6

Human Rights and Algorithmic Impact Assessment for Predictive Policing

Céline Castets-Renard[*]

6.1 INTRODUCTION

Artificial intelligence (AI) constitutes a major form of scientific and technological progress. For the first time in human history, it is possible to create autonomous systems capable of performing complex tasks, such as processing large quantities of information, calculating and predicting, learning and adapting responses to changing situations, and recognizing and classifying objects.[1] For instance, algorithms, or so-called Algorithmic Decision Systems (ADS),[2] are increasingly involved in systems used to support decision-making in many fields,[3] such as child welfare, criminal justice, school assignment, teacher evaluation, fire risk assessment, homelessness prioritization, Medicaid benefit, immigration decision systems or risk assessment, and predictive policing, among other things.

An Automated Decision(-making/-support) System (ADS) is a system that uses automated reasoning to facilitate or replace a decision-making process that would otherwise be performed by humans.[4] These systems rely on the analysis of large amounts of data from which they derive useful information to make

[*] Support from the Artificial and Natural Intelligence Toulouse Institute (ANITI), ANR-3IA, and the Civil Law Faculty of the University of Ottawa is gratefully acknowledged. I also thank law student Roxane Fraser and the attendees at the Conference on *Constitutional Challenges in the Algorithmic Society* for their helpful comments, and especially Professor Ryan Calo, Chair of the Panel. This text has been written in 2019 and does not take into account the EC proposal on AI published in April 2021.

[1] Preamble section of the Montréal Declaration, www.montrealdeclaration-responsibleai.com/the-declaration accessed 23 May 2019.

[2] Guido Noto La Diega, 'Against Algorithmic Decision-Making' (2018) https://papers.ssrn.com/abstract=3135357 accessed 23 May 2019.

[3] AINow Institute, 'Government Use Cases' https://ainowinstitute.org/nycadschart.pdf accessed on 22 December 2019.

[4] AINow Institute, 'Algorithmic Accountability Policy Toolkit' (October 2018) https://ainowinstitute.org/aap-toolkit.pdf accessed 23 May 2019 [Toolkit].

https://doi.org/10.1017/9781108914857.007 Published online by Cambridge University Press

decisions and to infer[5] correlations,[6] with or without artificial intelligence techniques.[7]

Law enforcement agencies are increasingly using algorithmic predictive policing systems to forecast criminal activity and allocate police resources. For instance, New York, Chicago, and Los Angeles use predictive policing systems built by private actors, such as PredPol, Palantir, and Hunchlab,[8] to assess crime risk and forecast its occurrence, in hope of mitigating it. More often, such systems predict the places *where* crimes are most likely to happen in a given time window (place-based) based on input data, such as the location and timing of previously reported crimes.[9] Other systems analyze *who* will be involved in a crime as either victim or perpetrator (person-based). Predictions can focus on variables such as places, people, groups, or incidents. The goal is also to better deploy officers in a time of declining budgets and staffing.[10] Such tools are mainly used in the United States, but European police forces have expressed an interest in using them to protect the largest cities.[11] Predictive policing systems and pilot projects have already been deployed,[12] such as PredPol, used by the Kent Police in the United Kingdom.

However, these predictive systems challenge fundamental rights and guarantees of the criminal procedure (Section 6.2). I will address these issues by taking into account the enactment of ethical norms to reinforce constitutional rights (Section 6.3),[13] as well as the use of a practical tool, namely Algorithmic Impact Assessment, to mitigate the risks of such systems (Section 6.4).

[5] Sandra Wachter and Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI* (2018) Columbia Business Law Review https://papers.ssrn.com/abstract=3248829 accessed 11 March 2019.

[6] European Parliamentary Research Service (EPRS) Study, 'Panel for the Future of Science and Technology, Understanding Algorithmic Decision-Making: Opportunities and Challenges, March 2019' (PE 624.261), 21 [PE 624.261].

[7] See, for instance, Florian Saurwein, Natascha Just and Michael Latzer, 'Governance of Algorithms: Options and Limitations' (2015) vol. 17 (6) info 35–49 https://ssrn.com/abstract=2710400 accessed 21 January 2020.

[8] Toolkit, supra note 4.

[9] PE 624.261, supra note 6.

[10] Walter L. Perry et al., 'Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations' (2013) www.rand.org/pubs/research_reports/RR233.html accessed 29 November 2018.

[11] Lubor Hruska et al., 'Maps of the Future, Research Project of the Czeck Republic' (2015) www.mvcr.cz/mvcren/file/maps-of-the-future-pdf.aspx accessed 23 May 2019 [Maps].

[12] Don Casey, Phillip Burrell, and Nick Sumner, 'Decision Support Systems in Policing' (2018 (4 SCE)) *European Law Enforcement Research Bulletin* https://bulletin.cepol.europa.eu/index.php/bulletin/article/view/345 accessed 23 May 2019.

[13] James Harrison, 'Measuring Human Rights: Reflections on the Practice of Human Rights Impact Assessment and Lessons for the Future' (2010) Warwick School of Law Research Paper 2010/26 https://papers.ssrn.com/abstract=1706742 accessed 23 May 2019.

6.2 HUMAN RIGHTS CHALLENGED BY PREDICTIVE POLICING SYSTEMS

In proactive policing, law enforcement uses data and analyzes patterns to understand the nature of a problem. Officers attempt to prevent crime and mitigate the risk of future harm. They refer to the power of information, geospatial technologies, and evidence-based intervention models to predict what and where something is likely to happen, and then deploy resources accordingly.[14]

6.2.1 *Reasons for Predictive Policing in the United States*

There are many reasons why predictive policing systems have been specifically deployed in the United States. First, the high level of urban gun violence pushed the police departments of Chicago,[15] New York, Los Angeles, and Miami, among others, to take preventative action.

