Ethics and Information Technology (2021) 23:165–176
https://doi.org/10.1007/s10676-020-09539-x

ORIGINAL PAPER

On the person-based predictive policing of AI

Tzu-Wei Hung[1] · Chun-Ping Yen[1]

Published online: 1 June 2020

What is predictive policing?

While using statistics in law enforcement is nothing new,[1] cutting-edge technology that uses big data is changing the face of law enforcement (Buchholtz 2020; Degeling and Berendt 2018; Egbert and Krasmann 2019; Ferguson 2017b; Sheehey 2019; Nissan 2017; Perry et al. 2013). As data and statistical tools have improved over time, a new police strategy called "predictive policing" (hereafter, PP) has come into practice (Saunders et al. 2016; Kreutzer and Sirrenberg 2020; Kulkarn and Akhilesh 2020). Based on the assumptions that certain aspects of the physical and social environment encourage predictable acts of criminal wrongdoing, and that interfering with that environment would deter the would-be crimes, PP aims to "forecast where and when the next crime or series of crimes will take place" by identifying trends and relationships that may not be readily apparent to us among the collected data (Uchida 2014, p. 3871; see also Ferguson 2017a; Moses and Chan 2018).

Techniques involving large quantities of digital information have been evolving at a rapid rate. As Ferguson (2017a) notes, while the social scientific research supports the insights behind PP, police adoption of the strategy has outpaced established scientific findings. More significantly, when a police department declares its adoption of PP, it could be doing things that vary greatly in their technical sophistication, effectiveness, and ethical concerns. As such, while there is an increasingly heated debate about the effectiveness and potential impacts of the emerging techniques involving large quantities of digital information, the discussion is easily conducted without careful awareness of the differences among various methods and practices of PP (Egbert and Krasmann 2019; Ferguson 2017b). Besides, myths and pitfalls may hinder proper evaluation of PP's development and deployment, such as assuming that AI actually knows the future, or focusing on prediction accuracy rather than tactical utility (Perry et al. 2013).

As a proactive policing model, the targeted units of crime predictions of PP can range from different sizes of geographical areas to individual people. Based on its focuses, PP can be divided into three subdivisions:

i. *Area-based* policing: targets on the time and place in which crimes are more likely to occur.
ii. *Person-based* policing: targets on the individual who is more likely to be involved in criminal acts.
iii. *Event-based* policing: targets on the type of activity that is more likely to occur.[2]

Among these, person-based PP is the most controversial, as it singles out individual names and faces (Ferguson 2017b). Although person-based PP can be applied to different types of crime, such as terrorism, mass shootings,[3] financial, and community crimes, this paper focuses primarily on *community policing*.

Challenge and opportunity

According to Ferguson, person-based PP rests on the insight that "negative social networks [of individuals], like environmental vulnerabilities [in the case of area-based PP], can encourage criminal activity" (2017a, p. 1142). It is believed

✉ Chun-Ping Yen
chunping.yen@gmail.com

Tzu-Wei Hung
htw@sinica.edu.tw

[1] Institute of European and American Studies, Academia Sinica, No. 128, Sec. 2, Academia Rd., Nankang District, Taipei 115, Taiwan

[1] Berk (2008) notes that researchers have applied statistical methods on crime for nearly a hundred years.

[2] This distinction is based on Ferguson (2017b), Egbert and Krasmann (2019), and a Hitachi (2019) report. Strictly speaking, while all labeled as "predictive policing," they share neither theoretical bases nor implications and consequences in practice (Ferguson 2017a, p. 1148).

[3] According to *USA TODAY* (Baig 2019), after the Parkland shooting, three US companies (Bark Technologies, Gaggle.Net, and Securly Inc.) claim that their AI systems can detect possible signs of cyber bullying and violence by scanning student emails, texts, documents, and social media activity.

that law enforcement interventions, such as developing "predictive profiles of individuals based on past criminal activity, current associations, and other factors that correlate with criminal propensity," would disrupt the continued pattern of crimes (Ibid., p. 1142). The importance of addressing the underlying social needs of the targeted population, however, is easily overlooked. This paper aims to explore the opportunities and the risks of person-based targeted policing for society. We argue that in the case of community policing, in order to break the pattern of crimes, we need not only spot the high-risk subgroups in the community but also provide them the social-service resources they require.

Person-based PP is already being implemented in the present, in cities like London, Amsterdam, Chicago, Kansas City, and New York (Amnesty International UK 2018; Ferguson 2017b; Oosterloo and van Schie 2018; Couchman 2019).[4] We are now at the beginning of a significant conversation about what we should do with the big data involved in PP. The situation depicted in the Hollywood film, *Minority Report,* has never been so close to becoming reality.

For a person-based policing project to succeed, the individuals involved must be accurately targeted, and the interventions justified. It also takes some upsides, such as increasing objectivity in policing and doing more with less, for such a project to be worth pursuing. However, while technique-centric enterprises (e.g., Israel's Faception and UK's WeSee) emphasize potential benefits of AI-based PP, rights-centric NGOs cast doubt on possible violations of civic rights (e.g., Amnesty International UK 2018; Human Rights Watch 2017, 2019). Recent studies also rightly indicate some negative impacts of PP, including issues of privacy (Couchman 2019), inequality (Ferguson 2017b), discrimination (Prince and Schwarcz 2019), and other such rights (Barocas et al. 2017; Richardson et al. 2019).

Instead of rejecting PP in light of these deficits, we hold that AI, like simpler technologies such as knives and fire, can be both beneficial and harmful.[5] We explore the necessary conditions of using PP to achieve social good. These conditions include *basic* moral principles and *further* requirements. These conditions are neither sufficient nor exhaustive of all possible necessary conditions, but they nevertheless offer a basis for properly use of PP.

To this end, "Limits of prediction technology" section shows that banning PP cannot eliminate epistemological concerns because some deficits (e.g., unreliability and blackbox) are also found in humans and some (prejudice and biased data) have long existed before the emergence of AI. What matters most is using the technology adequately. "Basic moral requirements" section offers three basic requirements specific to PP, which are refined from five common ideas of major ethics guidelines (IEEE, EU, and RIKEN, etc.). It also explains why these principles rule out China's AI totalitarianism. "Further requirements and case studies" section derives further requirements from case studies (New Orleans, New York City, San Francisco, Tokyo, and the UK), including *the right to know*, *informed consent*, *privacy rights*, and *the freedom of expression*. Finally, "Conclusions and further questions" section concludes that while risks are inevitable, a person-based PP could be helpful in community policing, especially if merged into the larger governance framework of the social safety net.

Limits of prediction technology

The PP applications currently used in European and American police departments include *Crime Anticipation System*, *PreCobs*, *PredPol*, and *Hunchlab*, etc. (Hardyns and Rummens 2018). Japan's Kanagawa Prefectural Police is also testing its Hitachi AI system, which integrates various sets of biometrics and data analytics for crime prediction and prevention (Hitachi 2019).

