BEYOND THE PREDICTION PARADIGM: CHALLENGES FOR AI IN THE STRUGGLE AGAINST ORGANIZED CRIME

DR. PAULA HELM* & DR. THILO HAGENDORFF**

I
INTRODUCTION

The use of Artificial Intelligence (AI) for policing is a hot topic. This is not only because of the hopes and promises placed in it, but also because it is sharply criticized by human rights activists, ethicists and social scientists.[1] In particular, the use and implementation of so-called predictive policing technologies (PPT) is being disapproved of by many entities. The criticism of these systems refers to some of its well-known aspects, such as lack of accountability, problematic biases in the data sets, intrusion into personal rights, and superficiality. Despite the criticism, PPTs are now regularly used across criminal justice and law enforcement institutions. Judges, parole boards, police commanders, and patrol officers make daily assessments, evaluations, and assignments based on these technologies. They insist that automated data analysis makes institutional decision-making more effective, consistent, neutral, and, most importantly, it makes policing smarter.[2]

This Article is also available online at http://lcp.law.duke.edu/.

* University of Tübingen, International Center for Ethics in the Sciences and Humanities. Dr. Helm was supported by the Federal Ministry of Education and Research (Germany), project "PEGASUS – Polizeiliche Gewinnung und Analyse heterogener Massendaten zur Bekämpfung organisierter Kriminalitätsstrukturen."

** University of Tübingen, International Center for Ethics in the Sciences and Humanities. Dr. Hagendorff was supported by the Cluster of Excellence "Machine Learning – New Perspectives for Science" funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Reference Number EXC 2064/1 – Project ID 390727645. The authors would like to thank all the participants of the symposium of *Law & Contemporary Problems* on "Black Box Artificial Intelligence and the Rule of Law" for helpful comments on the manuscript.

1. *See generally* ACLU et al., *Predictive Policing Today: A Shared Statement of Civil Rights Concern* (Aug. 13, 2016), https://www.aclu.org/other/statement-concern-about-predictive-policing-aclu-and-16-civil-rights-privacy-racial-justice [https://perma.cc/52PY-4Q57] (discussing concerns with predictive policing tools employed by law enforcement).

2. *See* Eric Siegel, *How to Fight Bias with Predictive Policing*, SCI. AM. (Feb. 19, 2018), https://blogs.scientificamerican.com/voices/how-to-fight-bias-with-predictive-policing/ [https://perma.cc/9SVD-255P] (explaining how predictive policing presents an opportunity to advance social justice if it is "done right").

What does the term "smart policing" mean in this context? While most definitions focus on the technical tools involved,[3] police practitioners Nola Joyce, Charles Ramsey, and James Stewart, as well as information scientist Walter Perry and others offer both a broader and deeper understanding by defining "smart policing" as investigations that, through the close collaboration between researchers, practitioners, and technology developers, apply an analytical lens to tackle the root causes of crime.[4] According to these authors, policing that is "smart" in this sense should be the goal that guides innovation and reform. What makes policing "smart" is not only its quantitative efficiency, but also its qualitative sensitivity. Smart policing should be attentive to its potentials, but also to its threats and risks, as well as its cultural and historical context. Smart policing should also not rely on exclusion, enforcement, and demarcation, but use technical tools to promote the mutual inclusion of competing groups, trust in government, and participation in society.[5]

In this conception, it is not simply the use of an algorithm that makes policing "smart." Rather, it is the embedding of the algorithm, its tasks, its applications, its utility functions, and its (long-term) impact on individuals and societies that make the difference. AI-based predictive policing as we know it today has little to do with "smart policing" in this broader sense. Current PPTs are not designed, let alone systematically deployable, to analyze data on the much more complex underlying structures of crime. They are only capable of performing rather simple tasks, such as designating areas for street patrols. Their aim is to predict and thus help prevent overt offenses such as street robbery, burglary, the use and sale of illegal drugs, or theft.[6] But what about organized crime, which, according to Interpol, remains one of the greatest threats to contemporary democracies because it systematically destroys small businesses and corrodes trust in democratic institutions?[7] This form of crime is mostly based on strategies of

3. *See* James R. Coldren, Alissa Huntoon & Michael Medaris, *Introducing Smart Policing: Foundations, Principles and Practice*, 16 POLICE Q. 275, 275–284 (2013) (defining smart policing and outlining characteristics in local smart policing sites).

4. *See generally* Nola M. Joyce, Charles Ramsey & James Stewart, *Commentary on Smart Policing*, 16 POLICE Q. 358 (2013) (focusing on research partnership and collaborative problem solving in smart policing); WALTER L. PERRY, BRIAN MCINNIS, CARTER C. PRICE, SUSAN C. SMITH & JOHN S. HOLLYWOOD, PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS (2013) (detailing the analytical techniques and tactical approaches available in predictive policing).

5. For a more in-depth analysis, see JACQUES RANCIÈRE, DISSENSUS: ON POLITICS AND AESTHETICS 36 (2010).

6. *See* Aaron Shapiro, *Predictive Policing for Reform? Indeterminacy and Intervention in Big Data Policing*, 17 SURVEILLANCE & SOC'Y 456, 456–72 (2019).