Second, it is an opportunity for American tech companies to deploy, within the national territory, products that have previously been developed and put into practice within the framework of international US military operations.

Third, beginning in 2007, within the context of the financial and economic crisis and ensuing budget cuts in police departments, predictive policing tools have been seen as a way 'to do more with less'.[16] Concomitantly, the National Institute of Justice (NIJ), an agency of the US Department of Justice, granted several police departments permission to conduct research and try these new technologies.[17]

Fourth, the emergence of predictive policing tools has been incited by the crisis of weakened public trust in law enforcement in numerous cities. Police violence, particularly towards young African Americans, has led to the research on more 'objective' methods to improve the social climate and conditions of law enforcement. Public outcry against the discrimination risks inherent to traditional methods has come from citizens, social movements such as 'Black Lives Matter', and even in an official capacity from the US Department of Justice (DOJ) investigations surrounding the actions of the Ferguson Police Department after the death of Michael Brown.[18] Following this incident, the goal was to find new and modern methods which are unbiased toward African Americans as much as possible. The unconstitutionality of methods,[19] such as Stop-and-Frisk in New York and Terry

[14] National Institute of Justice, 'Overview of Predictive Policing' (9 June 2014) www.nij.gov/topics/law-enforcement/strategies/predictive-policing/Pages/research.aspx accessed 23 May 2019 [NIJ].

[15] 'Tracking Chicago Shooting Victims' *Chicago Tribune* (16 December 2019) www.chicagotribune.com/news/data/ct-shooting-victims-map-charts-htmlstory.html accessed 16 December 2019.

[16] Andrew Fergurson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of the Law Enforcement* (2017), 21.

[17] NIJ, supra note 14.

[18] US Department of Justice, 'Investigation of the Ferguson Police Department' (2015) www.justice.gov/sites/default/files/opa/press-releases/attachments/2015/03/04/ferguson_police_department_report.pdf accessed 23 May 2019 [US DJ].

[19] *Floyd* v. *City of New York* (2013) 739 F Supp 2d 376.

https://doi.org/10.1017/9781108914857.007 Published online by Cambridge University Press

Stop,[20] based on the US Supreme Court's decision in the *Terry* v. *Ohio* case, converged with the rise of new, seemingly perfect technologies. The Fourth Amendment of the US Constitution prohibits 'unreasonable searches and seizures', and states, 'no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized'.

Fifth, the privacy laws are less stringent in the United States than in the European Union, due to a sectorial approach to protection within the United States. Such normative difference can explain why the deployment of predicting policing systems was easier in the United States.

6.2.2 *Cases Studies: PredPol and Palantir*

When working to predict crime, multiple methods and tools are available for use. I propose a closer analysis of two tools offered by the PredPol and Palantir companies.

6.2.2.1 PredPol

PredPol is a commercial software offered by the American company PredPol Inc. and was initially used in tests by the LAPD[21] and eventually used in Chicago and in Kent County in the United Kingdom. The tool's primary purpose is to predict, both accurately and in real time, the locations and times where crimes have the highest risk of occurring.[22] In other words, this tool identifies risk zones (*hotspots)* based on the same types of statistical models used in seismology. The input data include city and territorial police archives (reports, ensuing arrests, emergency calls), all applied in order to identify the locations where crimes occur most frequently, so as to 'predict' which locations should be prioritized. Here, the target is based on places, not people. The types of offenses can include robberies, automobile thefts, and thefts in public places. A US patent regarding the invention of an 'Event Forecasting System'[23] was approved on 3 February 2015 by the US Patent and Trademark Office (USPTO). The PredPol company claims that its product assists in improving the allocation of resources in patrol deployment. Finally, the tool also incorporates the position of all patrols in real time, which allows departments to not only know where patrols are located but also control their positions. Providing information on a variety of mobile tools such as tablets, smartphones, and laptops, in addition to desktop computers, was also a disruption from previously used methods.

[20] *Terry* v. *Ohio* (1968) 392 US 1.

[21] Issie Lapowsky, 'How the LAPD Uses Data to Predict Crime' (22 May 2018) www.wired.com/story/los-angeles-police-department-predictive-policing accessed 23 May 2019.

[22] 'PredPol Predicts Gun Violence' (2013) www.predpol.com/wp-content/uploads/2013/06/predpol_gun-violence.pdf accessed 23 May 2019.

[23] US Patent No. 8,949,164 (Application filed on 6 September 2012) https://patents.justia.com/patent/8949164 accessed 23 May 2019.

The patent's claims do not specify the manner in which data are used, calculated, or applied. The explanation provided in the patent is essentially based on the processes used by the predictive policing systems, particularly the organizational method used (the three types of data (place, time, offense), geographic division into cells, the transfer of information by a telecommunications system, the reception procedure of historic data, access to GPS data, the link with legal information from penal codes, etc.), rather than on any explanation of the technical aspects. The patent focuses more particularly on the various graphic interfaces and features available to users, such as hotspot maps (heatmaps), which display spatial-temporal smoothing models of historical crime data. It also allows for the use of the method in its entirety but does not relate to the predictive algorithm. The technical aspects are therefore not subject to ownership rights but are instead covered by trade secrets. Even if PredPol claims to provide transparency of its approach, the focus is on the procedure, rather than on the algorithm and mathematical methods used, despite the publication of several articles by the inventors.[24] Some technical studies[25] have been carried out by using publicly available data in cities, such as Chicago, and applying the data to models similar to that of PredPol. However, this tool remains opaque.