Despite these applications varying in details, skeptics may worry about two common limits of the technology. The first is that the predictive *processing* is not explicable, and may lead to accountability problems. The second is that the predictive *result* is not reliable; machines may create errors or duplicate human prejudice and discrimination (Barocas et al. 2017; Prince and Schwarcz 2019; Williams et al. 2018). Using Richardson et al's (2019) terms, AI-based policing is compromised by "bad prediction" and "dirty data".

Inexplicable processing and accountability

The first concern touches on issues of transparency and accountability. The way AI systems generate results is often described as opaque and inexplicable (Castelvecchi 2016; Wachter et al. 2017). This is because, during learning, machines automatically derive rules and models from large databases and then produce output accordingly. Human designers often do not know how these rules or models are

[4] In 2016, for example, the UK's National Police Chiefs' Council (NPCC), Association of Police and Crime Commissioners (APCC), and National Crime Agency released "The Policing Vision 2025" programme, setting out a ten-year plan to help the law enforcement transform and adapt to the modern policing environment. The aim is to employ innovative and transformative approaches for proactive and preventative policing. As was planned, a new national super-database, called the "National Law Enforcement Data Programme" (NLEDP), will be put in place to replace the existing separate systems by 2020.

[5] In the literature of dual use, *dual use technologies* refer to tools that can be used to achieve good or evil. AI seems to be dual use as by which malevolent individuals can perpetrate wrongful harms. However, AI is unlike guns and H-bombs in that it is not designed to harm. So, in what sense that AI is a dual use technology is an interesting question. Please see Miller (2018) for the analysis of the concept of dual use.

induced. It is also difficult for the public to rely entirely on the predictions made by the "blackbox." Thus, understanding how the machine works is a key agenda for current deep learning (Samek et al. 2017).

Where does this opacity come from? In fact, no matter how powerful and complex an algorithm is, it is still executed on computers conforming to Turing computability. Turing (1936, 1937) describes an abstract general device for executing a sequence of instructions. Each instruction is a clause or step of an algorithm and can be performed mechanically. A mathematical function is considered computable if the value derived from the function can be identified by effective (i.e., implementable in a finite number of steps, thus in finite time with finite processing resources) procedures. Therefore, the implementation procedures of an algorithm can be *theoretically* broken down into individual finite steps in finite time, no matter how complex. The practical problem here is that ascribing meaning to these, perhaps billions of, steps is extremely difficult because humans only have restricted cognitive power and resources.[6] Therefore, the opacity occurs not because the algorithm itself fails to provide mechanical steps, but because our limited cognitive system can hardly interpret them.

When assisting or replacing human decision-making, AI's opacity is thought to run into difficulties regarding responsibility and accountability. Responsibility refers to the capability to fulfill an obligation or duty, and accountability is about the liability to answer for one's performance of duties (*Oxford English Dictionary*). Using Miller's (2018) example, a student who completes his allotted task of conducting a routine experiment is responsible, but he is accountable for his performance as a student to his supervisor (who has the obligation to monitor and assess the student's performance). In other words, accountability presupposes responsibility, but is not identical with it (Miller 2018).

At least before the creation of fully autonomous AI, machines are still not considered moral agents (Allen et al. 2000) and hence cannot take responsibility. Therefore, a machine cannot be held accountable. Conversely, humans are moral agents and are able to take responsibility for our own actions. While the brain is also notorious for its blackbox nature,[7] accountability is not a problem for us. Thus far, to clarify, the accountability problem here is not a matter of blaming a machine due to its mysterious processing when something goes wrong because blameworthiness also presupposes responsibility (Miller 2018). Rather, the difficulty is about who, natural persons (e.g., users or programmers) or legal persons (governments or manufacturers), should be responsible and held accountable for wrong decision-making. The opacity of AI simply makes this investigation even harder.

Unreliable prediction

Second, regarding the problem of unreliability, the concern might be that AI prediction can easily be (considered) false because it essentially relies on probability inference. It is quite different from human thinking, which often combines deduction, induction, abduction, or heuristic methods to make a comprehensive judgment. Thus, although AI performs better than humans in some domains, it is still unreliable in the comprehensive judgment of the cross-task contexts.

However, recent cognitive sciences give us pause. According to the predictive coding account (PCA), the human brain is a powerful predictor which constantly generates and updates expectations of the external world. The brain's predictive processing has been shown to be consistent with Bayesian optimization (Brown and Friston 2012). The brain generates top-down predictions of sensory content. Its prior prediction is produced based on experience and calibrated against incoming stimuli to minimize errors (Friston 2019; Hohwy 2013; Orlandi 2018; Swanson 2016; Tamir and Thornton 2018). In this sense, both AI and human cognitive systems employ the same Bayesian predictive method. Unreliable predictive results are also observed in human police, prosecutors, and judges (e.g., the 1992 Los Angeles riots and the 2018 Dallas incident). Hence, it seems that appealing to unreliable inferences is not a good justification to reject AI in PP.

Nonetheless, one may be concerned that AI's predictive result may lead to discrimination (Hajian et al. 2016; Garcia

[6] Assigning meaning is crucial in computer science (e.g., mapping symbols onto actions). A Turing machine can initially be viewed as manipulating otherwise meaningless marks, which become symbols when they are linked with rules as to bear assignment of reference and conform to the rules of syntax. This happens, for example, when the marks are taken as 0s and 1s and construed as numerals and hence as symbols standing for binary numbers, and the same is true for standard construals of machine code. It is also standard practice to add further layers of symbols and representation by building up these up out of binary code. Thus we have higher levels of representations, which can be assigned different kinds of reference, subjected to further kinds of syntax. However, even if we can *construct* the meaning of individual computational procedure, it may be hard for us to *analyze* the meaning of billions of procedures in AI.

[7] The human brain is a two-way blackbox. On one hand, psychological behaviourists, holding that the mental states are hard to measure, suggest studying observable outer behaviours instead. On the other hand, Bayesian theorists of predictive coding argue that the brain only measures the sensory signal without directly measuring the external world (Swanson 2016). This creates a problem: how the brain only infers its "cause" in the external world based on the "effect" of the sensory signals. This puzzle is described as "view from inside the blackbox" (Clark 2013, p. 183) or "the skull-bound brain" (Hohwy 2013, p. 15).

2016; Suresh and Guttag 2019; Richardson et al. 2019). This discrimination is caused by both *Algorithmic bias* and *Big data bias*. The former refers to the circumstances where the algorithm developers, even with good intentions, emphasize certain factors in coding, but end up with unfair results (e.g., online advertisements for arrest records often show up when black names are searched on the web (Sweeney 2013)). The latter refers to the fact that our social behaviours are full of stereotypes and prejudices, and AI's prediction based on the data of these behaviours may reflect or even amplify these prejudices.[8]

Problematic data indeed pose a serious challenge to PP. This could happen when a set of contaminated data is merged into a larger database, or when hackers deliberately add data noise to cause an AI system to malfunction (Papernot et al. 2017). A mighty AI malfunction may turn these errors and prejudices into a disaster. So, should we avoid employing the technology, especially in person-based policing?

