postcode variables to lead to biased decisions, stating that:

> "Some of the predictors used in the model… (such as postcode) could be viewed as indirectly related to measures of community deprivation."[173]

They were also aware of the potential for the postcode variables to create 'feedback loops', reinforcing existing bias in policing and criminal justice:

> "one could argue that this variable risks a kind of feedback loop that may perpetuate or amplify existing patterns of offending. If the police respond to forecasts by targeting their efforts on the highest-risk postcode areas, then more people from these areas will come to police attention and be arrested than those living in lower-risk, untargeted neighbourhoods. These arrests then become outcomes that are used to generate later iterations of the same model, leading to an ever-deepening cycle of increased police attention."[174]

Despite this, they attempted to argue that due to the largely white demographic of Durham Constabulary's jurisdiction, the postcode stereotypes would not result in discriminatory predictions by HART:

> "due to the particular demographic of the force area, it is unlikely (although currently untested) that the residential predictors could currently be a proxy for race".[175]

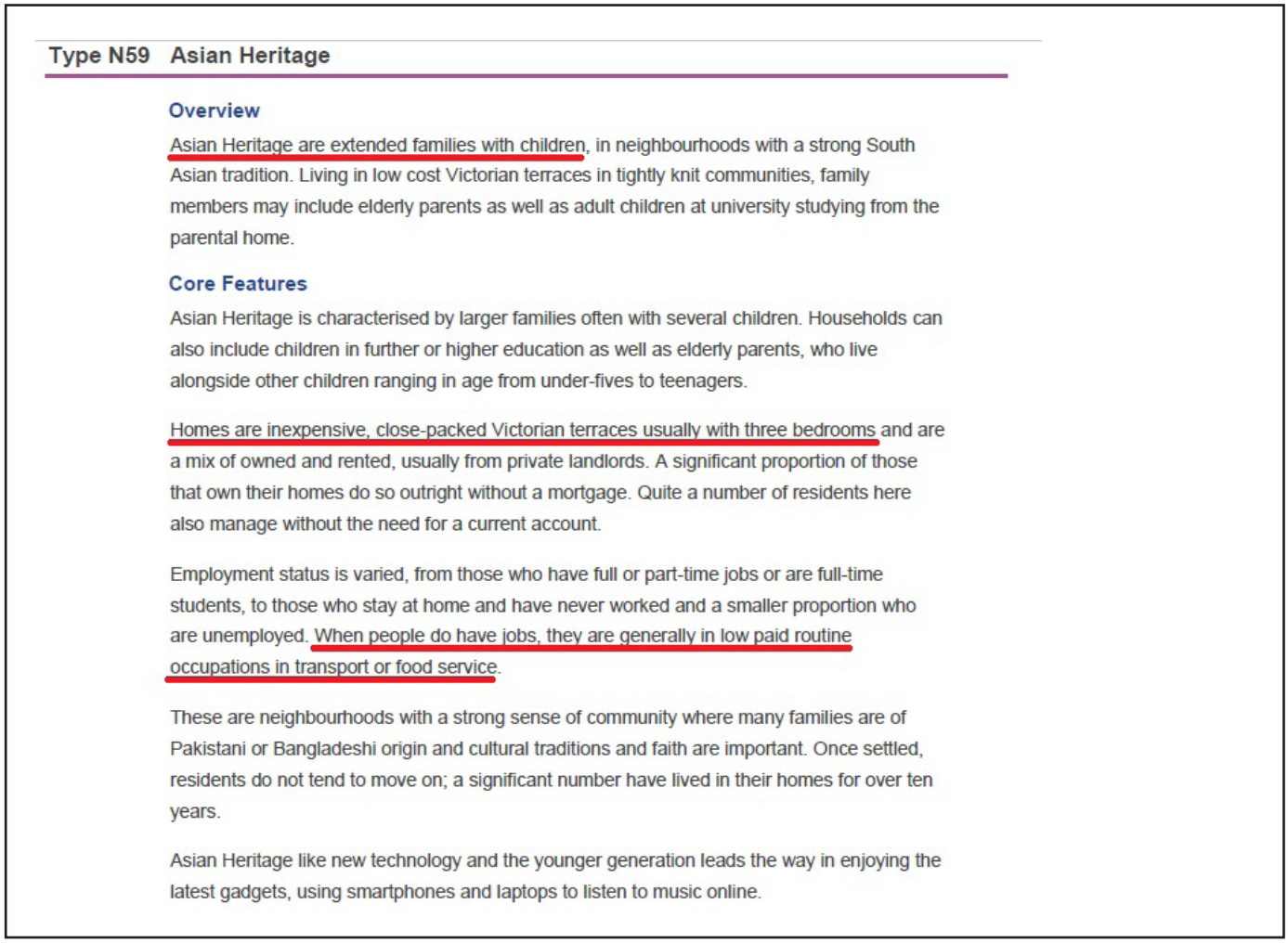
The organisation which exposed the use of this data within HART has also said that:

> "One of the academics instrumental to the development of HART stated to Big Brother Watch verbally that in their opinion the Experian Mosaic data was one of the strongest predictor variables and as such had a valid place in the tool."[176]

These assertions are deeply misguided. They show a complete lack of awareness of the difference between correlation and causation, ignoring the existence of people who do not fit such generalised stereotypes. Since the exposé of Durham Constabulary's use of this data, they have removed the Mosaic postcode variable. This does not stop potential issues stemming from the use of criminal history data or police 'intelligence' reports. Experian has also rebranded some of the Mosaic profiles: Asian Heritage has become Large Family Living and Crowded Kaleidoscope is now called City Diversity.[177] However, there is nothing to suggest that data used to create these profiles is in any way different, including the use of ethnicity and other proxy data. [178]

HART is also designed to over-estimate an individual's risk of re-offending in order to err on the side of caution, with "cautious errors, where the offenders' levels of risk are over-estimated" being intentionally favoured.[179] As a result, HART assesses a "sizeable proportion"[180] of people as being "high-risk", with the likelihood that innocent people, or people who fit a certain profile of offender, will be wrongly and unfairly profiled and subjected to prosecution. The accuracy of HART's model, calculated by the number of individuals assessed who did actually re-offend in line with HART's predictions, was just 62 per cent, little better than a guess. The 'high risk' predictions were just 52 per cent accurate.[181]

Durham Constabulary has announced that they expect HART's use to "expand", with its predictions "influencing all of the many other decisions that are made in the wake of bringing a suspected offender into police custody".[182]



Type N59 Asian Heritage

Overview

Asian Heritage are extended families with children, in neighbourhoods with a strong South Asian tradition. Living in low cost Victorian terraces in tightly knit communities, family members may include elderly parents as well as adult children at university studying from the parental home.

Core Features

Asian Heritage is characterised by larger families often with several children. Households can also include children in further or higher education as well as elderly parents, who live alongside other children ranging in age from under-fives to teenagers.

Homes are inexpensive, close-packed Victorian terraces usually with three bedrooms and are a mix of owned and rented, usually from private landlords. A significant proportion of those that own their homes do so outright without a mortgage. Quite a number of residents here also manage without the need for a current account.

Employment status is varied, from those who have full or part-time jobs or are full-time students, to those who stay at home and have never worked and a smaller proportion who are unemployed. When people do have jobs, they are generally in low paid routine occupations in transport or food service.

These are neighbourhoods with a strong sense of community where many families are of Pakistani or Bangladeshi origin and cultural traditions and faith are important. Once settled, residents do not tend to move on; a significant number have lived in their homes for over ten years.

Asian Heritage like new technology and the younger generation leads the way in enjoying the latest gadgets, using smartphones and laptops to listen to music online.

Figure 5: 'Experian's Mosaic 'Asian Heritage' profile (with added emphasis). Credit: Big Brother Watch.

1.4. Sentencing and Probation

There are several cases in which AI and ADM systems have been deployed to predict, profile and assess the risk of an individual's future re-offending, in order to influence and inform decisions about sentencing, and decisions about probation.

1.4.1 Offender Assessment System (OASys) and Offender Group Reconviction Scale (OGRS) – Prison and Probation Service (England and Wales, United Kingdom)

Used for: Risk assessment and re-offending prediction
Created: 2013

In England and Wales, the Prison and Probation Service use the Offender Assessment System (OASys), a hybrid of actuarial risk-assessment tools and human assessment, on adult offenders to make predictions and risk assessments which influence sentencing and probation outcomes.[183]

OASys combines actuarial methods of prediction with "structured professional judgement" to provide standardised assessments of offenders' risks and needs, in order to produce "individualised" sentence and "risk management" plans given to offenders, and "target interventions".[184] OASys is designed to assess the likelihood of reoffending, the risk of harm offenders pose to others and themselves, as well as assessing the individual's needs, producing a risk score: low, medium, high or very high risk.[185]

Originally developed and introduced in 2002 as a general risk assessment completed by prison or probation staff, the system was automated, and an electronic version of the tool was rolled out across both the prison and probation services in 2013.[186] The latest publicly available data shows that almost seven million prison and probation assessments have been collated within the central O-DEAT (OASys Data, Evaluation and Analysis Team) database for over one million offenders.[187]

The core OASys assessment considers offending-related needs, including assessment of the individual's personality, reasoning and temperament, and supposedly relevant 'external' societal factors.[188] This includes analysis of an individual offender's previous offences; their "social achievement" such as education, training and employment; their alcohol and drug misuse; as well as their "relationships", "lifestyle", "thinking and behaviour" and "attitudes".[189] This assessment is done by an assessor, usually a probation or prison officer, who assigns the offender a score based on each category.[190]

There are two key algorithmic risk predictions which form part of an OASys assessment: the OASys General reoffending Predictor (OGP), the alleged likelihood of non-violent offending, and the OASys Violence Predictor (OVP), the alleged likelihood of violent offending.[191] These predictors are calculated using the OASys assessment information, alongside data about an individual's offending record as well as "offender demographic information".[192] The Prison and Probation Service, which operates OASys, has previously found that the OGP and OVP predictive algorithms generated different predictions based on gender and race. It found that the relative predictive validity of OGP and OVP was "greater for female than male offenders, for white offenders than offenders of Asian, Black and Mixed ethnicity, and for older than younger offenders".[193] Both OGP1 and OVP1 predictors worked "less well for black offenders" and OGP1 also worked "less well for offenders of mixed ethnicity".[194]

Another predictive tool used in conjunction with OASys, or as an alternative predictor when the full OASys assessment has not been completed, is the Offender Group Reconviction Scale (OGRS). OGRS is another actuarial risk assessment tool, which attempts to predict an individual's likelihood of reoffending. OGRS uses data on an individual's age at sentence, gender, their number of previous criminal 'sanctions' (cautions and convictions), their age at first sanction and current offence, alongside a 'copas' rate (the volume and speed of an alleged criminal 'career'),[195] to predict an alleged probability of re-offending within two years, through a risk score of between 0 and 1.[196] OGRS includes separate predictors of general recidivism and violent recidivism.

The OASys assessment, and OGP, OVP and OGRS predictions are used within pre- and post-sentence reports used by the national Probation Service for England and Wales, informing decisions on both sentencing and probation. At the pre-sentence report stage, where the case is adjourned for 15 days to get a full report, these are based on an OASys assessment, but on-the-day reports, five-day reports and oral reports can be based upon an OGRS score (among other things).[197]

If on conviction the likely sentence for an individual for their offence is less than two years, the OASys assessment, as part of the pre-sentence report, can influence the choice between a custodial sentence, and a community order (such as unpaid work, fine or a curfew with a rehabilitative element). In probation hearings, a 'low' risk score can often mean an individual will be released 'on the papers' – an assessment based on written submissions and related documents – without a need for a hearing.

Lawyers have reported to Fair Trials that in practice, if the OASys assessment (and risk predictors contained within it) predicts someone as high-risk, probation services do not make further comment, but if someone is assessed as low-risk, probation services will comment on factors they believe the system should have taken into account, often countering 'low' risk assessments. Lawyers have also told Fair Trials that OASys reports are often poor copy and paste replications of previous assessments, with examples including one individual being wrongly labelled as having domestic violence issues due to names being mixed up and incorrect or out of date information being provided. Further, lawyers have reported to Fair Trials that OASys assessments are often biased, with, for example, members of the Traveller community often being ranked as higher risk than other people.

