Vol. 92 issue 1, 2021

RIDP

Gert Vermeulen, Nina Peršak & Nicola Recchia *(Eds.)*

Artificial Intelligence, Big Data and Automated Decision-Making in Criminal Justice

Revue Internationale de Droit Pénal
International Review of Penal Law
Revista internacional de Derecho Penal
Международное обозрение уголовного права
刑事法律国际评论
المجلة الدولية للقانون الجنائي
Revista Internacional de Direito Penal
Rivista internazionale di diritto penale
Internationale Revue für Strafrecht

MAKLU

AIDP – Association Internationale de Droit Pénal | The International Association of Penal Law is the oldest association of specialists in penal law in the world. Since 1924, it is dedicated to the scientific study of criminal law and covers: (1) criminal policy and codification of penal law, (2) comparative criminal law, (3) international criminal law (incl. specialization in international criminal justice) and (4) human rights in the administration of criminal justice. The Association's website provides further information (http://www.penal.org).

RIDP – Revue Internationale de Droit Pénal | The International Review of Penal Law is the primary publication medium and core scientific output of the Association. It seeks to contribute to the development of ideas, knowledge, and practices in the field of penal sciences. Combining international and comparative perspectives, the RIDP covers criminal law theory and philosophy, general principles of criminal law, special criminal law, criminal procedure, and international criminal law. The RIDP is published twice a year. Typically, issues are linked to the Association's core scientific activities, ie the AIDP conferences, Young Penalist conferences, world conferences or, every five years, the International Congress of Penal Law. Occasionally, issues will be dedicated to a single, topical scientific theme, validated by the Scientific Committee of the Association, comprising high-quality papers which have been either presented and discussed in small-scale expert colloquia or selected following an open call for papers. The RIDP is published in English only.

Peer review: All contributions are subject to double-layered peer review. The primary scientific and peer review responsibility for all issues lies with the designated Scientific Editor(s). The additional scientific quality control is carried out by the Executive Committee of the Editorial Board, which may turn to the Committee of Reviewers for supplementary peer review.

Disclaimer: The statements and opinions made in the RIDP contributions are solely those of the respective authors and not of the Association or MAKLU Publishers. Neither of them accepts legal responsibility or liability for any errors or omissions in the contributions nor makes any representation, express or implied, with respect to the accuracy of the material.

ISSN – 0223-5404
ISBN 978-90-466-1130-2
D/2021/1997/46
NUR 824
BISAC LAW026000

Maklu- Publishers
Somersstraat 13/15, 2018 Antwerpen, Belgium, info@maklu.be
Koninginnelaan 96, 7315 EB Apeldoorn, The Netherlands, info@maklu.nl
www.maklu.eu

USA & Canada
International Specialized Book Services
920 NE 58th Ave., Suite 300, Portland, OR 97213-3786, orders@isbs.com, www.isbs.com

Editorial Board

Summary

LOMBROSO 2.0: ON AI AND PREDICTIONS OF DANGEROUSNESS IN CRIMINAL JUSTICE

By Alice Giannini[*]

Abstract

The purpose of this paper is to analyze the legal and ethical issues raised by the use of artificial intelligence (AI) technologies in predicting criminal behavior. In fact, ever since Cesare Lombroso's L'uomo delinquente, scientists, on one side, and jurists, on the other, have been discussing the 'criminal brain'. Violence risk assessment tools have been applied in criminal courts for more than sixty years, yet the discussion has found once again importance thanks to the development of new AI techniques (such as machine learning) and to their application in both the medical and the criminal justice area. The new frontier is represented by the enhancement of these instruments through the combination of AI and neuropredictions. This paper presents critical reflections on the benefits and drawbacks of applying these technologies to predictions of violence in criminal justice. The inquiry concludes with a number of open questions which are hoped will contribute to the ongoing debate and work as a primer for future investigations on the matter.

1 Introduction

Dangerousness is a concept embedded in possibly every modern criminal legal system. Ever since Cesare Lombroso's *L'uomo delinquente*,[1] scientists and jurists have been discussing the 'criminal brain'. Classical questions of such debate include whether it is possible to scientifically map biological characteristics that lead to the commission of a crime, or if 'criminal science' can be used to identify in advance subjects which will incur in criminal behavior. In essence, we could claim that Lombrosianism boils down to one existential question: are we born criminals?

In point of fact, the assessment of violent behavior has been for long now the 'chief battlefield in the struggle between law and psychiatry'.[2] Accordingly, some argue that the most recent advances in behavioral genetics, neuroscience, psychiatry and criminological epidemiology, together with the emergence of neurocriminology, characterize the return

[*] PhD Student in Criminal Law, University of Florence and University of Maastricht (Double PhD Program). In 2021 her research project was awarded with the Giulia Cavallone price. For correspondence: <alice.giannini@unifi.it>.

[1] The book was first published in Italian in 1876 and then in English in 1911, after Lombroso's death, with the title 'Criminal Man'.

[2] Christopher D Webster, Mark H Ben-Aron and Stephen J Hucker, Dangerousness (Cambridge University Press 1987) 14.

of a Lombrosian vision of crime.[3] Better yet, a 'Lombroso 2.0'.[4] Conversely, more and more academics joined a newborn field of investigation named 'neurolaw', addressing the impact of neuroscience on the foundations of criminal responsibility.[5] This trend is being followed by the development of new artificial intelligence (AI) systems[6] and by their application in both the medical and the criminal justice area. The potential of using these technologies is enormous: they are capable of analyzing massive quantities of data at a very high speed, sometimes with little or no human supervision. AI technology is being combined with neuroscience to address future dangerousness and, consequently, never-before-seen correlations between violent behavior and a person's characteristics might be identified.

This paper will be structured as follows. First, it will briefly account for the relationship between science and criminal law, focusing specifically on the concept of dangerousness. Then, the investigation will touch upon algorithmic risk-assessment and AI neuroimaging. They represent two sides of the same medal: on one side, risk-assessment tools have been used in criminal justice since the 1930s and are seeing new potential today thanks to AI powered data collection and analysis. On the other, forensic neuroscience has been

[3] Christian Munthe and Susanna Radovic, 'The Return of Lombroso? Ethical Aspects of (Visions of) Preventive Forensic Screening' (2015) 8 Public Health Ethics 271.
Think for example of the experiment conducted at Cornell University in 2011, where a group of psychologists proved that individuals are capable of making rather accurate inferences about 'criminality' based on someone's facial appearance (ie on a static cropped image of a face); see Jeffrey M. Valla, Stephen J. Ceci and Wendy M. Williams, 'The Accuracy Of Inferences About Criminality Based on Facial Appearance.' (2011) 5 Journal of Social, Evolutionary, and Cultural Psychology. A few years later two researchers from the Shanghai Jiao Tong University developed an AI system which was able to recognize 'criminals' with an accuracy of 89.5%, based on the curvature of the upper lip, the distance between the inner corners of the eyes and the so-called nose-mouth angle. The publication of this study sparked a heated debate which were addressed by the researchers with a subsequent addendum; see Xiaolin Wu and Xi Zhang, 'Automated Inference on Criminality Using Face Images' [2017] arXiv:1611.04135. Everything considered, as it has been stated, '[i]f humans can spot criminals by looking at their faces, as psychologists found in 2011, it should come as no surprise that machines can do it, too', see 'Neural Network Learns to Identify Criminals by Their Faces' (MIT Technology Review, 2021) <https://www.technologyreview.com/2016/11/22/107128/neural-network-learns-to-identify-criminals-by-their-faces/> accessed 2 August 2021.