In fact, banning the technology would not sweep away such worries. These controversies are neither unique nor novel to the AI systems. Some of them existed long before the emergence of AI (e.g., prejudice), and others occur in the human brain too (e.g., blackbox and unreliable problems). While AI's powerful computation may make some cases worse, it is also true that AI, when used with caution, may bring us potential benefits. Just as other means adopted in the progress of human civilization (e.g., knives and fire), AI is instrumental. Whether it is favourable or dangerous depends on how people use it. What matters is how to use it properly to achieve social good. To this end, "Basic moral requirements" section investigates basic moral requirements specific to person-based PP. "Further requirements and case studies" section next offers further suggestions of employing AI-based PP.

Basic moral requirements

Recently, various ethical guidelines AI-based policing have been proposed for AI development and deployment by academies, governments, and NGOs. In North America, for example, there are IEEE's *Ethically Aligned Design* (2019), Asilomar AI principle (2017), and the Montreal Declaration for Responsible AI (2017). In Europe, the EU has approved its *Ethics Guidelines for Trustworthy AI* (High-Level Expert Group on Artificial Intelligence 2019). Amnesty International UK (2018) has also published its five overarching principles for an AI code. In Asia, Singapore announced its *Model Artificial Intelligence Governance Framework* (2019) and Taiwan unveiled the *Guidelines for the Research and Development of AI* (2019). Japan's Ministry of Internal Affairs and Communication drafted its AI Principles (2017), besides other ethics guidelines suggested by the RIKEN Center For AIP, Japan Deep Learning Association (JDLA), and Tokyo University.[9] Some principles focus on AI's research and design (R&D), whereas others are about its usage and impacts on stakeholders. Some guidelines aim to foster European values, and some were offered from Asian perspectives. While more than 115 principles have been proposed, they roughly converge around five main points:

(1) Respect for Autonomy: Decisions made by, or with the assistance of, AI should not undermine the freedom and control of humanity.
(2) Transparent and accountable AI: AI's processing should be explainable and fit into legal mechanisms of accountability.
(3) Data integrity and security: ensure data are correct and under proper protection over its entire life-cycle (e.g., to reduce bias, inaccuracy, and privacy breaches).
(4) Risk management: acknowledge that AI has its negative impacts, especially to vulnerable groups (e.g., the poor and ethnic minorities), and handle them fairly.
(5) Human-centric: the goal of developing and deploying AI is to improve human well-being (e.g., rights, democracy, prosperity and environmental protection).

These five points are crucial to the development and deployment of desirable AI, and thus, apply to the person-based PP as well. However, they are necessary, but not sufficient. Taking these principles into account in the domain of person-based PP, based on the lessons from cities employing the technology, the following three requirements should be applied.

a. *Execute in the context of social safety net*: as criminal records often link up with people of social-economic disadvantages, governments should help improve their social welfare, which conforms to principles (3) and (4).

[8] In addition to biases, there is also the *undecidable problem*. It has been proven to be impossible to construct an algorithm that can provide correct answers to all yes-or-no questions (Floridi 2016). For example, Kleene (1943) applies Gödel's incompleteness theorem to computation, and he shows that no effective system can correctly determine whether a program, if run with a given input, will finish running or continue to run (known as the halting problem). Therefore, biases and errors are somewhat inevitable (Lin et al. forthcoming).

[9] Interestingly, despite adopting a total social credit system, China also announced *the Beijing AI Principles* (2019, May) through the Beijing Academy of Artificial Intelligence (https://www.baai.ac.cn/blog/beijing-ai-principles)—an organization backed by the Chinese Ministry of Science and Technology and the Beijing municipal government.

b. *Ensure humans are the ultimate decision-makers*: responsibility and accountability are crucial in law enforcement, which also conforms to principles (1) and (2).
c. *Enhance the well-being of all stakeholders*: this is to make sure that the AI policing will not be abused or turned against (part of) the people, which is especially crucial in the case of anti-terrorism. This consideration also conforms to principle (5).

These requirements can be elaborated as follows. First, as already shown in cities applying person-based PP, while big data technologies can "inform strategies to reduce violence in a more targeted and cost-efficient manner," the technologies alone cannot efficiently reduce crime (Ferguson 2017b, p. 46).[10] Researchers note that simply identifying the high-risk subgroup in the community is not sufficient.[11] What matters to the "prevention" part of PP lies in the actions taken following up on the predictions to reduce crime. The goal is to reduce the environmental vulnerabilities which encourage crime, which need not necessarily involve police intervention. Police intervention, like any other available choice, is preferable only when it helps to achieve this goal. Mapping the social network of violence alone is not enough. The underlying causes of violence also need to be addressed.

The New Orleans Police Department has been relatively successful in this aspect. In its PP programme, law enforcement informs targeted individuals that they know of their past actions and will prosecute them to the fullest extent if they re-offend. If the subjects choose to cooperate, they are "called in" to a required meeting as part of their conditions of probation and parole, and are offered *job training*, *education*, *job placement*, and *health service*s (Corsaro and Engel 2015). "Records show the city hosted 10 [such call-in sessions] from October 2012 through November 2015, bringing in 285 participants. Since November 2015, only one call-in has been held—in March 2017" (Bullington and Lane 2018). It was reported that from 2011 to 2014, the city saw a 21.9% reduction in homicide and a 55% reduction in group or gang-involved murders (Ferguson 2017b, p. 42). The statics shows a significant difference whether resources are implemented to increase the targeted individuals' opportunities and chances to escape crime.[12]

However, caution is warranted. While big data-based PP appears to be objective and fair, "it may reflect subjective factors and structural inequalities" within the communities (Ferguson 2015, p. 402). Mistakes can occur at any point in the process and have real impacts on individuals' lives. Moreover, even with accurate data, there will still be false positives where predictions erroneously target individuals who should not be targeted, for the information in the database is incomplete. For example, although race was not included in the predictive algorithm, its variables (e.g., police contact and gang affiliation) often lead to the targeting of poor communities of color. According to Amnesty International UK (2018), a gang crime monitoring system used by London's Metropolitan Police (i.e., The Gangs Matrix) was reported to be racial and counterproductive, which often targeted youths, blacks, and immigrants. Logically, biases cannot be eliminated because data presuppose bias. Each data set contains numbers or symbols representing certain environmental states, but not certain others. If an AI system is designed to compute some data and ignore others, it is selective. While biases do not imply discrimination, they are highly relevant.[13]

Second, humans should be the ultimate decision-makers in PP, which means that a human individual or a group (e.g.,

[10] According to Ferguson (2017b), before the Kansas City Police Department introduced advanced social network analysis to spot at-risk suspects in 2012, Kansas City's homicide rate was two-to-four times the national rate. Although the number fell 26.5% after the new technology was employed, homicide and shooting rates dramatically climbed again in 2015. Likewise, in 2013, the Chicago police adopted different algorithms for focused deterrence, which located potential offenders based on their personal criminal record. At the beginning, the software generated numerous false-positive predictions, but its accuracy was significantly improved in 2016 (more than 70% shot people were on the list). However, this by no means implies the ending of violence because the technology only "identifies the disease but offers no cure" (Ferguson 2017b, p. 49). See also Saunders et al. (2016) for similar concerns.