It is difficult to estimate the value that these forecasts add in comparison to historic hotspot maps. The few works evaluating this approach that have been published do not concern the quality of the forecasting, but the crime statistics. Contrary to PredPol's claims,[26] the difference in efficiency is ultimately modest, depending on both the quantity of data available on a timescale and on the type of offense committed. The studies most often demonstrate that the prediction of crimes occurred most frequently in the historically most criminogenic areas within the city. Consequently, the software does not teach anything to the most experienced police officers who may be using it. While the Kent Police Department was the first to introduce 'predictive policing' in Europe in 2013, it has been officially recognized that it is difficult to prove whether the system has truly reduced crime. It was finally stopped in 2018[27] and replaced by a new internal tool, the NDAS (National Data Analytics Solution) project, to reduce costs and achieve a higher efficiency. It is likely that a tool developed in one context will not necessarily be relevant in another criminogenic context, as the populations, geographic configurations of cities, and the organization of criminal groups are different.

[24] George O. Mohler, 'Marked Point Process Hotspot Maps for Homicide and Gun Crime Prediction in Chicago' 2014 30(3) *International Journal of Forecasting*, 491–497; 'Does Predictive Policing Lead to Biased Arrests? Results from a Randomized Controlled Trial, Statistics and Public Policy' 5:1 1–6 10.1080/2330443X.2018.1438940 accessed 23 May 2019 [Mohler].

[25] Ismael Benslimane, 'Étude critique d'un système d'analyse prédictive appliqué à la criminalité: PredPol®' *CorteX Journal* https://cortecs.org/wp-content/uploads/2014/10/rapport_stage_Ismael_Benslimane.pdf accessed 23 May 2019.

[26] Mohler, supra note 24.

[27] BBC News, 'Kent Police Stop Using Crime Predicting Software' (28 November 2018) www.bbc.com /news/uk-england-kent-46345717 accessed 23 May 2019.

Moreover, the software tends to systematically send patrols into neighbourhoods that are considered as more criminogenic, which are mainly inhabited in the United States by African American and Latino/a populations.[28] Historical data certainly show high risk in these neighbourhoods, but most of the data were collected in the age of policies such as Terry Stop and Stop-and-Frisk, and were biased, discriminatory, and ultimately unconstitutional. The system, however, does not examine or question the trustworthiness of these types of data. Furthermore, the choice of the type of offense, primarily related to property crime (burglaries, car thefts), constitutes a type of crime that is more likely to be practiced by the poorest and most vulnerable populations, which are frequently composed of the aforementioned minority groups. The results would naturally be different if white-collar crimes were considered. These crimes are excluded from today's predictive policing due to the difficulties of modelling and the absence of significant data. The fact that law enforcement wants to prevent certain types of offenses rather than others, via the use of automated tools is not socially neutral and carries out discrimination against a part of the population. The founders of PredPol and its developers responded to these critiques of bias in several articles published in 2017 and 2018, in which they largely emphasize the auditing of learning data.[29] High-quality learning data are essential to avoid and reduce bias. But if the data used by PredPol are biased, this demonstrates that society itself is biased as a whole. PredPol simply emphasizes this fact, without actually being a point of origin of discrimination. Consequently, the bias present in the tool is no greater than the bias previously generated by the data collected by police officers on the ground.

6.2.2.2 Palantir

Crime Risk Forecasting is the patent held by the company Palantir Technologies Inc., based in California. This device has been deployed in Los Angeles, New York, and New Orleans, but the contracts are often kept secret.[30] Crime Risk Forecasting is an ensemble of software and material that constitutes an 'invention' outlined in US patent and obtained on 8 September 2015.[31] The patent combines several

[28] See the problem of algorithmic biases with COMPAS: Jeff Larson et al., 'How We Analyzed the COMPAS Recidivism Algorithm ProPublica' (2016) www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm accessed 12 August 2018.

[29] P. Jeffrey Brantingham, 'The Logic of Data Bias and Its Impact on Place-Based Predictive Policing' (2017) 15(2) *Ohio State Journal of Criminal Law* 473.

[30] For instance, Ali Winston, 'Palantir Has Secretly Been Using New Orleans to Test Its Predictive Policing Technology' (27 February 2018) www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd accessed 23 May 2019. However, New Orleans ended its Palantir predictive policing program in 2018, after the public's opposition regarding the secret nature of the agreement: Ali Winston, 'New Orleans Ends Its Palantir Predictive Policing Program' (15 March 2018) www.theverge.com/2018/3/15/17126174/new-orleans-palantir-predictive-policing-program-end accessed 23 May 2019.

[31] Crime Risk Forecasting, US Patent 9,129,219 (8 September 2015) https://patentimages.storage.googleapis.com/60/94/95/5dbde28fe6eea2/US9129219.pdf accessed 23 May 2019.

components and features, including a database manager, visualization tools (notably interactive geographic cartography), and criminal forecasts. The goal is to assist police in predicting when and where crime will take place in the future. The forecasts of criminal risk are established within a geographic and temporal grid, for example, of 250 square meters, during an eight-hour police patrol.

The data include:

- Crime history, classified by date, type, location, and more. The forecast can provide either a precise date and time, or a period of time over which risk is uniformly distributed. Similarly, the location can be more or less precise, either by address, GPS coordinates, or geographic zone. The offenses can be, for example, robberies, vehicle thefts (or thefts of belongings from within vehicles), and violence.
- Historical information which is not directly connected to crime: weather, presence of patrols within the grid or in proximity, distribution of emergency service personnel.
- Custody data indicating individuals who have been apprehended or who are in custody for certain types of crimes. These data can be used to decrease crime risk within a zone or to increase risk after the release of accused or convicted criminal.