[4] The term has appeared in a small number of recently published articles, but it has not (yet) been adopted in the legal academic discourse. See Simone Cosimi, 'Gay o Etero, un algoritmo "legge" l'orientamento sessuale sul volto. Il controverso studio di Stanford' (*la Repubblica*, 2021) <https://www.repubblica.it/tecnologia/2017/09/08/news>; Dario Ronzoni, 'Lombroso 2.0: Una Rete Neurale Per Riconoscere I Criminali Dai Tratti Somatici' (DDay.it, 2021) <https://www.dday.it/redazione/21681/lombroso-20-una-rete-neurale-per-riconoscere-i-criminali-dai-tratti-somatici> accessed 9 August 2021.

[5] See Stephen Morse, 'Neuroethics: Neurolaw', *Oxford Handbooks Online* (2017); Ariane Bigenwald and Valerian Chambon, 'Criminal Responsibility and Neuroscience: No Revolution Yet' (2019) 10 Frontiers in Psychology.

[6] For the purpose of this article, we will adopt the following definition of AI systems: 'An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy". See OECD, Recommendation OECD/LEGAL/0449 of 22 May 2019 R of the Council on Artificial Intelligence (2019).

depicted as a useful tool to improve the reliability of predictions of dangerousness, thanks to the application of AI to brain scanning and diagnosing. Such analysis will serve as a primer for the following sections: in a (not very) dystopian future AI neuroimaging and brain reading technologies could be combined with risk assessment tools in a technique called 'AI neuroprediction'.[7] Could this also lead to totally automated evaluations of dangerousness and the creation of virtual forensic experts?

As predictions of dangerousness are pivotal in many phases of a trial, this paper will not take into consideration the peculiarities of each and all applications of violence risk assessments. Furthermore, it will not account for the particularities distinctive of different jurisdictions. Rather, we will attempt at forming transversal observations which refer to common issues shared by most of these applications.

Within the paper the author will develop critical observations on ethical and legal questions raised in the previous sections. 'Classical' issues regarding the use of AI-systems such as bias and the perception of human vs. algorithmic error will be tackled. The reflection will be concluded with a number of open questions which, as it is hoped, will contribute to the ongoing debate and work as a primer for future investigations on the matter. Does the application of AI systems in these fields fall in the 'New Technology, Old Problems' saying?[8] Or do they pose new issues with regards to principles distinctive of criminal law?

2 Predicting Dangerousness

2.1 Lost in translation

Before embarking on an analysis of algorithmic risk assessment, it is relevant to briefly address the relationship between criminal law and science, focusing on predicting dangerous behavior. Dangerousness is, indeed, a dangerous concept and a label 'which is easy to attach but difficult to remove'.[9] If we take the American criminal justice system as an example, predictions of dangerousness are used as standards for sentencing matters (especially capital sentencing) and for criminal commitment following verdicts of not guilty by reason of insanity; they play an important role in sexual predator statutes and they are decisive in civil commitment procedures.[10] Indeed, in 2017 the American Law Institute in its first official amendment to the Model Penal Code, which is focused

[7] See Thomas Nadelhoffer and others, 'Neuroprediction, Violence, and the Law: Setting the Stage' (2010) 5 Neuroethics 67; Leda Tortora and others, 'Neuroprediction and A.I. in Forensic Psychiatry and Criminal Justice: A Neurolaw Perspective (2020) 11 Frontiers in Psychology 220.

[8] Stephen Morse, 'Neuroprediction: New Technology, Old Problems' (2015) Faculty Scholarship at Penn Law.

[9] P. D. Scott, 'Assessing Dangerousness in Criminals' (1977) 131 British Journal of Psychiatry quoting S. H. Shaw, 'The Dangerousness of Dangerousness' (1973) 13 Medicine, Science and the Law.

[10] See John Monahan and Jennifer L. Skeem, 'Risk Assessment in Criminal Sentencing' (2016) 12 Annu Rev Clin Psychol.

on sentencing, recommended that sentencing commissions 'develop actuarial instruments or processes, supported by current and ongoing research, that will estimate the relative risks that individual offenders pose to public safety'.[11] Said tools ought to be incorporated into sentencing guidelines.

With this in mind, it is important to stress that the legal and the clinical definitions of dangerousness do not match, and this is one of the reasons why definitive scientific answers to legal questions with regards to predicting dangerousness are hard to find – if not unattainable.[12] This issue can be referred to as the *lingua franca* problem.[13] As a matter of fact, in medicine the state of health of a patient can vary along a scale from extremely ill to completely healthy. On the contrary, legal language, especially in the field of criminal responsibility, is a strictly binary language. A person can be either guilty or not guilty, insane or sane, dangerous or not dangerous. The dialogue between these two languages is troublesome and forensic psychiatry has traditionally represented the bridge between the two disciplines when it comes to criminal liability.[14] Lately, it appears as said role is being taken over by neuroscience.

Indeed, even though dangerousness is a concept that is used in numerous legal contexts and that is strongly connected to decisions impacting on a person's freedom, the law does not define dangerous individuals with the same particularity as it is done, for example, by meteorologists when it comes to dangerous storms.[15] Questions like 'What level of violence do we expect from an individual in order for him to be considered dangerous? Is it just physical violence or also psychological violence?' or 'When we say that a person will be dangerous, will it happen tomorrow, in a month or in a year? What is the temporal scope of predictions of dangerousness?' remain unanswered. In other terms, magnitude, imminence, and frequency are not defined in the legal conception of dangerousness and no systematic effort has been done so far to fill these gaps. When it comes to dangerousness the law is permissive and discretionary because it *needs* predictions of violence: criminal law guards its own domains jealously.[16]

Another pivotal issues with regards to the relationship between science and criminal law is the so called 'group to individual problem (G2i)'.[17] What the law asks of science is to

[11] American Law Institute (ALI) Model Penal Code 2017 §6B.09(2).
[12] See N. Pollock and C. Webster, 'The clinical assessment of dangerousness', in Robert Bluglass and Paul Bowden (eds*), Principles and Practice of Forensic Psychiatry* (Churchill Livingstone 1990), 489.
[13] Joshua W. Buckholtz and David L. Faigman, 'Promises, promises for neuroscience and law' (2014) 24 Current Biology R864.
[14] See Zvi Zemishlany and Yuval Melamed, 'The Impossible Dialogue Between Psychiatry and the Judicial System: A Language Problem' (2006) 43 The Israel journal of psychiatry and related sciences 150.
[15] See John Monahan and Henry J. Steadman, 'Violent storms and violent people: How meteorology can inform risk communication in mental health law' (1996) 51 American Psychologist.
[16] See Nigel Eastman and Colin Campbell, 'Neuroscience and Legal Determination of Criminal Responsibility' (2006) 7 Nature Reviews Neuroscience 314.
[17] David L. Faigman and others, 'Group to individual (G2i) inference in scientific expert testimony' (2014) Univ. Chic. Law Rev.