[11] For example, the Chicago Police Department has used an algorithm to prioritize limited resources to focus on those at highest risk by rating every person arrested with a threat score from 1 to 500-plus. Due to the lack of specific guidance on what treatments to apply to the subjects on the list, however, most districts did not focus on intervening with these subjects (Saunders et al. 2016). Careful research shows that the list does not reduce homicides (Saunders et al. 2016). See also Ferguson (2017b, p. 40); Perry et al. (2013); Couchman (2019).

[12] The New Orleans Police Department has applied similar techniques to those employed by the Chicago Police Department since 2012. For a more integrated approach using predictive technologies to reduce crime, the city also supplemented the Group Violence Reduction Strategy as part of their broader NOLA for Life murder reduction strategy.

[13] As Ferguson observes, "[b]ig data collection will not count those whom it cannot see" (2017b, p. 179). Big-data-driven systems will overlook the populations who do not "engage in activities that big data and advanced analytics are designed to capture" (Lerman 2013, p. 56). In our case, those with criminal records or gang associations, as well as prior police contact, are most likely to be marked as suspicious. This creates the concern about "the initial selection bias" (Ferguson 2015, p. 402) of law enforcement data-collection systems that certain individuals will always be at risk to be future targets of suspicion, despite that they are not currently engaging in criminal activities. The danger is straightforward. The databases with "the initial selection bias" will make it easier for a police officer to justify her suspicion if she tends to believe that a particular type of person may

a police officer or congresspeople) should take charge of choosing action plans, or deciding to transfer the power of decision-making to machines. The argument for this view is as follows. While admitting a non-human juridical entity (e.g., legal person and animal), our legal systems primarily deal with human behaviours (e.g., prohibiting hollowing out a company and cruelty to dogs), no matter whether the behaviours are active (e.g., doing something) or passive (letting something happen). In democracies, we already have such legal systems to deter abuse and to hold someone accountable. Moreover, it is humans who find and collect data, interpret the results of the analyses, and take action in light of the findings of PP. It is also humans who participate in PP to minimize foreseen risks, fix mistakes, and bear responsibility. On one hand, we do not (and probably cannot) have such a legal system for machines. On the other hand, it is immoral to blame a person who has nothing to do with a wrong decision. Therefore, to ensure the balancing between power and responsibility, a human is better suited than an AI to be the ultimate decision maker in policing.

The human decisions in PP may interplay among distributed agents (e.g., engineers, police officers, politicians). It could happen that each decision made by different agents is morally neutral, or even good, but that the final outcome is evil. In this case, how should we allocate responsibility to agents in the network causally relevant for bringing about the decisions? Miller's (2017, 2018) analysis of three types of responsibility offers an answer here: A *natural responsibility* refers to an agent's causal role in a joint action, which neither relies on the agent's institutional role nor necessarily involves morality. An *institutional responsibility* refers to the responsibility of agents occupying institutional roles. As cooperative enterprise (science, policing, and politics) usually takes place in institutional settings and are shaped by institutional purposes, its members have such responsibility. If agents are naturally or institutionally responsible for a joint action and the action has moral significance, then the agents may have *collective moral responsibility* for the actions and may be praised or blamed for the actions. Agents with different complementary roles (e.g., designers, developers, and law enforcement officers) in PP have a collective end of the protection of moral rights. They are in what Miller calls *a chain of institutional responsibility* where each agent makes a different and distinct contribution, according to their roles in PP, to the collective end. Accordingly, if a police officer shot a wrong person due to prediction errors resulted from the bugs accidentally caused by a subcontractor software engineer, then not only the officer but also the engineer would share collectively morally responsible and thus blamed for the tragedy.[14]

Third, PP should aim to enhance the well-being of *all* stakeholders when developing and deploying big data technologies, including both targeted suspects as well as potential victims.[15] The unequal distribution of social resources, such as opportunities and wealth often fuels crime, violence, as well as drug and alcohol abuse (Fajnzylber et al. 2002; Room 2005). We should take all stakeholders' well-being into account because the support of the communities has positive impacts on crime reduction.

These requirements rule out China's model of PP, which employs PP for mass surveillance and detention. China has integrated CCTV, biometrics, and information from both government (ID number and data) and business company (hotel and flight records) to monitor its citizens, especially dissidents and Uyghur minority, in real-time (Human Rights Watch 2017, 2018, 2019; Amnesty International 2018; Shahbaz 2018).[16] While the UN Committee on the Elimination of Racial Discrimination urged China to release imprisoned Uighurs and Muslim minorities in 2018, more than one million Uyghurs are still held in 'political education' camps, which have no basis under Chinese law (Human Rights Watch 2019).

As China has no independent judicial system and rejects the principle of presumption of innocence, the rights of citizens targeted by AI cannot be guaranteed (Lewis 2011). For

Footnote 13 (continued)

be more likely to commit a crime (Saunders et al. 2016; Richardson et al. 2019).

[14] Also, according to Miller and Blackler's (2017) normative theory of policing, the protection of moral rights is the principal purpose of policing, constrained by democratically supported laws. The purpose in protecting these rights justifies policing. So we can, for example, claim that the police officers are justified to arrest and detain someone for assault. They possess the moral right to do so in virtue of their membership of a morally legitimate police institution. Police officers are individually institutionally responsible for at least some of their actions and omissions regarding the purpose of protecting moral rights.

[15] When used properly, the technologies may benefit law enforcement with increased accuracy. As Ferguson (2015) points out, big data enables not only a wealth of suspicious inferences, but also an equal number of potentially exculpatory facts. When big data is available, police should be required to use it in an exculpatory manner as well. It offers to search for more information and more precise information, including exculpatory information that reduces suspicion, and thus can make more reliable predictions than human investigators. It allows for a more focused use of police resources as well. Moreover, with a vast amount of information, the big data technologies allow collecting unexpected seemingly innocuous connections and correlations for future criminal activities. Take one of Ferguson's examples (2015, pp. 395–396), a drug dealer needs tiny plastic bags and a scale to package crack cocaine. It is considered that recent innovations can help to track the sale of these items and thus to help spot the drug dealer. Similarly, big data is useful to reveal patterns of national or transnational crimes which were difficult to track before.

