NEJ
Novos Estudos Jurídicos

Doi: 10.14210/nej.v26n1.p103-135

BIG DATA, INTELIGÊNCIA ARTIFICIAL E POLICIAMENTO PREDITIVO: BASES PARA UMA ADEQUADA REGULAÇÃO LEGAL QUE RESPEITE OS DIREITOS FUNDAMENTAIS

BIG DATA, ARTIFICIAL INTELLIGENCE AND PREDICTIVE POLICING: BASES FOR PROPER LEGAL REGULATION THAT RESPECTS FUNDAMENTAL RIGHTS

BIG DATA, INTELIGENCIA ARTIFICIAL Y POLICIAMIENTO PREDICTIVO: BASES PARA UNA ADECUADA REGULACIÓN LEGAL QUE RESPETE LOS DERECHOS FUNDAMENTALES

Cyntia Souza de Menezes[1]

José Ramon Agustina Sanllehí[2]

Resumo: Diante das novas ferramentas de análise policial se por um lado a polícia cumpre uma função essencial na prevenção, detecção e investigação de um delito, ela não está livre de ter o seu trabalho limitado e submetido ao princípio da legalidade. A eficácia do trabalho policial não pode, como sabemos, saltar as garantias próprias de um Estado de Direito e sua atividade deve submeter-se a critérios transparentes de razoabilidade e controle, entre outros princípios. Não é legítimo que, para garantir a segurança de todos, seja sacrificada a privacidade daqueles que apenas superficialmente parecem suspeitos. As denominadas investigações prospectivas (phishing expeditions) são proibidas quando é afetado um direito fundamental que requeira autorização judicial prévia. À luz das reflexões anteriores, entendemos que talvez seja precipitada a aplicação massiva de sistemas de IA a modelos de polícia preditiva, enquanto (i) estejam pendentes de refinação; (ii) não tenham uma base científica sólida, e; (iii) operem em um ambiente onde o marco legal, regulatório e de proteção cidadã é ainda limitado. Principalmente quando estes sistemas têm o potencial de afetar de maneira importante a vida de grupos e populações em situação de exclusão e que se veem cada dia mais vigiados.

Palavras-chave: Big Data; Inteligência artificial; Regulação; Policiamiento predictivo.

Abstract: In light of the new tools of police analysis, while on one hand, the police fulfill an essential role in the prevention, detection and investigation of an offence, on they other, they are not free to have its work limited and subject to the principle of legality. The effectiveness of police work cannot, as we know, ignore the guarantees inherent to a State of Law; their activities must be subject to transparent criteria of reasonableness and control, among other principles. It is not legitimate, for example, to sacrifice the privacy of those who only appear to be suspicious in order to ensure the safety of all. So-called phishing expeditions are prohibited when a fundamental right that requires prior judicial authorization is affected. In light of previous reflections, we understand that the widespread application of AI systems to predictive police models may be precipitated, while (i) still requiring further refinement ; (ii) not having a sound scientific basis, and; (iii) operating in an environment where the legal, regulatory and citizen protection framework is still limited. Especially when these systems have the potential to significantly affect the lives of groups and populations in situations of exclusion and that are increasingly monitored.

Keywords: Big Data; Artificial intelligence; Regulation; Predictive policing.

Resumen: Delante de las nuevas herramientas del análisis policial si por un lado la polícia cumple una función esencial en la prevención, detección e investigación de un delito, ella no está libre de tener su trabajo limitado y sometido al principio de la legalidad. La eficacia del trabajo policial no puede, como sabemos, saltar las garantias propias de un Estado de Derecho y su actividad debe someterse a criterios transparentes de razonabilidad y control, entre otros principios. No es legítimo que, para garantizar la seguridad de todos, sea sacrificada la privacidad de aquellos que apenas superficialmente parecen sospechosos. Las denominadas investigaciones prospectivas (phishing expeditions) son prohibidas cuando es afectado un derecho fundamental que requiera autorización judicial previa. Frente a las reflexiones anteriores, entendemos que tal vez sea precipitada la aplicación masiva de sistemas de IA a modelos de polícia predictiva, mientras (i) estén pendientes de refinación; (ii) no tengan una base científica sólida, y; (iii) operen en un ambiente donde el marco legal, regulatorio y de protección ciudadana es aún limitado. Principalmente cuando estos sistemas tienen el potencial de afectar de manera importante la vida de grupos y poblaciones en situación de exclusión y que se ven cada día más vigilados.

Palabras clave: Big Data; Inteligencia artificial; Regulación; Policiamiento predictivo.

INTRODUÇÃO

É fato que a evolução das técnicas de coleta, entrecruzamento e tratamento massivo de dados pessoais na era do Big Data deu à polícia acesso a potenciais meios de vigilância, investigação e controle sem precedentes. Por meio da utilização destas técnicas, hoje os corpos policiais e de segurança dispõem de um volume imenso de informações de caráter pessoal sobre os cidadãos. A partir destas informações coletadas, é possível traçar perfis, classificar pessoas em função de prognósticos de risco, cruzar dados sobre investigações em curso, reduzir o círculo de suspeitos e muito mais. A denominada teoria do mosaico (em contraste com a teoria das esferas)[3] já sinalizava que, por meio da inter-relação entre

3 Sobre a teoria do mosaico e sua implicação nas técnicas de prevenção e persecução do delito por meio de inteligência artificial, ver VALLS PRIETO, J. **Problemas jurídico penales asociados a las nuevas técnicas de prevención y persecución del crimen mediante inteligencia artificial**. Madrid: Dykinson, 2017.

Doi: 10.14210/nej.v26n1.p103-135

vários dados pessoais, é possível obter perfis com alta definição que afetam a privacidade do indivíduo e que, além disso, proporcionam indícios ou previsões sobre uma possível implicação em atividade delitiva no passado, presente ou no futuro. Esta intromissão sutil, mediante a soma e combinação de dados aparentemente irrelevantes, levanta uma questão de limites.

Diante das novas ferramentas de análise policial, convém não esquecer que, se por um lado a polícia cumpre uma função essencial na prevenção, detecção e investigação de um delito, ela não está livre de ter o seu trabalho limitado e submetido ao princípio da legalidade. A eficácia do trabalho policial não pode, como sabemos, saltar as garantias próprias de um Estado de Direito e sua atividade deve submeter-se a critérios transparentes de razoabilidade e controle, entre outros princípios. O trabalho policial não deve, portanto, aproveitar-se de qualquer informação pessoal obtida dos cidadãos com outras finalidades para exercer um controle preventivo inespecífico, sem que exista habilitação legal expressa.

Em uma sociedade com altas doses de intolerância ao risco, a pista escorregadia em direção a um estado de vigilância total é mais que possível. A tentação é conhecida, especialmente em situações de emergência social mais ou menos difusa[4]. No entanto, as razões de eficácia nunca podem passar por cima das razões de legitimidade. Entre as razões de legitimidade, para o que aqui nos interessa, emergem com particular importância o direito à privacidade e o direito a não sofrer discriminação por motivos de parcialidades e predições estatísticas ou algorítmicas.

A polícia não pode, portanto, ignorar as normas e utilizar dados pessoais sem controle apenas porque tem um pressentimento, ou por simples conjectura. Não é legítimo que, para garantir a segurança de todos, seja sacrificada a privacidade daqueles que apenas superficialmente parecem suspeitos. As denominadas investigações prospectivas (phishing expeditions) são proibidas quando é afetado um direito fundamental que requeira autorização judicial prévia. Ingerências menos invasivas devem vir precedidas de suspeita razoável (probable cause), sem prejuízo de que, em alguns contextos, tenha-se legitimado por lei determinadas atuações ad hoc, procedimentos rotineiros ou controles aleatórios. Também não pode a polícia, a modo de experimento e por causa de um determinado risco, realizar ingerências massivas em espaços de privacidade protegidos por lei. Por todas estas razões, é necessário que os corpos policiais se submetam a um programa de compliance. Um controle rigoroso poderá dar razão à origem do pressentimento e justificá-lo, porque o "olfato policial" não costuma ser espontâneo, senão baseado em dados. Caso contrário, sempre se poderia lavar, ex post facto, uma informação obtida de forma ilegal. A título de exemplo: com base em que informações e por quais motivos o serviço de imigração e controle de fronteiras seleciona determinados passageiros para um second examination? Poderia tratar-se de uma medida profilática que, limitando direitos fundamentais, tenha por base um juízo de prognóstico frágil ou pouco fundamentado? Alguém irá compensar o cidadão objeto de um exame desnecessário baseado em um algoritmo muitas vezes

4 Até mesmo em situações de emergência terrorista (ou sanitária: SARS-CoV-2) é necessário amparar-se legalmente e limitar a utilização desta cobertura legal ao lapso de tempo que justifique tal regime especial.

opaco e, em determinadas ocasiões, manifestamente discriminatório? Seria necessário informar ao cidadão sobre o motivo que justifica uma indagação extraordinária?

Em um terreno mais concreto, no informe Predictive Policing –The Role of Crime Forecasting in Law Enforcement Operations (2013) já se apresentava, no contexto de uma diversidade de métodos de policiamento preditivo, quatro grandes blocos de análises prospectiva: (1) métodos de predição de eventos delitivos; (2) métodos de predição de pessoas que cometerão delitos; (3) métodos de predição de pessoas que serão vítimas de um delito; (4) e métodos de predição da identidade dos autores de delitos cometidos no passado (methods for predicting perpetrators' identities)[5].

Contudo, os métodos modernos de predição de risco mediante algoritmos baseados em sistemas de inteligência artificial (**IA**) atualmente utilizados pela polícia não são tão distintos ao tradicional olfato-baseado-em-evidências, nem podem substituir a decisão humana. Convém observar a este respeito que, etimologicamente, IA expressa um conceito mediante a conjunção de duas palavras que vem a condensar uma contradictio in terminis. A palavra inteligência procede do latim: intus (dentro) legere (ler). Já artificial deriva de artificium, fazendo referência a algo criado ou fabricado manual ou artesanalmente. Ou seja, uma realidade não natural ou inata. Portanto, trata-se de um oxímoro, na medida em que uma máquina não pode captar a realidade interior a um objeto, nem tem capacidade de abstração para sintetizar um conjunto de dados empíricos que procedem dos sentidos ou da quantificação de aspectos mensuráveis. No entanto, esta discussão nos levaria muito longe. Teríamos que remontar a questões epistemológicas sobre a distinção kantiana entre fenômeno e número que, sem dúvida, excedem o propósito destas linhas. Seja como for, para o que aqui nos interessa, basta sinalizar que o agente de polícia que toma uma decisão pode fazê-lo utilizando ou não algoritmos. Afinal, a informação que recebe será uma simples pista para, a partir dela, verificar o prognóstico de periculosidade. A IA é tão somente um instrumento que não deveria decidir nada por si mesmo. O problema de base, portanto, já vem sendo planteado na jurisprudência há décadas, por exemplo, em relação às revistas policiais e suas práticas arbitrárias. Nihil novum sub sole.

Assim, na já distante STS (Sala 2ª) de 9 de abril de 1999 se razoava que "a chamada diligência de revista consiste na parada de uma pessoa para averiguar se oculta elementos que possam servir como prova de um delito (...) e está condicionada unicamente aos efeitos de sua legalidade, na medida em que não seja fruto da arbitrariedade ou da transgressão, mas seja racional e proporcional à situação, em cujo caso adquirem toda sua plenitude e operabilidade legal os artigos 11.1 f) e g) da LO de Forças e Corpos de Segurança e artigos 19 e 20 da LO de Proteção da Segurança Cidadã"[6].

5 RAND Corporation. **Predictive Policing** – The Role of Crime Forecasting in Law Enforcement Operations. 2013. Disponível em: https://www.rand.org

6 Tradução livre dos autores. O original lê: "la llamada diligencia de cacheo consiste en el registro de una persona para averiguar si oculta elementos que puedan servir para la prueba de un delito (...) y únicamente está condicionada a efectos de su legalidad a que la medida no sea fruto de la arbitrariedad o del desafuero, sino racional y proporcional a la situación, en cuyo caso adquieren toda su plenitud y operatividad legal los arts. 11.1 f) y g) de la LO de Fuerzas y Cuerpos de Seguridad y 19 y 20 de la LO de Protección de la Seguridad Ciudadana". STS (Sala 2ª) de 9 de abril de 1999.

Doi: 10.14210/nej.v26n1.p103-135

Neste sentido, assinala García Amado que a imagem jurisprudencial da polícia confere a ela uma "capacidade para ser ponderador de precisão"[7], quando nem sempre acontece assim.

Dado o anterior, o que há, então, de novo nesta discussão com o advento da IA e seu impacto no trabalho policial?

Muito se há escrito e discutido nos últimos anos, sobretudo nos meios de comunicação[8], a respeito da revolução causada e das infinitas possiblidades abertas pela aplicação de ferramentas de IA em sistemas de segurança pública e do Estado. Com efeito, os veículos de massa media nos prometem desde sistemas de vigilância e detecção de disparos[9], até a predição de delitos com alto grau de exatidão[10]. No entanto, o entusiasmo manifestado sobre estes sofisticados métodos de análise mediante IA podem ocultar déficits metodológicos importantes que, se não forem controlados e corrigidos, representam ameaças aos direitos humanos e liberdades fundamentais em geral, e ao direito à não discriminação em particular. Por outro lado – como aponta Miró com perspicácia, a própria utilização do termo IA induz à confusão, ao proporcionar uma mescla de eufemismo e desideratum, entre outras razões, porque dá a entender, ao aludir a sistemas de IA, que nos referimos a algo mais que um mero tratamento de informação. Além de uma operação automatizada, descobre-se uma vontade de que a máquina seja inteligente, capaz de ter ou imitar processos cognitivos próprios de seres humanos, ainda que, de momento, não conheçamos suficientemente o funcionamento real da mente e nem tampouco esteja claro o que é a inteligência humana[11]. A revolução industrial permitiu substituir o homem por máquinas porque suas capacidades físicas assim o permitiam. Com a chegada da IA, mesmo que um processo parecido já esteja acontecendo em termos de economia do trabalho, ainda aguardamos para ver até onde e em que medida as capacidades cognitivas serão igualmente substituíveis. Do mesmo modo, provavelmente muitas decisões mais ou menos transcendentais sobre nossas vidas passam por algoritmos informáticos[12]. No entanto, a tomada de decisões segue sendo, obviamente (ainda que alguns pareçam esquecê-lo), de competência exclusiva do ser humano: nenhum algoritmo decide por nós, senão que, em resumo, confiamos na sua capacidade para armazenar e selecionar dados como forma de economizar nossas energias cognitivas. Em termos de escalabilidade e economia de meios, a ajuda de sistemas de IA permite que agentes de polícia (e qualquer funcionário público ou privado) economizem tempo na

7 Tradução livre dos autores. O original lê: "capacidad para ser ponderador de precisión". A este respeito, ver interessante contribuição de GARCÍA AMADO, Juan Antonio. Anatomía de un imposible. La imagen jurisprudencial del policía. En AGRA, C. da; Domínguez, J.L.; GARCÍA AMADO, J.A.; HEBBERECHT, P.; RECASENS, A. (eds.) **La seguridad en la sociedad del riesgo**. Madrid: Atelier, 2003, pp. 181-200.