The National Offender Management Service also uses a prisoner categorisation algorithm to assess and decide which category prison, based on security level, prisoners should be held in. OASys assessments are one of the elements which inform these categorisations.[198]

1.4.2 Cassandra – Ukraine Ministry of Justice (Ukraine)

Used for: Risk assessment
Created: 2020

In September 2020, the Ministry of Justice of Ukraine announced that it had developed an automated risk assessment programme, which uses "elements of artificial intelligence", for use in its criminal justice system as part of supposed 'reforms' of penal institutions.[199]

Cassandra is said to 'automate' the process of providing pre-trial and pre-sentence reports. Cassandra analyses individual offenders to assess the potential risk of re-offending, producing assessments for pre-trial and pre-sentence reports, influencing judges' custodial decisions.[200] The Minister for Justice for Ukraine has said that Cassandra was designed to take an individualised approach to each person in the criminal justice system.[201] The Ukrainian Minister of Justice described Cassandra:

> "The programme analyses people's risks - whether they are addicted to alcohol or drugs, whether they have corrosive motives - they are greedy and can't afford the costs. We identify the person who caused the crime and then carry out individualized work with him or her. As a result, we will get a cheap system and a safe person."[202]

The probation officer inputs information into the system according to certain questions, with the system giving a certain score to each piece of information, and then summarises them, producing an estimate of the probability of committing a new crime from 0 to 97.[203] It appears that currently, Cassandra is in its early stages, with much of the risk assessment driven by a human operative. However, the Ukraine Minister for Justice has said that eventually Cassandra will automatically assess all the risk factors, as well as "analyse all the other data available in the state about the perpetrator."[204] Currently, there is no further information available in the public domain on the software, the specifics of the AI system or the data which will be used in Cassandra.

In February 2021, Ukraine's High Council of Justice also approved an "action plan to implement the concept of artificial intelligence in Ukraine" via a pilot project on AI in the court of first instance. This project is intended to "develop common standards for accounting for court decisions" in order to "identify unfair judicial practice" via textual analysis of judicial decisions.[205] The High Council of Justice believes that this system will allow the development of practical material on the most common minor administrative offences, inform the practice of law, eradicate difficulties in judicial interpretation, overcome errors, as well as reduce the workload of judges.[206]

1.4.3 RisCanvi – Catalonia Department of Justice (Spain)

Used for: Risk assessment
Created: 2009

RisCanvi is a risk assessment algorithm and software program developed in 2009 to address concerns within the Catalan prison system of violent re-offending by remanded and sentenced prisoners after release from prison following the completion of their sentence.[207]

The RisCanvi risk assessment is used to evaluate four outcomes: violent recidivism, self-directed harm/violence, violence within the prison facilities, and breaking of prison release conditions.208 To determine the probability of each of these outcomes, a predictive algorithm assesses each outcome, incorporating various risk factors and three additional variables: age, gender, and country of origin (Spanish or foreign).[209] An initial RisCanvi screening (RisCanvi-S) is carried out on all prisoners, incorporating only ten of these risk factors. If the screening results in a high risk score (out of either high or low), or if a prison official requires it, then the complete assessment (RisCanvi-C) using all 43 risk factors is carried out.[210] RisCanvi assessments are carried out on all prisoners entering Catalonian prisons and every six months, or at the discretion of a case manager.[211]

The risk factors used in RisCanvi include criminal record information, such as the start of the individual's 'criminal or violent activity', their age at the time of their offence, the length of their criminal convictions, time served in prison, escape attempts, and disciplinary reports. RisCanvi also uses information about an individual's education level, "low mental ability", "severe mental disorder", self-injury attempts, the criminal history of their family or parents, "criminal or antisocial friends", whether they are a "member of socially vulnerable groups", "sexual promiscuity", and problem drug or alcohol use, among other factors, to create its predictions.[212]

Each RisCanvi evaluation requires multiple interviews by several criminal justice representatives over several days.[213] The outcomes of these interviews are then input into RisCanvi, which generates a predictive risk score: low, medium or high. The committee deciding on what action should be taken in relation to the individual can accept the score or override this score with written justification, and then decide the next action, intervention, or program.[214] This can include allowing inmates who are assessed as low risk to get day-release or conditional release.[215] By February 2018, around 15,000 RisCanvi assessments had been carried out.[216]

It is very difficult for a prisoner to appeal a decision by RisCanvi, as judges would consider that the prison authorities, not RisCanvi, is responsible for any actual decision, even if the RisCanvi prediction has been a major influence or even 'the' major influence. One Spanish professor of law said:

> "As [the output of the algorithm] is taken into account by the administration but does not compel to anything, it doesn't matter that there is no regulation normalizing this situation (...) and art. 22 of the GDPR [which prohibits 'solely' automated decision-making] has a big hole for these cases."[217]

So far, RisCanvi has only been used in prisons. However, due to the close relation between prison authorities and other criminal justice authorities, RisCanvi evaluations are seen to have many external applications beyond prisons.[218] RisCanvi predictions have been used to create a procedure for the release of 'high-risk' inmates, with the RisCanvi score used as the basis for informing judicial authorities and police if a prisoner assessed as high-risk is about to be released, in order for them to prepare "community safety action to prevent or reduce the risk of recidivism".219 The system is also used to assess whether ex-prisoners are likely to re-offend.[220]

RisCanvi is not a fully automated system, given that many of the inputs are the result of in-person interviews. Five years ago, the Catalonian prison authorities considered introducing machine-learning techniques into RisCanvi, but decided against it on the basis that their analysis showed it would not substantially improve the result, nor have a more improved cost/benefit ratio.[221] However, the use of RisCanvi is an example of the kind of hybrid or partially automated algorithmic risk assessment systems that are used in criminal justice in Europe, which are increasingly becoming fully automated. Examples considered in this report include the OASys assessment system used by Prison and Probation in England and Wales and the newly developed and intended-to-be fully automated Cassandra system in Ukraine.

Another example of this, in Spain, is the Structured Assessment of Violence Risk in Youth (SAVRY), a non-automated violence risk assessment tool for juvenile offenders. A SAVRY assessment, undertaken by juvenile justice professionals, evaluates several risk factors about an individual, producing a risk score (low, moderate or high).[222] The final judgment is professional, not algorithmic. Studies of SAVRY's risk assessment outputs set against outputs from machine-learning re-offending risk assessment tools showed that the machine-learning system produced more accurate predictions. However, this came at the cost of fairness, as the machine-learning assessments were biased against non-Spanish 'foreigners' and people of specific national groups (classed in the study as Maghrebi,[223] Latin American, European and Other).[224] Foreigners who did not re-offend were almost twice as likely to be wrongly classified as high-risk by the machine-learning models than Spanish nationals and were less likely to be classed as unlikely to re-offend.[225] This shows the dangers of automating criminal justice risk assessment tools, with the likely result being that already existing bias and discrimination becomes integrated and reproduced by the automation.

2. AI & ADM SYSTEMS IN CRIMINAL JUSTICE: FUNDAMENTAL FLAWS

1.1 Discrimination

AI and ADM systems are created, trained, and operated using data. The system analyses that data and from that analysis, it produces an outcome or a decision. Some forms of AI and ADM systems, known as machine-learning, 'learn' how to make assessments or decisions based on their own analysis of data. Other more basic forms of ADM have specific forms of analysis coded in, in the form of algorithms. These automated tools, whether algorithmic or more advanced machine-learning systems, often require large amounts of data at the training stage, where the system or algorithm is trained to follow or recognise patterns in the data, or at the operation stage, where the system is actively operating in the field it was intended for, and making decisions or producing outcomes which are used or acted upon.

The way in which AI and ADM systems are designed, created, and operated can lead to biased and ultimately discriminatory outcomes. Bias can occur in the form of over-representations in the data that is used to train the system, or the data upon which the system carries out analysis. If the data used to train or operate the AI or ADM system is incomplete, inaccurate, or biased, this can lead to the system producing inaccurate or biased outcomes.

The type of AI or ADM designed or created for use in the criminal justice system will almost inevitably use data which is heavily reliant on or entirely made up of law enforcement data, crime records or other criminal justice authorities' data. These data and records do not represent an accurate record of criminality, but merely a record of law enforcement, prosecutorial or judicial decisions – the crimes, locations and groups that are policed, prosecuted and criminalised within that society, rather than the actual occurrence of crime. The data may not be categorised or deliberately manipulated to yield discriminatory results, but it will reflect the structural biases and inequalities in the society which the data represents. For example, policing actions resulting from or influenced by racial or ethnic profiling, or the targeting of people on low incomes, can result in biased data concerning certain groups in society.[226] The systematic under-reporting and systematic over-reporting of certain types of crime generally and in certain locations will also be represented in crime records and data.[227]

Although these systems are intended to predict an individual's likelihood of certain behaviour – criminal behaviour – in future, or the likelihood of criminal activity occurring in a specific area, these tools are, in reality, incapable of making individualised predictions. These systems use and analyse data from many individuals or areas, and then forecast aggregate group or area risk.[228] A person-oriented prediction or risk assessment 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.[229]

As law enforcement records are most accurately a record of law enforcement action, predictive or risk-modelling AI or ADM tools will most accurately provide predictions of future law enforcement action, such as the targeting of groups or areas that have historically been targeted by law enforcement.[230]

Discrimination in Europe

A significant challenge to assessing and analysing discriminatory police practices in Europe is that in most countries, the official collection of criminal justice data disaggregated by ethnicity, race and/or nationality is not available because it is either forbidden by law, or not standard practice, and there are not consistent practices for collecting and analysing such data across the Europe.

Discrimination in law enforcement

However, recent studies show that racially biased policing practices are prevalent throughout Europe. Data collected from a survey by the Fundamental Rights Agency (FRA) in 2018, showed that during a five year period, 66 per cent of individuals of Sub-Saharan African origin in Austria, and over half of respondents of South Asian origin in Greece were stopped and searched.[231] Over the same five year period, a huge proportion of Roma reported being stopped by the police because of their ethnicity: in Portugal (84%), Greece (63%), Czech Republic (57%), Romania (52%) and Spain (46%).[232] The same study carried out by FRA in 2010 found that in six out of ten Member states surveyed, ethnic minority respondents were stopped by police more often than majority (white) groups in the previous 12 months.[233]

A 2018 study by the JUSTICIA European Rights Network, a network co-ordinated by Fair Trials, found that in all twelve EU Member States that took part in the research, disparities exist for people of various ethnic, racial, and national origins, at least at some stages of their respective criminal justice systems and in some form.[234] Statistics showed a significant over-representation of various groups of 'foreigners' in crime rate statistics, pre-trial detention, and prison populations. In states where ethnicity data is available it was found that Black people are more likely to be arrested as a result of stop and search than white people, but less likely to be given out of court disposal, leading to higher prosecution rates for Black people.[235] Interviews with Italian, Hungarian and Romanian police officers, revealed a shared belief that all Roma have criminal characteristics.[236]

As several of the individual-oriented predictive policing case studies in this report come from the Netherlands, it is worth considering the levels of discrimination in policing and criminal justice in the Netherlands and specifically Amsterdam, where the Top400 and Top600 are used, to evidence how criminal justice data used in these systems will be imbued with bias. In the Netherlands, the FRA's 2018 study found that 20 per cent of those surveyed who had North African heritage reported being stopped, searched or questioned by the police due to perceived ethnic profiling, the second highest rate of all the countries surveyed. 13 per cent of those with Turkish heritage reported the same, also the highest rate of the countries surveyed.[237] An in-depth study of 'proactive' policing practices in the Amsterdam police force, carried out on behalf of the police themselves in 2012 found that the police's strategy involved the routine categorisation of the population into probable victims and probable perpetrators, with young Moroccan-Dutch and Eastern European men the prime targets.[238] The police study found that Dutch police widely use the term "target groups" (doelgroepen)—to refer to marginalised youths with Moroccan, Surinamese or Antillean (Caribbean) ancestry, people from Central and Eastern Europe, and parts of the white working class.[239] Dutch police were found to openly state that the decision to stop someone based on their ethnicity:

> "it depends on the district I am in and someone's colour. When I'm in West, I will stop and search more Moroccans and if I am in the Southeast of Amsterdam more Antilleans."[240]

Other examples recorded included stopping a Moroccan riding a particular type of bicycle:

> "Yesterday we saw this big guy on a chopper bicycle, with large wheels and huge handlebars. For a Moroccan, that's a weird combination. They normally drive on other bicycles, cheap ones"[241],

and a police officer making racial slurs about a Black man because he was riding a high-quality motorcycle.[242]

This outright racism and demonstrated disproportionate and discriminatory policing approach of "target groups" becomes a feedback loop within police tactics, a self-fulfilling prophecy. The police openly stated that they deliberately stopped and searched people from particular demographics on the basis that people from those demographics were over-represented in crime statistics.[243] This is the kind of information used in AI and ADM in the Netherlands to predict and profile.