Complex algorithms can be developed by aggregating methods associating hotspotting, histograms, criminology models, and learning algorithms. The combination possibilities and the aggregation of multiple models and algorithms, as well as the large numbers of variables, result in a highly complex system, with a considerable number of parameters to estimate and hyperparameters to optimize. The patent does not specify how these parameters are optimized, nor does it define the expected quality of the forecasts. It is difficult to imagine that any police force could actually use this tool regularly, without constant assistance from Palantir. Moreover, one can wonder: what are the risks of possible re-identification of victims from the historical data? What precautions are taken to anonymize and prevent re-identification? How about custody data, which are not only personal data, but are, in principle, only subject to treatment by law enforcement and government criminal justice services? Consequently, the features of these ADS remain opaque while the processed data are also unclear.

In this context, it would be a mistake to take predictive policing as a panacea to eradicate crime. Many concerns focus on inefficiency, risk of discrimination, as well as lack of transparency.

6.2.3 *Fundamental Rights Issues*

Algorithms are fallible human creations, and they are embedded with errors and bias, similar to human processes. More precisely, an algorithm is not neutral and depends

https://doi.org/10.1017/9781108914857.007 Published online by Cambridge University Press

notably on the data used. Many legal scholars have revealed bias and racial discrimination in algorithmic systems,[32] as well as their opacity.[33] When algorithmic tools are adopted by governmental agencies without adequate transparency, accountability, and oversight, their use can threaten civil liberties and exacerbate existing issues within government agencies. Most often, the data used to train automated decision-making systems will come from the agency's own databases, and existing bias in an agency's decisions will be carried over into new systems trained on biased agency data.[34] For instance, many data used by predictive policing systems come from the Stop-and-Frisk program in New York City and the Terry Stop policy. This historical data ('dirty data')[35] create a discriminatory pattern because data from 2004 to 2012 showed that 83 per cent of the stops were of black and Hispanic individuals and 33 per cent white. The overrepresentation of black and Hispanic people who were stopped may lead an algorithm to associate typically black and Hispanic traits with stops that lead to crime prevention.[36] Despite its over-inclusivity, inaccuracy, and disparate impact,[37] such data continue to be processed.[38] Consequently, the algorithms will consider African Americans as a high-risk population (resulting in a 'feedback loop' or a self-fulfilling prophecy),[39] as greater rates of police inspection lead to a higher rate of reported crimes, therefore reinforcing disproportionate and discriminatory policing practices.[40] Obviously, these tools may violate human rights protections in the United States, as well as in the European Union, both before or after their deployment.

A *priori*, predictive policing activities can violate the fundamental rights of individuals if certain precautions are not taken. Though predictive policing tools are useful for the prevention of offenses and the management of police forces, they should not be accepted as sufficient motive for stopping and/or questioning individuals. Several fundamental rights can be violated in case of abusive, disproportionate, or unjustified use of predictive policing tools: the right to physical and mental integrity (Charter of Fundamental Rights of the European Union, art. 3); the right to liberty and security (CFREU, art. 6); the right to respect for private and family life,

[32] Anupam Chander, 'The Racist Algorithm?' (2017)115 *Michigan Law Review* 1023–1045.

[33] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015).

[34] Kristian Lum and William Isaac, 'To Predict and Serve?' (7 October 2016) 13(5) *Significance* 14–19 https://doi.org/10.1111/j.1740-9713.2016.00960.x accessed 23 May 2019.

[35] Rashida Richardson, Jason Schultz, and Kate Crawford, 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice' (2019) https://papers.ssrn.com/abstract=3333423 accessed 15 February 2019.

[36] Solon Barocas and Andrew D. Selbst, 'Big Data's Disparate Impact' (2016) 104 *California Law Review* 671–732; Joshua Kroll et al., 'Accountable Algorithms' (2017) 165 *U Pa L Rev* 633.

[37] Solon Barocas and Andrew D. Selbst, 'Big Data's Disparate Impact' (2016) 104 *California Law Review* 671–732; Alexandra Chouldechova, 'Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments' (2016) http://arxiv.org/abs/1610.07524 accessed 12 August 2018.

[38] NYCLU, 'Stop and Frisk Data' (14 March 2019) www.nyclu.org/en/publications/stop-and-frisk-de-blasio-era-2019 accessed 23 May 2019.

[39] US DJ, supra note 18.

[40] PE 624.261, supra note 6.

home, and communications; the right to freedom of assembly and of association (CFREU, art. 12); the right to equality before the law (CFREU, art. 20); and the right to non-discrimination (CFREU, art. 21). The risks of infringing on these rights are greater if predictive policing tools target people, as opposed to places. The fact remains that the mere identification of a high-risk zone does not naturally lead to more rights for the police, who, in principle, must continue to operate within the framework of crime prevention and the maintenance of order.

In the United States, due process (the Fifth and Fourteenth Amendments)[41] and equal treatment clauses (the Fourteenth Amendment) could be infringed. Moreover, predictive policing could constitute a breach of privacy or infringe on citizens' rights to be secure in their persons, houses, papers, and effects against unreasonable searches and seizures without a warrant based on a 'probable cause' (the Fourth Amendment). Similar provisions have been enacted in the State Constitutions. Despite the presence of these *theoretical* precautions, some infringements of fundamental rights have been revealed *in practice*.[42]

A *posteriori*, these risks are higher when algorithms are involved in systems used to support decision-making by police departments. Law enforcement may find it needs to answer to the conditions of use of these tools on a case-by-case basis when decisions are reached involving individuals. To provide an example, the NYPD was taken to court for the use of the Palantir Gotham tool and its technical features.[43] The lack of information on the existence and use of predictive tools, the nature of the data in question, and the conditions of application of algorithmic results based on automated treatment were all contested on the basis of a lack of transparency and the resulting impossibility to enforce the defence's right to due process (the Fifth and Fourteenth Amendments).[44] Additionally, the media,[45] academics,[46] and civil rights defence organizations[47] have called out against the issues of bias and discrimination within these tools, which violate the Fourteenth Amendment principle of Equal Protection for all citizens under the law. In EU law, the Charter of Fundamental Rights also guarantees the right to an effective remedy and access to a fair trial (CFREU, art. 47), as well as the right to presumption of innocence and right of defence (CFREU, art.