8 Ver, entre muitas outras, "¿Puede predecirse un crimen antes de suceder con un algoritmo?", publicada em La Vanguardia de 18 de março de 2019, disponível online: https://www.lavanguardia.com; "Así trabaja el equipo de Interior que predice el perfil de los asesinos", publicada no El País de 29 de junho de 2019, disponível online: https://elpais.com

9 Para mais informações ver: https://www.shotspotter.com/technology/

10 Disponível para acesso em: https://www.predpol.com/

11 MIRÓ, Fernando. Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots. **Revista de Derecho Penal y Criminología**, vol. (20, 87-130. 2018.

12 A este respeito, ver DUPUY, Daniela. **Inteligencia aplicada al Derecho penal y proceso penal.** Em Cibercrimen II. BdeF, 2018. p. 281 y ss.

compilação e análise de dados, mas na decisão final somos (ou deveríamos ser) insubstituíveis. Como premissa para nossa análise, deve-se levar em conta dois fatores importantes. Por um lado, o contexto regulatório para a adoção de novas tecnologias pelas administrações públicas é significativamente insuficiente, tendo-se optado por normas pouco precisas e, em todo caso, não vinculantes (mediante a técnica do denominado soft law). Por outro lado, não podemos esquecer a opacidade no desenvolvimento e utilização desta tecnologia que, talvez apressadamente, vem sendo pouco a pouco implementada por estas mesmas entidades públicas[13] (em particular, pelas forças estatais de segurança e pelos sistemas de justiça criminal).

Tudo isso deveria conduzir-nos a não subestimar os riscos associados à tomada de decisões automatizadas na administração pública. Neste contexto, mediante a implementação de sistemas preditivos de justiça criminal, se há começado a trabalhar sobre hipóteses e inferências obtidas a partir da análise de quantidades imensas de dados, aos quais se aplicam algoritmos nem sempre transparentes. A tomada de decisões utilizando IA irrompe no âmbito de decisão tanto de juízes e magistrados, como das polícias (no que se denomina sistemas de "policiamento preditivo"), assim como em outras instâncias governamentais encarregadas de velar pela segurança pública.

A expectativa é de que a utilização de tais programas leve, entre outros objetivos, a uma melhor aplicação dos (sempre escassos) recursos policiais e, portanto, a uma maior eficiência e eficácia na prevenção e resolução de problemas delitivos. Sem dúvida, a IA permite otimizar os recursos humanos, ao dotá-los de novas ferramentas e tornar possível uma economia de custos (restringindo sua dedicação a aquelas tarefas nas quais seja imprescindível a intervenção humana). No entanto, uma análise mais detalhada de alguns destes programas nos alerta para um novo foco de preocupação: o reforço de estereótipos e vieses discriminatórios.

Diante do aparente conflito entre avanços tecnológicos e restrições regulatórias, pode-se observar como os Estados vem consistentemente optando pela via da auto-regulamentação (mesmo que às vezes se esforcem na divulgação e publicação de documentos de soft law), talvez na esperança de que os marcos legais existentes sejam suficientes para contrabalançar as falhas dos algoritmos. No entanto, em temas tão sensíveis e que afetam de maneira tão significativa ao indivíduo, como o são os temas relacionados com justiça criminal, entendemos que uma intervenção mais vigorosa do aparato legislativo para controlar a utilização destes sistemas é necessária, senão fundamental, para que o sistema poda aportar à sociedade os prometidos benefícios de prevenção da violência. Nas linhas que seguem, apresentamos um breve panorama do estado global atual em regulação de IA.

13 ENGSTROM, David F.; HO, Daniel E.; SHARKEY, Catherine M.; CUÉLLAR, Mariano-Florentino. **Government by Algorithm:** Artificial Intelligence in Federal Administrative Agencies. Fevereiro, 2020.

Doi: 10.14210/nej.v26n1.p103-135

1. MARCO LEGAL PRÉVIO: PERSPECTIVA GERAL DOS ESFORÇOS DE REGULAÇÃO DE IA.

Desde uma perspectiva ampla, apesar da importância dos avanços tecnológicos e suas crescentes implicações legais, a utilização de IA ainda não está significativamente regulada na Europa ou em escala global. Seguimos, portanto, à espera dos necessários marcos legais que, em prol da segurança jurídica e da proteção efetiva dos direitos individuais, cuidem dos problemas específicos da aplicação de IA na esfera da administração pública. Até o momento, os esforços, sejam eles nacionais ou regionais, seguem concentrados principalmente na elaboração e publicação de "guidelines", "white papers", códigos éticos e estratégias de pesquisa e desenvolvimento que, em seu conjunto, não são mais do que soft law, ou seja, guias orientativas com boas intenções, mas sem implicações legais efetivas.

Ainda que se possa contribuir com a definição de critérios de aplicabilidade da lei existente e seu alcance mediante a edição de soft law, esta técnica normativa sui generis não resolve questões mais relevantes, como por exemplo, a definição clara da responsabilidade legal por danos. Nos encontramos, assim, diante de um "accountability gap" de fato, mas muitas mais vezes de direito, pelo fato de que não se vislumbram normas às quais se possa acudir para resolver tais problemas. Até o momento, os Estados tem focado sua atenção e preocupações na regulação de relações consumo e na proteção dos direitos do cidadão-consumidor perante as novas tecnologias desenvolvidas e vendidas por empresas privadas. Este vazio legal deixa uma ampla margem de atuação às administrações públicas em geral – e às forças de segurança em particular -, no que se refere à utilização de sistemas de IA e decisões automatizadas, criando um ambiente favorável ao uso inconsequente ou abusivo destas tecnologias[14].

Além dos referidos esforços em soft law, duas questões vêm gerando mais atenção, tendo sido traduzidas em regulamentações específicas: os temas relacionados à privacidade de dados e, em alguns casos, o teste e/ou utilização de veículos autônomos. De sua parte, as Nações Unidas, além das discussões sobre os impactos da IA no âmbito dos direitos humanos nas comissões de praxe (como na Assembleia Geral e na respectiva Comissão de Direitos Humanos), criou alguns grupos de discussão específicos. Assim, o Centro para IA e Robótica em Haia trata da implantação de IA em temas relacionados à criminalidade e justiça[15], enquanto a International Telecommunications Union (**ITU**) discute a necessidade de estabelecer políticas e critérios técnicos na aplicação de IA[16]. No que se refere ao desenvolvimento e implantação de Lethal Autonomous Weapons Systems (**LAWS**), a Convention on Certain Conventional Weapons criou um grupo específico de expertos governamentais

14 BRUNDAGE, M.; AVIN, S.; CLARK, J.; TONER, H.; ECKERSLEY, P.; GARFINKEL, B.; AMODEI, D. **The Malicious Use of Artificial Intelligence:** Forecasting, Prevention, and Mitigation. February 2018. Disponível em: http://arxiv.org

15 Ver: http://www.unicri.it

16 Para mais informações ver: https://www.itu.int

que se reúne anualmente desde 2017[17].

Neste cenário, é interessante observar que as próprias Nações Unidas não conseguiram chegar a consensos sólidos a respeito da utilização e implantação de IA nos vários âmbitos de atuação dos Estados Membros. Nem sequer no âmbito das LAWS, no qual as Nações Unidas teriam, a priori, maior capacidade de influência, considerando os diversos tratados internacionais sobre o desenvolvimento e aplicação de armamentos por ela administrados, se há conseguido chegar a um acordo sobre os princípios mais básicos. Entre a proibição total de desenvolvimento e utilização de LAWS e a inação, há uma margem bastante extensa para a atuação dos Estados. Dado este panorama, à falta de um consenso global que limite ou impeça este desenvolvimento, a corrida armamentista seguirá avançando de livremente[18].

Os Estados Unidos tampouco desenvolveram regulação federal significativa sobre IA, além daquelas relativas à defesa nacional, ainda que várias propostas legislativas sobre IA e privacidade de dados tenham começado a tramitar (na legislatura de 2019 foram colocadas em tramitação até trinta e nove projetos de lei que se referiam à IA[19]). Uma das tentativas norte-americanas de regular especificamente as decisões automatizadas em âmbito nacional é o "Algorithim Accountability Act"[20] que, no momento em que escrevemos este artigo, segue em fase de discussão no Congresso. Definitivamente, ainda que alguns estados norte-americanos tenham avançado e regulado algumas questões relativas à privacidade dos dados e/ou veículos autônomos, o país segue sem ter uma estratégia nacional clara em matéria de IA.

Dado este cenário de soft laws e vazio regulatório, no âmbito da União Europeia publicam-se, em 2018 e 2019, dois estudos enfocados nas questões de ética e regulação de IA desenvolvidos por um grupo de expertos, o Grupo de Peritos de Alto Nível sobre a IA (GPAN AI)[21]. Em fevereiro de 2020, com a publicação da European Data Strategy, se estabelecem as prioridades europeias em matéria de regulação, o que inclui a criação de um "mercado comum de dados", a regulação das responsabilidades legais e o fomento de um desenvolvimento tecnológico democrático[22]. A tudo isso, soma-se a entrada em vigor do Regulamento Geral de Proteção de Dados (**RGPD**) e, antes disso, da Diretiva sobre o tratamento de dados pessoais por autoridades competentes para fins de prevenção, investigação, detecção ou ajuizamento de infrações penais (**Diretiva UE 2016/680**), que indicam um possível caminho legislativo para o desenvolvimento de um marco legal que seja efetivo na garantia dos objetivos éticos e humanos sempre destacados nos planos estratégicos e códigos de conduta gerais.

17 Sobre o 2017 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS), ver https://www.unog.ch

18 Como exemplo destacamos que, já em 2012, o Departamento de Defesa dos Estados Unidos aprovou a Diretiva 3000.09, autorizando e estabelecendo critérios para a investigação específica de autonomous weapons systems: https://www.esd.whs.mil

19 Neste sentido ver: https://www.loc.gov

20 Ver: https://www.congress.gov

21 Sobre o Grupo de Peritos de Alto Nível sobre a IA (GPAN AI), ver https://ec.europa.eu

22 Sobre a European Data Strategy: https://ec.europa.eu

Doi: 10.14210/nej.v26n1.p103-135

Cabe recordar que, obviamente, seguem aplicáveis ao desenvolvimento e utilização de sistemas de IA todo o marco legal já existente com relação à proteção dos direitos humanos, em geral, e à responsabilidade legal e de segurança do produto, em particular, dentre outras normativas de caráter geral ou setorial. No entanto, o que se observa é que o avanço tecnológico costuma ser bastante eficiente em ocultar os problemas derivados da parcialidade inerente aos dados ou ao próprio sistema e, ainda que seja possível detectar esta parcialidade, isto não significa necessariamente que haja uma atribuição clara de responsabilidade legal pelo dano causado, quando ocorre. Desta forma, fica óbvio que a incerteza na atribuição da responsabilidade jurídica causa significativo dano à aplicação eficaz da lei como instrumento de melhoria do marco legal e da proteção que se dispensa ao cidadão[23].

Na Espanha, os esforços regulatórios mais relevantes são resultado da aplicação das normativas procedente da União Europeia. Assim, junto com a participação nas iniciativas e acordos em âmbito europeu a respeito do desenvolvimento e implantação de IA, o marco legal espanhol também sofreu melhoria com o aperfeiçoamento da regulação do direito à privacidade de dados pessoais, mediante a aprovação da (nova) Lei Orgânica de Proteção de Dados Pessoais e Garantia de Direitos Digitais (**LOPDP**), em aplicação do RGPD, e com as recomendações feitas pela Agência Espanhola de Proteção de Dados (**AEPD**). No entanto, é necessário que se mencione uma significativa omissão do legislador espanhol: o (imperdoável) atraso na transposição da Diretiva UE 2016/680[24], norma que ainda espera incorporação ao direito interno espanhol, sem prejuízo dos efeitos diretos que possa ter a Diretiva em questão (no caso de que se invoque uma vulneração por parte do Estado Espanhol de um direito reconhecido na Diretiva).

Além disso, em 2019 o Ministério da Ciência, Inovação e Universidades publicou a "Estrategia Española de I+D+I en inteligencia artificial"[25]. O documento, além de estabelecer uma série de prioridades (tais como a utilização de infraestrutura existente, a formação e treinamento de pessoal, a identificação de áreas estratégicas e análises de ética em questões relativas à IA), reconhece o caráter multidisciplinar e transversal da tecnologia. Porém, para o que aqui nos interessa, deve-se ressaltar a relevância que este documento concede ao uso ético da IA, referindo-se aos princípios estabelecidos na "Declaración de Barcelona sobre el desarrollo y uso apropiado de la IA" e determinando, de forma explícita, como um dos objetivos críticos da estratégia "desenhar melhores sistemas de IA que incorporem um razoamento ético".

23 COSTANZO, P.; D'ONOFRIO, F., & FRIEDL, J. Big data and the Italian legal framework: Opportunities for police forces. In AKHGAR, B.; SAATHOFF, G.; ARABNIA, H. R.; HILL, R.; STANIFORTH, A. & BAYERL, P.S. (Eds.), **Application of big data for national security.** (pp. 238-249). UK: Oxford, 2015.

24 Em julho de 2019, a UE decidiu levar a Espanha perante o Tribunal de Justiça europeu pelo atraso na transposição da Diretiva ao ordenamento jurídico local (https://ec.europa.eu). Em 20 de março de 2020 o Governo espanhol finalmente enviou ao Congresso de Deputados o Projeto de Lei 121/000006 para completar tal transposição (http://www.congreso.es).