[16] China is also exporting its surveillance tech to the global. See Mozur et al. (2019).

instance, in China's social credit system, citizens with low social credit could lose certain rights, such as being denied the ability to buy travel tickets and pursue a college education. The fact that the system excludes targeted groups from the social safety net violates our first principle. China's arbitrary and mass detention in Xinjian violates the third principle too. Also, Hong Kong youths worry that China will use its cutting-edge surveillance technology to identify and arrest anti-government protesters (Bodeen 2019). While China's model of PP fulfills the requirement to have decisions made by human authority, there is no legal mechanism to balance power and responsibility, neither is the possibility of compensation and restorative justice in China's opaque law enforcement and judiciary system. It thus fails the principles (a) and (c) mentioned above and must be judged *untenable*.

To sum up, this section proposes three basic requirements specific to person-based PP, which is complementary to five common ethics guidelines for general AI. While these moral conditions (i.e. five guidelines and three requirements) are necessary but not sufficient in outlining what a theoretically desirable person-based PP should be, they help to filter undesirable models.

Further requirements and case studies

There are further details of these basic requirements that should be examined. One important aspect is—what does the term "well-being" in (c) refer to? Which human rights may be relevant in the context of PP? In this section, we discuss four further requirements about the well-being of stakeholders, including *the right to know*, *informed consent, privacy rights*, and *the freedom of expression*. They can be abstracted from the following case studies.

Right to know and informed consent

First, the public has the right to seek information relating to use of PP technologies by law enforcement as well as the policies, procedures, and guidelines governing such use, including, for instance, information about how the police use the system to make operational decisions and policies regarding the retention, sharing, and use of the collected data. Take New Orleans for instance. The New Orleans Police Department (NOPD) has applied big data techniques provided by Palantir Technologies since 2012. It is reported, however, that Palantir's collaboration with the NOPD was mostly unnoticed by the public. Neither the residents of New Orleans nor city council members were aware of the use of the PP program until the media exposed the news in 2018 (Stanley 2018). This case of the NOPD violates the public's right to know, and it is not the only one. Similar controversy has occurred in New York City as well. In 2017, the Brennan Center for Justice went to court to challenge the New York Police Department's (NYPD) refusal to produce crucial information about its use of PP technologies and won the case. In 2018, the New York State Supreme Court ordered the NYPD to produce substantially more records about their predictive policing program (Levinson-Waldman and Posey 2018).

Personal data protection is an important issue, especially after the Cambridge Analytica scandal in 2018. The well-established principles of collecting, storing, and using human biometrics in Bioethics serves as valuable references in considering further requirements for PP. The right to know and informed consent are two of these principles. In the context of PP, the former amounts to letting people know whether they are on the targeted list. Unless the police have secured a lawful interception from a judge, people spotted by algorithm should be informed. The latter is to ask for permission, either indirectly (democratic procedure of adopting AI-based PP) or directly (personal agreement of joining the police department program). This principle of informed consent also allows individual people to opt out at any time, which should be guaranteed through legislative and/or democratic processes.

Democratic procedures serve as a gatekeeper when adopting PP. In 2019, for example, the San Francisco council voted to ban the using of face recognition (hereafter, FR) in policing.[17] FR is a technology to measure and match unique facial characteristics for the purposes of biometric surveillance or identification. Advocates of the bill hold that the technology, either used in backtracking surveillance or predicting crime, will infringe on *privacy*. If FR is allowed to track people in public areas, almost everyone will be monitored. The massive surveillance will threaten free attendance of political protests and anonymous business activities (Conger et al. 2019). Recent studies also show that the accuracy of identifying white males is much better than identifying females with darker skin in the FR technology developed by IBM and Microsoft (Buolamwini and Gebru 2018). Thus, the premature

[17] A similar debate is currently ongoing in the UK, concerning the Metropolitan Police and the Home Secretary's trials of the facial recognition surveillance technology since 2016. According to a final report conducted by the London Policing Ethics Panel, an independent panel set up by the Mayor of London to provide ethical advice on policing issues that may impact on public confidence, '[m]arginal benefit would not be sufficient to justify [life facial recognition's] adoption in the face of the unease that it engenders in some, and hence the potential damage to policing by consent' (London Policing Ethics Panel 2019, p. 47). The panel suggests that the facial recognition surveillance technology should not be adopted unless it could be shown from the field trials that it could be able to significantly increase police efficiency and effectiveness in dealing with serious offences. Currently, human rights organisations Liberty and Big Brother Watch are challenging the use of FR cameras in the courts.

technology is banned to protect privacy and relevant rights (Big Brother Watch 2018).

It is worth noting that there are two issues to be discussed here. One issue is that, as technology of individual identification is ever-changing, what should be focused on is the systematic protection of human rights rather than banning of specific technology (e.g., FR). In fact, a large portion of the human visual cortex has evolved to deal with facial recognition because the human face is a crucial means of social and emotional communication (Haxby et al 2002; Tzourio-Mazoyer et al. 2002). The face is an open indicator to identify individuals in social cooperation, and covering the face (e.g., burqa) will also impede one's relationship to other social members. Thus, encouraging Muslim women to voluntarily uncover their faces is a way to resist humiliation (Lazreg 2009) and gender inequality (Bennoune 2006). However, the powerful FR may result in novel inequality between individuals and the government or company who own relevant technology and data. Unlike a face on a wanted poster in the nineteenth century, in which a target is identified before the search, the FR technology has to scan first to identify targets. No matter whether FR is combined with AI-based PP, this indifferent and active scanning threatens privacy and causes psychological pressure to *freedom of expression* (LGBT Pride or political protests). The presence of FR cameras affects people's behaviour by sending the message that they are being watched and can be tracked for further police action.

Moreover, what's worse is that FR is merely one among numerous techniques for identifying individuals from a mass in an unnoticed and distant manner. Japan's Hitachi Inc. announced an AI-based system of predictive policing that can identify a passenger in public, even if CCTVs only capture the back or side image of the person, without a clear face. According to its press release (Hitachi Inc. 2019; Nishida 2018), Hitachi's AI system can analyze more than 100 personal features and patterns (e.g., carried items, habitual wear, and gait analysis, etc.) to track down a suspect, besides face recognition. Hitachi announced that, under relevant law of privacy, it is now working with the Kanagawa Prefectural Police Department to maintain the safety of the 2020 Tokyo Olympics. Accordingly, face, gait, or iris recognition are merely some possible ways to identify individuals. In the future, there must be other biometric indicators, e.g., our unique functional brain fingerprint (Chen and Hu 2018). Therefore, while banning any of these technologies may buy us some time, what matters is to seek for a rights-oriented legal solution to protect universal rights while reflecting cultural and diverse values, with a supervised administration to avoid overconcentration of unchecked power.