Discrimination in detention

In those detained across Europe, there is also clear and widespread discrimination, with minoritised ethnic groups and 'non-nationals' disproportionately held in pre-trial detention. The UN Committee on the Elimination of Racial Discrimination (CERD) has raised concerns about disparities in who is held pre trial. It warns that, "persons held awaiting trial include an excessively high number of non-nationals" and:

> "persons belonging to racial or ethnic groups, in particular non-citizens – including immigrants, refugees, asylum-seekers, and stateless persons – Roma, indigenous peoples, displaced populations, persons discriminated against because of their descent, as well as other vulnerable groups which are particularly exposed to exclusion, marginalisation, and non-integration in society."[244]

In Council of Europe states, 40 per cent of all foreign nationals in prison were being held in pre-trial detention, compared to 25 per cent of all prisoners. In France, people born outside of France were three times more likely to be held in pre-trial detention (5.2%) than those born in France (1.8%);[245] in Belgium, 45.8 per cent of pre-trial detainees are non-Belgian nationals;[246] in Italy, 35 per cent of people held in pre-trial detention are non-Italian nationals;[247] and in Ireland, young people from Traveller communities are held in pre-trial detention four times more than their representation in the general population.[248] People held pre-trial—who are disproportionately minoritised ethnic people—are more likely to be sentenced to prison than similarly situated people released pre-trial.[249]

Data from across Europe shows that minoritised ethnic people are also disproportionately over-

represented in prison, relative to the percentage of the population they represent. Further research shows that these disparities cannot be attributed to increased criminality or other factors: the disparities are because of structural racism. Minoritised ethnic people face worse outcomes in their cases: they face longer prison sentences and are not granted non-custodial sanctions such as fines.[250] In the Netherlands, first-generation Antilleans and Turkish people received sentences 8 per cent and 11 per cent longer respectively than other people sentenced during the year of study, and researchers found evidence that "ethnic minority groups may thus receive different punishment because of their ethnicity".[251] In Belgium, people with a name perceived as Muslim were more likely to get convicted than a name perceived as white Belgian.[252] In France, a study found that one in four people born outside France were sent to prison, compared to one in six people from France,[253] with people born outside of France twice as likely to have their cases handled by fast-track proceedings, which provide fewer procedural rights.[254]

The manifestation of discrimination in AI systems

The fact that police action, including stop and searches, other forms of 'control', and arrest, as well as judicial and other criminal justice decisions, such as in pre-trial detention, bail, sentencing, and probation, are taken so disproportionately against minoritised ethnic people across Europe, means that they are significantly over-represented in police and criminal justice data.

The case studies discussed above illustrate how the over-representation of minoritised ethnic people in criminal justice data – and indeed, bias in other non-criminal justice data – result in discriminatory outcomes from AI and ADM systems. The same patterns are reproduced in many of these systems. Criminal justice data, containing significant over-representations is seen as having predictive value. Additional, non-criminal justice data, also containing significant over-representations, or with correlations drawn between factors representing deprivation and supposed criminality, is also used, further deepening and hard-wiring the discrimination, not just in terms of race or ethnicity but also nationality and socio-economic class.

For example, the NDAS uses stop and search data, a policing strategy which targets Black people almost ten times more than white people in England and Wales. KeyCrime's Delia system uses ethnicity data, at the developer's insistence. Roermond police's Sensing Project deliberately targets Roma and certain nationalities. HART uses commercial marketing profiles containing discriminatory profiles which stereotype South Asian people in the UK as poor and working class, as well as specifically using ethnicity data. RADAR-iTE is targeted at Muslims, while RisCanvi specifically includes data on nationality in making its risk assessments.

Police 'intelligence' reports are widely used as predictive indicators of criminality in these types of AI and ADM systems, including in many of those systems analysed in this report (the Top600, Top400, NDAS, HART, Delia, OASys, OGRS). The discriminatory potential of uncorroborated police reports, which will undoubtedly disproportionately represent minoritised ethnic people, is clear. These reports are allowed to influence police action and even prosecution decisions, despite the fact that in criminal proceedings such reports would be unlikely to be admitted as evidence and would likely be considered as nothing more than 'hearsay'. OASys also uses data on people's first contact with the criminal justice system, which will undoubtedly disadvantage those for whom increased attention from police and criminal justice authorities is a daily reality.

The significant over-representation of young Moroccan-Dutch men and children on the Top600; a clear example of the type of racial bias that all these factors lead to.[255]

The repeated use of socio-economic factors or data in many these systems may also act as proxies for ethnicity or race, and are problematic in their own right, given that they increase the likelihood of criminal justice action against people experiencing poverty. As an example, the Amsterdam Municipality has admitted that "a large number" of people profiled by the Top600 and Top400 "have outstanding fines" with the Dutch Central Fine Collection Agency.[256]

A common example used to illustrate how seemingly legitimate data can act as a proxy for other factors, such as race or ethnicity, is the use of home addresses or area codes.[257] AI and ADM systems will seek out correlations between area codes and the risk of re-offending – in other words, to identify which area codes have 'higher-risk' residents than others.[258] This approach overlooks the fact that there is very pronounced ethnic residential segregation in many European countries,[259] making it highly probable in practice, for AI and ADM systems to inadvertently establish a link between ethnic origin and risk. For example, Roma are vulnerable to this form of proxy discrimination, given that in many EU Member States, Roma are reported to live primarily in segregated areas.[260] Similarly, in Sweden, migrants from non-European countries face high levels of segregation.[261] Many forms of data can also act as proxy for race or ethnicity, such as financial information on income, data on access to welfare or benefits or other public services.

The use of financial or socio-economic data, such as in NDAS, CAS, HART and others, can also lead to targeting people based on their socio-economic background which is also problematic and discriminatory.

AI and ADM built and operated on data embedded with these types of biases and over-representations, which influence, inform or assist law enforcement or criminal justice decisions, can then re-enforce and re-entrench those biases.[262] When AI and ADM systems influence or inform criminal justice outcomes which repeat the discrimination inherent in historic criminal justice data, such as influencing police to repeatedly target individuals from the same demographics as those already over-represented in police records, those decisions and the resulting outcome will themselves be preserved in the data. This leads to a self-fulfilling prophecy whereby the predictions become true because police target those groups following the prediction or risk-assessment, which leads the system to strengthen its correlation between those groups and criminal justice outcomes, leading to self-perpetuating 'feedback loops' which reinforce patterns of inequality.[263]

Removing biased data is a widely touted solution to data bias in AI and ADM systems, but hidden biases can arise in AI systems in numerous ways. As considered above, the way in which AI & ADM systems are created and operated illustrate the difficulty, complexity, and sometimes impossibility, of preventing discriminatory outputs and effects of these systems.[264] In many cases, it might be difficult to pinpoint flaws either in the AI system itself, or in the training data that has caused the bias. It can also be difficult in practice to identify which variables are proxies for protected characteristics (and how they behave as such). Removing too many 'offending' variables might result in the AI system losing much of its functional utility and become unusable.[265]

These are fundamental, structural issues with the way AI and ADM in criminal justice are created, trained, operated and deployed, particularly those which predict, profile and make 'risk' assessments, leading to discrimination and contributing to a cycle of discrimination and inequality.

1.2 The right to a fair trial and the presumption of innocence

The right to be presumed innocent until proven guilty in criminal proceedings is a fundamental human right, a core part of the right to a fair trial, and one that is expressly recognised in, and safeguarded by the European Convention on Human Rights (ECHR)[266] and in EU law under Directive 2016/343 (the Presumption of Innocence Directive)[267] and the EU Charter of Fundamental Rights.[268]

The increasing use of AI and ADM in criminal justice threatens this right, and systems which predict, profile and risk assess individuals, before they have carried out the action for which they are being profiled, assessed, resulting in sanctions, completely undermine it. Individuals are labelled as criminal or criminalised following these predictions, profiles or risk assessments. The punishments and sanctions that follow, involving serious police action and criminal justice outcomes, including surveillance, stop and search, questioning, fines, arrest, prosecution, and a criminal record, as well as non-criminal justice sanctions such as the loss of benefits, or having children taken away, are carried out without proof of guilt according to law, and in the majority of examples, without even a formal charge.

Data-driven predictions, profiles and risk assessments as 'guilt'

Predictive, profiling and risk assessment AI and ADM systems like those considered above are part of a broader trend in law enforcement and criminal justice that is moving away from 'reactive' policing and criminal justice strategies, and towards 'preventative' or 'proactive' strategies.[269] These systems intend to pursue legitimate objectives of preventing, or reducing, crime or harm.[270] However, they do so in a way which targets individuals before they have carried out the action for which they are being profiled.

These systems operate in a grey area, using data to draw correlations between certain factors and 'guilt', data which is often directly or indirectly used as a proxy for other specific factors, such as race or socio-economic class, labelling people as criminals or making de facto findings of guilt as a result. These predictive and risk modelling AI and ADM systems designate individuals' risk levels based on their past actions; on information about their backgrounds, locations, and other aggregated data profiles; on protected characteristics and demographics, and also factors and information which include actions done by others, and even whether the individual is a victim of crime.

This undermines the presumption of innocence: people cannot and should not be pre-emptively judged as guilty or treated as akin to guilty of crimes that have not taken place, on the basis of their past actions - on their friendships, relationships and associations with others, or on the actions of other people who share similar characteristics or backgrounds, such as race or ethnicity, socio-economic status, neighbourhood or other factors. However, this is exactly what predictive, profiling and risk assessment AI and ADM systems are trained and operated to do.

Correlative links are drawn between this information, none of which are objective evidence of guilt, yet they are subjectively and unjustly seen as such by AI and ADM systems and those programming and operating them. As a result, these people or their neighbourhoods are then subjected to increased policing, stop and search, questioning, and even arrest, as well as threats of reports to public services, such as child protection services, or punitive sanctions from public services. These consequences and outcomes can occur even though the individual may not be formally suspected of any crime, without evidence that they have carried out the unlawful or wrongful action for which they have been profiled as likely to commit in future, and without any formal finding of guilt according to the law.

For example, ProKid uses data on the criminality of others in close proximity to a child against the child itself, as well as the child's own victimhood, and even the victimisation of others around the child, as indicators of that child's likelihood of future offending. The NDAS uses data on 'associates' to profile an individual and assess their likelihood of future criminality, while the CAS uses the proximity of known criminals as an indicator of likely criminality in an area. RisCanvi uses information on an individual's family or parental criminal history and "criminal or antisocial friends".[271] This is criminalisation by association, without any actual proof or finding of guilt.

Many AI and ADM systems in criminal justice also use uncorroborated police 'intelligence' reports, and data which records mere suspicion of involvement of crime without actual evidence. Of the case studies examined above, this data is used in the Top400, NDAS, HART and Delia. Dutch police even use an individual's Top400 designation as evidence of criminality, meaning that uncorroborated reports are washed via these systems, effectively coming out as a corroborated, evidential report of criminality.

The presumption of innocence is further undermined in the use of several AI and ADM systems which use of stop and search data in predictive, profiling and risk assessment. For example, West Midlands Police's NDAS system was trained using stop and search data. The use of this data in a crime prediction system assumes that being stopped and searched is an indicator of criminality or even likely criminality, when in actual fact it is merely a predictor of the demographics of people that police assume and project criminality onto, often in a racist and discriminatory way. In any case, stop and search data is not a valid or legitimate predictor of 'risk' given that the vast majority of stops and searches do not lead to arrest or any further criminal justice action.[272]

These systems also use socio-economic information as predictors of criminality. Socio-economic status and financial information are used in the CAS to indicate an area's propensity to criminal incidents. Similarly, HART uses commercial marketing profiles as a predictive value to assess individuals' risk of reoffending, which includes employment and occupation data, and benefits and welfare data. One of the indicators used in the Top600 to assess future criminality is appearances before a bankruptcy judge – a clear example of attempting to link criminality with socio-economic factors, merely resulting in direct targeting of people with financial issues: many of those on the Top600 have outstanding fines.[273] In addition, CAS profiles areas as criminal using information on incomes, access to benefits other public services, and even information on family structures – a clear targeting of people experiencing economic hardship.

No protection from legal conceptions of the presumption of innocence

While it is clear that predictive, profiling and risk assessment AI and ADM systems used in criminal justice and the actions they influence and justify can infringe the presumption of innocence from a moral and ethical viewpoint, it is unclear whether these systems also violate the legal presumption of innocence under European and international human rights law and EU law.

Although these systems do not technically or directly 'convict' people, they allow the police to treat legally innocent individuals as pseudo-criminals, resulting in increased police attention and actions. They can also result in further non-criminal justice action, such as denial of access to public services, or reporting to social services – effectively 'punishing' them on account of their profiles. This damages the fundamental human rights principle that such a matter of guilt or innocence can only be determined by means of a fair and lawful criminal justice process.[274]

The legal presumption of innocence generally only has effect when someone is charged with a crime. For example, the Presumption of Innocence Directive applies to natural persons who are "suspects" and "accused persons" in "criminal proceedings", from the moment they are suspected or accused of a crime.[275] However, there is some ambiguity about the exact stage at which an individual attains the status of a "suspect" under the Presumption of Innocence Directive,[276] and about whether the scope of the Presumption of Innocence Directive extends to decisions to profile or designate an individual as a suspect, or someone 'at risk' of committing a crime. The European Court of Human Rights (ECtHR) has so far taken the position that measures undertaken

pre-charge, as a general rule, fall outside the scope of the presumption of innocence.[277] It has also held that preventative measures, such as surveillance, do not amount to a criminal sanction for the purposes of Article 6 ECHR.[278]

However, ECtHR case law also evidences that the presumption of innocence can be a broader principle that goes beyond the strict boundaries of a criminal trial.[279] 'Reasonable suspicion' is defined by the ECtHR as:

> "(…) when a criminal offence has been committed with the presupposition that facts and information are able to satisfy an objective observer that the person concerned has committed such offence".[280]

'Suspect' is defined as:

> "(…) when the domestic authorities have plausible reasons for suspecting that person's involvement in a criminal offence".[281]

'Charge' is described as:

> "[T]he official notification given to an individual by the competent authority of an allegation that he has committed a criminal offence", a definition that also corresponds to the test whether "the situation of the [suspect] has been substantially affected"".[282]

In their use of AI and ADM systems to predict, profile and assess risk, leading to punitive sanctions and outcomes without any formal finding of guilt, 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.