25 Ministerio de Ciencia, Innovación y Universidades. **Estrategia Española de I+D+I en inteligencia artificial**. 2019. Disponível em: http://www.ciencia.gob.es

Então, em definitivo, à exceção das leis já existentes em matéria de privacidade de dados (RGPD e Diretiva UE 2016/680), pode-se afirmar que em geral os esforços no desenvolvimento de um marco legal efetivo seguem sendo ainda bastante tímidos. Contudo, os diversos informes, propostas, declarações e estudos publicados, ainda que se qualifiquem como soft law, sinalizam uma direção clara: os benefícios do desenvolvimento e implantação de IA não podem, ou ao menos não deveriam, ser perseguidos em detrimento dos direitos fundamentais do indivíduo e dos princípios éticos que devem fundamentar a convivência humana em uma sociedade livre. Nos indicam, ademais, que os países e organismos multilaterais são, em geral, bastante conscientes de que o uso malicioso da tecnologia é uma probabilidade real por enfrentar.

A este respeito, o risco do uso malicioso e abusivo da IA pode, obviamente, apresentar-se em qualquer âmbito, e as atividades do Estado, com todo seu aparato de administração pública, são cada vez mais cenário da aplicação ou experimentação com sistemas de IA.[26] No entanto, em termos concretos, desejamos nos concentrar nos desafios que a IA apresenta no exercício de poder das autoridades públicas encarregadas precisamente de proteger os cidadãos frente a tais abusos. Neste sentido, a utilização de tais tecnologias por forças de segurança pública nos coloca desafios muito específicos na defesa dos direitos da pessoa objeto de investigação criminal, seja esta preventiva ou reativa.

Por este motivo, nos parece urgente que os Estados e organismos internacionais se empenhem efetivamente na construção de um sólido entendimento comum, inclusive com o possível desenvolvimento de um marco legal específico que possa, simultaneamente, amparar o desenvolvimento de IA para utilização por forças de segurança e proteger o cidadão dos possíveis efeitos nocivos de sua aplicação.

Nas linhas que seguem, passaremos a descrever as características dos sistemas de policiamento preditivo, seus benefícios e possíveis carências. Igualmente, buscaremos identificar as razões pelas quais, neste momento, resulta preocupante a fragilidade tanto na base ou teoria científica, quando na compilação de dados com os quais se nutrem tais sistemas inteligentes. À luz de tal análise, proporemos as razões pelas quais o marco legal existente nos parece, de momento, insuficiente em sua tarefa de garantir aos cidadãos em geral, e aos indivíduos pertencentes aos coletivos mais vulneráveis em particular, um sistema de polícia preditiva algorítmica que seja efetivamente livre de preconceitos e discriminações indevidas.

26 ENGSTROM, David F.; HO, Daniel E.; SHARKEY, Catherine M.; CUÉLLAR, Mariano-Florentino. **Government by Algorithm:** Artificial Intelligence in Federal Administrative Agencies. Fevereiro, 2020. Ver também: CRAWFORD, K.; DOBBE, R.; BRYER, T.; FRIED, G.; GREEN, B.; KAZIUNAS, E.; WHITTAKER, M. **AI Now 2019 Report**. New York: AI Now Institute, 2019.

Doi: 10.14210/nej.v26n1.p103-135

3. ESCLARECIMENTOS CONCEITUAIS BÁSICOS.

Antes de seguir com a análise específica das implicações da IA no policiamento preditivo, é necessário fazer referência aos conceitos básicos indispensáveis para poder entrar em investigações jurídicas. Em pleno século XXI não é recomendável que o jurista examine, analise ou, claro, proponha um marco legal sem entender minimamente como funcionam, desde um ponto de vista técnico, as tecnologias sobre as quais se pretende trabalhar.

O caráter multidisciplinar da IA, tão mencionado nos foros que discutem tais questões, não é uma concessão ou uma expressão vazia de sentido prático. Se o propósito é prover a sociedade de um sistema legal que efetivamente funcione e proteja a cidadania e seus direitos humanos básicos, é cada vez mais necessário que juristas, filósofos, acadêmicos e membros da comunidade legal em geral olhem para além das leis e marcos regulatórios abstratos para entender o substrato das bases sobre as quais se constrói e desenha a tecnologia, suas possibilidades de aplicação, suas fortalezas e fragilidades. Mesmo que IA e decisões automatizadas sejam atualmente os termos da moda, é interessante notar, e talvez seja um sinal da complexidade do tema, que não existe consenso nem sequer acerca das definições destas (e de outras) expressões tecnológicas.

Os diferentes documentos oficiais publicados, sejam eles de soft ou hard law, trazem definições distintas e dão diferentes ênfases a este ou aquele aspecto da tecnologia. No entanto, é necessário fazer um esforço para compreender como os regulamentos existentes ou as entidades responsáveis por regular as novas tecnologias entendem e delimitam tais conceitos. Senão vejamos.

(i) Inteligência Artificial. Surpreendentemente não existe um consenso universal na definição do que se entende, especificamente, por inteligência artificial. Algumas definições se referem a sistemas mais sofisticados que incluem o conceito de "machine learning", enquanto outras aceitam que a simples capacidade do algoritmo de coletar os dados, analisá-los e decidir, mesmo que sem capacidade de automelhoria, já deveria ser qualificado como IA. No entanto, quase todas as definições de IA trazem um elemento comum: utilizam como parâmetro a inteligência humana para, em seguida, fazer uma distinção entre inteligência humana "de facto" e o comportamento do sistema artificial que simula ou imita a inteligência humana (o fato de que a própria definição de inteligência humana seja tema de debate até hoje talvez explique as razões pelas quais definir IA seja tão complexo). Algumas características comuns às diferentes definições disponíveis são: (1) os sistemas de IA "replicam" o comportamento humano de análise e tomada de decisões, incluindo a coleta de dados, seu processamento, a análise da informação coletada e a posterior toma de decisão com base em tal análise; (2) os sistemas têm algum grau de autonomia; e (3) têm capacidade de atuar para alcançar um objetivo, seja recomendando uma decisão/ação, seja atuando autonomamente para tal.

As definições podem ou não incluir o "machine learning": a capacidade do sistema de 'aprender' mediante sua própria observação do entorno (físico ou digital) e das mudanças causadas

em tal entorno pela própria atuação do sistema. Hoje em dia quase todas as definições de IA fazem referência ao "machine learning".[27]

(ii) Big Data. O que hoje chamamos de Big Data surge do desenvolvimento tecnológico na capacidade de geração, armazenamento e processamento de dados digitais, de maneira que a quantidade de dados é tão imensa que se faz necessário um supercomputador para processá-los[28]. Para que se tenha uma ideia, em 2018 foram gerados 33 zettabytes de dados e a projeção é de que em 2025 se alcancem 175 zettabytes[29]. O que transforma o Big Data no combustível básico da IA é justamente que esta quantidade extraordinária de dados pode ser processada de maneira cada vez mais rápida e barata, o que permite treinar os sistemas de IA para que obtenham cada vez mais precisão e eficácia em suas decisões.

O desafio que a geração de uma quantidade de dados desta magnitude implica, é a dificuldade de seu controle efetivo e, consequentemente, de sua regulação. Os dados já não se armazenam localmente, senão que em uma grande rede interconectada de servidores remotos[30]. Deste modo, o controle feito pelos cidadãos a respeito de que dados seus são armazenados e processados, ou onde e como se realiza tal processamento, se torna cada vez mais complexo e difícil. Isto gera um desafio não apenas para a proteção da privacidade, mas também para a transparência e a "explicabilidade" do programa, como detalharemos em seguida[31].

No que concerne ao marco regulatório existente, é importante entender que a geração de dados

27 Alguns exemplos de definições de IA: "Artificial Intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behavior by analyzing how the environment is affected by their previous actions." (EU High Level Expert Group: A Definition of AI.) https://ec.europa.eu "In this section, the term "artificial intelligence" includes the following: (1) Any artificial system that performs tasks under varying and unpredictable circumstance without significant human oversight, or that can learn from experience and improve performance when exposed to data sets. (2) An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action. (3) An artificial system designed to think or act like a human, including cognitive architectures and neural networks. (4) A set of techniques, including machine learning, that is designed to approximate a cognitive task. (5) An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision making, and acting." (US National Defense Authorization Act 2019: https://www.congress.gov).

28 CATE, Fred H.; KUNER, Christopher; MILLARD, Christopher; SVANTESSON, DAN JERKER, B. **"The Challenge of "Big Data" for Data Protection"**. 2012. Disponível em: http://www.repository.law.indiana.edu _

29 A este respeito, ver COUGHLIN, Tom. "**175 Zettabytes By 2025", publicado em Forbes**. 2018. Disponível em: https://www.forbes.com _

30 Ver CATE, Fred H.; KUNER, Christopher; MILLARD, Christopher; SVANTESSON, DAN JERKER, B. **"The Challenge of "Big Data" for Data Protection"**. 2012. Disponível em: http://www.repository.law.indiana.edu

31 Exemplo de definição de Big Data: "It is often characterised by the increased volume, velocity and variety of data being produced ("the three Vs"), and typically refers (but is not limited) to data from the internet. Big data comes from a variety of sources, including social media data or website metadata. The Internet of Things (IoT) contributes to big data, including behavioural location data from smartphones or fitness tracking devices. In addition, transaction data from the business world form a part of big data." (European Union Agency for Fundamental Rights. #BigData: Discrimination in data-supported decision making.) https://fra.europa.eu

Doi: 10.14210/nej.v26n1.p103-135

no contexto do Big Data não necessariamente implica no tratamento de dados pessoais, pelo menos da maneira em que este conceito é utilizado pelo RGPD e demais normas de privacidade em geral. De fato, para que um conjunto de dados seja legalmente considerado "dados pessoais" e, portanto, merecedor de proteção sob o RGPD, é indispensável a possibilidade de identificação inequívoca de um indivíduo[32]. No entanto, há uma infinidade de dados gerados diariamente pelo indivíduo que, ainda que não se qualifiquem legalmente como "dados pessoais", quando são analisados em conjuntou e/ou combinados com outras bases de dados e algoritmos, podem chegar a interferir de maneira significativa em um grupo de pessoas específico ou em um indivíduo em particular, sem que necessariamente as garantias dos regulamentos tais como o RGPD sejam aplicáveis. Conforme ensina Valls Prieto, as técnicas de Big Data permitem "ver e entender a relação entre peças de informação que por si só não nos dizem nada"[33]. Este entendimento sobre a natureza do Big Data e a teoria do mosaico são fundamentais para a construção da análise dos programas de IA e das possíveis parcialidades a eles inerentes.

(iii) Decisões automatizadas. Se dizem automatizadas as decisões tomadas primariamente por um sistema computacional automatizado. Tecnicamente, tal sistema pode ou não operar por meio de IA, ainda que se possa afirmar que esta seja a realidade em quase todos os casos de sistemas de decisões automatizadas[34].

As decisões automatizadas podem ter diferentes classificações, dependendo do nível de intervenção humana ou autonomia dos sistemas:

Human-out-of-the-loop (HOOL): sem intervenção humana. O sistema decide seu próprio objetivo e atua para alcançá-lo sem que haja interferência humana em qualquer momento do ciclo de decisão (e.g.: LAWS).

Human-in-the-loop (HITL): há intervenção humana em cada ciclo de decisão do sistema.

Human-on-the-loop (HOTL): há intervenção humana durante o desenho do ciclo de decisão do sistema e monitoramento da operação do sistema.

32 RGPD, Art. 4(1): "Dados pessoais: informação relativa a uma pessoa singular identificada ou identificável ("titular dos dados"); é considerada identificável uma pessoa singular que possa ser identificada, direta ou indiretamente, em especial por referência a um identificador, como por exemplo um nome, um número de identificação, dados de localização, identificadores por via eletrônica ou a um ou mais elementos específicos da identidade física, fisiológica, genética, mental, económica, cultural ou social dessa pessoa singular".

33 Tradução livre dos autores. O original lê: "ver y entender la relación entre piezas de información que por sí solas no nos dicen nada". VALLS PRIETO, J. **Problemas jurídico penales asociados a las nuevas técnicas de prevención y persecución del crimen mediante inteligencia artificial**. Madrid: Dykinson, 2017. p. 10.

34 Exemplos de definição de decisões automatizadas: "As decisões exclusivamente automatizadas correspondem à capacidade de tomar decisões através de meios tecnológicos e sem intervenção humana." RGPD - Grupo de Trabalho do Artículo 29 para a Proteção de Dados. Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento 2016/679. https://ec.europa.eu "The term 'automated decision system' means a computational process, including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques, that makes a decision or facilitates human decision making, that impacts consumers." US Algorithm Accountability Act. https://www.congress.gov

Human-in-command (HIC): existe a possibilidade de supervisão humana sobre toda a atividade do sistema (inclusive sobre seus impactos sociais, econômicos, legais e éticos) e capacidade de decidir quando e como utilizar o sistema em qualquer situação.