7. According to Interpol, the N'drangetha, with an annual turnover of 53 billion euros, and chapters in forty-three countries, is the most powerful criminal organization in the world. It is so immensely successful precisely because it avoids any kind of visible and hence easily detectable crime and instead builds on a combination of corruption, collusion, infiltration, and money laundering. INTERPOL, ITALY AND INTERPOL LAUNCH GLOBAL PROJECT TO COMBAT 'NDRANGETHA (Jan. 30, 2020), https://www.interpol.int/News-and-Events/News/2020/Italy-and-INTERPOL-launch-global-project-to-combat-Ndrangheta [https://perma.cc/J6WG-ZSJ6].

corruption and collusion and avoids all forms of open violence.[8] It is therefore unlikely to be combated by, for example, street patrols or hot-spot policing. Pattern recognition and crime series detection are not very useful here because these approaches so far address only acts of violent or well-reported and easy to detect crime, not subtle and complex ones such as corruption or money laundering.[9] Therefore, other systems are urgently needed to relieve criminal investigators of the overwhelming task of understanding, uncovering, and proving complicated structures of organized crime in which money flows, politics, logistics, international relations, blood ties, and cultural traditions are dynamically interwoven.

There is still a long way to go before we have machines capable of taking us significantly further in the struggle against organized crime. Recently, however, the pace of machine learning is accelerating. As more data is becoming available, the merging of heterogeneous mass data is working more smoothly than ever before. Moreover, visualization techniques that reduce complexity by translating machine language into comprehensible cues are making communication easier, and the development of more complex systems aimed not only at detecting patterns and correlations but also at analyzing causality and structure is on the horizon.[10] Applied to the police context, this shift from patterns to structures and from correlation to causality opens the potential for a shift from fighting symptoms and individuals to fighting relations and organizations.

The use of automation to assist in combating organized crime is certainly a laudable project. Nevertheless, there is reason to assume that many of the criticisms raised against existing PPTs could also apply against such an extension of the systems. This is because, although they are based on a different logic (instead of prediction, they rely on detection and understanding), they use similar statistical models, are commissioned and built by the same institutions, and are deployed within the same postcolonial societies. However, critical discussion and ethical evaluation of the potentially problematic consequences of organized

8. *See id*. The most powerful of the criminal organizations today can consistently be described as global players. In most cases, they have executive units on several continents, which often act autonomously. In this respect, criminal organizations can also be compared with social swarms, operating without centers but following a common logic. At the same time, however, many criminal organizations also have local roots, and their procedures are shaped by regional specifics across national borders. *See* Letizia Paoli, *The Paradoxes of Organized Crime,* 37 CRIME, L. & SOC. CHANGE 51, 51–97 (2002). In this paper, we take as our starting point criminal organizations that have their roots in Europe, but this does not mean that the findings discussed here do not apply to organized crime on other continents.

9. *See generally* Alex Chohlas-Wood & E.S. Levine, *A Recommendation Engine to Aid in Identifying Crime Patterns*, 49 INFORMS J. ON APPLIED ANALYTICS (2019) (implementing a recommendation algorithm called "Patternizr" to aid in identifying crime patterns); Tong Wang, Cynthia Rudin, Daniel Wagner & Rich Sevieri, *Learning to Detect Patterns of Crime*, *in* JOINT EUROPEAN CONFERENCE ON MACHINE LEARNING AND KNOWLEDGE DISCOVERY IN DATABASES (2013) (proposing a pattern detection algorithm).

10. *See generally* JONAS PETERS, DOMINIK JANZIG & BERNHARD SCHÖLKOPF, ELEMENTS OF CAUSAL INFERENCE: FOUNDATIONS AND LEARNING ALGORITHMS (2017) (focusing on causality and cause-effect models); JUDEA PEARL & DANA MACKENZIE, THE BOOK OF WHY: THE NEW SCIENCE OF CAUSE AND EFFECT (2018) (exploring causality and causal inference).

crime-oriented technologies is a blind spot in the current research landscape, which focuses almost exclusively on PPTs. This is because more complex organized crime-oriented technologies are still nascent; they are not yet in use, so their actual impacts cannot yet be empirically evaluated. However, we do not want to wait until they are mature and in use to evaluate them—not only because it is uneconomical to build these systems first and then change them again to comply with ethical standards, but also because the damage from discriminatory or intrusive automation can be very difficult to undo. Thus, since we have reason to believe that machine learning technologies for combating organized crime will mature and come into use sooner or later, they already deserve closer scrutiny. This can be realized by combining different approaches suggested within the field of anticipatory ethics.[11]

In this Article, we take a first step in this direction. In Parts II and III, we start with a discussion of the inherent logic and also the criticisms that have been raised against current PPTs so far. In Part IV, we continue by taking a more in-depth look at technologies which are "under development" that target more complex tasks related to combating organized crime. In this context, we explore the following questions: To what extent do the criticisms leveled against PPTs apply to these new systems? What other, potentially more far-reaching, ethical issues might be expected? What new challenges might arise and what potential lies in new approaches? In short, what will it take to make AI-based policing truly "smart" in a comprehensive sense?

II

THE PREDICTION PARADIGM – STATE OF THE ART

AI is being used in several ways to improve policing and criminal investigations. These include crime series detection, hotspot patrols, and suspect-based policing. They all target well-reported forms of crime for which sufficient past crime data are available. Most of the current approaches are subsumed and negotiated under the term "predictive policing". The term "predictive" can, however, be misleading because it suggests that specific predictions are being made, when in fact most systems are about prevention, deterrence, or even prosecution based on statistics, analysis, and probabilities. Nonetheless, the term "predictive policing" has become established, and when we talk about the predictive policing paradigm here, we are referring to a wide variety of types of AI-based policing approaches that are already in use. What unites them is that they address forms of physically violent or otherwise visible crime, but not the structural activities that occur beneath the surface of the iceberg, of which visible/direct forms of crime are often only the tip.

11. *See* Phillip Brey, *Anticipatory Ethics for Emerging Technologies*, 6 NANOETHICS 1, 1 (2012) (providing a method for technology impact assessment).