answer the question of whether a 'particular case is an instance of the general phenomenon',[18] where instead science 'is focused on characterizing generalizable phenomena to establish mechanistic explanations that apply within definable population groups and, hence, are generalizable to other members of those populations (who may not yet have been observed)'.[19]

Conclusively, the translation from science to law is not an easy path to walk. Some have described the process of going from science to law as a 'journey for which there is no map'.[20] What role does AI play in this scenario? As a matter of fact, following an expansive trend, AI systems are being used in healthcare in a number of areas such as diagnostics (radiology and medical imaging), surgery and clinical care.[21] What about AI applied to (forensic) psychiatry and risk assessment tools?

These questions will guide us in the following paragraphs.

2.2 Violence risk assessment tools

Violence risk assessment tools can be defined today as instruments 'designed to increase structure, consistency, and accuracy in the evaluation of the likelihood of violent recidivism through consideration of items associated with violence recidivism'.[22] These tools were first used in the 1960s for civil commitment hearings in the American criminal system to predict whether an individual with serious mental illness would be of danger to himself or to others. Subsequently, more tools were validated specifically for predicting the criminal recidivism of 'justice-involved persons with and without mental health problems'.[23]

Regardless of where they were applied, for a long time, scientific predictions of dangerousness shared one characteristic: their fallacy – no different, as some contended, from 'flipping coins in the courtroom'.[24] In the past twenty years, progress has been made and there now seems to be substantial evidence of the increase of the accuracy of predictions associated with the use of these tools.[25]

[18] David Faigman, Philip A. Dawid, and Steven E. Feinberg, 'Fitting Science into Legal Contexts: Assessing Effects of Causes or Causes of Effects?' (2014) 43 Soc. Methods & Res. 385.
[19]Russel A. Poldrack and others, 'Predicting Violent Behavior: What Can Neuroscience Add?' (2018) 22 Trends Cogn Sci. 2, 115.
[20] ibid.
[21] See WHO, Guidance on Ethics and governance of artificial intelligence for health (2021), 6.
[22] Sarah L. Desmarais and Samantha A. Zottola, 'Violence Risk Assessment: Current Status and Contemporary Issues' (2020) 794.
[23] ibid.
[24] Bruce J. Ennis and Thomas R. Litwack, 'Psychiatry and the Presumption of Expertise: Flipping Coins in the Courtroom' (1974) 62 California Law Review.
[25] Desmarais (n 22) 801.

Traditionally, scholars distinguished between two types of violence risk assessments: clinical and actuarial. This distinction is not clear-cut anymore, as recently these tools can present characteristics of both methods. We will enumerate here a number of paramount examples of the different types of violence risk assessments tools, as it is not in the scope of this analysis to provide an exhaustive account.[26]

Clinical predictions of dangerousness are based on observation, personal examination, history taking, and testing carried out by a clinician.[27] In a clinical approach the individual's past behavior is examined and the expert affirms whether the individual will likely act in the same manner in similar circumstances.[28] The factors (ie risk factors) assessed are combined in an intuitive manner in order to evaluate the risk of violence.[29] A risk factor can be defined as a 'variable that precedes and increases the likelihood of criminal behavior'.[30] In the field of predicting recidivism, risk factors have been classified into fixed markers (such as the early onset of antisocial behavior), variable markers (which can be changed over time but not through intervention, such as age), and variable factors (such as employment status).[31] The fourth type is causal risk factors, which are 'variable risk factors that, when changed through intervention, can be shown to change the risk of recidivism'.[32]

With this in mind, the main defect of clinical predictions is that they are highly discretionary and cannot be standardized. Since clinical predictions are not structured tools, each is subjective, and this entails that it may be based on erroneous stereotypes and prejudices.

On another note, actuarial methods are based on a number of pre-identified variables that are correlated statistically to risk and result in producing a probability (or a probability range) of risk.[33] The individual's future dangerousness can then be assessed based on the characteristics that he or she shares with other people for whom a base rate exists. The more factors in the assessment, the more complex the probability rate.

---

[26] For an exhaustive overview of risk assessment tools and of the most relevant literature see *inter alia* Georgia Zara, 'Tra il probabile e il certo' (2016) Diritto Penale Contemporaneo 13-23.

[27] See John Parry and Eric Y. Drogin, Criminal Law Handbook on Psychiatric and Psychological Evidence and Testing (ABA, 2000) 24.

[28] ibid 208.

[29] See John Monahan, 'A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients' (2009) 92 Virginia Law Review 405.

[30] Monahan and Skeem (n 10) 497.

[31] See Monahan, 'Violent storms and violent people: How meteorology can inform risk communication in mental health law' (n 14) 497.

[32] ibid.

[33] See Christopher Slobogin, *Proving the unprovable: the role of law, science, and speculation in adjudicating culpability and dangerousness* (Oxford University Press 2006) 101.

The main differences between actuarial and clinical violence risk assessment tools are efficiently explained by Hilton, Harris, and Rice,[34] who do so by identifying two conceptually distinct tasks. The first task is selecting the relevant characteristics, where the second task is combining said characteristics to obtain an interpretation. With regards to the first task, when adopting an actuarial method, the selection is typically based on one or more follow-up studies which map the factors that are related to the violent outcome. The purpose of the first task is to identify 'an optimum set of items on the basis of incremental validity—that is, selecting the most powerful predictors first and then adding items only when they improve prediction'.[35] Instead, when it comes to clinical tools this task is conducted based on 'intuition, non-empirical experience, and one's memory for empirical findings'.[36] In short, according to these authors 'it is the method of selection rather than the items attended to that distinguishes clinical from statistical prediction'.[37] With regards to the second task, clinical judgment leaves the combination rule unspecified and relies on 'gut-level' processes, where instead prototypical actuarial methods 'combine risk factors using item weights derived from empirically established relationship with violent recidivism'.[38]

The most famous actuarial tool is the Violence Risk Appraisal Guide (VRAG-R),[39] which focuses on 12 variables and aims at providing a score that indicates the probability of recidivism. It was developed from a research conducted in Canada on over 600 men committed to a maximum-security hospital. The variables were identified out of fifty predictors, which were coded from institutional files, and they were used to categorize the patients into nine groups based on their actuarial risk of future violence.