(iv) 'Explicabilidade'. A 'explicabilidade' é um conceito bastante amplo e genérico. A dificuldade em definir o que seria uma "explicação" do sistema e/ou de seu resultado final provavelmente deriva do fato de que a 'explicabilidade' não é exatamente um conceito em si mesmo, senão que um propósito: a intenção é que o resultado final de um determinado algoritmo (seja ele uma decisão, um perfil, uma recomendação ou, até mesmo, uma ação) possa ser explicado de uma maneira que permita que um humano entenda como o sistema – ou algoritmo – chegou a tal decisão. É o "porquê" e o "como" um sistema chega a uma determinada conclusão. Dependendo do tipo de decisão e suas implicações, isto pode criar uma necessidade de entender a lógica implícita na tomada de decisão, mas também pode demandar um entendimento sobre a base de dados utilizada, o processo técnico de análise de tais dados, ou até um entendimento do algoritmo em si mesmo e seu próprio source code.[35]

Este conceito ganha relevância quando grande parte dos documentos publicados (se não todos) em matéria de IA e decisões automatizadas – sejam eles de soft ou hard law – faz referência à necessidade de transparência dos sistemas e das decisões, e articula sobre este conceito de "transparência" a solução para todos os possíveis conflitos éticos e jurídicos que possa gerar a IA aplicada a decisões automatizadas: se o cidadão é capaz de entender a lógica existente por trás de uma determinada decisões automatizadas, seja pela administração pública, seja por uma entidade privada, este cidadão seria hipoteticamente capaz de se defender ante uma decisão injusta, incorreta ou até mesmo ilegal. Portanto a 'explicabilidade' se transforma em um dos pilares da eficácia dos instrumentos legais de controle relativos à IA. Neste sentido, o grupo independente de expertos europeus, no documento "Orientações Éticas para uma IA de Confiança", define do seguinte modo os princípios éticos que devem dirigir a investigação, o desenvolvimento e a utilização de sistemas de IA

> Os sistemas de IA devem melhorar o bem-estar individual e coletivo. Esta seção enumera quatro princípios éticos, enraizados nos direitos fundamentais, que devem ser respeitados para assegurar que os sistemas de IA são desenvolvidos, implantados e utilizados de forma confiável. São especificados como imperativos éticos, que os profissionais no domínio da IA devem esforçar-se sempre por respeitar. Sem impor uma hierarquia, enumeramos a seguir os princípios de

35 Exemplos de definição de "Explicabilidade": "A explicabilidade diz respeito à capacidade de explicar tanto os processos técnicos de um sistema de IA como as decisões humanas com eles relacionadas (p. ex., os domínios de aplicação de um sistema de IA). A explicabilidade técnica exige que as decisões tomadas por um sistema de IA possam ser compreendidas e rastreadas por seres humanos". (Grupo de Peritos de Alto Nível sobre IA. Orientações Éticas para uma IA de Confiança. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai); "…a study evaluating an automated decision system and the automated decision system's development process, including the design and training data of the automated decision system, for impacts on accuracy, fairness, bias, discrimination, privacy, and security that includes, at a minimum – (A) a detailed description of the automated decision system, its design, its training, data and its purpose" (US Algorithm Accountability Act).

Doi: 10.14210/nej.v26n1.p103-135

> modo a refletir a ordem em que os direitos fundamentais nos quais se baseiam são apresentados na Carta da UE. Estes são os princípios de: i) Respeito da autonomia humana ii) Prevenção de danos iii) Equidade iv) Explicabilidade.[36]

No entanto, o próprio RGPD serve como exemplo da dificuldade em definir este conceito. Ainda que em seu artigo 15(h) o Regulamento ancore, em parte, a proteção dos direitos do indivíduo no direito à explicação, o documento não propõe uma definição: apenas faz referência à necessidade de informar ao sujeito a respeito da "lógica subjacente" à decisão, sem detalhar em quê consistiria, precisamente, tal lógica. No Considerando 71, o RGPD menciona "uma explicação sobre a decisão tomada na sequência dessa avaliação", mas não especifica o nível de detalhe que se consideraria adequado. O Grupo de Trabalho do Artigo 29 para a proteção de dados (**GTA 29**)[37] tenta ir além ao explicar que:

> O responsável pelo tratamento deverá encontrar formas simples de comunicar ao titular dos dados a lógica subjacente, ou os critérios aplicados para tomar a decisão. O RGPD obriga o responsável pelo tratamento a fornecer informações úteis relativas à lógica subjacente, e não necessariamente a uma explicação complexa sobre os algoritmos utilizados ou a divulgação do algoritmo na íntegra. As informações prestadas devem, no entanto, ser suficientemente completas para permitir ao titular dos dados compreender os motivos da decisão.

No entanto, ainda que defina o objetivo da explicação, tanto o próprio RGPD quanto o documento do GTA 29 parecem deixar ampla discricionariedade a respeito de sua extensão.

(v) Polícia ou Policiamento Preditiva(o). Nos referimos anteriormente ao conceito de polícia e policiamento preditivo, e aos tipos de predições que se podem abarcar. Conforme Miró Llinares, o termo poderia ser definido do seguinte modo:

> (...) utilizar algoritmos para estimar riscos futuros associados a características criminológicas pessoais ou ambientais, de forma a melhorar a tomada de decisão para fins de prevenção, investigação e persecução do delito."[38]

Como exemplos de sistemas preditivos estão o programa PredPol, que utiliza análise de dados para indicar o risco da comissão de um delito por hora e região[39], ou o programa COMPAS[40], que analisa o risco de reincidência criminal de um indivíduo específico para avaliar, entre outras coisas, se é recomendável decretar prisão preventiva ou provisória. Assim, em geral, a predição delitiva por algoritmos costuma basear-se na localização/temporalidade (utilizando dados para determinar o onde e o quando se podem cometer delitos) ou no sujeito (utilizando dados para determinar quem

36 Grupo de Peritos de Alto Nível sobre IA. Orientações Éticas para uma IA de confiança. https://ec.europa.eu

37 Grupo de Trabalho do Artigo 29 para a Proteção de Dados. Orientações sobre decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento 2016/679. https://ec.europa.eu

38 Tradução livre dos autores. O original lê: "using algorithms to estimate future risks associated with crime-related personal or environmental characteristics and thereby improve decision-making for crime prevention, investigation and prosecution purposes." MIRÓ LLINARES, F. Predictive Policing: Utopia or Dystopia? On attitudes towards the use of Big Data algorithms for law enforcement. **Revista de Internet, Derecho y Política**, n. 30. 2020.

39 Mais sobre: https://www.predpol.com/about/

40 Neste sentido ver: https://www.equivant.com/

pode ser vítima ou autor de um ato delitivo)[41]. Outro sistema não diretamente relacionado com a justiça criminal, mas cujo impacto nos direitos humanos e liberdades individuais não pode ser subestimado, é o denominado iBorderCtrl, um sistema de controle fronteiriço atualmente em fase de teste na União Europeia, e que propõe identificar a fraude por meio da análise dos movimentos faciais do sujeito[42].

(vi) Parcialidade. Conforme o Dicionário Online de Português contemporâneo, o termo parcialidade pode ser definido como "qualidade de quem toma partido ao julgar a favor ou contra, tendo em conta sua preferência, sem se importar com a justiça ou com a verdade". No contexto social que aqui interessa, a presença ou utilização de parcialidade ou vieses preconceituosos em sistemas de IA em geral leva à inferência de padrões de conduta não a partir de dados e estatísticas objetivos que representem o comportamento individual, mas sim baseando-se em generalizações sem comprovação científica e em aspectos como gênero, origem, raça ou posição social. Ou seja, introduzir parcialidade leva à inserção de conceitos discriminatórios e estereotipados no processo de decisão. No âmbito da IA, a parcialidade presente em um algoritmo pode levar a resultados e decisões automatizadas que, ou bem não refletem a realidade, ou bem reforçam e/ou intensificam discriminações ocorridas fora do algoritmo, no universo social, e por isso os alimenta. É neste contexto de quantidades extraordinárias de dados, de algoritmos opacos, de decisões automatizadas e preconceitos sociais refletidos em sistemas artificiais, que queremos analisar a aplicação de IA e decisões automatizadas no trabalho da denominada polícia preditiva (predictive policing). Além disso, interessa-nos analisar como esta reprodução 'invisível' dos preconceitos pode levar a resultados contrários ao que é a intenção exteriorizada da construção de uma IA ética e centrada em valores humanos.

4. A COLETA DE INFORMAÇÃO RELEVANTE E OS PRECONCEITOS NOS SISTEMAS DE POLÍCIA PREDITIVA.

Ainda que cada sistema ou programa de polícia preditiva funcione de maneira distinta (utilizando dados de geolocalização ou antecedentes criminais, por exemplo) e tenha objetivos distintos (identificar possíveis zonas delitivas ou medir a probabilidade de reincidência criminal), todos convergem em um ponto crucial: a utilização da análise de Big Data (que pode ou não coincidir com a análise de dados pessoais) para encontrar padrões e fazer inferências que instruam decisões no sistema judicial (ou no sistema mais amplo de "segurança nacional", no caso do iBorderCtrl), e fazê-lo de modo mais eficiente que os modelos tradicionais de predição.

Aqui cabe destacar que, como nos recorda Miró Llinares, os sistemas automatizados de polícia preditiva não são capazes, no dia de hoje, de prever com exatidão a comissão de um delito em particular. Não estamos nesta fase utópica (ou distópica, dependendo do ponto de vista) da ficção científica, segundo a qual se supõe que poderemos indicar com precisão e suficiente

41 RICHARDSON, R.; SCHULTZ, Jason M.; & CRAWFORD, K. Dirty Data, Bad Predictions: How Civil Rights Violations impact police data, predictive policing systems, and justice. **New York University Law Review Online**. 2019.

42 Ver: https://www.iborderctrl.eu/The-project

Doi: 10.14210/nej.v26n1.p103-135

antecedência que um determinado indivíduo cometerá uma infração específica em um lugar e hora determinados, a fim de que a polícia possa intervir e impedi-lo ex ante. Na verdade, é questionável não apenas se um dia teremos tal capacidade, como também se é algo desejável. De fato, o que os sistemas existentes de predição delitiva costumam (ou prometem) fazer é simplesmente auxiliar na identificação de riscos de maneira mais eficiente por meio da análise de dados, baseados em eventos prévios e seus fatores condicionantes[43].

No entanto, neste momento não se pode afirmar de modo inequívoco que os sistemas de IA aplicados à polícia preditiva sejam mais eficientes, mais precisos e/ou menos custosos em comparação à atuação humana. Meijer e Wessels analisaram estudos empíricos da aplicação de tais sistemas, e as conclusões não conduzem ao entusiasmo: no máximo, os resultados se revelam ambíguos ou variados[44]. Além disso, fica claro que os sistemas de IA, quando aplicados a tarefas de polícia preditiva, não devem ser mais que instrumentos auxiliares à execução da função policial: de nenhuma maneira devem ser tratados como tabula rasa das estratégias preditivas. Isto porque, afinal, os erros em sistemas que afetam de maneira tão fundamental os direitos e liberdades individuais, como no âmbito da justiça criminal, podem ter um efeito perverso sobre a sociedade. Utilizando sistemas de IA podemos prever se vai chover ou não, se é ou não lucrativo investir em uma determinada empresa e comprar ações, ou se uma pessoa vai sofrer uma doença importante nos próximos 10-20 anos e confiar nessa previsão, arriscando-nos. Mas não podemos etiquetar uma pessoa tomando decisões que a prejulguem: não somos precogs (como no filme Minority Report) e, portanto, não podemos desprezar os padrões de certeza que são requeridos em cada fase da investigação. Os indícios racionais de criminalidade ou periculosidade devem estar solidamente fundamentados. Um dos problemas fundamentais na construção e operação dos sistemas de IA reside na qualidade dos dados coletados e analisados pelos algoritmos. Neste ponto, nos interessa analisar exemplos práticos no sistema de justiça criminal dos Estados Unidos.

Em 2019, um estudo sobre treze jurisdições norte-americanas que utilizam (ou utilizaram) sistemas de IA em polícia preditiva, produziu dados alarmantes[45]. Todas as jurisdições examinadas (que incluem comunidades tão diversas como Nova York, Seattle ou Phoenix) tiveram suas forças policiais sob investigação por práticas corruptas, discriminatórias e/ou ilegais em algum momento do passado recente (tão recente quanto 2018, por exemplo, no caso de Baltimore). No entanto, os dados produzidos por tais atuações policiais carregadas de erros seguiram, e em muitos casos seguem,

43 MIRÓ LLINARES, F. Predictive Policing: Utopia or Dystopia? On **attitudes towards the use of Big Data algorithms for law enforcement. Revista** de Internet, Derecho y Política, n. 30. 2020.

44 "Uma conclusão preliminar é que esta abordagem tem potencial, mas nem todos os tipos de crimes podem ser eficazmente reduzidos por meio de métodos de policiamento preditivo e, portanto, é necessário que os agentes que executam estas estratégias utilizem tais modelos adequadamente)". Tradução livre dos autores. O original lê: "A preliminary conclusion is that this approach has potential but not all types of crimes can be effectively reduced through predictive policing models and therefore the officers executing these strategies need to adequately use these models". VÉASE, Meijer; A. WESSELS, M. Predictive Policing: Review of Benefits and Drawbacks. **International Journal of Public Administration**, Vol. 42, no. 12, 1031-1039. 2019. Disponível em: https://www.tandfonline.com

45 RICHARDSON, R.; SCHULTZ, Jason M.; & CRAWFORD, K. Dirty Data, Bad Predictions: How Civil Rights Violations impact police data, predictive policing systems, and justice. **New York University Law Review Online**. 2019.

compiladas e arquivadas pelos diferentes sistemas aos quais a administração pública recorre em suas atividades rotineiras. Em outras palavras: mesmo que atuações policiais irregulares tenham sido identificadas, a informação (dados) resultante destas atuações não foi necessariamente eliminada do sistema. Denominamos este tipo de dados contaminados como "dirty data" [46].

Simultaneamente, estas mesmas jurisdições utilizavam (ou estavam desenvolvendo) sistemas de IA para predição policial. No entanto, ainda que se houvesse comprovado de maneira efetiva a má práxis policial em investigações e auditorias do próprio sistema judicial norte-americano (por exemplo, as cortes federais do distrito de Nova York constataram comportamento discriminatório sistêmico por parte do Departamento de Polícia de Nova York e impuseram sobre ele um monitoramento rigoroso de compliance; e a cidade de Baltimore firmou um consent decree com o Departamento de Justiça, ordenando uma série de reformas policiais a respeito deste tema), aqueles dados repletos de preconceitos que já haviam sido coletados seguiram sendo utilizados nos sistemas de polícia preditiva.

Como explicamos anteriormente, um sistema de IA somente funciona com dados: os dados são o seu alimento, o combustível fundamental de qualquer sistema de IA, por mais sofisticado que seja. Se a informação que alimenta o sistema (dados) não é confiável, é incorreta ou é, inclusive, resultado de práticas ilegais, não é razoável esperar que o resultado final deste sistema (neste caso, a "predição") o seja. Quer dizer, se uma força policial discrimina ilegalmente contra imigrantes, por exemplo, submetendo-os a mais revistas ou impondo sobre eles uma política de "no tolerance" em termos que não sejam equitativos com outros conjuntos sociais mais favorecidos, os dados gerados por esta força policial já contém preconceitos desde a sua origem. Portanto, se um algoritmo recebe dados informando sobre incidências delitivas significativamente mais altas em um determinado coletivo social (e.g., imigrantes de origem africano) ou uma determinada zona (bairros com alta presença de imigrantes), dificilmente este sistema enviará a força policial a zonas de outra categoria social. O sistema simplesmente reproduzirá o viés preconceituoso contido nos dados e seguirá enviando patrulhas a zonas já especialmente sobremonitoradas por razões discriminatórias que pouca (ou nenhuma) relação tem com o índice real de criminalidade.