There are several definitions of predictive policing. Most of them do not contradict each other, but they emphasize different aspects.[12] According to one of the most prominent and comprehensive definitions, predictive policing can be summarized as: "the application of analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions."[13] This definition is interesting because it introduces the possibility that PPTs could be used not only to prevent crime but also to solve "past crimes" based on predictions that are usually associated with the future, not the past. How is this possible? Following Professor Adrian Mackenzie's more general thoughts on the subject of "prediction," the new paradigm of predictive policing —including its application to past crimes—is best understood by placing it in the larger context of machine learning today. This logic is based on the idea of feeding statistical models with enough data and identifying sufficient features to arrive at generalizations that are accurate and reliable enough to work like predictions.[14] In this understanding, predictive policing is not about actually making predictions, but about identifying probabilities. These probabilities may concern not only who is likely to commit a crime in the future (prediction), but also who has already committed one in the past (prosecution).

By becoming increasingly accurate, machine learning-based forecasts generate what Mackenzie calls "the desire to predict."[15] Machine learning not only feeds this desire, but increasingly serves it through its self-reinforcing effects. Still, it is not 100% reliable, the system only indicates probabilities, like all other human decision-making mechanisms. Unlike purely human reasoning, however, many cases of automated decision-making suggest infallibility. Because of its imperfect accuracy and sometimes ideological exaggerations combined with self-reinforcing effects, AI-driven policing has been the subject of much controversy over the past decade of development, testing, and application.

III

CONTROVERSIES SURROUNDING PREDICTIVE POLICING

In this Part, we provide an overview of some of the major criticisms of the current predictive policing approaches. PPTs shortcomings include both ethical issues and limitations in scope.[16] Here, we have divided the criticism of PPTs into

12. *See generally* Albert Meijer & Martijn Wessels, *Predictive Policing: Review of Benefits and Drawbacks*, 42 INT'L J. PUB. ADMIN. 1031 (2019) (providing an overview of the "existing literature on the benefits and drawbacks of predictive policing).

13. PERRY, *supra* note 4, at 1–2.

14. *See* Adrian Mackenzie, *The Production of Prediction: What Does Machine Learning Want?*, 18 EUR. J. CULTURAL STUD. 329, 433 (2015) (the paper highlights several interlinked modes of machine generalization).

15. *Id.* at 432.

16. To this end, we conducted a literature review based on a 5-step process: selection of key papers for closer examination, identification of discursive patterns, analysis of key controversies, the search for empirical evidence/rebuttals of criticisms (discursive openings and closings), and the identification of

three major categories. First, PPTs rely on data taken from past police records which may present significant privacy concerns and result in perpetuating institutional discrimination within law enforcement practices. Second, PPT algorithms can be difficult for law enforcement practitioners and the public to understand and they can provide a misleading sense of certainty known as "mathwashing." And finally, the "paradox of prediction" introduces the problem that if PPTs successfully predict and prevent crime, then eventually the data they are based on will no longer provide accurate predictions. These criticisms serve as the basis for Part IV, in which we assess the possibilities and difficulties of applying these machine learning techniques to combat organized crime.

A. Difficult Data

The bulk of critical papers dealing with PPTs focus on the data that is being used to train algorithms and to produce predictions.[17] For place-based policing, for example, meta data of recorded emergency calls from recent decades is used to determine future areas for hot-spot patrolling. Critics argue that such police data is biased in numerous ways. First, it cannot represent the many cases that have gone undetected. Due to this blind spot, police data alone are poorly suited for modeling systems that can help police achieve new success in the very areas where their success rates are lowest and where performance improvement is most needed.[18] Instead, it helps police optimize their success rates in areas where their rates were already high (such as apprehending petty drug dealers), while further relegating the areas where they were least successful.[19]

Second, police officers and departments regularly face empirically demonstrable accusations of being biased (if not racist) against people of color.[20] One would then assume that the police data used to feed the PPT algorithms

derivations. To realize the review, known databases were searched, each combining three of the following keywords in different variations in a total of five queries: Machine Learning, Smart Policing, Predictive Policing, Prosecution, and Evidence.

17. *See* Claire Garvie, *Garbage In, Garbage Out: Face Recognition on Flawed Data*, GEO. L. CTR. OF PRIV. & TECH. (May 16, 2019), https://www.flawedfacedata.com/ [https://perma.cc/5GNY-5NR5].

18. *See* TRANSPARENCY INT'L, CORRUPTION PERCEPTION INDEX 2019: FULL SOURCE DESCRIPTION, 6 (2020), https://images.transparencycdn.org/images/2019_CPI_SourceDescription_EN-converted-merged.pdf [https://perma.cc/89DN-BPZP] (providing an estimation of the level of corruption in different countries and regions as well as an overview of known corruption cases and how they impact political integrity).

19. *See* Kristian Lum & William Isaac, *To Predict and Serve*?, SIGNIFICANCE, Oct. 2016, at 14–19 (documenting biased data used for predictive policing systems).

20. *See* Elizabeth E. Joh, *Feeding the Machine: Policing, Crime Data, & Algorithms*, WM. & MARY BILL RTS. J. 287, 289, 294 (2017) ("To be sure, there have already been concerns raised that the inputs for policing algorithms reflect racial biases."); William A. Geller & Hans Toch, AND JUSTICE FOR ALL: UNDERSTANDING AND CONTROLLING POLICE ABUSE OF FORCE 303–08 (1995) (studying the relationship between race and the use of force); E. Ashby Plant & B. Michelle Peruche, *The Consequences of Race for Police Officers' Responses to Criminal Suspects*, 16 PSYCH. SCI. 180, 180 (2005) ("Responses to the simulation revealed that upon initial exposure to the program, the officers were more likely to mistakenly shoot unarmed Black compared with unarmed White suspects."); Andrew Gelman et al., *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias*, 102 J. AM. STAT. ASS'N (2007) (analyzing racial bias in police stops).