Privacy and freedom of expression

The other issue is about privacy and the freedom of expression, which relates to our third and fourth requirements: there is an obligation to maintain our privacy rights and freedom of expression intact, but the extensions of, and relationship among, those rights should be carefully evaluated. On one hand, the use of PP technologies, such as automated FR with CCTV, would raise concerns about the possible chilling effect on the public. Even if people are informed and give consent to the employment of AI-based PP, people would not be comfortable going to an event if doing so meant being subjected to police surveillance. It is not desirable to have our daily activities disturbed by the state authorities with powerful AI. Thus, data rights should be considered as the new human rights and be protected at both national (Tisne 2018) and international (Guild 2019) levels.

On the other hand, the idea of rights, just as AI techniques, is diverse. Privacy is exchangeable, compared to the basic rights such as the right to life and prohibition of torture. Laws allow consumers, through informed consent, in many business activities, to exchange privacy (e.g., current location) for some services (GPS navigation), but laws do not allow people to voluntarily be killed for organ harvesting to earn money to save their families from poverty. So, when some basic rights are at stake (e.g., lives on airplanes), privacy may not be the priority. There are also occasions where big data technologies can help to improve public services and to reduce crimes by, for instance, scrubbing personal identifiers from raw data. Since anonymised data does not disclose privacy-sensitive information, it poses a lesser threat to either the privacy rights or the freedom of expression. Similarly, the notion of privacy changes constantly and reflects cultural diversity. For instance, a digital footprint (i.e., a person's distinctive and traceable record on digital devices or the Internet) and biometrics (i.e., unique, measurable identifiers of individual human species) are comparatively new notions that were not related to privacy in the past. Likewise, the criteria and boundary of privacy in Asia may not be the same as those in Europe and America. Therefore, the trade-off between security and privacy differs from place to place, depending on the negotiation among local residents, council, and government. The question of when the employment of the PP technologies violates our privacy rights and freedom of expression needs to be scrutinised on a case by case basis. To summarize, the above four requirements derived from the case study serve as a prerequisite for implementing PP.

Our proposed conditions (i.e., basic and further requirements) are *necessary* and not *sufficient*. So one may expect that some debates are inevitable in practice. For instance, the Home Office of UK announced a further £5 million Police Transformation Fund to support the National Data

Analytics Solution (NDAS), a trial of crime-predicting technology launched in 2018 by West Midlands Police. One area of focus of the NDAS involves analysis of the criminal and custody records and intelligence data of people with previous convictions for gun and knife offences. The aim is to derive from the analysis key indicators with which to identify patterns and common traits among the potential criminals so that the police or social services could intervene and offer support or guidance to prevent the crimes beforehand. While the government makes it clear that the programme is not to replace, but to support, police officers' professional judgments, criticism of NDAS and other similar projects by human rights groups persist.

A major concern is that the technologies could inherit pre-existing inequalities. In a report published earlier in 2019, Liberty, a human rights organization based in the UK, has protested the use of such PP technology and urged the government to terminate the AI-assisted PP programme (Couchman 2019). "At the very least," they demanded, the police should "fully disclose information about the use of predictive policing programs within their force." (Couchman 2019, p. 42) As was discussed in "Limits of prediction technology" section, we think it is too quick a conclusion to ban AI-assisted PP on the basis of the potential problematic data. As was pointed out in "Basic moral requirements" section, we think that the success of PP's crime reduction does not lie on the predictions made by the machine, but on the actions taken after the predictive outputs. Indeed, the technologies may reflect existing inequalities, but as long as we keep in mind the limits of the prediction technology and take appropriate actions, such as shifting resources into areas of need, changes are still possible. AI-assisted PP not only has the aforementioned limits, but must be constrained by law. The relevant regulations will be supplied. For instance, the UK has the Data Protection Act (of) 2018 (DPA), which updates previous data protection laws in the UK. It requires, subject to certain exemptions, that data subjects be told what information is held on them and how it is used.[18] While there is a tension between public safety and the possible intrusion of rights, it is subjected to rigorous conditions when the authorities are allowed to process personal data for law enforcement purposes. In practice, it would be challenging to disclose information involved in AI-assisted PP fully and unconditionally as required by Liberty. It is not a problem specific to AI-assisted PP, but a general one in criminal justice systems.

Conclusions and further questions

To summarize, "Limits of prediction technology" section examines epistemological limits of person-based PP and shows that these defects by no means preempt the application of the technology to PP; and yet, they are worse in humans. "Basic moral requirements" section then refines three moral principles specific to person-based PP based on versions of AI ethics guidelines (IEEE, EU, and RIKEN, etc.). "Further requirements and case studies" section derives further requirements from case studies, including PP in Chicago and New Orleans, New York City, San Francisco, Tokyo, and UK's NDAS.

Now, let us go back to the question we ask at the very beginning: "Should you be targeted for a crime that AI predicts that you will commit?" The answer is both yes, and no. AI-assisted PP could be helpful in handling potential threats if its usage satisfies the conditions proposed in "Basic moral requirements" section (three basic principles) and "Further requirements and case studies" section (four further requirements). Those conditions are only necessary, but not sufficient. On one hand, detention is not the only solution to handle individuals who are on the list. Since AI's predictive result may reflect social inequality, offering help through a social safety net is more crucial in reducing crime. As the statics reported by the New Orleans Police Department from 2011 to 2014 indicate, when the high-risk subgroups in the community are provided with the resources to improve, say, their job perspectives, there is indeed a significant reduction in homicide and gang-involved murders. The predictive technology should identify and reduce the environmental vulnerabilities which encourage crime in a larger governance framework.[19] On the other hand, detention and person-based identification can be used in the cases when the number of possible loss of lives might be large (e.g., airline security). In these cases, violating a right (e.g., privacy) can be justified in order to protect a more basic right (right to life). But even so, the usage needs to conform to the above conditions, as well as be legitimized and overseen through democratic procedures. Besides, if something goes wrong, the legal system should be able to hold someone accountable and avoid repeating the same mistake.

[18] Restrictions are placed on the rights of the data subjects, where necessary and proportionate, in order to avoid obstructing an investigation or inquiry, avoid prejudicing the prevention, detection, investigation or prosecution of criminal offences or the execution of criminal penalties, protect public security, protect national security, and protect the rights and freedoms of others. See the Guide to Law Enforcement Processing of DPA on the Information Commissioner's Office website (https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-law-enforcement-processing/individual-rights/the-right-of-access/).

[19] As our solution is not necessarily involving police intervention, it raises a question of whether the term "predictive policing" should be substituted or integrated into a larger framework of humanity security.

To conclude, while risks are somewhat inevitable, the person-based PP could have potentially positive impacts on human society. AI, just as knives and fire, can be used to do great good or great evil. The AI users must always be clear about what good is to be achieved when exploiting the tool. If our goal is to improve human well-being through fostering democracy and human rights (instead of consolidating regime and increasing national interests in China's model), then we need to stick to the moral principles that help achieve the goal.