A este respeito, convém recordar que os algoritmos trabalham de forma superficial (não podem abstrair ou avaliar em conjunto) e sem base em relações causais (apenas em correlações significativas). Não são capazes de diferenciar que peso causal uma determinada característica como raça, posição social ou gênero possui, dentre um conjunto de fatores. Parece óbvio dizer que a Criminologia atuarial[47] somente se preocupa com predições estatísticas quando, ao contrário, um sistema de justiça criminal garantista não pode operar com juízos estatísticos baseados em simples

46 RICHARDSON, R.; SCHULTZ, Jason M.; & CRAWFORD, K. Dirty Data, Bad Predictions: How Civil Rights Violations impact police data, predictive policing systems, and justice. **New York University Law Review Online**. 2019, p. 4.

47 Sobre as origens da denominada Criminologia atuarial, ver DOMÍNGUEZ FIGUEIRIDO, D.; BASANTA, A. Lógica actuarial, seguridad y sistema de justicia criminal. in **La seguridad en la sociedad del riesgo:** Un debate abierto. Barcelona: Atelier, 2003.

Doi: 10.14210/nej.v26n1.p103-135

números probabilísticos. Claramente, não há controvérsia a respeito do fato de que não se deve condenar ou decidir sobre a inocência de um acusado baseando-se apenas em fatores e estimações de probabilidade. O sistema requer a prática de provas que desgastem a presunção de inocência além de qualquer dúvida razoável: pode-se substituir "dúvida razoável" por "dúvida estatística"? O problema reside nas decisões do sistema de justiça criminal com caráter prospectivo, e não retrospectivo. Neste caso, a valoração probabilística se projeta ao futuro, sendo este sempre incerto por natureza. Imputar responsabilidade penal exige algo mais, e por isso também deve exigir-se algo mais de decisões que, sem imputar de modo definitivo, restrinjam a liberdade da pessoa com base em sólidos indícios racionais de criminalidade (prévia) e periculosidade (futura). A liberdade humana é um hiato inexplicável na lógica sequencial de uma concatenação de causas deterministas[48] e, portanto, a não-redução do homem a um conjunto de variáveis estatísticas deve projetar-se tanto em direção ao passado, quanto ao futuro.

O problema da parcialidade em um ambiente de políticas criminais e de segurança não é, tampouco, somente um problema da polícia ou do sistema judicial. Os resultados enviesados de um sistema de IA consistentemente mal aplicado levam a uma percepção também enviesada do entorno social sobre coletivos que, em muitos casos, já são objeto de discriminação mesmo sem o auxílio da tecnologia. Esta retroalimentação de discriminação sobre discriminação pode tornar-se não apenas nefasta para as populações estigmatizadas, como também resultar na ineficácia das pretendidas reformas policiais: os agentes de segurança, ao receber instruções de algoritmos que, ainda que utilizados legalmente, se alimentam de dados e fazem inferências enviesadas, atuarão em conformidade com tais instruções enviesadas, gerando mais dirty data e retroalimentando o sistema (um problema descrito como "runaway feedback loop"). Sem uma intervenção humana ativa e objetiva sobre a geração e o tratamento dos dados, e sobre a atuação policial derivada das recomendações do sistema, o ciclo discriminatório dificilmente será interrompido. O problema ganha complexidade quando contemplamos a realidade fora dos Estados Unidos. Apesar de todos os problemas identificados em suas bases de dados, os Estados Unidos avançam em direção a uma possível solução já que, ao menos, existem dados disponíveis sobre atuações policiais, políticas e justiças criminais, e seus possíveis vieses preconceituosos, discriminações e ilegalidades. Na Europa, ao contrário, os problemas parecem ser de ordem mais básica.

Em um estudo publicado em novembro de 2018 por Justicia European Rights Network, identificou-se uma série de desafios nas práticas de justiça criminal nos países europeus. Ainda que em pequena escala (cobre somente doze países, entre os quais, Espanha), o estudo evidenciou a existência de problemas como a inconsistência na coleta e análise de dados no contexto de justiça criminal, tanto em plano doméstico quanto entre os países. Ainda mais preocupante foi a constatação de uma significativa ausência de investigações relativas à atuação da justiça criminal e, em particular,

48 Sobre o conceito de imputação e sua relação com a liberdade, ver HRUSCHKA, J.; SÁNCHEZ-OSTIZ GUTIÉRREZ, P. **Imputación y derecho penal**: estudios sobre la teoría de la imputación. Thomson Aranzadi; Universidad de Navarra-Garrigues Cáterdra, 2005.

de investigações oficiais com o objetivo de identificar possíveis práticas discriminatórias. Se, como vimos no caso dos Estados Unidos, a existência e coleta de dados que evidenciam uma má prática policial não impedem a utilização persistente de dirty data no desenvolvimento dos algoritmos de policiamento preditivo, o problema se agrava substancialmente quando, de entrada, a administração pública nem sequer permite a identificação de tais práticas discriminatórias. Na Espanha, por exemplo, os dados públicos sobre paradas de identificação e revista pela polícia não detalham o perfil étnico, o que faz com que seja bastante mais complicado obter informação consistente sobre problemas de discriminação policial baseada em origem étnica ou raça, problema conhecido como racial profiling[49]. O perfilamento racial consiste na "prática de utilizar estereótipos étnicos ou raciais ao invés da conduta individual, da descrição de suspeitos ou do conhecimento acumulado para dirigir as atuações das forças da lei" [50].

Claro está que as paradas de identificação para posterior revista devem sempre estar baseadas em princípios legais amplamente aplicáveis à atuação policial. No caso da Espanha, tanto a Lei Orgânica de Proteção da Segurança Cidadã[51], como a Lei das Forças e Corpos de Segurança[52] exigem que a atuação policial seja proporcional e respeite os princípios de não discriminação por qualquer razão, inclusive, obviamente, razões de origem étnica ou racial.

No entanto, um estudo independente realizado em nível nacional em 2013 pela Universidade de Valência indicou que, nos dois anos anteriores, 60% de pessoas de origem cigana foram submetidas a paradas de identificação, contrastando com 6% de cidadãos de origem espanhola[53]. No caso de cidadãos do norte-africanos ou árabes esta proporção foi de 45% e para cidadãos latino-americanos, de 22%. Em um informe mais recente sobre a evolução da discriminação na Espanha entre os anos 2013 e 2016, 20% e 14% dos entrevistados, respectivamente, indicaram ter experimentado discriminação no trato com a polícia em razão de sua origem étnica ou racial[54].

---

49 Ver: https://www.justiceinitiative.org

50 "O perfil racial, ou seja, a prática de utilizar estereótipos étnicos ou raciais ao invés da conduta individual, da descrição de suspeitos ou do conhecimento acumulado para dirigir as atuações das forças da lei, é um problema compartilhado por inúmeros países europeus. Se manifesta de diferentes formas, entre as quais se contam as verificações de identidade desproporcionadas e arbitrárias, as paradas e revistas de membros de grupos étnicos minoritários e um incremento no patrulhamento de bairros de minorias étnicas. O perfil racial contribui desnecessariamente a tensionar ainda mais as já frequentemente tensas relações entre a polícia e a comunidade, e é, além disso, uma prática policial ineficiente e ilegal." Tradução livre dos autores. O original lê: "El 'perfil racial', es decir, la práctica de utilizar estereotipos étnicos o raciales en lugar de la conducta individual, la descripción de sospechosos o el conocimiento acumulado para dirigir los actos de las fuerzas de la ley, es un problema que comparten numerosos países europeos. Se manifiesta de diferentes formas, entre las que se cuentan los chequeos de identidad desproporcionados y arbitrarios, las paradas y cacheos de miembros de grupos étnicos minoritarios y un incremento en el patrullaje en barrios de minorías étnicas. El perfil racial contribuye a innecesariamente tensar más las ya frecuentemente tensas relaciones policía-comunidad, y es además una práctica policial inefectiva e ilegal." WEGMAN, D.; PERNAS, B. **Perfil Racial en España:** investigaciones y recomendaciones. Grupo de Estudios y Alternativas 21, Open Society Justice Initiative, 2005.

51 Ley Orgánica 4/2015, de 15 de marzo, de protección de la seguridad ciudadana. Ver Artigo 16(1).

52 Ley Orgánica 2/1986, de 13 de marzo, de fuerzas y cuerpos de seguridad. Ver Artigo 5.

53 AÑON, José G.; BRADFORD, B.; SÁEZ, José Antonio G.; CUENCA, Andrés G. & FERRERES, Antoni L. **Identificación Policial por Perfil Étnico en España**. Informe sobre experiencias y actitudes en relación con las actuaciones policiales. Tirant Lo Blanch. Valencia 2013. Disponível em: https://www.uv.es.

54 CEA D'ANCONA, M.A.; VALLÉS MARTÍNEZ, M.S. **Evolución de la Discriminación en España**. Informe de las encuestas IMIO-CIS de 2013 y 2016. Instituto de la Mujer para la Igualdad de Oportunidades (IMIO). 2018. Disponível em: http://www.inmujer.gob.es

Doi: 10.14210/nej.v26n1.p103-135

Igualmente, a European Union Agency for Fundamental Rights (**FRA**) em seu segundo estudo amplo sobre minorias e discriminação na União Europeia publicado em 2018, informou que 24% dos entrevistados de origem afrodescendente sofreu paradas de identificação e revista nos cinco anos anteriores à pesquisa, e 11% nos doze meses anteriores. Destes últimos, até 44% manifestou acreditar que a parada foi motivada por racial profiling[55]. Assim, sem que nos aprofundemos mais na discussão sobre a ilegalidade e o caráter discriminatório do racial profiling, temos que (i) oficialmente, os dados da polícia espanhola (e de grande parte dos países da União Europeia) não separam os sujeitos por sua etnia, e; (ii) no entanto, o viés de racial profiling existe na atuação policial, como confirmam os números de investigações independentes. Em outras palavras, a discriminação/racial profiling é uma realidade da atuação policial espanhola e europeia, mas não é tratada de maneira específica pelos dados oficiais. Tudo isto sem que esteja definitivamente comprovada a eficácia de paradas de identificação que sejam baseadas primordialmente em características raciais, senão que ao revés[56].

Não pretendemos aqui negar que exista sensibilidade a respeito do tratamento de dados de categorias especiais, como o são todos os dados sobre origem étnico ou racial, por exemplo. Na realidade, proteção específica é dispensada a tais dados tanto no RGPD quanto na Diretiva UE 2016/680, em seus artigos 9 e 10, respectivamente, onde se estabelecem restrições para o tratamento de categorias especiais de dados. No entanto, ambas normas permitem tal tratamento sob condições específicas, entre as quais se incluem o "interesse público essencial" ou o tratamento "para fins estatísticos"[57]. Portanto, nos parece que cumpridas as condições técnicas de proteção dos dados, inclusive a sua possível pseudonimização[58], ambas normas legais permitiriam a coleta e o tratamento

55 Second European Union Minorities and Discrimination Survey. **Being Black in the EU**. European Union Agency for Fundamental Rights. 2018. https://fra.europa.eu

56 "Como outros informes já demostraram (ej. FRA 2010b: 34), quando o critério que se adota para deter e identificar pessoas é um critério étnico, o índice de acertos é consideravelmente baixo. Isto indica, como trataremos de demonstrar, que as identificações por perfil étnico, além de vulnerar os direitos das pessoas, são uma prática pouco efetiva em termos de garantia da segurança pública. Por outro lado, a experiencia STEPSS na Espanha demonstrou que quando a polícia se baseia na inteligência ou nos indícios objetivos de suspeita razoável ao invés da aparência étnica, o viés discriminatório se reduz e aumenta a eficiência (OSI, 2009:77-78)." Tradução livre dos autores. O original lê: "Como han demostrado ya otros informes (ej. FRA 2010b: 34), cuando el criterio que se adopta para detener e identificar a las personas es un criterio étnico, el índice de aciertos es considerablemente bajo. Esto indica, como trataremos de demostrar, que las identificaciones por perfil étnico, además de vulnerar los derechos de las personas, son una práctica poco efectiva en términos de garantía de la seguridad pública. Por otro lado, la experiencia STEPSS en España demostró que, cuando la policía se basa en la inteligencia o en los indicios objetivos de sospecha razonable en lugar de en la apariencia etnica de las personas, el sesgo discriminatorio se reduce y aumenta la efectividad (OSI, 2009:77-78)." WEGMAN, D.; PERNAS, B. **Perfil Racial en España:** investigaciones y recomendaciones. Grupo de Estudios y Alternativas 21, Open Society Justice Initiative, 2005.

57 Artigo 9 (g) e (j) do RGPD. Ver também Artigo 10 da Diretiva UE 2016/680, onde se permite o tratamento de categorias especiais de dados sempre e quando "for autorizado pelo direito da União ou do Estado-Membro".

58 Ainda que seja necessário levar em conta que a possibilidade técnica de anonimização ou eliminação real de dados pessoais no entorno digital seja controversa, a possibilidade legal existe nos regulamentos citados. A própria AEPD reconhece as dificuldades técnicas para a anonimização permanente de dados pessoais, e estabelece critérios para o desenvolvimento e implantação de tecnologias desta natureza no documento "Orientaciones y garantías en los procedimientos de anonimización de datos personales" (https://www.aepd.es). Ver também: "The Right to be Forgotten: between expectation and practice" publicado por European Network and Information Security Agency (https://www.enisa.europa.eu). Sobre as possibilidades técnicas para minimizar os problemas da anonimização digital de dados pessoais, ver: MAJEED, Abdul; ULLAH, Farman; LEE, Sungchang. **"Vulnerability- and Diversity-Aware Anonymization of Personally Identifiable Information for Improving User Privacy and Utility of Publishing Data**". Sensors 2017, vol. 17, 1059; doi:10.3390/s17051059. Disponível em: https://www.mdpi.com.

de dados de categorias étnicas, por exemplo, para efeitos de detecção e investigação da prevalência de paradas de identificação e revista sobre determinadas etnias, e que podem indicar uma possível parcialidade na atuação policial.