These types of high impact, fact-sensitive decisions in criminal justice should never be delegated to automated processes, particularly those which operate in such deeply pernicious and problematic ways, identifying correlations rather than causal links between an individual's characteristics and their likely behaviour. In order to properly protect individuals and their fundamental rights, specifically the right to a fair trial and the presumption of innocence, the use of AI and ADM systems in criminal justice to predict, profile and assess risk must be prohibited.

1.3 Transparency and accountability

AI and ADM systems can have a significant influence over criminal justice decisions, including those that result in surveillance, arrest, the deprivation of liberty, prosecution and sentencing. It is crucial that individuals affected by these systems' decisions are able to challenge them meaningfully and effectively. The right to a fair trial and the right to liberty can only be exercised effectively in practice, where suspects and defendants have the facilities and capabilities to challenge decisions regarding them.

Transparency is a fundamental aspect of these rights, and the adversarial process that underpins the right to a fair trial. Individuals subject to police and criminal justice action must be able to understand and contest decisions made about them and have access to materials that inform decisions regarding them. Defendants should be notified[283] of a decision which has a meaningful or legal impact on them and be given unrestricted access to information about their case,[284] so that they can contest their case and challenge the accuracy and lawfulness of those decisions, providing meaningful accountability. Criminal justice decision-makers must also give reasons for their decisions, as well as stating how and in what way decisions were influenced, through fully reasoned, case-specific, written decisions.

However, AI and ADM systems that are used to influence, inform and assist law enforcement and criminal justice decisions through individualised predictions and risk-assessments often have technological barriers that prevent effective and meaningful scrutiny, transparency, and accountability.

This lack of transparency is both as a result of deliberate efforts to conceal the inner workings of AI and ADM systems for legal or profit-driven reasons, and of the nature of the technology used to build these systems that can be uninterpretable to non-experts, and some forms which are uninterpretable even to experts. This makes it difficult, if not outright impossible, to subject these systems, and the decisions they make and influence, to meaningful impartial analysis and criticism.

This opacity of the automated decision-making process can occur where these systems are developed by for-profit companies, whose primary motives are financial. This process involves a design of the AI or ADM product which has little to no meaningful input from the public or public institutions, and the AI or ADM system is seen as a product, to be sold for profit. Often, key elements of the system, if not the entire system itself, uses or contains software which is proprietary. Details of how the system or

software utilised are designed and how they make decisions and assessments are, in many cases, closely guarded as intellectual property and/or trade secrets that are given protection by laws common to private and free-market capitalist economies.[285] Many of the details about the systems considered as case studies above have only been discovered or exposed following extensive investigation and research, and many details of these systems and their inner workings are still unknown, deliberately kept as such for profit-driven reasons by technology and data companies. For example, many of the algorithmic systems developed by major tech companies such as Google or Facebook,[286] or other technology companies with law enforcement and criminal justice contracts across Europe (including Europol) and the US such as Palantir,[287] are proprietary, legally shielded as trade secrets.288 In the examples considered in this report, several systems involve or are made up of proprietary software, with the full details of their inner workings not publicly available, including the Sensing Project, Delia, SKALA, Precobs, KrimPro and HART. However, the lack of publicly available information on many AI and ADM systems means that several more of those case studies considered may also involve proprietary software.

The other fundamental reason for the lack of transparency within some forms of AI or ADM systems are that they are sometimes based on certain algorithmic and computational processes, such as machine-learning algorithms, and 'neural networks', a form of machine-learning, which learn, analyse and compute information, producing outputs and decisions in ways which cannot be analysed and sometimes even understood by its operators, let alone those subject to their outputs.[289] Some machine-learning algorithms are simply too complex to be understood to a reasonable degree of precision,[290] especially where machine-learning systems incorporate neural networks. Decision-making processes of this kind have been described as 'intuitive' because they do not follow a defined logical method, making it impossible to analyse the exact process by which a particular decision is reached.[291] It has also been suggested that some AI systems are uninterpretable to humans because the machine-learning algorithms that support them are able to identify and rely on geometric relationships that humans cannot visualise. Certain machine-learning algorithms are able to make decisions by analysing many variables at once, and by finding correlations and geometric patterns between them in ways that are beyond human capabilities.[292]

This opacity and lack of accountability poses a serious challenge to fairness and justice. Individuals cannot respond to or contest a decision made about them if they are not notified or otherwise made aware of the decision, if they are unable to understand how a decision has been made, or if they cannot understand why it has been made, such as is often the case with AI and ADM systems' decisions. People without little or no digital literacy, let alone the level of technical expertise required to analyse AI and ADM outputs, are unable to interpret such output and contest it, or prepare a case or defence in relation to it. Research has shown that the lack of case-specific reasoning in pre-trial detention decisions is already a serious challenge in many EU Member States, and AI and ADM systems risk worsening the standardisation of such decision-making processes by obscuring or concealing key elements of the process. Defence rights are also undermined if defendants are not aware or able to access meaningful routes to challenge or appeal the decision-making process and the decision itself. The effective exercise of the rights of the defence must be recognised as a crucial test for determining whether an AI or ADM system is sufficiently explainable and intelligible.

These fundamental issues AI and ADM systems have with meeting basic criminal justice transparency and accountability requirements are a clear indicator of their incompatibility with criminal justice. Where full transparency and accountability cannot be guaranteed, AI and ADM systems should not be used in criminal justice. This includes the ability to provide clear explanations of AI or ADM system processes and outcomes understandable by a lay-person, as well as meaningful accountability, through notification of any decision influenced, informed or taken by an AI or ADM system, and clear routes to contest or challenge that decision as well as any impact or outcomes the decision leads to.

3. OTHER KEY ISSUES

1.1 Human input and oversight: automation bias and legal loopholes

One of the main challenges of AI, automated, or semi-automated decision-making systems is that of 'automation bias' – the tendency to over-rely on automated processes in ways that can cause or errors in decision-making or legitimise errors or bias produced by an AI or ADM system.

Automation bias occurs primarily due to the perception that AI and ADM systems are generally neutral, reliable and, therefore, trustworthy. Automated evaluations have been found to be particularly salient to decision-makers, and research has shown that users of automated decision-making systems have a tendency to place greater weight on automated assessments over other sources of advice.[294]

The disproportionate influence of automated systems can undermine the quality of decision-making, by discouraging its users from consulting a wider range of factors that could inform more accurate decisions.

Most AI and ADM systems currently being used to inform or assist criminal justice decision-making do not completely replace human decision-making. They are instead designed and deployed to be used as decision aids, whose outputs are factored into consideration for the purposes of human decision-making. The phenomenon of automation bias however, raises questions about whether AI and ADM systems are being used in reality in accordance with their intended purpose as decision aids, and not as de facto replacements for human decision-making processes.

A simple requirement to have a human decision-maker 'in the loop' or to have a human decision-maker review or check the automated decision is insufficient, because this risks overestimating the capacity or willingness of human decision-makers to question and overrule automated decisions. A mere requirement to have an automated decision reviewed by a human, on its own, could reduce the human review into a rubber-stamping exercise which, in practice, is no oversight at all.

It is clear from the many examples considered above that AI and ADM systems currently being used in Europe often produce inaccurate, discriminatory outputs, despite the fact that there are human decision-makers involved in these processes. This evidences how human intervention cannot compensate or correct for the problematic outcomes of an AI or ADM system.

Loopholes in EU data protection and privacy law

There are notable gaps in the existing legislative framework governing ADM systems under both the General Data Protection Regulation (GDPR) and the Law Enforcement Directive (LED). These ambiguities and potential loopholes can be exploited in ways that seriously undermine the general prohibition of ADM processes, and adversely impact human rights.

Article 22 of the GDPR provides that data subjects have the right not to be subject to decisions 'solely' based on automated processes, where they produce 'legal effects' concerning them, or where they 'similarly significantly affect' them. The LED, the EU legislation that governs the processing of data for criminal justice purposes, has a very similar provision at Article 11, which requires Member States to prohibit decisions based solely on automated processing, where they produce 'adverse legal effects' on the individual, or effects that are 'similarly significant'.

However, neither the GDPR or LED address situations where automated processing is not the sole basis of a decision, but a primary influencer. In reality the difference between a fully automated decision and a decision made with a 'human-in-the-loop' is not always clear, and because of this strict classification, AI and ADM systems are able to be used and have significant legal effects without the corresponding safeguards. Stronger legal standards are needed to make sure that semi-automated decision-making processes do not become de facto automated processes.

Secondly, the prohibition on automated decision-making is subject to two very broad exceptions. Automated decisions are prohibited under the GDPR and LED, "unless authorised by Union or Member State law" and there need to be "appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention".[295] These provisions give extremely wide discretion on Member States to override the general prohibition.

In recognition of this challenge, the European Data Protection Board has recommended that in order for decisions to be regarded as not 'based solely' on automated processing for the purposes of Article 22 GDPR, there has to be 'meaningful' human oversight, rather than just a token gesture.[296]

1.2 Children's rights

In all actions concerning children, the best interests of the child must always be a primary consideration.[297] The EU has recognised this in the context of criminal justice, with Directive 2016/800 (procedural safeguards for children in criminal proceedings)[298]

setting out minimum rules concerning the rights of children in conflict with the law. The right to private and family life is also protected under the ECHR.[299]

It is of serious concern that predictive and profiling AI and ADM systems in criminal justice are targeted at children, with several examples considered in this report. ProKid, Top400 and Top600 are being used inform decisions regarding children under the age of 18, with ProKid used to assess children as young as 12. These are new and 'experimental' ways of policing, but they are being trialled on children, and making sensitive decisions that could have a profound impact on their lives.

Ostensibly, these systems and the models of social care they partially serve are intended to address complex social issues and to reduce harm. However, these aims are best achieved through sensitive, case-specific decisions, rather than through data-driven policing tools that channel children through 'care' or 'punishment' based on correlations. While some of the policies that these AI and ADM models support have positive intentions in terms of preventing recidivism, it is doubtful that they are, in reality, effective at promoting the best interests of children. Children have been wrongly removed from parents on the basis of faulty risk assessments, been put under police surveillance and repeatedly harassed for extensive periods of time, and have even suffered deprivations of liberty for crimes that they have not committed.

Where possible and appropriate, children should be kept out of the formal justice system to promote their best interests, and law enforcement and other authorities should be focused on diverting children away from the criminal justice system. These systems operate by designating criminal suspicion on children via these predictions and profiles, and as such, they are effectively designed to facilitate the entry of children into the criminal justice system. Further, given that these systems often result in increased surveillance and general harassment by police, and they facilitate the police's role in matters relating to 'care', they promote direct contact between children and the criminal justice system.

The fact that these AI and ADM systems all use information and indicators about children that are non-criminal justice related, and often even non-crime related, is also extremely troubling. It is particularly perverse that ProKid and Top400 both include child victims and child witnesses in their assessments of criminal 'risk', treating them as potential criminals. The Top400 also uses education information, including if a child has attended more than three schools or has been absent, and deliberately targets the 'brothers and sisters' of other children, who themselves may have been caught up for non-criminal reasons, with potential criminal justice consequences. The potential for structural discrimination and inequality to find its way into AI and ADM assessment results, as well as feedback loops that re-entrench discrimination and inequality, are likely to lead to unchallengeable narratives of criminal families, with children unable to shake off the stigma and designation of criminality.

Children who are suspects or accused persons in criminal justice should be given particular attention in order to preserve their potential for development and reintegration into society.[300] The use of AI and ADM systems which profile and designate children as criminals; which does so based on statistical prediction, and uses non-criminal data to do so, does not facilitate this aim, and it not only does not promote the best interests of the child, but actively works against them.

4. RECOMMENDATIONS: PROHIBITION AND OTHER SAFEGUARDS

4.1 Prohibition on predictive, profiling and risk assessment AI and ADM

- Ban the use of AI and ADM by law enforcement and judicial and criminal justice authorities to predict, profile or assess people's risk or likelihood of 'criminal' behaviour.