[41] Danielle Keats Citron, 'Technological Due Process' (2008) 85 *Washington University Law Review*.

[42] David Robinson and Logan Koepke, 'Stuck in a Pattern: Early Evidence on "Predictive Policing" and Civil Rights' *Upturn* (August 2016) www.stuckinapattern.org accessed 23 May 2019.

[43] *Brennan Center for Justice at New York University, School of Law* v. *NYPD*, Case n. 160541/2016, December 22nd, 2017 (FOIA request (Freedom of Information Law Act)). The judge approved the request and granted access to the *Palantir Gotham* system used by the NYPD: https://law.justia.com/cases/new-york/other-courts/2017/2017-ny-slip-op-32716-u.html.

[44] *State of Wisconsin* v. *Loomis*, 371 Wis 2d 235, 2016 WI 68, 881 N W 2d 749 (13 July 2016).

[45] For example, Ben Dickson, 'What Is Algorithmic Bias?' (26 March 2018) https://bdtechtalks.com/2018/03/26/racist-sexist-ai-deep-learning-algorithms accessed 23 May 2019.

[46] For example, AINow Institute https://ainowinstitute.org.

[47] For example, Vera Eidelman, 'Secret Algorithms Are Deciding Criminal Trials and We're Not Even Allowed to Test Their Accuracy' (ACLU 15 September 2017) www.aclu.org/blog/privacy-technology/surveillance-technologies/secret-algorithms-are-deciding-criminal-trials-and accessed 23 May 2019.

48). All of these rights can be threatened if the implementation of predictive policing tools is not coupled with sufficient legal and technical requirements.

The necessity of protecting fundamental rights has to be reiterated in the algorithmic society. To achieve this, adapted tools must be deployed to ensure proper enforcement of fundamental rights. Some ethical principles need to be put in place in order to effectively protect fundamental rights and reinforce them. The goal is not substituting human rights with ethical principles but adding new ethical considerations focused on risks generated by ADS. These ethical principles must be accompanied by practical tools that will make it possible to provide designers and users with concrete information regarding what is expected when making or using automated decision-making tools. Algorithmic Impact Assessment (AIA) constitutes an interesting way to provide a concrete governance of ADS. I argue that while the European constitutional and ethical framework is *theoretically* sufficient, other tools must be adopted to guarantee the enforcement of Fundamental Rights and Ethical Principles *in practice* to provide a robust *framework* for putting human rights at the centre.

6.3 HUMAN RIGHTS REINFORCED BY ETHICAL PRINCIPLES TO GOVERN AI

Before considering the enactment of ethical principles to reinforce fundamental rights in the use of ADS, one needs to identify whether or not efficient legal provisions are already enacted.

6.3.1 *Statutory Provisions in the European Law*

At this time, very few statutory provisions in European Law are capable of reinforcing the respect and protection of fundamental rights with the use of ADS. ADS are algorithmic processes which require data in order to perform. Predictive policing systems do not automatically use personal data, but some of them do. In this case, if the processed personal data concerns some data subjects within the European Union, the General Data Protection Regulation (GDPR) may be applied by the private companies. Moreover, police services are subject to the Data Protection Law Enforcement Directive. It provides for several rights in favour of the data subject, especially the 'right to receive a meaningful information concerning the logic involved' (art. 13–15) and the right 'not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning one or similarly significantly affects one' (art. 22),[48] in addition to a Data Protection Impact Assessment (DPIA) tool (art. 35).[49]

[48] Margot E. Kaminski, 'The Right to Explanation, Explained' (2018) *Berkeley Technology Law Journal* 34(1).

[49] Margot E. Kaminski and Malgieri, Gianclaudio, 'Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations' (2019). U of Colorado Law Legal Studies Research Paper No. 19–28. Available at SSRN: https://ssrn.com/abstract=3456224.

https://doi.org/10.1017/9781108914857.007 Published online by Cambridge University Press

However, these provisions fail to provide adequate protection against the violation of human rights. First, several exceptions restrict the impact of these rights. Article 22 paragraph 1 is limited by paragraph 2, according to which the right 'not to be subject to an automated decision' is excluded, when consent has been given or a contract concluded. This right is also excluded if exceptions have been enacted by the member states.[50] For instance, French Law[51] provides an exception in favour of the governmental use of ADS. Consequently, Article 22 is insufficient per se to protect data subjects. Second, ADS can produce biased decisions without processing personal data, especially when a group is targeted in the decision-making process. Even if the GDPR attempts to consider the profiling of data subjects and decisions that affect groups of people, for instance, through collective representation, such provisions are insufficient to prevent group discrimination.[52] Third, other risks against fundamental rights have to be considered, such as procedural guarantees related to the presumption of innocence and due process. The protection of such rights is not, or at least not directly, within the scope of the GDPR. The personal data protection regulations cannot address all the social and ethical risks associated with ADS. Consequently, such provisions are insufficient, and because other specific statutory provisions have not yet been enacted,[53] ethical guidelines could be helpful as a first step.[54]

6.3.2 *European Ethics Guidelines for Trustworthy AI*

In the EU, the *Ethics Guidelines for Trustworthy Artificial Intelligence (AI)* is a document prepared by the High-Level Experts Group on Artificial Intelligence (AI HLEG). This group was set up by the European Commission in June 2018 as part of the AI strategy announced earlier that year. The AI HLEG presented a first draft of the *Guidelines* in December 2018. Following further deliberations, the *Guidelines*

[50] Céline Castets-Renard, 'Accountability of Algorithms: A European Legal Framework on Automated Decision-Making' (2019) *Fordham Intell. Prop., Media & Ent. Law Journal* 30(1). Available at https://ir.lawnet.fordham.edu/iplj/vol30/iss1/3.