would reflect these prejudices.[21] When such biased data is being used to determine future police work, this inevitably leads to compounding injustice because the built-in biases are being solidified through the self-reinforcing effects of automation.[22] The more they patrol, the more likely they are to discover crime and feel their prejudices confirmed. This confirmation bias has been described as the vicious circle of "automated inequality."[23]

Based on the above arguments, the criticism against PPTs went viral in both scientific discourse and popular media. However, criticism was based only on theory, simulation, and anticipation. To disprove it, technology developers and researchers conducted several randomized field trials to test the advantages and disadvantages of PPTs.[24] Particularly the study carried out by Jeffrey Brantingham himself, owner of PredPol, showed that there is actually no evidence that PPTs reinforce existing biases.[25] However, because unfair biases existed before these systems were implemented, the argument remains that automation based on unfairly biased data sets stabilizes and confirms existing inequities.

Privacy concerns pose another problem with data mining in such a sensitive field like criminal investigation. Police technologies attempt to protect *individual* privacy by outputting only generalizations that are approximate enough to be considered predictions. But data-based profiling practices have also been criticized because they can provide the backdrop against which people are discriminated based on their membership in certain groups. For example, people might be regularly frisked and inspected based on their algorithmic assignment to a group that the algorithm has identified as particularly likely to possess illegal weapons, traffic in illegal drugs, or commit burglary and theft.[26] These interventions can have devastating effects not only on the social lives of the suspects, but on entire neighborhoods and communities.[27]

---

21. *See* Michael Townsley et al., *The Missing Link of Crime Analysis: A Systematic Approach to Testing Competing Hypotheses*, 5 POLICING 158, 163–64 (2011) (presenting a method for systematically analyzing multiple explanations for crime problems).

22. *See generally* Tim O'Brien, *Compounding Injustice: The Cascading Effect of Algorithmic Bias in Risk Assessments*, 13 GEO. J. L. & MOD. CRITICAL RACE PERSP. 1, 38–39 (2020) ("The criminal justice process consists of a chain of decisions that are each being increasingly influenced or automated with statistics and algorithms, creating a cascading effect that continues long after release and reentry.").

23. VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR, 190 (2018) (investigating the impact of policy algorithms on poor and working-class Americans).

24. George Mohler, M.B. Short, Sean Malinowski, Mark Johnson, George Tita, Andrea L. Bertozzi & P. Jeffrey Brantingham, *Randomized Controlled Field Trials of Predictive Policing*, 110 J. OF THE AM. STAT. ASS'N 1399 (2015); Jerry H. Ratcliffe, Ralph B. Taylor, Amber P. Askey, Kevin Thomas, John Grasso, Kevin J. Bethel, Ryan Fisher & Josh Koehnlein, *The Philadelphia Predictive Policing Experiment*, 17 J. EXPERIMENTAL CRIMINOLOGY 15 (2020).

25. P. Jeffrey Brantingham, Matthew Valasik & George Mohler, *Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial*, 5 STAT. & PUB. POL'Y 1, 1–6 (2018).

26. *See* Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 139, 139–40 (2017) (discussing this practice of "Suspect-Based Predictive Policing").

27. *See* PERRY, *supra* note 4.

Whether people are assigned to a suspicious group is largely outside their control. Even if they are very careful to protect their privacy, they change the fact that others in the group reveal much about themselves. Due to this interdependence of data subjects, the concept of privacy as an individual right is reaching its limits in the digital age.[28] As a result, scholars concerned with privacy and surveillance are increasingly turning their attention to group privacy.[29] In doing so, they strive to move beyond an understanding of group privacy in the traditional sense, according to which individuals deserve privacy protection even when they act within a group. Instead, they propose to shift the boundary between an understanding of group privacy as the privacy of individuals within a group and an understanding of group privacy as the privacy of the group itself.[30] Such boundary shifting is important to address the challenges posed by algorithmic profiling and typecasting.

B. Opacity and Conflicts of Authority

This strand of criticism addresses the problem of algorithmic accountability. While the black box architecture of many machine learning systems is often the focus in this context,[31] predictive policing is usually based on statistical models that an expert in the field can easily understand.[32] However, police officers, prosecutors, and investigators are generally not trained machine learning experts. Therefore, it is difficult for them to fathom how the systems they use generate their recommendations. This is a problem because practitioners are deprived of the ability to question recommendations. Also, the algorithms employed by commercial systems can be subject to trade secret protections. This black box architecture thus creates a conflict because the public has a right to transparency, and the way public officials interact with citizens is a matter of public interest.

Another strand of critique deals with the phenomenon of "mathwashing." Unlike the black box discourse, the "mathwashing" discourse focuses on the problem of opacity not primarily as a technological problem, but as an ideological one. "Mathwashing" refers to a worldview that classifies the results of algorithmic models as truths that cannot be discussed because they are mathematical.[33] As

28. *See* Alice E. Marwick & Danah Boyd, *Networked Privacy: How Teenagers Negotiate Context in Social Media*, 16 NEW MEDIA & SOC., 1051, 1051–67 (2014) ("Although models of data sharing are typically understood through the lens of individual rights and controls, the networked nature of social media means that individuals' experiences with their data are consistently imbricated with others.").

29. *See generally* Linnet Taylor, *Introduction: A New Perspective on Privacy, in* GROUP PRIVACY: NEW CHALLENGES OF DATA TECHNOLOGIES, 1, 1–12 (Linnet Taylor, Luciano Floridi & Bart van der Sloot eds., 2017); *see also* Paula Helm, *Treating Sensitive Topics Online: A Privacy Dilemma,* 20 ETHICS & INFO. TECH. 303, 303–313 (2018); *see also* Brent Mittelstadt, *From Individual to Group Privacy in Big Data Analytics,* 30 PHILOS. TECH. 475, 475–94 (2017).