Nonetheless, there are other challenges to be dealt with. While basic and further requirements are proposed to prevent abuse, there are potential conflicts *between principles and practices*, as well as conflicts *among principles*. For instance, in 2018 the UK government proposed the Law Enforcement Data Service (LEDS), as part of the NDAS, aiming to enhance the efficiency of AI-based PP through integrating two existing but disconnected databases by 2025.[20] While the UK Home Office (and West Midlands Police) commissioned scholars and professionals (e.g., Alan Turing Institute) to evaluate privacy impact and publish an ethics advisory report, the programme is strongly opposed by NGOs such as Liberty and Big Brother Watch. The controversy here reveals not only the clash between theory and practice, but also that among theories.

In the former case, for instance, what should we do if a policy of AI-assisted PP introduced by a democratically legislative process violates the proposed moral principles? We answer that, at *prima facie*, principles should outweigh practice because democracy is fragile and may fail for various reasons (Chomsky 2006; Devarajan and Khemani 2018; Myerson 2006). Democracy sometimes is ineffective in protecting basic rights, and sometimes even turns against its people (e.g., Weimarer Republik). In contrast, although ethics changes over time, some rights are ancient and relatively invariant (e.g., property rights, right to life, freedom from unlawful torture). Hence, the technology should follow the principles which uphold these basic rights, instead of political practice.

The latter case, in which conflicts occur among principles, is more difficult. For instance, one can imagine the situation in which some people's autonomy is disregarded to enhance the overall well-being of all stakeholders, making principle (1) conflict with the utilitarianism consideration (c). Here, we do not have a good solution at hand. The EU's *The Ethics Guidelines for Trustworthy AI* only state(s) that, when there is a conflict, people "should approach ethical dilemmas and trade-offs via reasoned, evidence-based reflection rather than intuition or random discretion" (High-Level Expert Group on Artificial Intelligence 2019, p. 13) and that "[i]n situations in which no ethically acceptable trade-offs can be identified, the development, deployment and use of the AI system should not proceed in that form." (High-Level Expert Group on Artificial Intelligence 2019, p. 20) In other words, this is a hard question, and the EU offers no concrete solution. However, the conflict among principles, especially in the domain of predictive policing, constitutes a valuable theme for further study.

[20] The databases are the Police National Computer (PNC) and the Police National Database (PND). For further details, please refer to the UK government's the *NLEDP Privacy Impact Assessment Report* (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/721542/NLEDP_Privacy_Impact_Assessment_Report.pdf).

References

ACLU. (2016). Community control over police surveillance—Guiding principles. Retrieved June 10, 2019, from https://reurl.cc/M7EdKX.

Allen, C., Varner, G., & Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence, 12*(3), 251–261.

Amnesty International. (2018). *Amnesty international report 2017/18: The state of the world's human rights*. Retrieved March 3, 2019, from https://reurl.cc/Ylz6Ko.

Amnesty International United Kingdom. (2018). *Trapped in the matrix: Secrecy, stigma, and bias in the Met's gangs database*. Retrieved March 3, 2019, from https://reurl.cc/8lmnzy.

Baig, E. C. (2019). Can artificial intelligence prevent the next Parkland shooting? *USA TODAY* (Feb 13, 2019). Retrieved July 10, 2019, from https://reurl.cc/ObWqzy.

Barocas, S, Bradley, E, Honavar, B., & Provost, F. (2017). Big data, data science, and civil rights. arXiv preprint http://arxiv.org/abs/1706.03102.

Bennoune, K. (2006). A contextual analysis of headscarves, religious expression, and women's equality under international law. *Columbia Journal of Transnational Law, 45*, 367–426.

Berk, R. (2008). Forecasting methods in crime and justice. *The Annual Review of Law and Social Science, 4*, 219–238.

Big Brother Watch. (2018). *Face off: The lawless growth of facial recognition in UK policing*. Retrieved July 10, 2019, from https://reurl.cc/xDq0XL.

Brown, H. R., & Friston, K. J. (2012). Dynamic causal modelling of precision and synaptic gain in visual perception—An EEG study. *Neuroimage, 63*(1), 223–231.

Buchholtz, G. (2020). Artificial intelligence and legal tech: Challenges to the rule of law. In *Regulating artificial intelligence* (pp. 175–198). Cham: Springer.

Bodeen, C. (2019). Hong Kong protesters wary of Chinese surveillance technology. *The Associated Press* (June 14, 2019). Retrieved July 8, 2019, from https://reurl.cc/24qg3O.

Bullington, J., & Lane, E. (2018). How a tech firm brought data and worry to New Orleans crime fighting. *The New Orleans Times-Picayune* (Mar 1, 2018). Retrieved June 9, 2019, from https://reurl.cc/D156DR.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st conference on fairness, accountability and transparency*, *PMLR* (Vol. 81, pp. 77–91).

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News, 538*(7623), 20–23.