[51] Loi n 78–66 'Informatique et Libertés' enacted on 6 January 1978 and modified by the Law n 2018–493, enacted on 20 June 2018: www.legifrance.gouv.fr/eli/loi/2018/6/20/JUSC1732261L/jo/texte.

[52] However, we also have to consider antidiscrimination directives: Directive 2000/43/EC against discrimination on grounds of race and ethnic origin; Directive 2000/78/EC against discrimination at work on grounds of religion or belief, disability, age, or sexual orientation; Directive 2006/54/EC equal treatment for men and women in matters of employment and occupation; Directive 2004/113/EC equal treatment for men and women in the access to and supply of goods and services.

[53] The situation is similar in the United States, except the adoption of the NYC Local Law n 2018/049 concerning automated decision systems used by the local agencies. In the state of Idaho, the Bill n 118 concerning the pretrial risk assessment algorithms and the risk to civil rights of automated pretrial tools in criminal justice was enacted on 4 March 2019: www.muckrock.com/news/archives/2019/mar/05/algorithms-idaho-legislation.

[54] See Luciano Floridi et al., 'AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations' (2018) 28 *Minds & Machines* 689–707.

were revised and published in April 2019, the same day as a European Commission Communication on *Building Trust in Human-Centric Artificial Intelligence*.[55]

Guidelines are based on the fundamental rights enshrined in the EU Treaties, with reference to dignity, freedoms, equality and solidarity, citizens' rights, and justice, such as the right to a fair trial and the presumption of innocence. These fundamental rights are at the top of the hierarchy of norms of many States and international texts. Consequently, they are non-negotiable and even less optional. However, the concept of 'fundamental rights' is integrated with the concept of 'ethical purpose' in these *Guidelines*, which creates a normative confusion.[56] According to the Experts Group, while fundamental human rights legislation is binding, it still does not provide comprehensive legal protection in the use of ADS. Therefore, the *AI Ethics Principles* have to be understood both within and beyond these fundamental rights. Consequently, trustworthy AI should be (1) lawful – respecting all applicable laws and regulations; (2) ethical – respecting ethical principles and values; and (3) robust – both from a technical perspective while taking into account its social environment.

The key principles are the principle of respect for human autonomy, the principle of prevention of harm, the principle of fairness, and the principle of explicability.[57] However, an explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible.[58] These cases are referred to as 'black box' algorithms and require special attention. In those circumstances, other explicability measures (e.g., traceability, auditability, and transparent communication on system capabilities) may be required, provided that the system as a whole respects fundamental rights.

In addition to the four principles, the Expert Group established a set of seven key requirements that AI systems should meet in order to be deemed trustworthy: (1) Human Agency and Oversight; (2) Technical Robustness and Safety; (3) Privacy and Data Governance; (4) Transparency; (5) Diversity, Non-Discrimination, and Fairness; (6) Societal and Environmental Well-Being; and (7) Accountability.

Such principles and requirements certainly push us in the right direction, but they are not concrete enough to indicate to ADS designers and users how they can ensure the respect of fundamental rights and ethical principles. Back to the predictive policing activity, the risks against fundamental rights have been identified but

[55] European Commission, 'Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Building Trust in Human-Centric Artificial Intelligence' COM (2019) 168 final.

[56] B. Wagner and S. Delacroix, 'Constructing a Mutually Supportive Interface between Ethics and Regulation' (14 June 2019): https://ssrn.com/abstract=3404179.

[57] Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2018) *IEEE Security & Privacy* 16(3) https://papers.ssrn.com/abstract=3052831 accessed 5 December 2018.

[58] Paul B. de Laat, 'Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?' (2017) *Philos Technol* 1–17.

not yet addressed. The recognition of ethical principles adapted to ADS is useful for highlighting specific risks but nothing more. It is insufficient to protect human rights, and they must be accompanied by practical tools to guarantee their respect on the ground.

6.4 HUMAN RIGHTS REINFORCED BY PRACTICAL TOOLS TO GOVERN ADS

In order to identify solutions and practical tools, excluding the instruments of self-regulation,[59] the 'Trustworthy AI Assessment List' proposed by the Group of Experts can first be considered. Aiming to operationalize the ethical principles and requirements, the Guidelines present an assessment list that offers guidance on the practical implementation of each requirement. This assessment list will undergo a piloting process in which all interested stakeholders can participate, in order to gather feedback for its improvement. In addition, a forum to exchange best practices for the implementation of Trustworthy AI has been created. However, the goal of these Guidelines and the List is to regulate the activities linked with AI technologies via a general approach. Consequently, the measures proposed are broad enough to cover many situations and different applications of AI, such as climate action and sustainable infrastructure, health and well-being, quality education and digital transformation, tracking and scoring individuals, and lethal autonomous weapon systems (LAWS). But while our study concerns predictive policing activities, it is more relevant to consider specific, practical tools which regulate the governmental activities and ADS.[60] In this sense, the Canadian government enacted in February 2019 a Directive on Automated Decision-Making[61] and a method on AIA.[62] These tools pursue the goal of offering governmental institutions a practical method to comply with fundamental rights, laws, and ethical principles. I argue that these methods are relevant to assess the activity of predictive policing in theory.