30. TAYLOR ET AL., *supra* note 29, at 1–15.

31. *See generally*, FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015).

32. *See, e.g.*, Chohlas-Wood & Levine, *supra* note 9, at 93–171 (elaborating on the concept of predictive policing).

33. *See* TAYLOR WOODS, *'Mathwashing,' Facebook and the Zeitgeist of Data Worship,*

such, they are seen as neutral—free of biases, errors, and irrational decisions. To counter the myth of the flawless machine, scholars in the field of Critical Algorithm Studies are drawing attention to the various forms of subjectivity and intention that are woven into the technologies.[34] After all, the machines producing these predictions are the result of collective human decisions. These decisions include questions about what data to use, what features to select, what machine learning techniques to use, how to design the user interface so that it can be operated by non-tech savvy end users, what affordances to implement, how to visualize the results, and so on. Given the collective work and design decisions on which it is based, machine learning seems anything but neutral. Instead, it must be understood and treated as a power-saturated technology that serves the interests of those who commission it. In such an understanding, mathwashing appears to be a dangerous ideology. When people trust the neutrality of power-saturated tools that serve particular interests, they are stripped of their ability to question the results. This "erasure of doubt" is deeply problematic.[35]

C. Symptom-Oriented Policing: The Paradox of Prediction

This line of criticism is the most fundamental. It attacks PPTs for their conservatism and for treating only symptoms, not causes. Generalizations, and thus predictions, are made by statistically analyzing data from the past. Therefore, under this line of criticism, PPT can only be "reactionary." It relies on stability while reproducing that stability through its self-reinforcing effects.[36] It is a technology designed to identify effects, sequences, and correlations. It is not a technology designed to identify and change causes, because if it did, it would stop working. After all, if the basis on which PPTs operate changes, then the assumptions from the past that are used as the foundation for its analysis no longer applies. We call this *the paradox of predictive policing*: once it produces significant changes in criminal behavior patterns, it loses its power.

This shortcoming, namely the inability of this kind of technology to deal with instability or change, has been well demonstrated recently by AI systems used for marketing, inventory management, or fraud detection.[37] They failed to deal with rapidly changing consumer behavior during the COVID-19 pandemic.[38] Machine

---

TECHNICAL.LY (June 8, 2016), https://technical.ly/brooklyn/2016/06/08/fred-benenson-mathwashing-facebook-data-worship/ [https://perma.cc/4JR5-MPX] ("Mathwashing can be thought of using math terms (algorithm, model, etc.) to paper over a more subjective reality. For example, a lot of people believed Facebook was using an unbiased algorithm to determine its trending topics, even if Facebook had previously admitted that humans were involved in the process.").

34. *See generally* David Beer, *The Social Power of Algorithms*, 20 INFO., COMM. & SOC. (2017); Rob Kitchin, *Thinking Critically about and Researching Algorithms*, 20 INFO., COMM. & SOC. (2017).

35. *See* Louise Amoore, *Doubt and the Algorithm: On the Partial Accounts of Machine Learning,* 36 THEORY, CULTURE & SOC. 1, 1–23 (2019) (analyzing the status of doubt in humans in view of its calculability in machines).

36. *See* PERRY, *supra* note 4, at 674–92.

37. Thilo Hagendorff & Katharina Wezel, *15 Challenges for AI: or What AI (currently) Can't Do*, 35 AI & SOC. 355, 355–65 (2019).

38. Will Douglas Heaven, *Our Weird Behavior During the Pandemic Is Messing with AI Models*,

learning models work well with "normal" or "stable" behavior, but irregularities throw automated systems off their game. In policing, this is unacceptable. After all, for PPT to be socially valuable, positive change, such as safer and fairer communities, must be part of its goal. If this change occurs gradually, then the PPT systems must be constantly updated and fed with new data. However, this service is not often part of the contracts between police stations and the private companies that are selling the software.

IV

MACHINE LEARNING-BASED PROSECUTION OF ORGANIZED CRIME: OVERLAPS OF THE CRITIQUE?

AI-based prosecution of organized crime seems to be the logical next step to predictive policing techniques. While predictive policing techniques addresses incidents of violent or well-reported, visible forms of crimes, AI-based prosecution of organized crime is focused on identifying criminal networks and structures. Having discussed the major points of criticism regarding PPTs in Part III, we now scrutinize whether those same criticisms hold true for AI-based criminal prosecution techniques targeted at organized crime. We are aware of the fact that a difference exists between predicting and detecting crime. While prediction aims at prevention, detection aims at conviction. Hence, tools for predictive policing and criminal prosecution may in part follow different dynamics. This becomes obvious when keeping in mind that algorithmic predictions deal with the possible state of affairs, while prosecution aims at uncovering facts. However, as mentioned before, it would be a mistake to treat both fields completely separately, because PPTs are also about solving past crimes.

It is true that most of the criticism directed at PPTs addresses the uncertainties associated with the act of predicting the future. AI-based law enforcement tools, however, are not fact-finding assistants, but rather "fact reconstructors." These tools do not only follow probabilistic methods. They also incorporate subjective decisions that result from the technical measures used in crime analysis. This is the connecting link, the probabilistic as well as subjective logic that is either used to unveil potential future crimes or to infer unknown attributes of past crimes. Ultimately, both fields deal with uncertainty in law enforcement, so it makes sense to look for overlaps in the criticism.