4.2 Safeguards for other forms of AI and ADM systems in criminal justice

In relation to all other forms of AI and ADM used by law enforcement and criminal justice authorities (which do not carry out predictive, profiling or risk assessment functions), strict legal safeguards are needed.

Safeguards against discrimination

- Implement mandatory, independent testing for biases in the design and pre-deployment phase, as well as continuously post-deployment.

- In order to carry out this testing, data on criminal justice must be available, and such data must be separated by race, ethnicity and nationality.

In relation to AI and ADM in criminal justice which do not carry out predictive, profiling or risk assessment functions, a rigorous testing regime is the bare minimum required to lessen the risk of discrimination and ensure equality before the law and non-discrimination. AI or ADM systems should never be used or even 'trialled' in real-world situations where they have actual effects on individuals or criminal justice outcomes, before they have been tested in this way. Impacted groups and individuals must also be involved in this process. If these tests are not carried out, and/or if an AI or ADM system cannot be proven to be non-discriminatory, it should be legally precluded from deployment.

To carry out the necessary mandatory bias testing, data on criminal justice must be available, and such data must be separated by race, ethnicity and nationality. However, such tests are not feasible in much of Europe, as in many European states, including most EU Member States, the official collection of criminal justice data disaggregated by ethnicity, race and nationality is not available because it is either forbidden by law, or not standard practice, and there are not consistent practices for collecting and analysing such data across the EU.[301] Without the relevant data to analyse AI and ADM system outcomes, there can be no way of detecting whether there is bias or discrimination. The absence of racial and ethnic data could also prevent pre-emptive measures to combat racial bias. It is doubtful that developers will be able to design systems free from racial bias, if they have no data against which to measure their performance. Data needed for monitoring and evaluation purposes will also need to have been collected starting from well before the introduction of the AI and ADM system, so that a proper pre- and post-analysis comparison can be made.

Ensuring transparency and accountability

- AI and ADM used in criminal justice must be transparent, with system processes open-source and not subject to legal protections such as trade secrecy or intellectual property requirements.

- Their outputs must be able to be understood and scrutinised by their controllers, subjects of decisions (such as suspects and accused persons), as well as the general public.

- Individuals must be notified whenever there has been an AI or ADM system involved, assistive or otherwise, that has or may have impacted a criminal justice decision.

- Human decision-makers must evidence how and in what way decisions were influenced, through fully reasoned, case-specific, written decisions, including what factors influenced a decision, and whether this involved AI or ADM system outputs.

- There must also be clear routes for challenge or redress for individuals attempting to contest or challenge AI and ADM decisions, or the systems themselves.

5. FOOTNOTES

1. Council of Europe, 'Unboxing Artificial Intelligence: 10 steps to protect Human Rights', May 2019. https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64

2. High Level Expert Group on Artificial Intelligence (AI HLEG), 'Ethics Guidelines for Trustworthy AI', 8 April 2019, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

3. Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM (2021) 206, 21.04.2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX-%3A52021PC0206

4. CAHAI, 'Feasibility study on potential elements of a legal framework for the development, design and application of artificial intelligence', Strasbourg, 17 December 2020 https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da

5. AlgorithmWatch, 'Automating Society: Taking Stock of Automated Decision Making in the EU', January 2019. https://algorithmwatch.org/en/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf

6. These are the product of open-source research and investigation, freedom of information requests, reviews of published literature, and interviews with impacted people, lawyers and practitioners, and experts in the field.

7. Abraham et al, 'ProKid 12- identification tool evaluated', (2011) WODC DSP-groep, https://english.wodc.nl/onderzoeksdatabase/evaluatie-signaleringsinstrumenten-prokid.aspx

8. Ibid.

9. La Fors, Karolina 'Legal Remedies For a Forgiving Society: Children's rights, data protection rights and the value of forgiveness in AI-mediated risk profiling of children by Dutch authorities' (2020) Computer Law & Security Review vol. 38, https://www.sciencedirect.com/science/article/pii/S0267364920300352.

10. Wientjes et al, 'Identifying potential offenders on the basis of police records: development and validation of the ProKid risk assessment tool', October 2017, Journal of Criminological Research Policy and Practice 3(1):00-00

11. Ibid.

12. Wientjes et al, 'Identifying potential offenders on the basis of police records: development and validation of the ProKid risk assessment tool', October 2017, Journal of Criminological Research Policy and Practice 3(1):00-00

13. Government of the Netherlands, 'National information system puts an end to police's ICT problems' (2011), https://www.government.nl/latest/news/2011/09/19/national-information-system-puts-an-end-to-police-s-ict-problems; La Fors, Karolina 'Legal Remedies For a Forgiving Society: Children's rights, data protection rights and the value of forgiveness in AI-mediated risk profiling of children by Dutch authorities' (2020) Computer Law & Security Review vol. 38, https://www.sciencedirect.com/science/article/pii/S0267364920300352

14. Wientjes et al, 'Identifying potential offenders on the basis of police records: development and validation of the ProKid risk assessment tool', October 2017, Journal of Criminological Research Policy and Practice 3(1):00-00.

15. La Fors-Owyczynik, Karolina 'Prevention strategies, vulnerable positions and risking the 'identity trap': digitalized risk assessments and their legal and socio-technical implications on children and migrants' (2016), https://www.tandfonline.com/doi/full/10.1080/13600834.2016.1183307

16. Abraham et al, "ProKid 2- identification tool evaluated" (2011) WODC DSP-groep, https://english.wodc.nl/onderzoeksdatabase/evaluatie-signaleringsinstrumenten-prokid.aspx

17. La Fors, Karolina 'Legal Remedies For a Forgiving Society: Children's rights, data protection rights and the value of forgiveness in AI-mediated risk profiling of children by Dutch authorities' (2020) Computer Law & Security Review vol. 38, https://www.sciencedirect.com/science/article/pii/S0267364920300352

18. Ibid.

19. La Fors-Owyczynik, Karolina 'Prevention strategies, vulnerable positions and risking the 'identity trap': digitalized risk assessments and their legal and socio-technical implications on children and migrants' (2016), https://www.tandfonline.com/doi/full/10.1080/13600834.2016.1183307

20. La Fors-Owyczynik, Karolina, 'Profiling 'Anomalies' and the Anomalies of Profiling: Digitalized Risk Assessments of Dutch Youth and the New European Data Protection Regime' (2016), https://link.springer.com/chapter/10.1007/978-3-319-48342-9_7)

21. La Fors, Karolina 'Legal Remedies For a Forgiving Society: Children's rights, data protection rights and the value of forgiveness in AI-mediated risk profiling of children by Dutch authorities' (2020) Computer Law & Security Review vol. 38, https://www.sciencedirect.com/science/article/pii/S0267364920300352

22. Abraham et al. 'ProKid 2- identification tool evaluated' (2011) WODC DSP-groep. https://english.wodc.nl/onderzoeksdatabase/evaluatie-signaleringsinstrumenten-prokid.aspx

23. van der Meer, Marije 'Ik ben bang en wantrouwend geworden' (2013), https://www.ouders.nl/artikelen/ik-ben-bang-en-wantrouwend-geworden (translated from Dutch)

24. La Fors, Karolina 'Legal Remedies For a Forgiving Society: Children's rights, data protection rights and the value of forgiveness in AI-mediated risk profiling of

children by Dutch authorities' (2020) Computer Law & Security Review vol. 38, https://www.sciencedirect.com/science/article/pii/S0267364920300352

25. Abraham et al. 'ProKid 2- identification tool evaluated' (2011) WODC DSP-groep. https://english.wodc.nl/onderzoeksdatabase/evaluatie-signaleringsinstrument-en-prokid.aspx

26. Gemeente Amsterdam 'Top600 – Informatie voor professionals, https://www.amsterdam.nl/bestuur-organisatie/organisatie/overige/acvz/top600/ (translated from Dutch). The 'core participants' of the Top600 are: Action Center for Safety and Care (AcVZ), the Municipality of Amsterdam (city districts, Work, Participation & Income (WPI), Preventive Intervention Team (PIT), GGD and Education Compulsory Plus Office), Police, Public Prosecution Service, Netherlands Probation Service (RN), Youth Protection Region Amsterdam (JBRA), William Schrikker Group (WSG), Probation Service Inforsa and the Salvation Army (LdH). Open Research Amsterdam 'Halfjaarmonitor Top 400/600 Actiecentrum Veiligheid en Zorg (2017), https://open-research.amsterdam/nl/page/61334/halfjaarmonitor-top-400-600-actiecentrum-veiligheid-en-zorg

27. Politie 'Persoonsgerichte Aanpak (PGA)', https://www.politie.nl/themas/persoonsgerichte-aanpak-pga.html

28. Ibid.

29. Ibid.

30. Gemeente Amsterdam 'Top600', https://www.amsterdam.nl/wonen-leefomgeving/veiligheid/top600/#h87a00c6d-a600-453e-aa2d-5298a596f56e (translated from Dutch)

31. Ibid.

32. Ibid.

33. Openbaar Ministrie 'Top 600 aanpak', https://www.om.nl/organisatie/arrondissementsparket-amsterdam/top-600 (translated from Dutch)

34. Ibid.

35. Groenendaal, Eline and Van de Venn, Anke 'Top of flop?' Advocatenblad (6 September 2012): 18–20, https://assets.budh.nl/advocatenblad/article_pdf/20090318/basis_pdf_oid.pdf (translated from Dutch)

36. Ibid.

37. A reference to all those profiled on both the Top600 as well as the Top400, another risk profiling system considered in more detail below. Gemeente Amsterdam 'Monitor Top1000', https://www.amsterdam.nl/bestuur-organisatie/organisatie/overige/acvz/monitor-top1000/ (translated from Dutch)

38. Gemeente Amsterdam 'Top 600 – Informatie voor professionals', https://www.amsterdam.nl/bestuur-organisatie/organisatie/overige/acvz/top600/ (translated from Dutch).

39. Ibid.

40. Ibid.

41. Groenendaal, Eline 'Top 1000 zal averechts werken' (2014), https://www.parool.nl/nieuws/top-1000-zal-averechts-werken~b1b2874f/ (translated from Dutch)

42. Groenendaal, Eline 'Top 1000 zal averechts werken' (2014), https://www.parool.nl/nieuws/top-1000-zal-averechts-werken~b1b2874f/ (translated from Dutch)

43. Ibid

44. AD, 'Amsterdamse Top 600-crimineel vecht groepsverbod aan' (2013), https://www.ad.nl/binnenland/amsterdamse-top-600-crimineel-vecht-groepsverbod-aan~a2f000f9/ (translated from Dutch)

45. Ibid.

46. Vugts, Paul 'Naar rechter om van Top 600-lijst te komen' (2013), https://www.parool.nl/nieuws/naar-rechter-om-van-top-600-lijst-te-komen~b3bef3d7/ (translated from Dutch)

47. Ibid.

48. Vugts, Paul 'Een zoon van Diana Sardjoe belandde in de Top 600: "Ik was in shock"' (2017), https://www.parool.nl/nieuws/een-zoon-van-diana-sardjoe-beland-de-in-de-top-600-ik-was-in-shock~b65a98ed/ (translated from Dutch)

49. Ibid.

50. Vugts, Paul 'Een zoon van Diana Sardjoe belandde in de Top 600: "Ik was in shock"' (2017), https://www.parool.nl/nieuws/een-zoon-van-diana-sardjoe-beland-de-in-de-top-600-ik-was-in-shock~b65a98ed/

51. Vugts, Paul 'Een zoon van Diana Sardjoe belandde in de Top 600: "Ik was in shock"' (2017), https://www.parool.nl/nieuws/een-zoon-van-diana-sardjoe-beland-de-in-de-top-600-ik-was-in-shock~b65a98ed/

52. de Koning, Anouk 'Handled with care': Diffuse policing and the production of inequality in Amsterdam' (2017) Ethnography vol. 18(4) 535 – 555, https://journals.sagepub.com/doi/pdf/10.1177/1466138117696107

53. Ibid.

54. Ibid.

55. Ibid.

56. Ibid.

57. Ibid.

58. Gemeente Amsterdam 'Top 600'. https://www.amsterdam.nl/wonen-leefomgeving/veiligheid/top600/#h4afca487-4824-4bc7-83ca-cc9079bd27f1 (translated from Dutch)

59. NPO 3 TV 'Ajouad en de Top 600', https://www.youtube.com/watch?v=Yrycgb0m8PE&list=PLY3N-4g1TbGHsygLQURizfVjeTZmPcBOvc&index=9 (translated from Dutch)

60. NL Times 'Amsterdam to focus more on drug crimes

in approach to young repeat offenders' (2019), https://nltimes.nl/2019/02/12/amsterdam-focus-drug-crimes-approach-young-repeat-offenders

61. de Koning, Anouk 'Handled with care': Diffuse policing and the production of inequality in Amsterdam' (2017) Ethnography vol. 18(4) 535 – 555, https://journals.sagepub.com/doi/pdf/10.1177/1466138117696107

62. Anouk De Koning, 'Tracing anxious politics in Amsterdam', Patterns of Prejudice, 50:2, 109-128, 31 March 206https://www.tandfonline.com/doi/full/10.1080/0031322X.2016.1161387

63. Amsterdam Municipality 'Tweede halfjaarmonitor Top400' (2016) https://docplayer.nl/113213400-Tweede-halfjaarmonitor-top400.html (translated from Dutch)

64. Gemeente Amsterdam, 'Top400 – Informatie voor professionals'. https://www.amsterdam.nl/bestuur-organisatie/organisatie/overige/acvz/top400/ (translated from Dutch).