[59] European Parliamentary Research Service (EPRS), 'Panel for the Future of Science and Technology, A Governance Framework of Algorithmic Accountability and Transparency' April 2019 (PE 624.262) [PE 624.262]. I exclude the self-regulation solutions, such as ethics committees, because they may, in fact, be a way to manage public image and avoid government regulation. See Ben Wagner, *Ethics as an Escape from Regulation: From Ethics-Washing to Ethics-Shopping?* (Amsterdam University Press, 2018); Yeung Karen et al., *AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing* (Oxford University Press, 2019). Luciano Floridi, 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical' (2019) *Philosophy & Technology* 32(2).

[60] For instance , Marion Oswald et al., 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and "Experimental" Proportionality' (2017) *Information & Communications Technology Law* https://papers.ssrn.com/abstract=3029345 accessed 23 May 2019.

[61] Directive on Automated Decision-Making (2019) www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

[62] Government of Canada, Algorithmic Impact Assessment (8 March 2019) https://open.canada.ca/data/en/dataset/748a97fb-6714-41ef-9fb8-637a0b8e0da1 accessed 23 May 2019.

6.4.1 *Methods: Canadian Directive on Algorithmic Decision-Making and the Algorithmic Impact Assessment Tool*

The Canadian Government announced its intention to increasingly look to utilize artificial intelligence to make, or assist in making, administrative decisions to improve the delivery of social and governmental services. This government is committed to doing so in a manner that is compatible with core administrative legal principles such as transparency, accountability, legality, and procedural fairness, as based on the directive, and an AIA. An AIA is a framework to help institutions better understand and reduce the risks associated with ADS and to provide the appropriate governance, oversight, and reporting/audit requirements that best match the type of application being designed. The Canadian AIA is a questionnaire designed to assist the administration in assessing and mitigating the risks associated with deploying an ADS. The AIA also helps identify the impact level of the ADS under the proposed Directive on Automated Decision-Making. The questions are focused on the business processes, the data, and the systems to make decisions.

The Directive took effect on 1 April 2019, with compliance required by no later than 1 April 2020. It applies to any ADS developed or procured after 1 April 2020 and to any system, tool, or statistical model used to recommend or make an administrative decision about a client (the recipient of a service). Consequently, this provision does not apply in the criminal justice system or criminal proceedings. This Directive is divided into eleven parts and three appendices on Purpose, Authorities, Definitions, Objectives and Expected Results, Scope, Requirements, Consequences, Roles and Responsibilities of Treasury Board of Canada Secretariat, Application, References, and Enquiries. The three appendices concern the Definitions (appendix A), the Impact Assessment Levels (appendix B), and the Impact Level Requirements (appendix C).

The objective of this Directive is to ensure that ADS are deployed in a manner that reduces risks to Canadians and federal institutions, leading to more efficient, accurate, consistent, and interpretable decisions made pursuant to Canadian law. The expected results of this Directive are as follows:

- Decisions made by federal government departments are data-driven, responsible, and comply with procedural fairness and due process requirements.
- Impacts of algorithms on administrative decisions are assessed, and negative outcomes are reduced, when encountered.
- Data and information on the use of ADS in federal institutions are made available to the public, where appropriate.

Concerning the requirements, the Assistant Deputy Minister responsible for the program using the ADS, or any other person named by the Deputy Head, is responsible for AIA, transparency, quality assurance, recourse, and reporting. He has to provide with any applicable recourse options that are available to them to

challenge the administrative decision, and to complete an AIA prior to the production of any ADS. He can use the AIA tool to assess and mitigate the risks associated with deploying an ADS based on a questionnaire.

6.4.2 *Application of These Methods to Predictive Policing Activities*

Though such measures specifically concern the Government of Canada and do not apply to criminal proceedings, I propose to use this method both abroad and more extensively. It can be relevant for any governmental decision-making, especially for predictive policing activities. I will consider the requirements that should be respected by people responsible for predictive policing programs. Those responsible should be appointed to perform their work on the ground, for each predictive tool used. This would be done using a case-by-case approach.

The first step is to assess the impact in consideration of the 'impact assessment levels' provided by appendix B of the Canadian Directive.

Appendix B: Impact Assessment Levels

Level	Description
I	The decision will likely have little to no impact on: • the rights of individuals or communities, • the health or well-being of individuals or communities, • the economic interests of individuals, entities, or communities, • the ongoing sustainability of an ecosystem. Level I decisions will often lead to impacts that are reversible and brief.
II	The decision will likely have moderate impacts on: • the rights of individuals or communities, • the health or well-being of individuals or communities, • the economic interests of individuals, entities, or communities, • the ongoing sustainability of an ecosystem. Level II decisions will often lead to impacts that are likely reversible and short-term.
III	The decision will likely have high impacts on: • the rights of individuals or communities, • the health or well-being of individuals or communities, • the economic interests of individuals, entities, or communities, • the ongoing sustainability of an ecosystem. Level III decisions will often lead to impacts that can be difficult to reverse, and are ongoing.
IV	The decision will likely have very high impacts on: • the rights of individuals or communities, • the health or well-being of individuals or communities, • the economic interests of individuals, entities, or communities, • the ongoing sustainability of an ecosystem. Level IV decisions will often lead to impacts that are irreversible, and are perpetual.

https://doi.org/10.1017/9781108914857.007 Published online by Cambridge University Press

At least level III would be probably reached for predictive policing activities in consideration of the high impact on the freedoms and rights of individuals and communities previously highlighted.