A third type of prediction of dangerousness is the adjusted actuarial assessment (or structured professional judgment, SPJ) which, similar to the actuarial approach, is based on a finite number of pre-identified variables connected to risk. Thus, differently from an actuarial tool, the factors and the conclusions are not reached through a mathematical process. Following this kind of approach, the expert will assess a probability of dangerousness using an actuarial method and then adjust the result based on other factors, such as those connected to the offender. An example of this kind of approach is the Historical Clinical Risk Management (HCR-20 V3). It is a violence risk assessment scheme which is made of twenty different ratings based on historical (such as previous violence, age, em-

[34] See N. Zoe Hilton, Grant T. Harris and Marnie E. Rice, 'Sixty-Six Years of Research on the Clinical Versus Actuarial Prediction of Violence' (2006) 34 The Counseling Psychologist 401.
[35] ibid.
[36] ibid.
[37] ibid.
[38] ibid.
[39] See Grant T Harris and others, *Violent Offenders: Appraising and Managing Risk (3rd Ed.)* (American Psychological Association 2015).

ployment, substance use, relational instability...), clinical (lack of insight, active symptoms of major mental illness, unresponsive to treatment...) and risk management variables (lack of personal support, exposure to destabilizers…).[40]

One of the most relevant deficiencies of actuarial methods is that, regardless of the number of factors which can be included in the evaluation, their aggregate determination will never be sufficient to account wholly for the individual involved and '[a]ccordingly, no matter how carefully a forensic expert assembles the available actuarial information, there still is going to be a significant leap in reaching conclusions about a particular individual unless it can be shown that that individual has recently engaged in the behavior at issue or tried to do so, but was denied the opportunity'.[41] In other words, actuarial predictions neglect some characteristics of the individual evaluated which might be deemed relevant, obscuring the person's individuality.

Recently, a new tool was developed by Monahan and his colleagues: the Classification of Violence Risk Software (COVR).[42] It is an interactive software aimed at estimating the risk that a person hospitalized for mental disorder will be violent to others based on 40 risk factors. Its goal is 'offering clinicians an actuarial ''tool''' to assist in their predictive decision making'.[43] The method developed by Monahan is based on classification trees, which means that the sequence of the questions asked after the first is based on the answer given in the previous question. This methodological choice is significant: it entails that the team was able to develop the 'first software application for actuarial risk assessment'[44]. The tool does not replace the clinical decision: in fact, the inventors of the COVR software recommend themselves a review by the clinician responsible for the risk assessment in order to avoid mistakes. Moreover, as the software was trained and validated on data pertaining exclusively to psychiatric inpatients in the US, its reliability in a criminal justice setting remains uncertain.

It is possible to identify a fourth generation of violence risk assessment tools which combine individual risk factors with community-level risk variables. Said tools integrate the prediction of risk with risk management and therefore assist the doctor to develop a case management plan.[45]

To conclude, even though it is well established that the reliability of actuarial methods outperforms clinical evaluations, they are still not clinicians' nor judges' preferred tool and therefore their use is not as pervasive as one ought to think.[46] Why? One possible

[40] See Christopher D. Webster and others, HCR-20 V3: Assessing Risk for Violence. User Guide (Burnaby 2013).
[41] Parry (n 26) 24.
[42] See John Monahan and others, 'The Classification of Violence Risk' (2019) 24 Behavioral Sciences & The Law.
[43] ibid 721.
[44] ibid.
[45] An example is the Italian C-VRR (Checklist di Valutazione del Rischio di Recidiva). See Zara (n 26) 20.
[46] See Hilton (n 32), noting that '[i]f forensic decisions mirror unaided clinical judgment and not actuarial assessments after a well-validated actuarial system has been available for a decade, we must concede that

answer to the question is connected to the G2i problem and to the fact that actuarial assessments are seen as 'too impersonal for the purposes of the law'.[47] Consequently, according to some, '[r]eliable and accurate assessment of violence risk remains an elusive goal for forensic psychiatrists'.[48]

It is in this scenario that neuroprediction first, and then AI, enter the picture.

2.3 Enter: AI

2.3.1 Neuroprediction and AI

The lion's share of the debate on the impact of neurosciences on criminal law is centered on ascertaining criminal responsibility,[49] rather than predictions of violent behavior.[50] Nevertheless, in the last ten years scientists started turning to neuroscience in a quest to improve the reliability of forecasting future violent behavior, by linking the structure of the brain to dangerousness. One of the reasons guiding the inclusion of neuroscientific data in risk assessment is that by adding 'important personalized information about the brain of offenders to the risk assessment equation' said studies might 'make it more likely that legal decision makers rely on the best available tools of violence risk assessment'.[51] Some argue that the future use of neuroprediction is 'not morally worse than the present use of clinical risk assessments', since it regards 'objectionable features of the use of statistical, unlike individualized, evidence' and because it is 'more accurate'.[52] Others harshly criticize arguments in favor of moral permissibility of this practice, based on the fact that they will present the same issues that all actuarial tools present.[53]

The most recent scientific studies take an even forward steps and focus on 'the use of structural or functional brain parameters coupled with machine learning methods to make clinical or behavioral predictions', namely AI neuroprediction.[54] As a matter of fact, Tortora and others predict that AI neuroprediction of recidivism 'is likely to become available in the near feature'.[55] These studies include using Functional Magnetic Resonance Imaging (fMRI) data to predict recidivism based on potential correlations between low activation of the portion of the brain deputed to impulses and error processing (the dorsal anterior cingulate cortex) and recidivism;[56] using machine learning to study

---

the mere availability of actuarial methods—and the careful analysis and dissemination of their consistent advantage—is fundamentally insufficient for their adoption' 404.

[47] Nadelhoffer (n 6) 79.

[48] Richard G. Cockerill, 'Ethics Implications of the Use of Artificial Intelligence in Violence Risk Assessment' (2020) 48 J Am Acad Psychiatry Law 1.

[49] Tortora (n 6) 3.

[50] Nadelhoffer (n 6) 634.

[51] ibid 86.

[52] Lippert-Rasmussen (n 61) 125 summarizing '*Nadelhoffer's Argument*', Nadelhoffer (n 6).

[53] ibid 135.

[54] Tortora (n 6) 2.

[55] ibid.

[56] See Alexandre Abraham and others, 'Machine Learning for Neuroimaging with Scikit-Learn' (2014) 8 Frontiers in neuroinformatics 14.

whether brain age can be used as a predictive factor for rearrest;[57] and evaluating whether including the levels of cerebral blood flow in risk assessment could improve predictions of violent behavior in long-term follow-up forensic psychiatric patients.[58]

One critique to (AI) neuroprediction which is hardly surmountable at this point – but might be in the future – is that current neuropredictions are developed on a very low base rate which often comprises of a homogenous population, such as inmates or psychiatric inpatients. In this regard, finding plausible solutions to the G2i problem becomes troublesome. Consider the following example:

> A neuroscientist uses fMRI to scan 100 participants who are instructed to either lie or tell the truth about a set of facts. Contrasting brain activity during lying with truth telling reveals statistically significant activation in dorsolateral prefrontal cortex (DLPFC). This permits the valid group-level inference that lying is associated with DLPFC engagement. However, examination of each individual's data reveals that, while most subjects exhibited higher DLPFC activity during lying, some participants showed no difference and still others demonstrated lower DLPFC activity during lying compared with truth telling.[59]

Simply put, the risk of false positive and false negatives is prominent.