But despite the similarities between the two fields, there are also differences. While PPTs target violent or well-reported crimes (including in the case of crime series detection), organized crime-oriented law enforcement tools seek to unravel complex structures of interconnected, covert activity by multiple

MIT TECH. REV. (Jan. 26, 2020, 2:44 PM), https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop/ [https://perma.cc/9PYT-UCLP] (describing how changing behavior routines due to COVID-19 impact the reliability of algorithmic decision making).

offenders and intricate, sometimes international, social networks. While predictive policing techniques on overt crime can build on a trove of well-populated databases of past police records, organized crime prosecution success rates have been extremely low, mostly due to a lack of available data. Therefore, AI-based law enforcement targeting organized crime must inductively infer unknown information based on heterogeneous mass data. It also must respond to real world limitations, such as by producing and using synthetic data to train algorithms where real-life data is unavailable. This is where machine learning comes into play.

The initial situation is that police departments are collecting more and more data traces about organized crime suspects. This unstructured data comes from, among others, seized smartphones, hard drives, lawful wiretaps and the like. These resources yield, for example, data on call logs, call duration, and cellular geo-information. In addition to this metadata, the authorities can also make use of simple audio data. If this data comes from wiretaps at locations where multiple people are present and speaking at the same time, different speakers can be separated and verified, and noise can be eliminated. Frequent targets are also cars, from which geodata is obtained. In addition, image and video files are of interest. Here, among other things, time stamps, geodata or information on the camera model are evaluated. By means of facial recognition techniques as well as other methods for checking biometric features, persons depicted can be identified, and various features can be determined, including their age and gender, whether they are wearing glasses, or whether they have "micro features" such as patches and tattoos. In addition, pieces of luggage or vehicle types can be evaluated. Furthermore, text files are evaluated with regard to their type, source, time stamp, and mentioned persons, places, or objects. Also, forensic text analysis is used to evaluate the semantics in confession letters, extortion letters, or hate mail, which can ultimately lead to probabilistic findings about the respective authors.

Information extracted from text, audio, images, graph structures, and other visual representations can be used to create representations designed to uncover crime connections. Systems translate unstructured big data into visualizations of suspected criminal structures. To begin with, seized smartphones, hard drives, and other data sources are transformed into a graphical interface that provides an overview of relevant individuals, their relationships, role assignments in the organized crime context, whereabouts, residences, headquarters and the like. These insights are further supplemented with open-source intelligence data obtained using tools such as web crawlers, meaning a bot that is systematically browsing the Internet. Ultimately, police software should make it easier to identify person-related "micro features" in heterogeneous data, extract relationships between people and places, and it should include timeline functionality that puts relevant events in chronological order. In summary, machine learning helps data preparation by converting audio files to text, translating text files, or extracting geospatial information from data. It helps in relevance assessments regarding the importance of certain data traces. Also,

machine learning is used to connect the dots to find recurring identifiers across different types of data traces. Finally, it is used to generate criminal hypotheses, for example by generating motion profiles of suspects, social networks of relevant individuals, or other visual representations for police officers.

A. Difficult Data

AI-based prosecution software can suffer from machine bias due to problematic biases in data, just like PPTs. Machine biases may for instance be an issue when generating synthetic data.[39] With synthetic data we here refer to data that is artificial, fictitious, yet realistic and which is generated by generative adversarial networks. Its purpose is to provide additional training data where real historic data are sparse or where available data cannot be used for reasons of privacy protection. For organized crime, synthetic data is sometimes the only choice not just to avoid reproducing racism baked in to past crime data or to circumvent privacy issues but also because available data on organized crime is sparse compared to small-scale or obviously violent offenses. Many areas of organized crime are still "dark fields", which can be shown via statistics for instance from Germany, where only very few investigation procedures are conducted.[40] Until today, few cases have been resolved.

Further complicating matters, organized crime is very context and culture specific, and patterns are therefore hard to generalize: they differ not only between organizations, but—due to chapter-like structure of most criminal organizations—also between regions within the same organization.[41] However, a sufficient amount of heterogenous data on past offenses must be available in order to train machine learning models. Therefore, apart from synthetic data, other sources of big data need to be taken into account.

The larger the data sets, the more possibilities can be considered. Apart from raising privacy issues, data taken from CCTV, social media, or other kinds of unstructured big data can bring its own problems. This kind of data can be both a blessing and a curse, as data sets can contain dirty or biased data as well as data with low accuracy. They can lead to spurious correlations that make it difficult to identify the "needles in the haystack" alongside criminally irrelevant information.

The data-collecting apparatuses generally capture what is easy to ensnare, meaning data that are made easily available or data that are a by-product of a

39. *See* Sachit Menon, Alexander Damian, Shijia Hu, Nikhil Ravi & Cynthia Rudin, *PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models*. In ARXIV: 2003.03808V3, 1–20 (2020) (discussing the potential of biases in techniques for generating synthetic, high-resolution, realistic images).

40. Bundeskriminalamt, *Organisierte Kriminalität. Bundeslagebild 2019,* BUNDESKRIMINALAMT (Nov. 25, 2020, 2:36 PM), https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/JahresberichteUndLagebilder/OrganisierteKriminalitaet/organisierteKriminalitaetBundeslagebild2019.pdf?__blob=publicationFile&v=2 [https://perma.cc/PX2A-VTAK].

41. *See* Letizia Paoli, *The Paradoxes of Organized Crime,* 37 CRIME, L. & SOC. CHANGE 51, 51–97 (2002) (arguing that most criminal organization can not be reduced to illegal market activities).

primary output.[42] Policing software that relies on unstructured big data does so by repurposing different types of data that were initially created as a by-product of a random activity, like riding a car or using a phone.[43] Hence, the logic of the data collection and analyses determines the way prosecution methods are shaped. Individuals become suspects because they occupy a certain position in artificial sociograms or in spatial patterns or because they are affected by other (arbitrary) correlations. Here, form can be mistaken for substance. That is not to say that machine learning cannot be a useful tool in making sense from large datasets on past crimes, quite the contrary. It can lead to insights that were previously impossible to achieve. But these insights must be interpreted with the mentioned methodological shortcomings, as well as privacy issues, in mind.