Keeping these levels III and IV in mind, they reveal in a second step the level of risks and requirements. Defined in appendix C, it indicates the 'requirements', concerning especially the notice, the explanation, and the human-in-loop process. The 'notice requirements' are focus on more transparency, which is particularly relevant to address the opacity problem of predictive policing systems.

Appendix C: Impact level requirements

Requirement	Level I	Level II	Level III	Level IV
Notice	None	Plain language notice posted on the program or service website.	Publish documentation on relevant websites about the automated decision system, in plain language, describing: • How the components work; • How it supports the administrative decision; and • Results of any reviews or audits; and • A description of the training data, or a link to the anonymized training data if these data are publicly available.	

These provisions allow one to know if the algorithmic system makes or supports the decision at levels III and IV. They also inform the public about the data used, especially from the start of the training process. This point is particularly relevant, in consideration of the historical and biased data mainly used in predictive policing systems. These requirements could help solve the discriminatory problem.

Moreover, AIAs usually provide a pre-procurement step that gives the public authority the opportunity to engage in a public debate and proactively identify concerns, establish expectations, and draw on expertise and understanding from relevant stakeholders. This is also when the public and elected officials can push back against deployment before potential harms occur. In implementing AIAs, authorities should consider incorporating them into the consultation procedures that they already use for procuring algorithmic systems or for assessing their pre-acquisition.[63] It would be a way to address the lack of transparency of predictive policing systems which should be addressed at levels III and IV.

Besides, other requirements concern the 'explanation'.

[63] PE 624.262, supra note 60.

https://doi.org/10.1017/9781108914857.007 Published online by Cambridge University Press

Requirement	Level I	Level II	Level III	Level IV
Explanation	In addition to any applicable legislative requirement, ensuring that a meaningful explanation is provided for common decision results. This can include providing the explanation via a Frequently Asked Questions section on a website.	In addition to any applicable legislative requirement, ensuring that a meaningful explanation is provided upon request for any decision that resulted in the denial of a benefit, a service, or other regulatory action.	In addition to any applicable legislative requirement, ensuring that a meaningful explanation is provided with any decision that resulted in the denial of a benefit, a service, or other regulatory action.	

At levels III and IV, each regulatory action that impacts a person or a group requires the provision of a meaningful explanation. Concretely, if these provisions were made applicable to police services, the police departments who use some predictive policing tools should be able to give an explanation of the decisions made and the way of reasoning, especially in the case of using personal data. The place or a person targeted by predictive policing should also be explained.

Concerning the 'human-in-loop for decisions' requirement, levels III and IV impose a human intervention during the decision-making process. That is also relevant for predictive policing activities which require that the police officers keep their free will and self-judgment. Moreover, the human decision has to prevail over the machine-decision. That is crucial to preserve the legitimacy and autonomy of the law enforcement authorities, as well as their responsibility.

Requirement	**Level I**	**Level II**	**Level III**	**Level IV**
Human-in-the-loop for decisions	Decisions may be rendered without direct human involvement.		Decisions cannot be made without having specific human intervention points during the decision-making process, and the final decision must be made by a human.	

Furthermore, if infringement on human rights has to be prevented, additional requirements on testing, monitoring, and training have to be respected at all levels. Before going into production, the person in charge of the program has to develop the appropriate processes to ensure that training data are tested for unintended data biases and other factors that may unfairly impact the outcomes. Moreover, he has to

ensure that data being used by the ADS are routinely tested to verify that it is still relevant, accurate, and up-to-date. He also has to monitor the outcomes of ADS on an ongoing basis to safeguard against unintentional outcomes and to ensure compliance with legislations.

Finally, the 'training' requirement for level III concerns the documentation on the design and functionality of the system. Training courses must be completed, but contrary to level IV, there is surprisingly no obligation to verify that it has been done.

The sum of these requirements is relevant to mitigate the risks of opacity and discrimination. However, alternately, it does not address the problem of efficiency. Such criteria should also be considered in the future, as the example of predictive policing activities reveals a weakness regarding the efficiency and social utility of this kind of algorithmic tool at this step. It is important not to consider that an ADS is necessarily efficient *by principle*. Public authorities should provide evidence of it.

6.5 CONCLUSION

Human rights are a representation of the fundamental values of a society and are universal. However, in an algorithmic society, even if a European lawmaker pretends to reinforce the protection of these rights through ethical principles, I have demonstrated that the current system is not good enough when it comes to guaranteeing their respect *in practice*. Constitutional rights must be reinforced not only by ethical principles but even more by specific practical tools taking into account the risks involved in ADS, especially when the decision-making concerns sensitive issues such as predictive policing. Beyond the *Ethics Guidelines for Trustworthy AI*, I argue that the European lawmaker should consider enacting similar tools as the *Canadian Directive on Automated Decision Making* and AIAs policies that must be made applicable to police services to make them accountable.[64] AIAs will not solve all of the problems that algorithmic systems might raise, but they do provide an important mechanism to inform the public and to engage policymakers and researchers in productive conversation.[65] Even if this tool is certainly not perfect, it constitutes a good starting point. Moreover, I argue this policy should come from the European Union and not its member states. The protection of human rights in an algorithmic society may be considered globally as a whole system integrating human rights. The final result is providing a robust theoretical and practical framework, while human rights keep a central place within this broad system.

[64] See a similar recommendation in EPRS Study PE 624.262, supra note 60.
[65] Ibid.

https://doi.org/10.1017/9781108914857.007 Published online by Cambridge University Press