Dadas estas circunstâncias, é razoável esperar que qualquer sistema de polícia preditiva que se pretenda desenvolver e que termine por utilizar os dados oficiais existentes, mas sem considerar sua possível natureza discriminatória, leve a resultados que, além de enviesados, podem até mesmo ser contraproducentes. Isso porque a discriminação não traz problemas apenas de integração e violação de direitos humanos dos grupos sujeitos a paradas mais frequentes, mas também porque implica em uma perda da legitimidade das próprias forças e corpos policiais, como aponta o relator especial das Nações Unidas para a discriminação em uma de suas visitas à Espanha[59]. Além disso, a ausência de dados oficiais confiáveis sobre a atuação policial cria dificuldades para a eventual documentação e correção das atuações que sejam ilegais ou discriminatórias. Ou seja, se oficialmente o sistema não permite identificar ações discriminatórias, é razoável inferir que será cada vez mais difícil que, a partir de qualquer demanda cidadã, se obtenha uma explicação clara a respeito de uma determinada decisão de um sistema automatizado. Ainda mais quando o tratamento dos dados, feito pelo próprio sistema, não é totalmente transparente e implica em uma aparente "imparcialidade" quando, na verdade, opera com generalizações estereotipadas.

Obviamente, o problema da discriminação em sistemas de IA não é exclusivo dos âmbitos relativos à segurança cidadã, atuação policial ou justiça criminal. Também é possível reforçar a discriminação mediante sistemas ou algoritmos imprecisos em entrevistas de emprego, decisões da administração pública sobre assistência social, ou oferta de crédito por instituições financeiras, por exemplo. Não por outra razão, o considerando 71 do RGPD, ao explicar as limitações impostas pelo regulamento sobre a utilização de decisões automatizadas e elaboração de perfis, esclarece que:

> (...) o responsável pelo tratamento deverá utilizar procedimentos matemáticos e estatísticos adequados à definição de perfis, aplicar medidas técnicas e organizativas que garantam designadamente que os fatores que introduzem imprecisões nos dados pessoais são corrigidos e que o risco de erros é minimizado, e proteger os dados pessoais de modo a que sejam tidos em conta os potenciais riscos para os interesses e direitos do titular dos dados, e evitar, por exemplo, efeitos discriminatórios contra pessoas singulares em razão da sua origem racial ou étnica, opinião política, religião ou convicções, filiação sindical, estado genético ou de saúde ou orientação sexual, ou que o tratamento dos dados resulte em medidas que venham a ter tais efeitos.

59 "De fato, o uso de perfis raciais pelos agentes da ordem segue sendo um problema persistente e generalizados, que repercute negativamente na relação entre a polícia e a comunidade, e no usufruto dos direitos das pessoas afetadas". Tradução livre dos autores. O original lê: "De hecho, el uso de perfiles raciales por los agentes del orden sigue siendo un problema persistente y generalizado, que repercute negativamente en la relación entre la policía y la comunidad y en el disfrute de los derechos de las personas afectadas." Informe do Relator Especial sobre as formas contemporâneas de racismo, discriminação racial, xenofobia e formas conexas de intolerância, Mutuma Ruteere, sobre sua visita à Espanha. Assembléia Geral, Nações Unidas. 2013. https://documents-dds-ny.un.org

Doi: 10.14210/nej.v26n1.p103-135

Igualmente, a Diretiva UE 2016/680 não apenas proíbe a utilização de decisões totalmente automatizadas, inclusive a elaboração de perfis em geral[60], senão que proíbe especificamente a elaboração de perfis que "conduzam à discriminação de pessoas singulares com base nas categorias especiais de dados pessoais"[61]. No entanto, ainda que tais esforços regulatórios em limitar a utilização indiscriminada de decisões automatizadas sejam necessários, tais instrumentos legais têm suas próprias limitações. Tanto o RGPD quanto a Diretiva UE 2016/680, por sua natureza, se aplicam somente às instâncias nas quais a decisão se baseia em dados pessoais, ou seja, dados observáveis, informados pelo próprio sujeito no momento de sua coleta, e que permitam sua identificação inequívoca. Isto exclui as decisões automatizadas baseadas em Big Data analytics, cuja base em geral não se enquadra na definição de "dados pessoais" de tais marcos legais[62]. Como corretamente nos recordam Valls Prieto e Gómez Romero, o processamento de bases de dados em grande escala por largos períodos, e a possibilidade de análise de dados de diferentes origens e fontes, ao combinar-se, podem conduzir a inferências sobre a personalidade do sujeito, suas ideias políticas e/ou religiosas, e até mesmo sobre sua identificação individual, sem que haja sido necessário o processamento de dados pessoais na forma prevista pelo RGPD e a Diretiva UE 2016/680[63]. Assim, os algoritmos que utilizam dados de grupos e pessoas que, no entanto, não são identificáveis individualmente, não estão sujeitos a tais limitações legais. Em outras palavras: tanto o RGPD quanto a Diretiva UE 2016/680 proveem uma proteção ao indivíduo, mas são potencialmente ineficazes na proteção a grupos ou coletivos[64].

Desta forma, aqueles sistemas que agrupam dados estatísticos por geografia ou idade (como, por exemplo, um sistema que agrupe residentes de um determinado bairro ou distrito, e assinale uma determinada zona geográfica como de alto potencial delitivo), podem afetar diretamente a um indivíduo específico (sujeitá-lo a revistas mais frequentes, por exemplo), sem que o indivíduo sequer saiba que está sendo assim classificado. Além disso, os algoritmos de Big Data analytics costumam identificar padrões que uma análise humana não conseguiria encontrar, gerando muitas vezes inferências e decisões contraintuitivas que são extremamente difíceis de explicar e, por isso mesmo, de contestar. Com frequência, tais decisões não são completamente explicáveis nem mesmo para os que desenham, controlam ou operam o sistema, seja porque a complexidade do próprio sistema o impede (algorithm black box), seja porque os sistemas de IA que incorporam o processo conhecido

60 Diretiva UE 2016/680 Art. 11 (1) e (3). Além disso, o Considerando 38 esclarece que: "A definição de perfis que conduza a discriminação contra pessoas singulares com base em dados pessoais que sejam, pela sua natureza, especialmente sensíveis do ponto de vista dos direitos e liberdades fundamentais, deverá ser proibida (…)".

61 São categorias especiais de dados os "dados pessoais que revelem a origem racial ou étnica, as opiniões políticas, as convicções religiosas ou filosóficas, a filiação sindical, o tratamento de dados genéticos, dados biométricos destinados a identificar uma pessoa singular de forma inequívoca, dados relativos à saúde ou dados relativos à vida sexual ou à orientação sexual (...)." Diretiva UE 2016/680, Art. 10.

62 WACHTER, S.; MITTELSTADT, B. A right to reasonable inferences. Re-thinking data protection law in the age of big data and AI. **Columbia Business Law Review**. Vol. 2019, n. 2. 2019. Disponível em: https://papers.ssrn.com

63 VALLS PRIETO, Javier; GÓMEZ-ROMERO, Juan. "**Use of big-data and the prediction of organized crime". Building a European Digital Space**. Huygens Editoria. 2016.

64 Ver DREYER, Stephan; SCHULZ, Wolfgang. The GDPR and algorithm decision-making. Safeguarding individual rights, but forgetting society. **Völkerrechtsblog**, 3 June 2019, doi: 10.17176/20190603-235442-0. Disponível em: https://voelkerrechtsblog.org.

como machine learning podem, por meio de sua própria observação do ambiente a ação sobre o mesmo, executar alterações e melhorias em si mesmos sem a necessidade de ação ou aprovação prévia por um humano. Tais alterações, por vezes, tampouco são previsíveis ou explicáveis nem mesmo pelos profissionais envolvidos em seu desenho e operação. O próprio Parlamento Europeu, em resolução de 12 de fevereiro de 2020 na qual trata especificamente de processos automatizados de tomada de decisões "reconhece que a emergência de produtos com capacidade de decisão automatizada coloca novos desafios, uma vez que esses produtos podem evoluir e agir de forma imprevista aquando da sua primeira colocação no mercado" [65]. Por tanto, a incapacidade técnica de explicar o funcionamento intrínseco do próprio algoritmo pode limitar significativamente o direito do sujeito de questionar de maneira eficaz as decisões (profiling) automatizadas e, em consequência, dificultar a já complicada tarefa de determinar uma possível discriminação na atuação policial.

De todo modo, ainda que fosse possível obter a explicação de uma decisão, perfil ou inferência, tal explicação não impede a decisão equivocada ou discriminatória, por ser um remédio ex post: logicamente, a explicação só pode existir posteriormente à tomada de decisão, como, por exemplo, do perfilamento do sujeito. Assim, a possível parcialidade é intrínseca ao processamento dos dados e à própria decisão e, a posteriori, qualquer contestação, oposição ou a própria obtenção de uma explicação dependerá necessariamente de uma objeção ativa por parte do sujeito afetado. Tendo em vista que em geral se tratam de coletivos vulneráveis ou pessoas em situação prévia de exclusão social, nem sempre será possível encontrar capacidade de (ou interesse em) objetar, ainda mais em se tratando de relações com forças policiais e de segurança. Até agora, nos referimos aos problemas criados por algoritmos que derivam inferências partindo de bases de dados parciais e que contém incorreções. Contudo, é possível esperar que, uma vez que o procedimento de obtenção e coleta de dados seja corrigido em sua origem, a aplicação de um sistema de IA em policiamento preditivo torne a atuação policial mais eficaz.

Um segundo problema identificado no desenho de algoritmos de predição em sistemas de segurança cidadã se refere ao risco da implantação de sistemas de IA que, ainda que utilizem dados "limpos", apresentam inconsistências em sua própria permissa. Como exemplo, fazemos referência ao sistema iBorderCtrl, atualmente em teste na União Europeia. A ideia básica do iBorderCtrl é identificar uma possível fraude por meio de sistemas de "affect recognition" [66].

65 Resolução do parlamento Europeu, de 12 de fevereiro de 2019, sobre processos automatizados de tomada de decisões: assegurar a proteção dos consumidores e a livre circulação de bens e serviços. https://www.europarl.europa.eu

66 A versão anterior deste artigo incluía a seguinte descrição do sistema, que constava do site do iBorderCtrl e que foi subsequentemente tirado do ar: "O Sistema Automatizado de Detecção de Fraude (Automatic Deception Detection System - ADDS) efetua, controla e avalia a entrevista de pré-registro através do sequenciamento de uma série de questões feitas ao viajante por um Avatar. O ADDS quantifica a probabilidade de mentira nas entrevistas por meio da análise de microgestos não-verbais do entrevistado. Isto, combinado com o avatar, move essa nova abordagem para a detecção de fraude na fase de pré-registro, resultando na aplicação do programa sem impactar no tempo gasto no cruze fronteiriço". No entanto, uma nova descrição do sistema pode ser encontrada no site www.cordis.europa.eu e, embora não faça referência explícita à "análise de microgestos não-verbais" o novo texto explica que a pesquisa "combinou em um único sistema tecnologias state of the art (de verificação biométrica a detecção de fraude, autenticação de documentos e análise de riscos)". Tradução livre dos autores. O original lê: "This was researched combining in one system state of the art technologies (from biometric verification to deception detection, document authentication and risk assessment)."

Doi: 10.14210/nej.v26n1.p103-135

O denominado affect recognition é definido como "uma tecnologia de IA que alega ser capaz de detectar o estado emocional de um indivíduo com base na utilização de algoritmos de visão computadorizada para analisar microexpressões faciais, tom de voz ou até mesmo seu modo de andar"[67]. Pode parecer simples e eficiente, exceto pelo fato de que este tipo de análise carece de base científica consolidada. Pesquisas independentes de sistemas que utilizam affect recognition detectaram erros significativos na identificação de sinais de agressão em áudios, vieses raciais na identificação de emoções por meio de reconhecimento facial[68], além de problemas básicos tais como confundir expressão com emoção[69]. A detecção de emoções e, especificamente no caso do iBorderCtrl, da fraude, depende de algo mais que uma análise superficial do tom de voz ou de micro expressões faciais. Estudo da universidade de Berkeley[70] mostra que o affect recognition não funciona no vácuo, senão que requer um contexto para obter resultados confiáveis.

De fato, em julho de 2019 um jornalista do The Intercept testou o iBorderCtrl[71] e gerou imediatamente um falso positivo, ou seja, o sistema sinalizou a presença de fraude mesmo que o jornalista tenha respondido honestamente a todas as perguntas. Por estar em período de teste, a avaliação do sistema não foi utilizada no controle fronteiriço, e o jornalista pôde cruzar a fronteira sem maiores dificuldades. No entanto, tendo em vista que um sistema desta natureza seria empregado oficialmente em grande escala, a geração de falsos positivos que resultem na recusa da entrada de um visitante pode desencadear situações de enorme gravidade e, às vezes, de difícil reversão (como, por exemplo, uma deportação). Portanto, considerando os cenários e desafios tecnológicos acima descritos, nos parece que são insuficientes os recursos legais hoje disponíveis ao cidadão para garantir um tratamento legítimo por parte das autoridades de segurança quando da utilização de sistemas de IA (ou seja, os mecanismos descritos no RGPD, na Diretiva UE 2016/680, e os princípios gerais de direitos humanos e direito à não-discriminação).

Em realidade, é insuficiente a existência de remédios exclusivamente ex post facto em situações como, por exemplo, uma deportação equivocada ou uma revista injustificada que tenham sido motivadas pela recomendação de um algoritmo. Além de requerer um questionamento ativo do sujeito, um conhecimento razoável de seus direitos (em especial daqueles relativos à privacidade de dados), e a ativação dos necessários trâmites burocráticos, a solução ex post dificilmente reparará o dano já causado.

---

67 Tradução livre dos autores. O original lê: "An AI-driven technology that claims to be able to detect an individual's emotional state based on the use of computer-vision algorithms to analyze their facial microexpressions, tone of voice, or even their gait". Ver CRAWFORD, K.; DOBBE, R.; BRYER, T.; FRIED, G.; GREEN, B.; KAZIUNAS, E.; WHITTAKER, M. **AI Now 2019 Report**. New York: AI Now Institute, 2019.