B. Opacity and Authority Conflicts

Machine learning-based crime analysis tools are opaque, just like any other machine learning-technique whose interpretability (the degree of human comprehensibility) and explainability (the understanding of criteria for decisions) are significantly restricted.[44] To solve explainability issues, visualization techniques are being used in order to make computational operations and outputs understandable for humans. But that process can further obfuscate the complexity and initial contextuality of what actually happened.[45] Besides visualization, opacity results from the statistical methods that are in use and the combinations thereof.[46]

In many areas of organized crime analysis, techniques for dimensionality reduction are put to use.[47] Heterogenous data are at the same time often high-dimensional data, hence data points have to be rendered understandable for humans, who can only interpret two, three, or maybe four-dimensional plots. Even if descriptive data analysis tools for dimensionality reduction increase interpretability while trying to minimize information loss, some loss is unavoidable. This can happen for instance by conflating several categories (such as similar crime cases) into one. Of course, domain and implicit knowledge are used when doing so, but, as outlined in Part III, one has to assume that most

42. Rob Kitchin & Tracey P. Lauriault, *Small Data in the Era of Big Data*, 80 GEOJOURNAL 463, 463–75 (2015).

43. *See* Klara Došlo, Conrad Meyer, Andreas Attenberger & Jessica Steinberger, *Driver Identification Using In-Vehicle Digital Data in the Forensic Context of a Hit and Run Accident*, 35 FORENSIC SCI. INT'L: DIGIT. INVEST. 1 (2020).

44. Brent Mittelstadt, Chris Russell & Sandra Wachter, *Explaining Explanations in AI* 1–10 (Conf. on Fairness, Accountability & Transparency, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3278331 [https://perma.cc/2L6W-BXUV].

45. *See* Matthias Leese, '*Seeing Futures:' Politics of Visuality and Affect*, in ALGORITHMIC LIFE: CALCULATIVE DEVICES IN THE AGE OF BIG DATA 143, 143–158 (Louise Amoore & Volha Piotukh eds., 2016) (describing how difficult it is to to make sense of data-driven environments).

46. Tasks like speech pattern recognition or visual analytics are tackled via recurrent neural nets or Transformers, but also classic methods like decision trees or k-nearest neighbor algorithms.

47. Principal component analysis or t-SNE, but also other methods like UMAP for clustering, or multidimensional scaling.

criminal analysts are not familiar with the complex technical processes that went into the tools that are supposed to assist them.

All in all, the vast majority of machine learning-methods that are currently used for tackling organized crime are black boxes, meaning their interpretability and transparency are non-existent or rather low. In this regard, it might be better to create simpler models that are interpretable in the first place and that in some cases can achieve the same degree of accuracy, instead of using more complex algorithms that lack interpretability.[48] But current approaches to AI-based detection of organized crime work mainly with unstructured big data sets, which greatly complicate the use of interpretable machine learning. Therefore, presently, end-to-end learning is the way to go—where brute-force-like techniques are used until the networks perform as required, skipping over more pipelined architectures where each module has to be tuned. This renders the machine learning-models hard or impossible to validate.

Moreover, modifying the system by applying structural changes retrospectively is impossible without starting the training from scratch. Ultimately, while some choices may be trivial, for instance deciding whether a task can be solved via regression or classification, other design choices for machine learning architectures can have far-reaching consequences that must be considered when applying respective software solutions to real-world policing. There is "no free lunch."[49] Especially for unsupervised learning tasks, it is not just difficult to find the right algorithm in order to be able to interpret data meaningfully; the choice of algorithms can also have significant backlashes against the course of criminal investigations.[50]

Furthermore, criticizing the effects of mathwashing does not only make sense in the context of PPTs, but also with regard to organized crime-oriented systems. Both come with the semblance of a flawless, objectively functioning, neutral, and value-free machine, while the opposite may be true.[51] Pattern-based policing strategies become an epistemological authority where the data doubles of suspects can seem more real than the actual person behind it.[52] Moreover, machines are far away from grasping non-rational, spontaneously-working human systems in all their complexity. Software developers and advocates must acknowledge that large parts of organized crime are riddled with unforeseeable circumstances like clan feuds, stock market fluctuations, health-crises, natural

48. *See* Cynthia Rudin, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 NATURE MACHINE INTEL. 206, 206–15 (2019).

49. *See* David H. Wolpert, William G. Macready, *No Free Lunch Theorem for Optimization*, 1 IEEE TRANS. ON EVOLUTIONARY COMPUTATION 67 (1997).

50. *See* David H. Wolpert, William G. Macready, *No Free Lunch Theorem for Optimization*, 1 IEEE TRANS. ON EVOLUTIONARY COMPUTATION 67, 67–82 (1997) (discussing "the dangers of comparing algorithms by their performance on a small sample of problems" and of selecting algorithms without "incorporating problem-specific knowledge").

51. TRANSPARENCY INT'L, *supra* note 18, at 287.

52. *See* Maria Los, *Looking Into the Future: Surveillance, Globalization and the Totalitarian Potential,* in THEORIZING SURVEILLANCE: THE PANOPTICON AND BEYOND 69, 69–94 (David Lyon ed., 2006) (exploring the totalitarian potential of modern forms of surveillance).

disasters, internal agreements, individual affects and emotions, and others.