All things considered, discoveries in neuroscience have traditionally been accompanied by a great level of enthusiasm regarding their potential for explaining causal processes of violence behavior, ie of opening the 'black box' of the human brain.[60] However, by applying complex AI to neuroprediction we might be, in fact, introducing a new black box in the system. We will develop this argument further on in this analysis.

2.3.2 AI: friend or foe?

One common characteristic of the risk assessment tools which were mentioned in section 2.2. is that they are static, not dynamic, systems. They are based on rather simple algorithms which simply weight and combine information to produce the likelihood of future violent behavior.[61] This dynamicity can be explained as follows: by applying deep learning[62] to violence risk assessments 'new data are constantly incorporated to improve and

---

[57] See Kent A. Kiehl and others, 'Age of gray mattes: Neuroprediction of recidivism' (2018) 19 NeuroImagie: Clinical.

[58] See Carl Delfin and others, 'Prediction of recidivism in a long-term follow-up of forensic psychiatric patients: Incremental effects of neuroimaging data' (2019) PLoS One.

[59] Poldrack (n 17) 115.

[60] ibid.

[61] Desmarais (n 22) 813.

[62] Deep learning (DL) is a subset of ML. An algorithm based on Machine Learning (ML) techniques teaches itself rules by learning from the training data through statistical analysis, detecting patterns in large amounts of information and generating outputs. Deep Learning consists of layers of artificial neural networks (ANNs). Neural networks' engineering was inspired by the functioning of biological neurons, hence their basic function is to establish features from input. ANNs are made of layers of functions ("nodes" or "neurons") that perform various operations on the data that they are fed. The main difference

refine a predictive model' as 'correct predictions reinforce the model, while incorrect predictions cause it to recalibrate'.[63] In other words, risk assessment tools based on AI represent an 'enhanced version of … the empirical actuarial approach', as these systems have the potential of combining 'countless data points in complex ways to identify persons at risk of violence', developing models of 'unfathomable complexity'.[64] As such, AI systems could 'amplify or mitigate both the strengths and the weaknesses associated with existing actuarial prediction techniques'.[65]

Let us consider an example. Between 2015 and 2018, a study on how to detect risks of school violence was conducted on 131 students.[66] The aim of the study was to develop an AI system based on natural language processing[67] and machine learning capable of automatically analyzing the contents of interviews and information on the student's household in order to identify risk factors and predict the risk of violent behavior of the single student. The study proved that linguistic patterns are relevant indicators of a student's risk of school violence. The deployment of this system would facilitate immensely school violence risk assessment, which is a costly procedure (each assessment conducted by a clinician costs from $1000 to $3500 dollars).

What does this entail? If currently applied tools for predicting dangerousness are 'merely' an instrument in the hand of judges and psychiatrists, in the future AI systems could be programmed to replace judicial and clinical decision-making. We will analyze this perspective in the following sections.

It should be noted at this point that one of the benefits of applying AI to risk assessment is that they process massive amounts of data in an incredibly fast and accurate way, saving time in often very lengthy criminal trials. Indeed, the revolutionary aspect of machine learning algorithms is that they are capable of identifying unforeseen patterns and correlations (ie predictions) between factors that are not perceived by the human agent (as it would be with a traditional actuarial risk assessment tools), finding new solutions for a given task. The question that rises, then, is the following: should all the links identified by AI through these massive data analysis operations be considered significant links?

---

between ML and DL is that in DL the algorithm is fed raw (unlabeled) data and then identifies by itself which features are relevant. In ML learning, instead, the algorithm is given an established of relevant features to analyze.

[63] Cockerill (n 22) 1.

[64] Neil R. Hogan, Ethan Q. Davidge, Gabriela Corabian, 'On the Ethics and Practicalities of Artificial Intelligence, Risk Assessment, and Race' (2021) 49 J Am Acad Psychiatry Law 3, 2.

[65] ibid.

[66] See Ni Yizhaho and others, 'Finding warning markers: Leveraging natural language processing and machine learning technologies to detect risk of school violence' (2020) 139 International Journal of Medical Informatics.

[67] Process by which the system extracts data from human language and makes decisions based on that information. It enables clear human-to-machine communication. Examples of NLP systems are voice-activated digital assistants such as Alexa, Siri, Cortana and Google Assistant.

Who should decide which kind prediction is acceptable from a criminal law standpoint and which one is not?

If we think of 'traditional' actuarial risk assessment tools, the choice of which risk factors to include in the analysis has always been deemed critical *per se*. Think of gender, race or life history variables such as the level of education or employment or the family environment: all in all, the use of these technologies cannot be perpetrated without taking into account the principles paramount of criminal law. Applying these principles to risk factors entails that in order for a variable to be relevant from a criminal law standpoint, the parameter has to be connected to the blameworthiness of the individual. As it was clarified,

> Demographic and life history variables that characterize an offender may have significant predictive validity in assessing his or her likelihood of recidivism, but no bearing on the ascription of blame for the crime of which he or she was convicted'. Both race and gender correlate significantly with criminal recidivism … However, neither race nor gender is seen as bearing on an offender's blameworthiness for having committed crime—as a class, offenders who are women are seen as no more (or no less) blameworthy than offenders who are men, and offenders who are African American are seen as no more (or no less) blameworthy than offenders who are white.[68]

We can now apply the same kind of argument used as a critique to "classical" violence risk assessment tools to AI based systems. Imagine an AI-system that analyzes huge amounts of data contained in criminal records at a very fast pace. Say that the task given to the system is to identify the factors which show the highest correlation with convictions of rape, as the intended purpose is to prevent the commission of such crimes by felons on a national basis. Imagine now that the algorithm identifies with a certainty of 99.9% that out of 100.000 individuals convicted for rape, 60% had red hair. Should police forces focus more on red haired individuals in their crime prevention activities? Taking this example even further, say that according to the output of the algorithm out of the aforementioned 60% red haired sex offenders, 35% of them reoffended and they were all taller than 1.80 m. Is this connection noteworthy? Should it affect a judge deciding on whether the (poor) red haired tall defendant can be admitted to parole? The issue becomes even thornier if we take into consideration other traits of an individual.

Think of mental illness, which has been carrying the stigma of dangerousness for decades: it follows that there has been a lot more attention (*rectius*, data collection) on dangerous behaviors by mentally ill offenders than there has been on sane offenders, making available databases unbalanced.[69] This is surely an overly simplified case scenario but it

[68] Monahan, 'Risk Assessment in Criminal Sentencing' (n 9) 503.

[69] It has been empirically proven that symptoms or diagnoses of serious mental disorders are not related or inversely related to subsequent violence in a variety of clinical populations such as civil psychiatric patients, forensic patients and mentally disordered offenders, sex offenders and violent offenders in general. See the studies cited in Hilton (n 35) 402.

brings to light important ethical questions which will have to be confronted in the future in order to truly avoid Lombroso 2.0 criminal policies.