65. Ibid.

66. Ibid.

67. Public Interest Litigation Project (PIPL), https://pilpnjcm.nl/en/.

68. Gemeente Amsterdam, 'Top400 – Informatie voor professionals'. https://www.amsterdam.nl/bestuur-organisatie/organisatie/overige/acvz/top400/ (translated from Dutch).

69. Ibid.

70. Gemeente Amsterdam, 'Top 400'. https://www.amsterdam.nl/wonen-leefomgeving/veiligheid/top400/ (translated from Dutch).

71. Ibid.

72. Gemeente Amsterdam, 'Top 400'. https://www.amsterdam.nl/wonen-leefomgeving/veiligheid/top400/ (translated from Dutch).

73. van de Klundert, Mitchell 'Moeders kritisch over hulp criminele jongeren: "Alsof je Holleedertjes hebt"' (2019), https://nos.nl/artikel/2303633-moeders-kritisch-over-hulp-criminele-jongeren-alsof-je-holleedertjes-hebt.html (translated from Dutch).

74. Gemeente Amsterdam, 'Top 400'. https://www.amsterdam.nl/wonen-leefomgeving/veiligheid/top400/ (translated from Dutch).

75. Founding members of NDAS are West Midlands Police; Warwickshire Police; West Mercia Police; West Yorkshire Police; Greater Manchester Police; Merseyside Police, the Metropolitan Police Service; Staffordshire Police, and the National Crime Agency.

76. Police Transformation Fund, National Analytics Solution, Final Business Case v6.0, 29 March 2017 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf). See also: West Midlands Police, National Data Analytics Solution Privacy Notice. https://west-midlands.police.uk/about-us/privacy-notice/national-data-analytics-solution# and NDAS Data Protection Impact Assessment, 2019. https://foi.west-midlands.police.uk/wp-content/uploads/2019/09/DPIA_NDAS.pdf

77. Ibid.

78. Police Transformation Fund investments 2018-19. https://www.gov.uk/government/publications/police-transformation-fund-investments-in-2018-to-2019; Police Transformation Fund investments 2019-20. https://www.gov.uk/government/publications/police-transformation-fund-investments-in-2019-to-2020

79. West Midlands Police, National Data Analytics Solution Privacy Notice. https://west-midlands.police.uk/about-us/privacy-notice/national-data-analytics-solution#

80. Ibid.

81. West Midlands police, 'National Data Analytics Solution', February 2021. https://www.spa.police.uk/spa-media/1t0ngjyt/presentation-3-ndas-supt-matt-tite-and-mandeep-dhensa.pdf

82. Ibid.

83. Police Transformation Fund, National Analytics Solution, Final Business Case v6.0, 29 March 2017 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf).

84. Intelligence Management System (IMS) – police intelligence reports about events, locations and offenders; ICIS – custody information data; Corvus – an intelligence, briefing and tasking system; Police National Computer (PNC) – major police database containing information on individuals, crimes, vehicles and property; Drug Intervention Programme (DiP) data; OASIS – event logging system; among others. See: Police Transformation Fund, National Analytics Solution, Final Business Case v6.0, 29 March 2017 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf)

85. Police Transformation Fund, National Analytics Solution, Final Business Case v6.0, 29 March 2017 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf).

86. West Midlands police, National Data Analytics Solution Privacy Notice. https://west-midlands.police.uk/about-us/privacy-notice/national-data-analytics-solution#

87. Police Transformation Fund, National Analytics Solution, Final Business Case v6.0, 29 March 2017 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf)

88. Ibid.

89. Big Brother Watch, 'A closer look at Experian big data and artificial intelligence in Durham police', 6 April 2018. https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

90. West Midlands police, 'National Data Analytics Solu-

tion Privacy Notice'. https://west-midlands.police.uk/about-us/privacy-notice/national-data-analytics-solution#

91. West Midlands police, 'NDAS Submission to the WMP Ethics Committee', September 2020. https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2020/11/25092020-EC-Agenda-Item-5-NDAS-Update.pdf?x16458

92. West Midlands police, 'Data Driven Insight & Data Science Capability for UK Law Enforcement'. http://www.excellenceinpolicing.org.uk/wp-content/uploads/2017/10/EIP17_2-5_Utilising_Data_Science.pdf

93. Ibid.

94. See Section 2.1 for detailed discussion. See also: Fundamental Rights Agency, 'EU-MIDIS II Second European Union Minorities and Discrimination Survey' (2018), https://ec.europa.eu/knowledge4policy/dataset/ds00141_en; European Union Agency for Fundamental Rights 'Data in Focus Report – Police Stops and Minorities' (2010), https://fra.europa.eu/sites/default/files/fra_uploads/1132-EU-MIDIS-police.pdf

95. StopWatch, Your area - West Midlands Police. https://www.stop-watch.org/your-area/area/west-midlands

96. Ibid.

97. Ibid.

98. Gov.uk, 'Stop and search', 22 February 2021. https://www.ethnicity-facts-figures.service.gov.uk/crime-justice-and-the-law/policing/stop-and-search/latest#-by-ethnicity-over-time

99. House of Commons Library, 'Police powers: stop and search', Briefing paper, 10 March 2021. https://researchbriefings.files.parliament.uk/documents/SN03878/SN03878.pdf

100. NDAS Data Protection Impact Assessment, 2019. https://foi.west-midlands.police.uk/wp-content/uploads/2019/09/DPIA_NDAS.pdf

101. Logistic Regression, Gradient Boosting Machines and Random Forest algorithms. The Random Forest algorithm was used as the main classifier. See: West Midlands police, 'WMP Ethics Committee briefing note, NDAS – Most Serious Violence (MSV)', 20 November 2019. https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/12/27112019-EC-Item-3-Briefing-Note-NDAS-MSV.pdf

102. Ibid.

103. West Midlands police, 'National Data Analytics Solution', February 2021. https://www.spa.police.uk/spa-media/1t0ngjyt/presentation-3-ndas-supt-matt-tite-and-mandeep-dhensa.pdf

104. West Midlands police, WMP Ethics Committee briefing note, NDAS – Most Serious Violence (MSV), 20 November 2019. https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/12/27112019-EC-Item-3-Briefing-Note-NDAS-MSV.pdf

105. Ibid.

106. Ibid.

107. Police Transformation Fund, National Analytics Solution, Final Business Case v6.0, 29 March 2017 (http://foi.west-midlands.police.uk/wp-content/uploads/2019/01/report1_.pdf)

108. West Midlands police, WMP Ethics Committee briefing note, NDAS – Most Serious Violence (MSV), 20 November 2019. https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/12/27112019-EC-Item-3-Briefing-Note-NDAS-MSV.pdf

109. West Midlands police, WMP Ethics Committee briefing note, NDAS – Most Serious Violence (MSV), 20 November 2019. https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/12/27112019-EC-Item-3-Briefing-Note-NDAS-MSV.pdf

110. Ibid.

111. Alan Turing Institute Data Ethics Group (ATI DEG) and Independent Digital Ethics Panel for Policing (IDEPP), 'Ethics Advisory Report for West Midlands Police', 28 July 2017. https://www.turing.ac.uk/sites/default/files/2018-11/turing_idepp_ethics_advisory_report_to_wmp.pdf

112. Ibid.

113. Ibid.

114. Ibid.

115. West Midlands police and crime commissioner Ethics Committee, June 2021 – NDAS VC Case studies. https://www.westmidlands-pcc.gov.uk/ethics-committee/ethics-committee-reports-and-minutes/

116. Amnesty International 'We Sense Trouble – Automated Discrimination and Mass Surveillance in Predictive Policing in the Netherlands', 2020. https://www.amnesty.org/download/Documents/EUR3529712020ENGLISH.PDF.

117. Ibid.

118. Ibid

119. Ibid.

120. Dutch National Police 'Mobiel banditisme - Mobiele bendes aan het Roer: Een exploratief onderzoek naar aard, omvang en aanpak naar mobiel banditisme in de gemeente Roermond', https://www.politie.nl/binaries/content/assets/politie/wob/00-landelijk/programma-mobiel-banditisme-%E2%80%93-proeftuin-roermond/002---eindversie-mobiele-bendes-aan-het-roer_def.pdf.

121. Amnesty International 'We Sense Trouble – Automated Discrimination and Mass Surveillance in Predictive Policing in the Netherlands', 2020. https://www.amnesty.org/download/Documents/EUR3529712020ENGLISH.PDF.

122. Ibid.

123. Ibid.

124. Ibid.

125. Ibid.

126. Ibid.

127. Bundeskriminalamt, 'Presseinformation: Neues Instrument zur Risikobewertung von potentiellen Gewaltstraftätern', 2 February 2017 https://www.bka.de/DE/Presse/Listenseite_Pressemitteilungen/2017/Presse2017/170202_Radar.html (translated from German)

128. https://www.bka.de/DE/Presse/Listenseite_Pressemitteilungen/2017/Presse2017/170202_Radar.html (translated from German)

129. Nos Nieuws 'Politie gaat misdaad voorspellen met nieuw systeem' (2017), https://nos.nl/artikel/2173288-politie-gaat-misdaad-voorspellen-met-nieuw-systeem.html.

130. Oosterloo, Serena and van Schie, Gerwin 'The Politics and Biases of the "Crime Anticipation Systems" of the Dutch Police' (2018) Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems co-located with 13th International Conference on Transforming Digital Worlds (iConference 2018), http://ceur-ws.org/Vol-2103/paper_6.pdf.

131. Ibid.

132. Oosterloo, Serena and van Schie, Gerwin 'The Politics and Biases of the "Crime Anticipation Systems" of the Dutch Police' (2018) Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems co-located with 13th International Conference on Transforming Digital Worlds (iConference 2018), http://ceur-ws.org/Vol-2103/paper_6.pdf

133. Ibid.

134. Ibid.

135. Ibid.

136. Ibid.

137. KeyCrime, 'About us'. https://www.keycrime.com/about-us

138. KeyCrime, 'Delia'. https://www.keycrime.com/delia

139. Polizia Moderna, 'La chiave del crimine', July 2015 https://www.poliziadistato.it/statics/16/la-chiave-del-crimine.pdf

140. KeyCrime, 'Crime Linking'. https://www.keycrime.com/crime-linking

141. KeyCrime website. https://www.keycrime.com/

142. KeyCrime website. https://www.keycrime.com/

143. Polizia Moderna, 'La chiave del crimine', July 2015. https://www.poliziadistato.it/statics/16/la-chiave-del-crimine.pdf; AlgorithmWatch, 'Automating Society 2019 – Taking stock of automated decision-making in the EU', 2019. https://algorithmwatch.org/en/automating-society-2019/italy/

144. Essex University, 'Impact: imagine being able to predict a crime in the future'. https://www.essex.ac.uk/research/showcase/imagine-being-able-to-predict-a-crime-in-the-future; KeyCrime, 'The Italian software that changed the world of predictive policing", 18 May 2019. https://www.keycrime.com/stampa1

145. Ibid.

146. KeyCrime, 'Decision support software'. https://www.keycrime.com/decision-support-software

147. KeyCrime, 'The Italian software that changed the world of predictive policing", 18 May 2019. https://www.keycrime.com/stampa1 (Translated from Italian)

148. KeyCrime, 'The Italian software that changed the world of predictive policing", 18 May 2019. https://www.keycrime.com/stampa1

149. KeyCrime, 'About us'. https://www.keycrime.com/about-us; KeyCrime, 'Crime Linking'. https://www.keycrime.com/crime-linking

150. Ibid.

151. Kai Seidensticker et al, 'Predictive Policing in Germany', State Office of Criminal Investigation, North Rhine Westphalia, August 2018. https://www.researchgate.net/publication/332170526_Predictive_Policing_in_Germany

152. State Office of Criminal Investigations in North Rhine Westphalia, 'Project SKALA', 2018. Dusseldorf. https://polizei.nrw/sites/default/files/2018-07/180628_Abschlussbericht_SKALA.PDF (Translated from German).