68 RHUE, L. **Racial Influence on Automated Detection of Emotions**. November 9, 2018. Disponible en SSRN: https://ssrn.com or http://dx.doi.org

69 BARRETT, Lisa F.; ADOLPHS, R.; MARSELLA, S.; MARTINEZ, Aleix M.; POLLAK, Seth D. **Emotional Expressions Reconsidered:** Challenges to inferring emotion from human facial movements. Psychological Science in the. 2019.

70 CHEN, Z.; WHITNEY, D.; **Tracking the Affective State of Unseen Persons**. April 9, 2019. Disponível em: https://www.pnas.org

71 Ver: https://theintercept.com

Na hipótese de que o sujeito afetado tente questionar uma decisão equivocada de um sistema de IA, os meios disponíveis parecem frágeis. Assim, no que tange ao tratamento de dados pelas forças e corpos de segurança no âmbito de investigações policiais, a Diretiva EU 2016/680 parece reconhecer em seus artigos 11 a 18, alguns direitos do investigado, entre os quais se destacam principalmente: a proibição da elaboração de perfis que deem razão à discriminação baseando-se em categorias de dados especiais (art. 11 (3)), o direito à informação (arts. 12 y 13), e o direito de retificação ou supressão (art. 16). Não obstante, observamos que a proteção do art. 11(3) muito provavelmente não seria aplicável aos exemplos práticos de polícia preditiva aqui discutidos. Ao contrário, como explicamos anteriormente, os sistemas de IA que utilizam técnicas de Big Data analytics tratam dados que, em grande parte, não podem ser tecnicamente qualificados como "dados pessoais" segundo o conceito da própria Diretiva e, por tanto, não estão sujeitos à regulação por este marco legal. Mais que isso, de modo contraditório, é justamente por não utilizar dados pessoais e, portanto, não levar em conta o caráter potencialmente discriminatório dos dados analisados em sua origem, que tais sistemas de IA podem levar à elaboração de perfis enviesados, sem proteções legais claras e efetivas ao cidadão. Neste contexto, se a ideia preconizada para o desenvolvimento e aplicação de sistemas de IA é tornar a administração pública e os sistemas de polícia e justiça criminal mais eficientes, nos parece extremamente contraproducente a utilização de tecnologias que, ao final, reproduzam exatamente os erros dos modelos tradicionais, quando não os intensificam.

Contudo, a questão de fundo é antiga. A discriminação racial na atividade policial não surgiu com a implantação de sistemas de IA. Cabe recordar o caso de Rosalind Williams, cidadã espanhola cuja parada de identificação baseada em critérios exclusivamente étnicos ocorreu em 1992 e que, após ter visto seu processo indeferido em todas as instâncias judiciais espanholas, finalmente conseguiu, em 2009, o reconhecimento da causa pelo Comitê de Direitos Humanos das Nações Unidas, que entendeu que características físicas ou étnicas não podem ser consideradas como um indício de residência ilegal[72]. No entanto, até hoje o Estado espanhol ainda não cumpriu com a determinação de tal comitê de proporcionar uma solução efetiva para o caso e formalizar um pedido

72 Ver https://www.justiceinitiative.org. Ver também, Sentença do Tribunal Constitucional Espanhol núm. 13/2001, de 29 de janeiro. Disponível online: http://hj.tribunalconstitucional.es. Em Rosalind Williams Lecraft v. Spain, o Comitê de Direitos Humanos das Nações Unidas expõe o problema intrínseco da discriminação racial em paradas de identificação e seus efeitos prejudiciais: "O Comitê considera que checagens de identificação feitas por razões de segurança pública ou prevenção de crime em geral, ou para controle de imigração ilegal, serve a um propósito legítimo. Porém, quando as autoridades conduzem tais checagens, as características físicas ou étnicas das pessoas sujeitas à tais checagens não deveriam por si só ser consideradas indicativos de uma possível presença ilegal no país. Nem deveriam ser conduzidas de maneira a selecionar apenas pessoas com uma característica física ou étnica específica. Agir de outro modo não apenas afeta negativamente a dignidade das pessoas envolvidas, mas também contribui para a propagação de atitudes xenófobas entre a população em geral e vai de encontro a uma política efetiva de combate à discriminação racial." Tradução livre dos autores. O original lê: "The Committee considers that identity checks carried out for public security or crime prevention purposes in general, or to control illegal immigration, serve a legitimate purpose. However, when the authorities carry out such checks, the physical or ethnic characteristics of the persons subjected thereto should not by themselves be deemed indicative of their possible illegal presence in the country. Nor should they be carried out in such a way as to target only persons with specific physical or ethnic characteristics. To act otherwise would not only negatively affect the dignity of the persons concerned, but would also contribute to the spread of xenophobic attitudes in the public at large and would run counter to an effective policy aimed at combating racial discrimination." Rosalind Williams Lecraft v. Spain, UN Human Rights Committee, CCPR/C/96/D/1493/2006 (2009), 7.2. http://www.worldcourts.com

Doi: 10.14210/nej.v26n1.p103-135

oficial de desculpas à demandante. Como indicamos ao nos referirmos ao marco de soft law, o que se espera dos sistemas de IA é que trabalhem com padrões éticos e respeitem os valores humanos de forma que possam auxiliar-nos, como seres humanos ao mesmo tempo agentes e sujeitos da lei, a tomar decisões e atuar de maneira mais eficaz, justa e imparcial. Um sistema de IA que não alcance estas metas não será mais que simples terceirização dos erros e disfunções existentes no modelo tradicional. Em outras palavras, ou somos capazes de desenvolver sistemas que sejam melhores que nós, ou não conseguiremos tirar real proveito dos prometidos avanços tecnológicos.

CONSIDERAÇÕES FINAIS

Não pretendemos nestas linhas negar que os sistemas de IA podem ser capazes de, eventualmente, tornar-se ferramentas efetivas e eficientes para os fins de prevenção do delito. Ao contrário, uma vez que estes sistemas de IA com grande capacidade de processamento e análise de dados sejam alimentados e treinados com dados que reproduzam corretamente a realidade e se encontrem livres de vieses preconceituosos inerentes ao comportamento humano, sim podemos esperar que os efeitos benéficos superem os problemas. À luz das reflexões anteriores, entendemos que talvez seja precipitada a aplicação massiva de sistemas de IA a modelos de polícia preditiva, enquanto (i) estejam pendentes de refinação; (ii) não tenham uma base científica sólida, e; (iii) operem em um ambiente onde o marco legal, regulatório e de proteção cidadã é ainda limitado. Principalmente quando estes sistemas tem o potencial de afetar de maneira importante a vida de grupos e populações em situação de exclusão e que se veem cada dia mais vigiados. Ademais, além das leis e normas eventualmente vigentes, o que proporcionará legitimidade à atuação policial baseada em IA e decisões automatizadas será, em particular, (i) a coleta e o tratamento cuidadoso das bases de dados utilizadas e a transparência no desenvolvimento dos sistemas algorítmicos, (ii) o treinamento extensivo dos usuários destes sistemas e, (iii) avaliações periódicas e testes independentes.

No que tange à coleta de dados, e conforme já sinalizamos, os dados "crus" ("raw data") nunca são verdadeiramente "crus" [73]: dados são, em geral, um reflexo da sociedade e, como tal, refletem também seus preconceitos. Desta maneira, se a intenção é criar um sistema que seja mais preciso e eficiente que os métodos tradicionais, então precisamos inventar algum modo pelo qual a própria

73 Tecnicamente, denomina-se "raw data" ("dados crus") os dados coletados, mas ainda não processados. O termo pode conduzir a entender que o dado não-processado, o dado em seu estado "puro", equivale necessariamente a um dado neutro, vazio de juízo ou da interferência humana, e que seria correspondente à "verdade". Esta ideia pode, no entanto, revelar-se enganosa já que o preconceito será inerente ao próprio dado que reflete o comportamento humano (como o é grande parte dos dados utilizados nos sistemas de polícia preditiva), sem necessidade do seu posterior processamento por um sistema computacional. Como nos explica Nick Barrowman: "A forma como dados são entendidos, armazenados e coletados é o resultado de decisões humanas – decisões sobre o quê, exatamente, medir, como e quando fazê-lo, e por quais métodos. Inevitavelmente, o que é medido e armazenado tem impacto nas conclusões às quais chegamos." Tradução livre dos autores. O original lê: "How data are construed, recorded, and collected is the result of human decisions — decisions about what exactly to measure, when and where to do so, and by what methods. Inevitably, what gets measured and recorded has an impact on the conclusions that are drawn." Barrowman, N. Why Data is Never Raw. The New Atlantis. 2018. https://www.thenewatlantis.com Ver também a interessante entrevista de Catherine D'Ignazio para o The Guardian: https://www.theguardian.com

coleta de dados, e a análise deles derivada, nos ajudem a alcançar uma sociedade melhor e não somente replicar ou reforçar seus erros.

O mesmo cuidado deveria ser aplicado à premissa científica utilizada: o desenvolvimento de sistemas de IA deve partir de bases científicas consolidadas e transparentes, de maneira a evitar a precipitação na implantação de sistemas que, ainda que possam funcionar tecnicamente, não trazem resultados que correspondam à realidade. Em outras palavras: ainda que seja possível desenvolver um programa de análises de micro expressões faciais, por exemplo, não necessariamente tal programa deve ser desenvolvido e implementado sem que haja comprovação suficiente de que a premissa científica é válida. Menos ainda em ambientes e situações que podem afetar de maneira significativa os direitos fundamentais do indivíduo[74].

Além disso, ainda que um sistema de predição policial possa recomendar uma determinada estratégia, em última instância deve ser a autoridade policial a tomar a decisão de maneira definitiva sobre a efetiva atuação.

Portanto, se o desenvolvimento tecnológico aplicado no âmbito da segurança cidadã não deve prescindir da agência humana, o treinamento extensivo da força policial em matéria de tecnologia se torna condição sine qua non para sua implantação. Será necessário que os agentes saibam interpretar as recomendações derivadas dos sistemas de IA e entendam minimamente como funciona o sistema que utilizam, já que o desconhecimento poderá facilmente resultar em um corpo policial que não faz mais que cumprir ordens de um sistema computacional de maneira autômata, quando deveria fazê-lo de maneira autônoma. Mais ainda, o controle social efetuado pela sociedade civil sobre a atuação policial deverá, de agora em diante, estender-se também aos sistemas de IA que sejam implantados. Como apontamos, no caso de alguns países tais como Espanha, por exemplo, isto demandará uma mudança importante no próprio sistema de controle social da atuação policial como um todo: impõe-se com urgência desde uma revisão detalhada dos dados estatísticos compilados por fontes oficiais, até um incentivo à investigação aprofundada sobre a atividade policial, tanto pelos órgãos estatais como por agentes independentes.

O simples propósito de reduzir ou eliminar vieses discriminatórios e ilegais deveria ser suficiente para justificar uma revisão aprofundada da atuação policial e o desenvolvimento de um marco legal adequado às novas tecnologias. Junto a isto, sem dúvida, o objetivo de dotar o sistema de justiça criminal de maior legitimidade seria um motivo adicional relevante para avançar em tal propósito: um sistema que não é capaz de rastrear, identificar e corrigir práticas discriminatórias e ilegais dentro de suas próprias forças de segurança corre o risco de colocar em xeque a sua própria

74 Na data de fechamento deste artigo foi publicada a notícia no jornal La Vanguardia (11 de junho de 2020) relativamente ao rechaço (veremos se real ou estético) de algumas corporações aos sistemas de reconhecimento facial: ver “IBM y Amazon ajuran de la tecnologia de reconocimiento facial por su sesgo racista”. https://www.lavanguardia.com.

Doi: 10.14210/nej.v26n1.p103-135

legitimidade, e tal legitimidade é a base fundamental sobre a qual a cidadania se dispõe a aceitar o monopólio do uso da força por parte do Estado. Afinal, se a polícia (uma das instituições públicas mais visíveis encarregadas da aplicação da lei e do poder estatal) utiliza seu poder e autoridade de forma abusiva, este abuso de poder pode não apenas prejudicar a aceitação cidadã da aplicação da lei e da necessidade de observar as normas impostas, senão também levar a um questionamento sobre a própria autoridade moral do corpo policial[75].

Além disso, há evidências de que as atuações policiais guiadas por critérios preconceituosos, e não por dados objetivos, tendem a ser menos eficazes na prevenção e detecção de delitos[76]. Apesar de suas limitações, o império da lei e o acesso democrático e amplo à informação e educação são, até o momento, as melhores ferramentas encontradas pela sociedade para tentar compensar o permanente desequilíbrio entre os amplos poderes estatais e os poderes limitados dos cidadãos[77].

O que sugerimos é, portanto, a adoção simultânea de duas estratégias para atingir este objetivo: por um lado, a criação de leis e regulamentos efetivos para a proteção dos cidadãos perante o desenvolvimento tecnológico no âmbito da justiça criminal; e por outro lado, a criação de programas educativos e de treinamento, tanto para a sociedade quanto para as administrações públicas e forças de segurança, vistas como um todo.

Em primeiro lugar, a lei deve exigir um alto grau de transparência na comissão e aquisição de programas de IA por órgãos públicos estatais, em geral, e pelas polícias e forças de segurança em particular. Aqueles sistemas que possam afetar de maneira importante os direitos fundamentais do indivíduo (tais como costumam ser os sistemas relacionados à atividade policial, à justiça criminal ou migrações), devem por esta razão mesma ser objeto de escrutínio permanente e detalhado tanto por parte do Estado quanto pela sociedade em geral. A identidade dos contratantes, os termos e condições contratuais, o tipo de sistemas utilizados e os detalhes operacionais do sistema são apenas algumas das informações que deveriam ser postas à disposição para uma análise adequada por organismos independentes e pelo público em geral, tanto antes quanto depois de sua comissão e aquisição.

---

75 HOUGH, Mike; JACKSON, Jonathan; BRADFORD, Ben; MYHILL, Andy; QUINTON, Paul. Procedural justice, trust and institutional legitimacy. **Policing: a journal of policy and practice**, vol. 4, n. 3. pp. 203-210. 2010. DOI: 10.1093/police/paq027

76 WEGMAN, D.; PERNAS, B. **Perfil Racial en España:** investigaciones y recomendaciones. Grupo de Estudios y Alternativas 21, Open Society Justice Initiative, 2005..