To continue, the use of AI and neurodata does not supersede on of the classical issues related to violence risk assessments: the asymmetrical perception of statistic (or actuarial) evidence versus individualized evidence. As humans, we are bound to think that machines cannot make mistakes ('to err is human, not algorithmic'[70]), yet we trust more (fallible) human decisions. This problem is best explained through the famous 'gatecrasher's case',

> It is certain that 1,000 people attended a football match and certain that only 10 people bought a ticket. In one case, John is convicted for gatecrashing solely on the basis of statistical evidence that he, undeniably, went to the football match and accordingly, on the basis of this information, there is a 99% probability that he gatecrashed. In another case, John is convicted solely on the basis of a piece of individualized evidence consisting in an eyewitness report of John's gatecrashing. There is a 99 % probability that the eyewitness report is true.[71]

The critical issue is that most people will believe that John should be acquitted in the first scenario but convicted in the second. Nevertheless, the evidence in the two scenarios is exactly the same with regards as the likelihood that John gatecrashed.

Moreover, the application of AI and the introduction of neuroprediction does not seem to solve the problem of the lack of individualization of actuarial assessments: criminal judgments have to be specific and tailored to the distinctive features of the single perpetrator and cannot be reduced to pointing out that an individual X due to certain characteristics belongs to a general category Y of people who recidivated in the past. Yet, some say that this objection to actuarial assessments can be overcome, since clinicians also apply a G2i logic in their evaluations: their clinical assessment relies on their training and on comparing the individual case to their past experiences with similar cases, hence 'while clinicians look at individual patterns, they do not so in a vacuum'.[72] One thing is for certain: judges and clinicians could be influenced in their decisions by a number of (criminally) irrelevant and unaccounted factors, where instead an AI system might not. In other words: an AI system will not deliver a more lenient judgment based on what it had for lunch. A (human) judge might.[73]

---

[70] Laetitia A. Renier, Marianne Schmid Mast and Anely Bekbergenova, 'To Err Is Human, Not Algorithmic – Robust Reactions to Erring Algorithms' (2021) Computers in Human Behavior.

[71] Kasper Lippert-Rasmussen (n 62) 124.

[72] Christopher Slobogin, 'Dangerousness and Expertise' (1984) 133 University of Pennsylvania Law Review, 126.

[73] See Myles Udland, 'Want a Favorable Ruling in Court? Catch A Judge Right After Lunch.' (Business Insider, 2021) <https://www.businessinsider.com/court-leniancy-improves-after-judges-eat-2015-11?r=US&IR=> accessed 19 July 2021.

Further concerns regarding AI-based risk assessments include the 'GIGO' problem (Garbage In, Garbage Out)[74] and bias: biased data leads to biased predictions. Be that as it may, one must also account for two common misconceptions with regards to bias in statistical predictions systems which are paramount of criminal justice. First, that these systems will produce a biased output only if they are trained on inaccurate or incomplete datasets.[75] The second misconception is referred to as 'fairness through unawareness' and indicates the belief that 'predictions can be made unbiased by avoiding the use of variables indicating race, gender, or other protected classes'.[76]

Admittedly, there is no such thing as an 'easy fix'.[77] If we take race as an example, scholars argue that the root of racial inequality in risk assessment lies in the 'nature of the prediction itself', [78] rather than in the databases or in the structure of the algorithm. Since predictions 'look to the past to make guesses about future events', it is only natural that in 'a racially stratified world, any method of prediction will project the inequalities of the past into the future'.[79] In this regard, some claim that one of the benefits of including neuroimaging in risk assessment evaluations would represent as a way to decrease bias in risk assessments, provided that neuroprediction is not just incorporated in (already biased) existing risk assessment tools.[80]

A problematic feature which is distinctive of AI systems – and not of traditional actuarial systems – is the one of black box: the process of creation of outputs produced by complex AI systems, such as the one based on deep learning, might not be explainable (to the laymen, but even to its creator) since the algorithm teaches itself a rule without any kind of instruction. In other words, we are dealing with a 'system [that] is so complicated that even the engineers who designed it may struggle to isolate the reason for any single action. And you can't ask it: there is no obvious way to design such a system so that it could always explain why it did what it did.' [81]

For example, a recent study on the application of deep learning to detect Covid-19 in chest radiographs proved that certain AI systems produced a diagnosis by relying on 'confounding factors rather than medical pathology' so that the results of the systems appear accurate at first sight, but they are not when tested in a different medical facility. Simply put, these systems used 'shortcuts': instead of learning the real underlying pathology which could prove the presence of COVID-19, they used 'spurious associations

[74] See, amongst others, 'Report On Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System - The Partnership On AI' (The Partnership on AI, 2021) <https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/> accessed 19 July 2021.
[75] ibid.
[76] ibid.
[77] Sandra G. Mayson, 'Bias In, Bias Out' (2018) 128 The Yale Law Journal 8, 2218.
[78] ibid.
[79] ibid.
[80] Tortora (n 6) 5.
[81] Will Knight, 'The Dark Secret at the Heart of AI' (MIT Technology Review, 2021) <https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/> accessed 19 July 2021.

between the presence or absence of COVID-19 and radiographic features that reflect variations in image acquisition'.[82] Notably, the study proved that this kind of shortcut learning might occur even when data of better quality (and not data that 'confuse' the system) is used to train the system. The perils of these 'shortcuts' and the lack of transparency, were the AI be applied for risk assessment in a field as delicate as criminal justice, are self-evident.

3 A Glance to the Future: Towards Virtual Forensic Experts?

According to a global survey conducted in 2020, psychiatrist are not that worried that AI could steal their jobs: '[t]he mental status examination, evaluation of dangerous behavior and formulation of a personalized treatment plan, all essential roles of a psychiatrist, were … felt to be tasks that a future AI/ML technology would be unlikely to perform'.[83] As a matter of fact, the work of a forensic psychiatrist is 'rife with ethical dilemmas'.[84] Differently from other branches of medicine, psychiatrists have to balance the duty of care of their patient with issues regarding public safety almost on a daily basis. How could AI assist psychiatrists through this minefield? Possible solutions might lie ahead.