One must keep in mind that the kinds of insights from prosecution software are determined by decisions taken by software manufacturers and programmers. It is not just training data that introduces problematic bias into software, but also individual presuppositions. Here, these presuppositions concern criminal behavioral patterns, for instance, theories on repeating victimizations, movement patterns, the relevance of certain nodes in social networks, methods for dimensionality reduction, and more. But as soon as these subjective presuppositions are "filtered" through a complex machine, they vanish in favor of computer outputs that come along with a sheer undoubted epistemological authority. This can cause police officers to come under pressure when they do not follow machine decisions but want to stick to deviating intuitions in going after certain clues in the process of criminal prosecution.

Although AI-based policing tools are generally termed "assistance systems," the relationship between machine assistant and sovereign could reverse, rendering police detectives to mere assistants of machine decisions. This effect is akin to defensive decision-making.[53] In an increasingly data-driven policing culture, police officers and other officials sometimes deliberately make suboptimal decisions by following machine outputs to protect themselves against negative consequences or to secure themselves against redress claims in case something goes wrong. Especially in organized crime, expertise and experience are important because organized crime contains complex socio-cultural factors. Consequently, investigations should not be subjugated to algorithmic authority.

C. Symptom-Oriented Policing

Last but not least, PPTs are criticized for only addressing symptoms of crime and not its causes. This is where organized crime-oriented law enforcement tools come to the fore. They certainly carry the potential to make criminal networks visible, to detect suspicious irregularities, and to identify complex criminal operations whose isolated acts are not conspicuous in themselves because their criminal background only becomes apparent when viewed in the larger context. But do they really address the root causes? Law enforcement techniques based on the analysis of unstructured Big Data seem to aim at finding causal inferences for criminal acts, but they too operate within a "post-Newtonian technorationality,"[54] just like PPTs. This technorationality follows a logic of tinkering, exploring ways to create patterns of correlations in the data.

53. *See* Rocio Garcia-Retamero & Mirta Galesic, *On Defensive Decision Making: How Doctors Make Decisions for Their Patients*, 17 HEALTH EXPECTATIONS 664, 664–69 (2014); *see also* Florian M. Artinger, Sabrina Artinger & Gerd Gigerenzer, *C. Y. A.: Frequency and Causes of Defensive Decisions in Public Administration*, 12 BUS. RESEARCH 9, 9–25 (2013).

54. *See* Jutta Weber, *Keep Adding. On Kill Lists, Drone Warfare and the Politics of Databases*, 34 ENV'T. & PLAN. D: SOC'Y & SPACE 107, 107–25 (2016) ("In contrast to modern scientific rationality, [post-Newtonian technorationality] is not interested in the analysis of intrinsic properties of entities (organisms, machines) but focuses instead on their behaviour, on their (inter)relations and on possible recombination of modules, fragments of code and building blocks of systems.").

Postrelational databases, which form the basis for law enforcement searches, shift police work away from directly observed clues or causal assessments of specific events toward a "politics of possibility," in which algorithmic compilations generate uncertainties that must be combined in ways that are just good enough to be valid or court-proof for legal purposes.[55] The challenge is that, because of the complexity of organized crime, robust evidence is rarely found in isolated data sets, but suspicions must be derived by combining different data sets or data mining methods. Here again, human expertise and experience are indispensable.

V

CONCLUSION

There is no doubt that AI-based police systems promise much potential for solving criminal cases. But they suffer from a fundamental paradox: In order to function, they depend on stability, otherwise the projection of knowledge from the past into the present or future would not work. The dependence on stability makes them very costly to apply meaningfully in dynamic fields. In these dynamic fields, they only work if they are constantly updated. But this is rarely the case because it requires a constant stream of new data, which moreover must be constantly priced into the system. This is very labor-intensive. Therefore, most systems work best in practice when patterns remain stable. But one of the central goals of smart policing should be to break up entrenched structures in order to end destructive spirals of poverty, corruption, and crime. To resolve this paradox of current AI-based policing, solutions are needed that make technical systems less dependent on past crime records and more compatible with causal reasoning. In this way, they could help identify, understand, and address the complex relationships and structures that put people in situations that lead them to commit crime in the first place.

At present, most tools are targeted at combating well-reported forms of overt, violent, or petty crime that are easy to detect and obviously definable in themselves as a criminal act. Organized crime, however, operates in hidden, complex structures. These structures consist of a combination of corruption, money laundering, and the infiltration of cultural traditions, local and international politics, and business.[56] In order to prove anything within these structures, a wide variety of data sets must be combined and put into relation with each other in order to detect not only patterns but also suspicious irregularities and coalitions of people. Because these are complex tasks that require knowledge in various fields including culture, law and politics, human experts are still urgently needed for this, and preferably those familiar with both criminal structures and machine learning technologies.

---

55. *See generally* LOUISE AMOORE, THE POLITICS OF POSSIBILITY: RISK AND SECURITY BEYOND PROBABILITY (2013).

56. Interpol, *Global Strategy on Organized and Emerging Crime*, INTERPOL GENERAL SECRETARIAT, 1–4 (2017).

To sum up: In order to support a kind of policing that goes beyond the stability-reliant, self-reinforcing and symptom-based logic of predictive policing, heterogeneous data and profound human expertise are the keys. Thereby, it must be ensured that individual *and* group privacy is protected, as well as discriminatory bias avoided, in both the data sets and the features used for the models to train the algorithms. In doing so, a policing and programming culture is needed where practitioners and developers are working together in close collaboration and where practitioners are not competing with the machines they are using, but where these machines are really serving them as assistants in the literal sense. This, however, is only possible when the value of doubt is being cultivated and not undermined by mathwashing ideologies. Even though these may sound to be unrealistic demands, they are nevertheless all necessary to facilitate an application of AI which would make policing truly "smart."