77 "Há uma tensão inerente e inevitável entre forças policiais e de segurança pública, e direitos humanos. São necessários poderes adequados que permitam que as forças policiais/de segurança nacional cumpram suas funções; mas o exercício destes poderes irá necessariamente interferir com o direito ao respeito à privacidade e com o direito à proteção de dados pessoais e devem, portanto, ser proporcionais ao objetivo a ser atingido." Tradução livre dos autores. O original lê: "There is an inherent and inevitable tension between law enforcement and public security powers, and human rights. Adequate powers are necessary to allow the law enforcement/ national security agencies to fulfil their tasks; but exercise of such powers will necessarily interfere with the right of respect for private life as well as the right to protection of personal data, and must therefore be proportionate to the aim to be achieved." Caruana, Mireille M. "The reform of the EU data protection framework in the context of the police and criminal justice sector: harmonisation, scope, oversight and enforcement". International Review of Law, Computers & Technology. 2017. DOI: 10.1080/13600869.2017.1370224

Os sistemas de natureza similar aos discutidos neste artigo deverão, além disso, ser exaustivamente testados por diferentes partes interessadas e organizações independentes antes de ser implantados em grande escala. Uma vez implantados, deveria levar-se a cabo testes e análises detalhados de forma periódica. Este alto nível de transparência e escrutínio independente por diferentes grupos de interesse na fase de desenvolvimento do sistema será particularmente importante se queremos evitar a utilização de bases de dados enviesadas ou incompletas e a implantação de programas construídos sobre bases científicas frágeis ou incipientes.

Adicionalmente, tanto as empresas que trabalham na pesquisa e desenvolvimento de programas de IA, quanto órgãos estatais e agentes públicos envolvidos no programa, devem ser legalmente responsáveis por manter a transparência requerida por lei durante o desenho e implantação de todos os programas de IA que possam ser utilizados para determinar e/ou recomendar políticas públicas relacionadas ao sistema de justiça criminal.

Em segundo lugar, como detalhamos anteriormente, as análises de Big Data que não conduzam ao processamento de dados pessoais na forma definida por lei não necessariamente estão sujeitas às restrições impostas pelo RGPD e/ou marcos legais similares atualmente aplicáveis ao tratamento de dados pessoais. Por esta razão, uma vez que sistemas de IA aplicados por forças de segurança façam uso deste tipo de análise ampla de dados não-pessoais, toda a informação relacionada à coleta, utilização e armazenamento dos dados, além das categorias de dados tratados, as inferências e os perfis resultantes de tal análise, devem ser claras, amplamente divulgadas e postas à disposição para exame pela cidadania e organismos independentes. Não somente porque cada indivíduo deve ter o direito a saber se está sujeito a algum tipo de análise ou perfilamento de grupo, senão também porque é fundamental garantir que os dados que operam no sistema não são/estão enviesados.

Além disso, em virtude do constante aumento na geração massiva de dados (quase sempre voluntária), ainda não está claro se faz sentido, ou até mesmo se seria útil, impor algum tipo de restrição sobre o uso de Big Data (entendido como o tratamento massivo de dados não-pessoais) para os propósitos aqui discutidos. No entanto, a divulgação de informações claras e transparentes a respeito da utilização de tal tipo de análise será chave, se não essencial, para a aplicação da lei e a responsabilização legal das corporações e autoridades públicas perante uma eventual violação dos direitos do indivíduo.

Em terceiro lugar, a lei deve estabelecer requerimentos de treinamentos periódicos e substanciais para todos os agentes envolvidos na operação do sistema, em todos os níveis. Desde altos funcionários encarregados do desenvolvimento de políticas públicas, do planejamento e aquisição de novas tecnologias, até o agente na última ponta da atuação do sistema (seja um oficial dedicado à patrulha de vias ou um agente de fronteira), todos devem ser periodicamente treinados e sujeitos a testes e avaliações, de maneira a permitir a identificação de possíveis elementos débeis

Doi: 10.14210/nej.v26n1.p103-135

na cadeia de geração e análise dos dados e sua aplicação prática. Ao mesmo tempo, deve-se fazer revisões periódicas dos resultados da aplicação do sistema, tanto internamente pelo Estado quanto por organizações independentes, permitindo identificar possíveis feedback loops que reforcem discriminações e parcialidade, e que não correspondam à realidade da atividade criminal.

Finalmente, em quarto lugar, a sociedade necessita urgentemente de um sistema claro de responsabilidade legal e de mecanismos eficientes de aplicação da lei, no que diz respeito ao desenho, implantação e utilização de sistemas de IA. O que temos visto até o momento é que os legisladores (e as leis), ainda que bem intencionados, não propõem soluções eficazes para prevenir o uso malicioso da tecnologia em grande escala.

Na realidade, o marco legal existente e hoje aplicável também se vem mostrando pouco efetivo inclusive nos casos nos quais se há identificado una violação normativa. Desde os abusos por parte do Google ou Facebook que, ainda que tenham sido sancionados, não foram capazes de impedir a reincidência por parte destas empresas, até os problemas mais recentes na plataforma Zoom durante a pandemia do SARS-CoV-2, o que se tem visto é que as sanções, quando aplicadas, não solucionam o problema original, e gigabytes de dados privados dos cidadãos seguem sendo tratados de forma ilegal, tornando sua recuperação ou eliminação pela vítima virtualmente impossível. As sanções eventualmente impostas sobre estas corporações não produziram mais que um pequeno arranhão na fortaleza de sua envergadura financeira. Não é absurdo afirmar que as violações ao nosso direito fundamental à privacidade e à não discriminação vão seguir ocorrendo já que tais sanções são, na realidade, muito pouco dissuasórias.

Assim as coisas, os métodos à disposição da cidadania para proteger-se de tais abusos nos parecem, de momento, pouco claros, frágeis e de pouca utilidade prática. Além disso, neste entorno tecnológico, as medidas ex post tendem a ser consistentemente ineficazes em restaurar ao indivíduo seu status quo ante: uma vez que a privacidade tenha sido violada, que os dados pessoais tenham sido obtidos sem autorização ou consentimento apropriados e/ou que se tenha obtido um perfil ou inferência de forma enviesada ou ilegal, é improvável que tal informação possa ser eficazmente eliminada.

Finalmente, se por um lado, é verdade que as leis que regulam os direitos humanos e a não-discriminação seguem sendo aplicáveis, a ausência de regras claras a respeito da implantação e utilização de sistemas de IA que reforcem preconceitos pode levar à ineficácia destes preceitos legais gerais. Especialmente quando sabemos que o conjunto social mais afetado pela discriminação por parte das forças policiais são os coletivos mais vulneráveis, com pouco ou nenhum acesso ou recurso ao sistema judicial. Assim, os sistemas desenhados para limitar ou restringir ações da cidadania, como costumam ser os sistemas de justiça criminal, quando não estão sujeitos a controles efetivos, perdem rapidamente sua legitimidade. No entanto, quando são desenhados de maneira apropriada, sujeitos à supervisão e exame externos, e utilizados em um entorno de transparência e responsabilização legal clara, a IA poderia tornar-se um instrumento efetivo para o desenvolvimento de um sistema

de justiça que consiga ampliar os níveis de segurança, construir comunidade e, afinal, alcançar uma convivência social pacífica.

REFERÊNCIAS DAS FONTES CITADAS

AÑON, José G.; BRADFORD, B.; SÁEZ, José Antonio G.; CUENCA, Andrés G. & FERRERES, Antoni L. **Identificación Policial por Perfil Étnico en España**. Informe sobre experiencias y actitudes en relación con las actuaciones policiales. Tirant Lo Blanch. Valencia 2013. Disponível em: https://www.uv.es.

BARRETT, Lisa F.; ADOLPHS, R.; MARSELLA, S.; MARTINEZ, Aleix M.; POLLAK, Seth D. **Emotional Expressions Reconsidered:** Challenges to inferring emotion from human facial movements. Psychological Science in the. 2019.

BRUNDAGE, M.; AVIN, S.; CLARK, J.; TONER, H.; ECKERSLEY, P.; GARFINKEL, B.; AMODEI, D. **The Malicious Use of Artificial Intelligence:** Forecasting, Prevention, and Mitigation. February 2018. Disponível em: http://arxiv.org

CATE, Fred H.; KUNER, Christopher; MILLARD, Christopher; SVANTESSON, DAN JERKER, B. **"The Challenge of "Big Data" for Data Protection"**. 2012. Disponível em: http://www.repository.law.indiana.edu

CEA D'ANCONA, M.A.; VALLÉS MARTÍNEZ, M.S. **Evolución de la Discriminación en España**. Informe de las encuestas IMIO-CIS de 2013 y 2016. Instituto de la Mujer para la Igualdad de Oportunidades (IMIO). 2018. Disponível em: http://www.inmujer.gob.es

CHEN, Z.; WHITNEY, D.; **Tracking the Affective State of Unseen Persons**. April 9, 2019. Disponível em: https://www.pnas.org

COSTANZO, P.; D'ONOFRIO, F., & FRIEDL, J. Big data and the Italian legal framework: Opportunities for police forces. In AKHGAR, B.; SAATHOFF, G.; ARABNIA, H. R.; HILL, R.; STANIFORTH, A. & BAYERL, P.S. (Eds.), **Application of big data for national security.** (pp. 238-249). UK: Oxford, 2015.

CRAWFORD, K.; DOBBE, R.; BRYER, T.; FRIED, G.; GREEN, B.; KAZIUNAS, E.; WHITTAKER, M. **AI Now 2019 Report**. New York: AI Now Institute, 2019.

DOMÍNGUEZ FIGUEIRIDO, D.; BASANTA, A. Lógica actuarial, seguridad y sistema de justicia criminal. in **La seguridad en la sociedad del riesgo:** Un debate abierto. Barcelona: Atelier, 2003.

DREYER, Stephan; SCHULZ, Wolfgang. The GDPR and algorithm decision-making. Safeguarding individual rights, but forgetting society. **Völkerrechtsblog**, 3 June 2019, doi: 10.17176/20190603-235442-0. Disponível em: https://voelkerrechtsblog.org

DUPUY, Daniela. **Inteligencia aplicada al Derecho penal y proceso penal.** Em Cibercrimen II. BdeF, 2018. p. 281 y ss.

ENGSTROM, David F.; HO, Daniel E.; SHARKEY, Catherine M.; CUÉLLAR, Mariano-Florentino. **Government by Algorithm:** Artificial Intelligence in Federal Administrative Agencies. Fevereiro, 2020.

GARCÍA AMADO, Juan Antonio. Anatomía de un imposible. La imagen jurisprudencial del policía. En AGRA, C. da; Domínguez, J.L.; GARCÍA AMADO, J.A.; HEBBERECHT, P.; RECASENS, A. (eds.) **La seguridad en la sociedad del riesgo**. Madrid: Atelier, 2003, pp. 181-200.

HOUGH, Mike; JACKSON, Jonathan; BRADFORD, Ben; MYHILL, Andy; QUINTON, Paul. Procedural justice, trust and institutional legitimacy. **Policing: a journal of policy and practice**, vol. 4, n. 3. pp. 203-210. 2010. DOI: 10.1093/police/paq027

HRUSCHKA, J.; SÁNCHEZ-OSTIZ GUTIÉRREZ, P. **Imputación y derecho penal**: estudios sobre la teoría de la imputación. Thomson Aranzadi; Universidad de Navarra-Garrigues Cáterdra, 2005.

MAJEED, Abdul; ULLAH, Farman; LEE, Sungchang. **"Vulnerability- and Diversity-Aware Anonymization of**

Doi: 10.14210/nej.v26n1.p103-135

Personally Identifiable Information for Improving User Privacy and Utility of Publishing Data". Sensors 2017, vol. 17, 1059; doi:10.3390/s17051059. Disponível em: https://www.mdpi.com

Ministerio de Ciencia, Innovación y Universidades. **Estrategia Española de I+D+I en inteligencia artificial**. 2019. Disponível em: http://www.ciencia.gob.es

MIRÓ LLINARES, F. **Predictive Policing:** Utopia or Dystopia? On attitudes towards the use of Big Data algorithms for law enforcement. Revista de Internet, Derecho y Política, n. 30. 2020

MIRÓ, Fernando. Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots. **Revista de Derecho Penal y Criminología**, vol. (20, 87-130. 2018.

RAND Corporation. **Predictive Policing** – The Role of Crime Forecasting in Law Enforcement Operations. 2013. Disponível em: https://www.rand.org

RHUE, L. **Racial Influence on Automated Detection of Emotions**. November 9, 2018. Disponible en SSRN: https://ssrn.com or http://dx.doi.org

RICHARDSON, R.; SCHULTZ, Jason M.; & CRAWFORD, K. Dirty Data, Bad Predictions: How Civil Rights Violations impact police data, predictive policing systems, and justice. **New York University Law Review Online**. 2019

Second European Union Minorities and Discrimination Survey. **Being Black in the EU**. European Union Agency for Fundamental Rights. 2018. https://fra.europa.eu

COUGHLIN, Tom. "**175 Zettabytes By 2025", publicado em Forbes**. 2018. Disponível em: https://www.forbes.com

VALLS PRIETO, J. **Problemas jurídico penales asociados a las nuevas técnicas de prevención y persecución del crimen mediante inteligencia artificial**. Madrid: Dykinson, 2017.

VALLS PRIETO, Javier; GÓMEZ-ROMERO, Juan. "**Use of big-data and the prediction of organized crime". Building a European Digital Space**. Huygens Editoria. 2016.

VÉASE, Meijer; A. WESSELS, M. Predictive Policing: Review of Benefits and Drawbacks. **International Journal of Public Administration**, Vol. 42, no. 12, 1031-1039. 2019.

WACHTER, S.; MITTELSTADT, B. A right to reasonable inferences. Re-thinking data protection law in the age of big data and AI. **Columbia Business Law Review**. Vol. 2019, n. 2. 2019. Disponível em: https://papers.ssrn.com

WEGMAN, D.; PERNAS, B. **Perfil Racial en España:** investigaciones y recomendaciones. Grupo de Estudios y Alternativas 21, Open Society Justice Initiative, 2005.

Recebido em:14/10/2020

Aprovado em: 30/11/2020