For instance, in January 2021 the Department of Neuroscience of the University La Sapienza in Rome published an application for a research grant titled 'Artificial intelligence in forensic psychiatry: the development of a new algorithm to guide and structure forensic evaluations of criminal responsibility and social dangerousness'. The aim of this project is to overcome ethical issues related to the application of machine learning techniques in the forensic field by 'developing models and an (Explainable & Trustworthy) AI-based Decision Support System (DSS) (Virtual Forensic Experts) to guide and support forensic psychiatric evaluations of criminal responsibility and social dangerousness, in order to make them more objective, transparent, and reliable'.[85]

The themes which emerge from these few sentences are innumerable. We will only mention a few here: What would be the role of Virtual Forensic Experts in future society? We have touched upon some of the most futuristic developments of violence risk assessment (AI neuroprediction) and, as we know, these systems could produce a reliable output without being able to provide an explanation of the underlying logical process. How will a forensic psychiatrist interact with the machine? Will the AI systems be faithful assistants of human forensic experts or will they take their place completely? We mentioned earlier in this paper that AI systems could be fairer than a human judge or a human

[82] Alex J. DeGrave, Joseph D. Janizek and Su-In Lee, 'AI for radiographic COVID-19 detection selects shortcuts over signal (2021) 3 Nature Machine Intelligence 610-619.
[83] P. Murali Doraiswamy, Charlotte Blease and Kaylee Bodner, 'Artificial Intelligence and the Future of Psychiatry: Insights from a Global Physician Survey' (2020) 102 Artificial Intelligence in Medicine, 12.
[84] Hogan (n 67) 6.
[85] 'Artificial Intelligence in Forensic Psychiatry: The Development of New Algorithm to Guide and Structure Forensic Evaluations of Criminal Responsibility and Social Dangerousness' (*EURAXESS*, 2021) <https://euraxess.ec.europa.eu/jobs/598745> accessed 19 July 2021.

psychiatrist. However, can AI systems make ethical decisions? Could we think of a Hippocratic Oath for 'artificial psychiatrists'? These will be the topics of future conversations between psychiatrists and criminal legal scholars.

4 Conclusion

In this paper we have attempted at breaking down the issues related to applying AI to risk assessment tools for predicting future dangerous behavior. We started by addressing a classical debate, namely the dialogue between criminal law and science. Subsequently, we provided a short state of the art with regards to violence risk assessment tools. Then, we introduced AI to our reflection: we first focused on a niche area, namely the intersection between AI and neurosciences applied to predictions of recidivism. Building on these notions, we conducted a number of critical reflections on the benefits and drawbacks of applying AI to predictions of violence in criminal justice. We concluded this investigation with the introduction of a futuristic scenario such as the creation of virtual forensic experts and its possible impact from a legal and an ethical point of view.

Many questions remain open. Hopefully, they will work as a primer for the debate that is coming. For example, it will be relevant to analyze what the role of judges would be in a future where AI neuroprediction is applied. How binding will AI neuropredictions be in a criminal trial? There is no transversal solution to the issues that emerged in this short analysis. The answers to the questions we asked will not only depend on the possible legal implications, but also on social concerns. The key factor in the future discussion will be, for example, the degree of development of AI techniques in a certain country, the degree of trust placed in such techniques by the population (and therefore by the legislator) and the degree of confidence that the judge will have when having to decide on AI-related matters. We end where we began, as we will be bound to ask ourselves one final (recurrent) question: are we criminals because of what we do or because of who we are?

References

Abraham A, and others, 'Machine Learning for Neuroimaging with Scikit-Learn' (2014) 8 Frontiers in neuroinformatics

American Law Institute (ALI) Model Penal Code [2017]

'Artificial Intelligence in Forensic Psychiatry: The Development of a New algorithm to Guide and Structure Forensic Evaluations of Criminal Responsibility and Social Dangerousness' (*EURAXESS*, 2021) <https://euraxess.ec.europa.eu/jobs/598745> accessed 19 July 2021

Bigenwald A, and Chambon V, 'Criminal Responsibility and Neuroscience: No Revolution Yet' (2019) 10 Frontiers in Psychology

Bluglass R, and Bowden P (eds), *Principles and Practice of Forensic Psychiatry* (Churchill Livingstone 1990)

Buckholtz J W, and Faigman L D, 'Promises, promises for neuroscience and law' (2014) 24 Current Biology R864

Cockerill R, 'Ethics Implications of the Use of Artificial Intelligence in Violence Risk Assessment' (2020) 48 J Am Acad Psychiatry Law

Cosimi S, 'Gay o etero, un algoritmo "legge" l'orientamento sessuale sul volto. Il controverso studio di Stanford' (la Repubblica, 2021) <https://www.repubblica.it/tecnologia/2017/09/08/news/gay_o_etero_un_algoritmo_legge_l_orientamento_sessuale_sul_volto_il_controverso_studio_di_stanford-174923406/> accessed 9 August 2021

DeGrave A J, Janizek J D, and Lee S, 'AI for radiographic COVID-19 detection selects shortcuts over signal (2021) 3 Nature Machine Intelligence

Delfin C, and others, 'Prediction of recidivism in a long-term follow-up of forensic psychiatric patients: Incremental effects of neuroimaging data' (2019) PLoS One

Desmarais S L, and Zottola S A, 'Violence Risk Assessment: Current Status and Contemporary Issues' (2020) 794

Doraiswamy P, Blease C, and Bodner K, 'Artificial Intelligence and the Future of Psychiatry: Insights from a Global Physician Survey' (2020) 102 Artificial Intelligence in Medicine

Douglas T, and others, 'Risk Assessment Tools in Criminal Justice and Forensic Psychiatry: The Need for Better Data' (2017) 42 European Psychiatry

Eastman N, and Campbell C, 'Neuroscience and Legal Determination of Criminal Responsibility' (2006) 7 Nature Reviews Neuroscience

Ennis B, and Litwack T, 'Psychiatry and the Presumption of Expertise: Flipping Coins in the Courtroom' (1974) 62 California Law Review

Faigman D, Dawid P A, and Feinberg S E, 'Fitting Science into Legal Contexts: Assessing Effects of Causes or Causes of Effects?' (2014) 43 Soc. Methods & Res.

—— and others, 'Group to individual (G2i) inference in scientific expert testimony' (2014) Univ Chic Law Rev

Glenn A, and Raine A, 'Neurocriminology: Implications for the Punishment, Prediction and Prevention of Criminal Behaviour' (2013) 15 Nature Reviews Neuroscience

Harris G T and others, 'Violent Recidivism of Mentally Disordered Offenders: The Development of a Statistical Prediction Instrument' (1993) 20 Crim Just & Behav

—— *Violent Offenders: Appraising and Managing Risk (*3Rd Ed., American Psychological Association 2015)

Hilton N, Harris G T, and Rice M, 'Sixty-Six Years of Research on the Clinical Versus Actuarial Prediction of Violence' (2006) 34 The Counseling Psychologist

Hoffman M, 'Nine Neurolaw Predictions' (2018) 21 New Criminal Law Review

Hoga N R, Davidge E Q, Corabian G, 'On the Ethics and Practicalities of Artificial Intelligence, Risk Assessment, and Race' (2021) 49 J Am Acad Psychiatry Law

Kiehl A K, and others 'Age of gray mattes: Neuroprediction of recidivism' (2018) 19 NeuroImagie: Clinical

Lippert-Rasmussen K, 'Neuroprediction, Truth-Sensitivity, and the Law' (2014) 18 The Journal of Ethics 2

Monahan J, and Steadman H J, 'Violent storms and violent people: How meteorology can inform risk communication in mental health law' (1996) 51 American Psychologist

—— 'A Jurisprudence of Risk Assessment: Forecasting Harm among Prisoners, Predators, and Patients' (2009) 92 Virginia Law Review 405

—— and others 'The Classification of Violence Risk' (2006) 24 Behavioral Sciences & the Law.