153. Dr Tobias Knobloch, 'Predictive Policing in Germany: opportunities and dangers of data analytic forecasting techniques and recommendations for the use in police work', Bertelsmann-stiftung, August 2018. https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/predictive.policing.pdf. (Translated from German)

154. Nexiga website. https://nexiga.com/english/

155. State Office of Criminal Investigations in North Rhine Westphalia, 'Project SKALA', 2018. Dusseldorf. https://polizei.nrw/sites/default/files/2018-07/180628_Abschlussbericht_SKALA.PDF (Translated from German).

156. Ibid.

157. Dr Tobias Knobloch, 'Predictive Policing in Germany: opportunities and dangers of data analytic forecasting techniques and recommendations for the use in police work', Bertelsmann-stiftung, August 2018. https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/predictive.policing.pdf (Translated from German)

158. P. Hunt et al, 'Evaluation of the Shreveport Predictive Policing Experiment', 2014 Santa Monica: RAND Corporation. https://www.ncjrs.gov/pdffiles1/nij/grants/248883.pdf

159. Kai Seidensticker et al, 'Predictive Policing in Ger-

many', State Office of Criminal Investigation, North Rhine Westphalia, August 2018. https://www.researchgate.net/publication/332170526_Predictive_Policing_in_Germany

160. Ibid.

161. Oswald et al, 'Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model', 31 August 2017, Information & Communications Technology Law. https://www.tandfonline.com/doi/full/10.1080/13600834.2018.1458455

162. The Checkpoint programme last four months and individuals have to agree to several conditions, including agreeing not to re-offend during the programme, taking part in restorative justice, addressing relevant personal issues, and completing voluntary community work or wearing a GPS tag. See: Centre for Justice Innovation, 'Checkpoint'. https://justiceinnovation.org/project/checkpoint

163. Oswald et al, 'Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model', 31 August 2017, Information & Communications Technology Law. https://www.tandfonline.com/doi/full/10.1080/13600834.2018.1458455

164. Ibid

165. Paul Cresswell et al., 'Under the bonnet: Mosaic data, methodology and build', Experian Marketing Services, 1 April 2014. https://docplayer.net/4444904-Under-the-bonnet-mosaic-data-methodology-and-build.html. See also: Big Brother Watch, 'A closer look at Experian big data and artificial intelligence in Durham police', 6 April 2018. https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

166. Experian, 'Mosaic: the consumer classification solution for consistent cross-channel marketing'. https://www.experianintact.com/content/uk/documents/productSheets/MosaicConsumerUK.pdf

167. Paul Cresswell et al., 'Under the bonnet: Mosaic data, methodology and build', Experian Marketing Services, 1 April 2014. https://docplayer.net/4444904-Under-the-bonnet-mosaic-data-methodology-and-build.html. and Experian, 'Mosaic Public Sector', 2014. https://www.whatdotheyknow.com/request/499023/response/1211711/attach/2/Mosaic%20Public%20Sector%20Brochure.pdf. See also: Big Brother Watch, 'A closer look at Experian big data and artificial intelligence in Durham police', 6 April 2018. https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

168. Big Brother Watch, 'A closer look at Experian big data and artificial intelligence in Durham police', 6 April 2018. https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

169. Ibid.

170. Big Brother Watch, 'A closer look at Experian big data and artificial intelligence in Durham police', 6 April 2018. https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

171. Big Brother Watch, 'A closer look at Experian big data and artificial intelligence in Durham police', 6 April 2018. https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

172. Sheena Urwin, 'Algorithmic Forecasting of Offender Dangerousness for Police Custody Officers: An Assessment of Accuracy for the Durham Constabulary Model', University of Cambridge, 2016. https://www.crim.cam.ac.uk/system/files/documents/sheena-urwin-thesis-12-12-2016.pdf

173. Oswald et al, 'Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model', 31 August 2017, Information & Communications Technology Law. https://www.tandfonline.com/doi/full/10.1080/13600834.2018.1458455

174. Oswald et al, 'Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model', 31 August 2017, Information & Communications Technology Law https://www.tandfonline.com/doi/full/10.1080/13600834.2018.1458455

175. Oswald et al, 'Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model', 31 August 2017, Information & Communications Technology Law. https://www.tandfonline.com/doi/full/10.1080/13600834.2018.1458455

176. Big Brother Watch, 'Briefing on Algorithmic Decision-Making in the Criminal Justice System', February 2020. https://bigbrotherwatch.org.uk/wp-content/uploads/2020/02/Big-Brother-Watch-Briefing-on-Algorithmic-Decision-Making-in-the-Criminal-Justice-System-February-2020.pdf

177. Experian, 'Mosaic Public Sector', Marketing Brochure. https://www.experian.co.uk/assets/marketing-services/brochures/mosaic-ps-brochure.pdf

178. Paul Cresswell et al., 'Under the bonnet: Mosaic data, methodology and build', Experian Marketing Services, 1 April 2014. https://docplayer.net/4444904-Under-the-bonnet-mosaic-data-methodology-and-build.html

179. Oswald et al, 'Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model', 31 August 2017, Information & Communications Technology Law https://www.tandfonline.com/doi/full/10.1080/13600834.2018.1458455

180. Ibid.

181. Ibid.

182. Ibid.

183. National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2015). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/re-

search-analysis-offender-assessment-system.pdf

184. Ibid, See also: Prison Service Order, 'Offender Assessment and Sentence Management – OASys', (2005) https://www.justice.gov.uk/downloads/offenders/psipso/pso/PSO_2205_offender_assessment_and_sentence_management.doc.

185. Ibid.

186. Ibid.

187. This data is significantly out of date (2015), but gives some idea of the scale of the system's use. See: National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2015). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf

188. Ibid.

189. Ibid.

190. Non-scored categories: Health and other, emotional wellbeing, financial management.

191. National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2015). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf; Phillip Howard, 'Hazards of different types of reoffending', Ministry of Justice Research Series 3/11, May 2011. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/217377/research-reoffending-hazards.pdf

192. Ibid.

193. National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2015). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf

194. Ibid.

195. Copas rate: The length in years of an individual's known criminal career and their total number of convictions. For example, the 'copas rate' of an individual is higher when they have more criminal appearances within a short 'criminal career' (from their first through to their most recent offence). See: National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2015) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf

196. National Offender Management Service, 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013', (2015). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf

197. Ibid.

198. Justice.gov.uk, 'Categorisation and recategorization of adult male prisoners', 2011. https://www.justice.gov.uk/downloads/offenders/psipso/psi-2011/psi-40-2011-categorisation-adult-males.doc

199. Hromadske, 'The Ministry of Justice has created an artificial intelligence that will analyze criminals. He determines if they can do something again', 13 September 2020. https://hromadske.ua/posts/minyust-stvoriv-shtuchnij-intelekt-yakij-analizuvatime-zlochinciv-vin-viznachaye-chi-mozhut-voni-znovu-shos-skoyiti (translated from Ukrainian). Announced by the Minister for Justice via a Facebook video, 12 September 2020: https://www.facebook.com/100011121947008/videos/1230395760674477/

200. Olga Shenk and Roman Hryshyn-Hryshchuk, 'Digital Justice - Can artificial intelligence replace a judge?', 17 March 2021. https://www.lexology.com/library/detail.aspx?g=f478f3a2-f1f9-4846-875b-cf648215182b; https://hromadske.ua/posts/minyust-stvoriv-shtuchnij-intelekt-yakij-analizuvatime-zlochinciv-vin-viznachaye-chi-mozhut-voni-znovu-shos-skoyiti (translated from Ukrainian).

201. Kharkiv Human Rights Protection Group, 'Prisoners' rights in Ukraine 2014-2021: KHPG report', 16 June 2021. https://khpg.org/en/1608809217

202. Suspilne Media, 'Sale of prisons and reform of the penitentiary system', 3 February 2021. https://suspilne.media/101586-prodaz-vaznic-ta-reforma-sistemi-pokaran-maluska-rozpoviv-cim-zajmatimetsa-minust/ (translated from Ukrainian)

203. Olena Yara et al, 'Legal Regulation of the Use of Artificial Intelligence: Problems and Development Prospects', European Journal of Sustainable Development (2021), 10, 1, 281-289

204. Hromadske, 'The Ministry of Justice has created an artificial intelligence that will analyze criminals. He determines if they can do something again', 13 September 2020. https://hromadske.ua/posts/minyust-stvoriv-shtuchnij-intelekt-yakij-analizuvatime-zlochinciv-vin-viznachaye-chi-mozhut-voni-znovu-shos-skoyiti. (translated from Ukrainian). Announced by the Minister for Justice via a Facebook video, 12 September 2020: https://www.facebook.com/100011121947008/videos/1230395760674477/

205. Ukraine High Court of Justice, 'GRP initiates launch of a pilot project on the use of artificial intelligence on the basis of the court of first instance', 210 February 2021. https://hcj.gov.ua/news/vrp-iniciyuye-zapusk-pilotnogo-proyektu-shchodo-zastosuvannya-shtuchnogo-intelektu-na-bazi-sudu (translated from Ukrainian)

206. Ibid.

207. Jay Singh (ed) et al, 'Handbook of Recidivism Risk / Needs Assessment Tools', February 2018, Wiley-Blackwell. https://www.wiley.com/en-gb/Handbook+of+Recidivism+Risk+Needs+Assessment+Tools-p-9781119184294 Marzieh Karimi-

Haghighi & Carlos Castillo, 'Efficiency and Fairness in Recurring Data-Driven Risk Assessments of Violent Recidivism, SIGAPP Symposium on Applied Computing (SAC), pp. 994-1002, ACM Press. https://chato.cl/papers/karimi_haghighi_castillo_2021_efficiency_fairness_temporal_recurring.pdf

208. Ibid.; Karin Arbach-Lucioni & Antonio Andres-Pueyo, 'Violence Risk Assessment Practices in Spain', https://ri.conicet.gov.ar/bitstream/handle/11336/118001/CONICET_Digital_Nro.5c4abd46-f49c-4762-8896-9dc43dd807e5_A.pdf

209. Ibid.

210. Ibid.

211. Ibid.

212. Ibid.

213. Ibid.

214. Ibid.

215. http://www.fbg.ub.edu/en/news/antonio-andres-pueyo-the-challenge-of-riscanvi-is-to-detect-which-inmates-pose-high-risk-of-recidivism/

216. Antonio Andres-Pueyo et al, 'The RisCanvi: A New Tool for Assessing Risk for Violence in Prison and Recidivism', in Jay Singh (ed) et al, 'Handbook of Recidivism Risk / Needs Assessment Tools', February 2018, Wiley-Blackwell. https://www.wiley.com/en-gb/Handbook+of+Recidivism+Risk+Needs+Assessment+Tools-p-9781119184294

217. AlgorithmWatch, 'In Catalonia, the RisCanvi algorithm helps decide whether inmates are paroled', 25 May 2021. https://algorithmwatch.org/en/riscanvi/

218. Antonio Andres-Pueyo et al, 'The RisCanvi: A New Tool for Assessing Risk for Violence in Prison and Recidivism', in Jay Singh (ed) et al, 'Handbook of Recidivism Risk / Needs Assessment Tools', February 2018, Wiley-Blackwell. https://www.wiley.com/en-gb/Handbook+of+Recidivism+Risk+Needs+Assessment+Tools-p-9781119184294

219. Ibid.

220. Ibid.

221. AlgorithmWatch, 'In Catalonia, the RisCanvi algorithm helps decide whether inmates are paroled', 25 May 2021. https://algorithmwatch.org/en/riscanvi/

222. S. Tolan et al, 'Why machine learning may lead to unfairness: evidence from risk assessment for juvenile justice in Catalonia', 2019. https://chato.cl/papers/miron_tolan_gomez_castillo_2019_machine_learning_risk_assessment_savry.pdf

223. People from the region of north Africa which includes Morocco, Algeria, Tunisia, Libya, Western Sahara and Mauritania.

224. Ibid.

225. Ibid.

226. Oosterloo, Serena and van Schie, Gerwin 'The Politics and Biases of the "Crime Anticipation Systems" of the Dutch Police' (2018) Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems co-located with 13th International Conference on Transforming Digital Worlds (iConference 2018), http://ceur-ws.org/Vol-2103/paper_6.pdf

227. Lum, Kristian and William, Isaac 'To Predict and Serve?' (2016) Significance 13 (5): 14–19 https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x; Bennett Moses, Lyria and Chan, Janet, 'Algorithmic prediction in policing: assumptions, evaluation, and accountability' (2016) Policing and Society 28:7, https://www.tandfonline.com/doi/10.1080/10439463.2016.1253695; Barocas, Solon and Selbst, AndrewD., 'Big Data's disparate impact' (2016) California Law Review, 104, 671, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899

228. Buskey, Brandon and Woods, Andrea 'Making Sense of Pretrial Risk Assessments' (2018) The Champion Issue June 2018, https://www.nacdl.org/Article/June2018-MakingSenseofPretrialRiskAsses.