Chen, S., & Hu, X. (2018). Individual identification using the functional brain fingerprint detected by the recurrent neural network. *Brain Connectivity, 8*(4), 197–204.
Chomsky, N. (2006). *Failed States: The abuse of power and the assault on democracy*. New York: Metropolitan Books.
Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.
Conger, K., Fausset, R., & Kovaleski, S. F. (2019). San Francisco bans facial recognition technology. *The New York Times* (May 14, 2019). Retrieved June 25, 2019, from https://reurl.cc/1QR4pV.
Corsaro, N., & Engel, R. S. (2015). Most challenging of contexts: Assessing the impact of focused deterrence on serious violence in New Orleans. *Criminology and Public Policy, 14*(3), 471–505.
Couchman, H. (2019). *Policing by machine: Predictive policing and the threat to our rights*. Retrieved July 10, 2019, from https://reurl.cc/RdM1Er.
Degeling, M., & Berendt, B. (2018). What is wrong about robocops as consultants? A technology-centric critique of predictive policing. *AI & Society, 33*(3), 347–356.
Devarajan, S., & Khemani, S. (2018). If politics is the problem, how can external actors be part of the solution? In K. Basu & T. Cordella (Eds.), *Institutions, governance and the control of corruption* (pp. 209–251). Cham: Palgrave Macmillan.
Egbert, S, & Krasmann, S. (2019). Predictive policing: Not yet, but soon preemptive? *Policing and Society*.
Fajnzylber, P., Lederman, D., & Loayza, N. (2002). Inequality and violent crime. *The Journal of Law and Economics, 45*(1), 1–40.
Ferguson, A. G. (2015). Big data and predictive reasonable suspicion. *University of Pennsylvania Law Review, 163*(2), 327–410.
Ferguson, A. G. (2017a). Policing predictive policing. *Washington University Law Review, 94*(5), 1115–1194.
Ferguson, A. G. (2017b). *The rise of big data policing: Surveillance, race, and the future of law enforcement*. New York: New York University Press.
Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 374(2083).
Friston, K. (2019). Publisher correction: Does predictive coding have a future? *Nature Neuroscience, 22*(1), 144.
Garcia, M. (2016). Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal, 33*(4), 111–117.
Guild, E. (2019). Data rights: Searching for privacy rights through international institutions. In D. Bigo, E. Isinb, & E. Ruppert (Eds.), *Data politics: Worlds, subjects, rights* (pp. 230–245). London: Routeldge.
Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 2125–2126). New York: ACM.
Hardyns, W., & Rummens, A. (2018). Predictive policing as a new tool for law enforcement? Recent developments and challenges. *European Journal of Criminal Policy Research, 24*, 201–218.
Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry, 51*(1), 59–67.
High-Level Expert Group on Artificial Intelligence. (2019). *The ethics guidelines for trustworthy AI*. Retrieved March 3, 2019, from https://reurl.cc/RdM1gG.
Hitachi Inc. (2019). Hitachi provides an AI environment in research on Kanagawa prefecture police's crime and traffic accident prediction techniques. Retrieved January 16, 2020, from https://reurl.cc/lL6d2E.
Hohwy, J. (2013). *The predictive mind*. New York: OUP.
Human Rights Watch. (2017). China: Police 'big data' systems violate privacy, target dissent. Retrieved June 25, 2019, from https://reurl.cc/A1Z8ld.
Human Rights Watch. (2018). China: Big data fuels crackdown in minority region. Retrieved June 25, 2019, from https://reurl.cc/Nae6om.
Human Rights Watch. (2019). *World report 2019*. Retrieved June 25, 2019, from https://reurl.cc/6g641d.
Kleene, S. C. (1943). Recursive predicates and quantifiers. *Transactions of the American Mathematical Society, 53*(1), 41–73.
Kreutzer, R. T., & Sirrenberg, M. (2020). Fields of application of artificial intelligence—Security sector and military sector. *Understanding artificial intelligence* (pp. 225–233). Cham: Springer.
Kulkarni, P., & Akhilesh, K. B. (2020). Big data analytics as an enabler in smart governance for the future smart cities. In *Smart technologies* (pp. 53–65). Singapore: Springer.
Lazreg, M. (2009). *Questioning the veil: Open letters to Muslim women*. Princeton: Princeton University Press.
Lerman, J. (2013). Big data and its exclusions. *Stanford Law Review Online, 66*, 55–63.
Levinson-Waldman, R., & Posey, E. (2018). Court: Public deserves to know how NYPD uses predictive policing software. Retrieved July 16, 2019, from https://reurl.cc/A1Z8Wd.
Lewis, M. K. (2011). Presuming innocence, or corruption, in China. *Columbia Journal of Transnational Law, 50*, 287–369.
Lin, Y., Hung, T., & Huang, T. L. (forthcoming). Engineering equity: How AI can help reduce the harm of implicit bias. *Philosophy & Technology*.
London Policing Ethics Panel. (2019). *Final report on live facial recognition*. Retrieved July 22, 2019, from https://reurl.cc/RdM17G.
Miller, S. (2017). Institutional responsibility. In M. Jankovic & K. Ludwig (Eds.), *The Routledge handbook of collective intentionality* (pp. 338–348). New York: Routledge.
Miller, S. (2018). *Dual use science and technology, ethics and weapons of mass destruction*. New York: Springer.
Miller, S., & Blackler, J. (2017). *Ethical issues in policing*. New York: Routledge.
Moses, L. B., & Chan, J. (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing and Society, 28*(7), 806–822.
Mozur, P., Kessel, J. M., & Chan, M. (2019). Made in China, exported to the world: The surveillance state. The New York Times (April 24, 2019). Retrieved Jan 4, 2020, from https://reurl.cc/zy9zje.
Myerson, R. B. (2006). Federalism and incentives for success in democracy. *Quarterly Journal of Political Science, 1*, 3–23.
Nishida, T. (2018). Kanagawa police to launch AI-based predictive policing system before olympics. *Australasian Policing, 10*(1), 43.
Nissan, E. (2017). Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement. *AI & Society, 32*(3), 441–464.
Oosterloo, S., & van Schie, G. (2018). The politics and biases of the 'crime anticipation system' of the Dutch police. In *Proceedings of the international workshop on bias in information, algorithms, and systems* (BIAS 2018).
Orlandi, N. (2018). Predictive perceptual systems. *Synthese, 195*(6), 2367–2386.
Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celikand, Z. B., & Swami, A. (2017). Practical black-box attacks against machine learning. In *Proceedings of the 2017 ACM on Asia conference on computer and communications security* (pp. 506–519). New York: ACM.
Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. Rand Corporation. Retrieved Jan 16, 2020, from https://reurl.cc/QpQ3k0.

Prince, A., Schwarcz, D. (2019). Proxy discrimination in the age of artificial intelligence and big data. *Iowa Law Review, 105*, 1257–1318.

Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review, 94*, 192–233.

Room, R. (2005). Stigma, social inequality and alcohol and drug use. *Drug and Alcohol Review, 24*(2), 143–155.

Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint http://arxiv.org/abs/1708.08296.

Saunders, J., Hunt, P., & Hollywood, J. S. (2016). Predictions put into practice: A quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology, 12*(3), 347–371.

Sheehey, B. (2019). Algorithmic paranoia: The temporal governmentality of predictive policing. *Ethics and Information Technology, 21*(1), 49–58.

Shahbaz, A. (2018). *The rise of digital authoritarianism: Fake news, data collection and the challenge to democracy*. Retrieved July 1, 2019, from https://reurl.cc/vnN1Oa.

Stanley, J. (2018). New Orleans program offers lessons in pitfalls of predictive policing. Retrieved Jan 15, 2020, from https://reurl.cc/Gk0r6d.

Suresh, H., & Guttag, J. V. (2019). A framework for understanding unintended consequences of machine learning. arXiv preprint http://arxiv.org/abs/1901.10002.

Swanson, L. R. (2016). The predictive processing paradigm has roots in Kant. *Frontiers in Systems Neuroscience, 10*, 79.

Sweeney, L. (2013). Discrimination in online Ad delivery. *Queue, 11*(3), 10.

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences, 22*(3), 201–212.

Tisne, M. (2018). It's time for a bill of data rights. *MIT Technology Review* (Dec 14, 2018). Retrieved Jan 6, 2020, from https://reurl.cc/vnN1zA.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungs problem. *Proceedings of the London Mathematical Society (Series 2), 2*(42), 230–265.

Turing, A. M. (1937). Computability and λ-Definability. *Journal of Symbolic Logic, 2*(4), 153–163.

Tzourio-Mazoyer, N., De Schonen, S., Crivello, F., Reutter, B., Aujard, Y., & Mazoyer, B. (2002). Neural correlates of woman face processing by 2-month-old infants. *Neuroimage, 15*(2), 454–461.

Uchida, C. (2014). Predictive policing. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 3871–3880). New York: Springer.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. *Science Robotics, 2*(6), eaan6080.

Williams, B. A., Brooks, C. F., & Shmargad, Y. (2018). How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *Journal of Information Policy, 8*, 78–115.