—— and Skeem J L, 'Risk Assessment in Criminal Sentencing' (2016) 12 Annu Rev Clin Psychol

Morse S, 'Neuroprediction: New Technology, Old Problems' (2015) Faculty Scholarship at Penn Law

—— 'Neuroethics: Neurolaw', Oxford Handbooks Online (2017)

Munthe C, and Radovic S, 'The Return of Lombroso? Ethical Aspects of (Visions of) Preventive Forensic Screening' (2015) 8 Public Health Ethics

Musumeci E, 'Against the Rising Tide of Crime: Cesare Lombroso and Control of the "Dangerous Classes" in Italy, 1861-1940' (2018) Crime, Histoire & Sociétés

Nadelhoffer T, and Sinnott-Armstrong W, 'Neurolaw and Neuroprediction: Potential Promises and Perils' (2012) Philosophy Compass 7/9

'Neural Network Learns to Identify Criminals by Their Faces' (MIT Technology Review, 2021) <https://www.technologyreview.com/2016/11/22/107128/neural-network-learns-to-identify-criminals-by-their-faces/> accessed 2 August 2021

OECD, Recommendation OECD/LEGAL/0449 of 22 May 2019 R of the Council on Artificial Intelligence (2019)

Parry J, and Drogin E Y, *Criminal Law Handbook on Psychiatric and Psychological Evidence and Testing* (ABA, 2000)

Renier L, Schmid Mast M, and Bekbergenova A, 'To err is human, not algorithmic – Robust reactions to erring algorithms' [2021] Computers in Human Behavior

Poldrack R A, and others, 'Predicting Violent Behavior: What can Neuroscience add?' (2018) 22 Trends Cogn Sci 2

'Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System - The Partnership on AI' (The Partnership on AI, 2021) <https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-us-criminal-justice-system/> accessed 19 July 2021

Ronzoni D, 'Lombroso 2.0: una rete neurale per riconoscere i criminali dai tratti somatici' (DDay.it, 2021) < https://www.dday.it/redazione/21681/lombroso-20-una-rete-neurale-per-riconoscere-i-criminali-dai-tratti-somatici> accessed 9 August 2021

Scott P, 'Assessing Dangerousness in Criminals' (1977) 131 British Journal of Psychiatry

Slobogin C, 'Dangerousness and Expertise' (1984) 133 University of Pennsylvania Law Review

—— *Proving the unprovable: the role of law, science, and speculation in adjudicating culpability and dangerousness* (Oxford University Press 2006)

Stone A, and Stromberg D C, *Mental Health and Law: A System in Transition* (National Institute of Mental Health, Center for Studies of Crime and Delinquency 1976)

Strano M, 'A Neural Network Applied to Criminal Psychological Profiling: An Italian Initiative' (2004) 48 Int J Offender Therapy & Comp Criminology 495

Tortora L, and others, 'Neuroprediction and A.I. in Forensic Psychiatry and Criminal Justice: A Neurolaw Perspective' (2020) 11 Frontiers in Psychology

Udland M, 'Want A Favorable Ruling in Court? Catch a Judge Right After Lunch.' (Business Insider, 2021) <https://www.businessinsider.com/court-leniancy-improves-after-judges-eat-2015-11?r=US&IR=> accessed 19 July 2021

Walsh C G, Ribeiro D J, Franklin J C, 'Predicting Risk of Suicide Attempts Over Time Through Machine Learning' (2017) 5 Clinical Psychological Science 3

Webster C, Ben-Aron M, and Hucker S, Dangerousness (Cambridge University Press 1987)

——and others, *HCR-20 V3: Assessing Risk for Violence. User Guide (*Burnaby 2013)

WHO, Guidance on Ethics and governance of artificial intelligence for health (2021)

Will Knight, 'The Dark Secret at the Heart of AI' (MIT Technology Review, 2021) <https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/> accessed 19 July 2021

Yizhaho N, and others, 'Finding warning markers: Leveraging natural language processing and machine learning technologies to detect risk of school violence' (2020) 139 International Journal of Medical Informatics

Zara G, 'Tra il probabile e il certo' (2016) Diritto Penale Contemporaneo

Zemishlany Z, and Melamed Y, 'The impossible dialogue between psychiatry and the judicial system: a language problem' (2006) 43 The Israel journal of psychiatry and related sciences

Mayson G S, 'Bias In, Bias Out' (2018) 128 The Yale Law Journal 8

Valla J, Ceci S, and Williams W, 'The accuracy of inferences about criminality based on facial appearance.' (2011) 5 Journal of Social, Evolutionary, and Cultural Psychology

Wu X, and Zhang X, 'Automated Inference on Criminality using Face Images' [2017] arXiv:1611.04135v1

Artificial intelligence (AI) is impacting our everyday lives in a myriad of ways. The use of algorithms, AI agents and big data techniques also creates unprecedented opportunities for the prevention, investigation, detection or prosecution of criminal offences and the efficiency of the criminal justice system. Equally, however, the rapid increase of AI and big data in criminal justice raises a plethora of criminological, ethical, legal and technological questions and concerns, eg about enhanced surveillance and control in a pre-crime society and the risk of bias or even manipulation in (automated) decision-making. In view of the stakes involved, the need for regulation of AI and its alignment with human rights, democracy and the rule of law standards has been amply recognised, both globally and regionally. The lawfulness, social acceptance and overall legitimacy of AI, big data and automated decision-making in criminal justice will depend on a range of factors, including (algorithmic) transparency, trustworthiness, non-discrimination, accountability, responsibility, effective over-sight, data protection, due process, fair trial, access to justice, effective redress and remedy. Addressing these issues and raising awareness on AI systems' capabilities and limitations within criminal justice is needed to be better prepared for the future that is now upon us.

This special issue on 'Artificial intelligence, big data and automated decision-making in criminal justice' comprises topical and innovative papers on the above issues, centred around AI and big data in predictive detection and policing, liability issues and jurisdictional challenges prompted by crimes involving AI, and AI-assisted and automated actuarial justice or adjudication of criminal cases.

Gert Vermeulen is Senior Full Professor of European and international Criminal Law and Data Protection Law, Director of the Institute for International Research on Criminal Policy (IRCP), Di-rector of the Knowledge and Research Platform on Privacy, Information Exchange, Law Enforcement and Surveillance (PIXLES) and Director of the Smart Solutions for Secure Societies (i4S) business development center, all at Ghent University, Belgium. He is also General Director Publications of the AIDP and Editor-in-Chief of the RIDP.

Nina Peršak is Scientific Director and Senior Research Fellow, Institute for Criminal-Law Ethics and Criminology (Ljubljana), Advanced Academia Fellow (CAS Sofia), Member of the European Commission's Expert Group on EU Criminal Policy, Independent Ethics Adviser, and Co-Editor-in-Chief of the RIDP.

Nicola Recchia is Postdoc Researcher in Criminal Law at the Goethe-University Frankfurt, Germany. He is also member of the Young Penalists Committee and of the Scientific Committee of the AIDP.

www.maklu.be
ISBN 978-90-466-1130-2