229. Ibid.

230. Ibid.

231. Fundamental Rights Agency, 'EU-MIDIS II Second European Union Minorities and Discrimination Survey' (2018), https://ec.europa.eu/knowledge4policy/dataset/ds00141_en.

232. Ibid.

233. European Union Agency for Fundamental Rights 'Data in Focus Report – Police Stops and Minorities' (2010), https://fra.europa.eu/sites/default/files/fra_uploads/1132-EU-MIDIS-police.pdf

234. Justicia European Rights Network 'Disparities in Criminal Justice Systems for Individuals of Different Ethnic, Racial, and National Background in the European Union' (November 2018)

235. Justicia European Rights Network 'Disparities in Criminal Justice Systems for Individuals of Different Ethnic, Racial, and National Background in the European Union' (November 2018)

236. Namoradze, Zaza and Pacho, Irmina 'When It Comes to Race, European Justice Is Not Blind' (2018) Open Society Justice Initiative, https://www.justiceinitiative.org/voices/when-it-comes-race-european-justice-not-blind

237. Fundamental Rights Agency, 'EU-MIDIS II Second European Union Minorities and Discrimination Survey' (2018), https://ec.europa.eu/knowledge4policy/dataset/ds00141_en

238. Cankaya, S. (2012). De Controle van Marsmannetjes en ander Schorriemorrie. Den Haag: Boom Lemma. https://research.vu.nl/en/publications/de-controle-van-marsmannetjes-en-ander-schorriemorrie

239. Cankaya, Sinan 'De Controle van Marsmannetjes en ander Schorriemorrie' (2012) Den Haag: Boom Lem-

ma, https://research.vu.nl/en/publications/de-controle-van-marsmannetjes-en-ander-schorriemorrie

240. Cankaya, Sinan 'Geopolicing Race, Gender, and Class: How the Police Immobilise Urban Allochthones' (2020) A Radical Journal of Geography, https://onlinelibrary.wiley.com/doi/full/10.1111/anti.12613.

241. Ibid.

242. Ibid.

243. Cankaya, Sinan 'Geopolicing Race, Gender, and Class: How the Police Immobilise Urban Allochthones' (2020) A Radical Journal of Geography, https://onlinelibrary.wiley.com/doi/full/10.1111/anti.12613

244. Committee on the Elimination of Racial Discrimination, General Comment 31 on the Prevention of Racial Discrimination in the Administration and Functioning of the Criminal Justice System preamble and paragraph 1.III.2, quoted in The Justice Initiative, A Global Campaign for Pre-trial Justice Report: The Socioeconomic Impact of Pre-trial Detention (2011).

245. Virginie Gautron & Jean-Noël Retière, La décision judiciaire : jugements pénaux ou jugements sociaux?, 88 Mouvements 11, 11-18 (2016).

246. Dieter Burssens, Carrol Tange, and Eric Maes (2015) A la recherche de determinants du recours a la detention preventive et de sa duree, Institut National de Criminalistique et de Criminologie (2015).

247. Patrizio Gonnella, Detenuti Stranieri in Italia, Antigone (2015).

248. David Perry & Mary Rogan, DETOUR – Towards Pre-trial Detention as Ultima Ratio, 1st National Report of Ireland, § 4, (2016). 5.6% of respondents self-identified as Traveller. At the time of the study, Traveller youth were 1.15% of the population. Department of Children and Youth Affairs, Young Travellers in Ireland, Population Overview, (2020).

249. Philippe Robert, 'A lawmaker's headache: Pretrial detention'. Paris: CESDIP (1994). See also Fair Trials, 'A Measure of Last Resort?' The practice of pre-trial detention decision making in the EU (2016) (summarising research in this area). https://www.fairtrials.org/publication/measure-last-resort

250. Fair Trials, 'Disparities and Discrimination in the European Union's Criminal Legal Systems', January 2021. https://www.fairtrials.org/sites/default/files/publication_pdf/Disparities-and-Discrimination-in-the-European-Unions-Criminal-Legal-Systems.pdf

251. Hilde Wermink, Sigrid van Wingerden, Johan van Wilsem & Paul Nieuwbeerta, Studying Ethnic Disparities in Sentencing: The Importance of Refining Ethnic Minority Measures, in Handbook on Punishment Decisions 239-264 (L Ulmer & M Bradley 1ed 2019).

252. Samantha Bielen, Peter Grajzl and Wim Marneffe, Blame Based on One's Name? Extralegal Disparities in Criminal Conviction and Sentencing, SSRN Electronic Journal, (2008).

253. Virginie Gautron & Jean-Noël Retière, La décision judiciaire : jugements pénaux ou jugements sociaux?, 88 Mouvements 11-18 (2016).

254. Virginie Gautron & Jean-Noël Retière, La décision judiciaire : jugements pénaux ou jugements sociaux?, 88 Mouvements 11-18 (2016).

255. NPO 3 TV 'Ajouad en de Top 600', https://www.youtube.com/watch?v=Yrycgb0m8PE&list=PLY3N4g1T-bGHsygLQURizfVjeTZmPcBOvc&index=9.

256. Gemeente Amsterdam 'Top 600 – Informatie voor professionals', https://www.amsterdam.nl/bestuur-organisatie/organisatie/overige/acvz/top600/ (translated from Dutch).

257. Zuiderveen Borgesius, Fredreik 'Discrimination, artificial intelligence, and algorithmic decision-making' (2018) Directorate General of Democracy, Council of Europe.

258. Oswald, Marion et al. 'Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality' (2018) Information & Communications Technology Law. 27:2, 223-250.

259. E.g. Sweden. See Malberg, Bo 'Residential Segregation of European and Non-European Migrants in Sweden' (2018) Eur J Popul 34(2): 169-193 https://link.springer.com/article/10.1007/s10680-018-9478-0.

260. FRA, 'Summary Report – The State of Roma and Traveller Housing in the European Union – Steps towards Equality' (2010).

261. Malberg, Bo 'Residential Segregation of European and Non-European Migrants in Sweden' (2018) Eur J Popul 34(2): 169-193 https://link.springer.com/article/10.1007/s10680-018-9478-0.

262. Lum, Kristian and William, Isaac 'To Predict and Serve?' (2016) Significance 13 (5): 14–19 https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x

263. Ibid; Ensign, Danielle et al. 'Runaway Feedback Loops in Predictive Policing' (2017) Cornell University Library, 29 June 2019, https://arxiv.org/abs/1706.09847; Bennett Moses, Lyria and Chan, Janet, 'Algorithmic prediction in policing: assumptions, evaluation, and accountability' (2016) Policing and Society 28:7, https://www.tandfonline.com/doi/10.1080/10439463.2016.1253695

264. A full analysis can be found in Zuiderveen Borgerius, Frederik 'Discrimination, artificial intelligence, and algorithmic decision-making' (2018) Council of Europe.

265. Information Commissioner's Office (ICO), 'Human bias and discrimination in AI systems' (2019), https://ico.org.uk/about-the-ico/news-and-events/ai-blog-human-bias-and-discrimination-in-ai-systems/

266. Article 6(2), ECHR.

267. Directive (EU) 2016/343 of the European Parliament and of the Council of 9 March 2016 on the strengthening of certain aspects of the presumption of innocence

and of the right to be present at the trial in criminal proceedings; Article 6(2), ECHR.

268. Ibid., Article 48.

269. Alan Turing Institute Data Ethics Group (ATI DEG) and Independent Digital Ethics Panel for Policing (IDEPP), 'Ethics Advisory Report for West Midlands Police', 28 July 2017. https://www.turing.ac.uk/sites/default/files/2018-11/turing_idepp_ethics_advisory_report_to_wmp.pdf

270. Alan Turing Institute Data Ethics Group (ATI DEG) and Independent Digital Ethics Panel for Policing (IDEPP), 'Ethics Advisory Report for West Midlands Police', 28 July 2017. https://www.turing.ac.uk/sites/default/files/2018-11/turing_idepp_ethics_advisory_report_to_wmp.pdf

271. Marzieh Karimi-Haghighi & Carlos Castillo, 'Efficiency and Fairness in Recurring Data-Driven Risk Assessments of Violent Recidivism, SIGAPP Symposium on Applied Computing (SAC), pp. 994-1002, ACM Press. https://chato.cl/papers/karimi_haghighi_castillo_2021_efficiency_fairness_temporal_recurring.pdf

272. House of Commons Library, 'Police powers: stop and search', Briefing paper, 10 March 2021. https://researchbriefings.files.parliament.uk/documents/SN03878/SN03878.pdf

273. Gemeente Amsterdam 'Top 600 – Informatie voor professionals', https://www.amsterdam.nl/bestuur-organisatie/organisatie/overige/acvz/top600/ (translated from Dutch)

274. ECHR, Article 6(2).

275. Ibid., Article 2.

276. cf. ECtHR's definition of 'suspect' and 'charge' in Mikolajova v. Slovakia, App No. 4479/02 (Judgment of 18 January 2011), paras 40-41 and Bandeltov v. Ukraine, App No. 23180/06 (Judgment of 31 October 2013), para. 56.

277. ECtHR, Gogitizde and Others v. Georgia, App. No. 36862/05 (Judgment of 12 May 2015).

278. ECtHR, Raimondo v. Italy, App. No. 12954/87 (Judgment of 22 February 1994).

279. Mendola, Marco 'One Step Further in the "Surveillance Society": The Case of Predictive Policing' (2016) Tech and Law Center, http://techandlaw.net/wp-content/uploads/2016/10/One-Step-Further-in-the-Surveillance-Society_The-Case-of-Predictive-Policing.pdf.

280. Ilgar Mammadov v. Azerbaijan (judgement of 22 May 2014).

281. Bandeltov v. Ukraine, App No. 23180/06 (Judgment of 31 October 2013), para. 56; Brusco v. France, no. 1466/07, § 47, 14 October 2010.

282. Mikolajova v. Slovakia, App No. 4479/02 (Judgment of 18 January 2011), paras 40-41.

283. https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a

284. ECtHR, Beraru v Romania App. No. 40107/04 (Judgment of 18 March 2014); ECtHR, Kuopila v Finland, App. No. 27752/95 (Judgment of 27 April 2000); Directive 2012/13/EU of the European Parliament and of the Council of 22 May 2012 on the right to information in criminal proceedings (Right to Information Directive, Article 7(1); ECtHR, Wloch v Poland, App. No. 27785/95 (Judgment of 19 October 2000).

285. Moore, Taylor 'Trade Secrets & Algorithms as Barriers to Social Justice', Center for Democracy & Technology (2017), https://cdt.org/files/2017/08/2017-07-31-Trade-Secret-Algorithms-as-Barriers-to-Social-Justice.pdf.

286. https://www.mediapost.com/publications/article/327256/google-defends-right-to-protect-search-algorithm-i.html; https://europeanlawblog.eu/2021/01/27/algorithm-transparency-how-to-eat-the-cake-and-have-it-too/

287. The Guardian, 'Seeing stones: pandemic reveals Palantir's troubling reach in Europe', 21 April 2021. https://www.theguardian.com/world/2021/apr/02/seeing-stones-pandemic-reveals-palantirs-troubling-reach-in-europe; Privacy International & No Tech for Tyrants, 'All roads lead to Palantir: A review of how the data analytics company has embedded itself throughout the UK', 29 October 2020. https://privacyinternational.org/sites/default/files/2020-11/All%20roads%20lead%20to%20Palantir%20with%20Palantir%20response%20v3.pdf

288. For example, by the EU Trade Secrets Directive - (Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure).

289. Article 32, Toronto Declaration. https://www.torontodeclaration.org/declaration-text/english/

290. Bathaee, Yavar, 'The Artificial Intelligence Black Box and the Failure of Intent and Causation', Harvard Journal of Law & Technology, vol 31, no. 2, 890 (2018).

291. Ibid.

292. Ibid.

293. Fair Trials, 'Measure of Last Resort' (2016) https://www.fairtrials.org/sites/default/files/publication_pdf/A-Measure-of-Last-Resort-Full-Version.pdf

294.

295. Raja Parasuraman and Dietrich Manzey, 'Complacency and Bias in Human Use of Automation: An Attentional Integration', Human Factors, The Journal of Human Factors and Ergonomics Society (2010)

296. Article 11(1), LED; Article 22(2)(c) and (3), GDPR

297. Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and profiling for the purposes of Regulation 2016/679' (3 October 2017)

298. UN Convention on the Rights of the Child, Article 3

299. Directive (EU) 2016/800 of the European Parliament and of the Council of 11 May 2016 on procedural safeguards for children who are suspects or accused persons in criminal proceedings, OJ L 132, 21.5.2016, p. 1–20

300. ECHR, Article 8.

301. Ibid, Preamble (9)

302. Justicia, 'Comparative Report – Ethnic, Racial Disparities in Criminal Justice' (2018), http://www.eujusticia.net/images/uploads/pdf/Justicia_Network_Disparities_in_Criminal_Justice_Comparative_Report_2018-1.pdf

Fair Trials