UNIVERSITATEA DIN ORADEA
FACULTATEA DE DREPT

In honorem VALENTIN MIRIŞAN

Gânduri, Studii şi Instituţii

Editat de **Universul Juridic SRL**.
Editura **Universul Juridic** este acreditată CNATDCU (lista A2) şi este considerată editură cu prestigiu recunoscut.

Descrierea CIP a Bibliotecii Naţionale a României
OMAGIU. MIRIŞAN, VALENTIN
In honorem Valentin Mirişan: Gânduri, Studii şi Instituţii / ed.: conf. univ. dr. Cristian Dumitru Miheş - Bucureşti : Universul Juridic, 2020
Conţine bibliografie
ISBN 978-606-39-0615-2

34
082.2

Redacţie:
tel.: 0732.320.666
e-mail: redactie@universuljuridic.ro

Distribuţie:
tel.: 021.314.93.15
fax: 021.314.93.16
e-mail: distributie@universuljuridic.ro
edituraunversuljuridic.ro

Editura Universul Juridic

Universul Juridic.ro
nihil sine iure

Portal:
tel.: 0725.683.560
e-mail: portal@universuljuridic.ro
universuljuridic.ro

Universul Juridic

ujmag.ro®
UNIVERSUL JURIDIC MAGAZIN

Librăria UJmag:
tel.: 0733.673.555; 021.312.22.21
e-mail: comenzi@ujmag.ro
ujmag.ro

Ujmag.ro

ALGORITHMIC DECISION MAKING AND ISSUES OF CRIMINAL JUSTICE – A GENERAL APPROACH

Prof. dr. Krisztina KARSAI[1]

Abstract:

In this paper, I intend to outline the basic concepts concerning algorithmic decision-making and fundamentals necessary to understand its use within the criminal justice system. The paper deals with central definitional issues and identifies the key terms and risks of any regulation and presents a cross-cultural approach in order to demonstrate the recent – diverging – legal narratives of relevant geographic actors. This paper touches the social-legal environment of criminal justice as identifying and defining the different needs for and possibilities of deploying algorithmic decision-making solutions in the distinct stages of the criminal procedure. Further research will be dedicated to specific criminal justice issues and the challenges that have arisen from the human right protection regimes (mainly within the EU).

Keywords: *criminal justice, algorithmic decision making,human rights, disruptive technologies, data protection law*

Introduction and definitions

While algorithms are hardly a recent invention, they are nevertheless increasingly involved in systems used to support decision-making. Known as 'ADM' or 'ADMS' (algorithmic decision-making systems), ADMS often rely on the analysis of large amounts of personal data to infer correlations or, more generally, to derive information deemed useful to make decisions.

Human intervention in the decision-making process may vary and may even be completely out of the loop in entirely automated systems. In many situations, the impact of the decision on people can be significant, such as on access to credit, employment (ALFTER – EISELT – SPIELKAMP, 2019),[2] medical treatment (ALFTER et al., 88),[3] and judicial sentences (MIHES, 2018, 208-209)[4], among other things (LISCHKA – KLINGEL, 2017; MIRIŞAN-ONICA CHIPEA, 2014).[5] Advances in data

[1] University of Szeged, Faculty of Law and Political Science, Head of Institute of Criminal Law & Criminal Science, Jean Monnet Professor

[2] Sifting through personal emails for personality profiling in Finland. See more Automating Society: Taking Stock of Automated Decision-Making in the EU. A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations. 1st edition, January 2019, 62-63

[3] Allocating treatment for patients in the public health system in Italy.

[4] Police officers apply facial recognition algorithms to identify suspects (or victims) appearing in recording from a crime scene, judges use risk assessment ADM solutions on bail, on sentencing, and parole decisions based on an individual's demographic characteristics and criminal history (and in order to predict recidivism). Also please see:

[5] Further examples are the ADM solutions for airport security by assessing risks for airline passengers (no-flight lists); the automated processing of traffic offences in France or algorithmic identification of children possibly vulnerable to be neglected (Denmark).

collection, the development of algorithms and accelerations in computational power have enabled data analytics processes to proliferate into new fields (ALFTER et al. 8). Entrusting ADMS with the power to influence or make such decisions raises an array of different ethical, political, legal, and technical issues, and thus all due care must be applied to properly analyze and address these. If these issues are neglected, then the expected benefits of these systems may be negated by the variety of risks for the individuals (discrimination, unfair practices, loss of autonomy, etc.), for the economy (unfair practices, limited access to markets, etc.), and for society as a whole (manipulation, threat to democracy, etc.; CASTELLUCCIA – LE MÉTAYER, 2019).

Considerable literature in statistics, mathematics and computer science is available concerning the use of algorithms and machine learning. In this work, I use the basic and most important terms and concepts of these other disciplines to upon the elaborate legal-criminal and human rights issues that arise, but simplification and generalization are necessary. However, the proportions need to be appropriately defined, as legal experts and lawyers also need to understand the conceptual building blocks of algorithmic computation in order to properly address the evolving legal issues.

Algorithmically controlled, automated decision-making or decision support systems are procedures in which decisions are initially—partially or completely—delegated to another person or corporate entity. The recipient then in turn uses automatically executed decision-making models to perform an action (ALFTER et al. 9). “The algorithm itself is the expression of the sum of the objectives and perspectives of those for whose objectives the algorithm is deployed.” (BACKER 2018a, 19-20)

In the case of traditional rule-based systems, the relationship between inputs and outputs are “crafted by hand” and are often complicated, with hundreds or even thousands of steps, but generally represent pre-existing rules or theories (VEALE et al., 2019). More advanced systems are able to learn through the accumulation of data and by machine learning procedures. In case of these algorithms, “pre-existing rules or theories do not capture the desired input-output relationships well. As a result, machines craft the relationship between inputs and outputs backwards from the data, usually without regard for human interpretability. In some cases, this can allow machines to make much more effective input-output connections – which computer scientists call predictions – than hand-crafted rule-based systems could.” (VEALE et al., 2019). “Machine learning is nonparametric in that it does not require the researcher to specify any particular functional form of a mathematical model in advance. Instead, these algorithms allow the data themselves to dictate how information contained in input variables is put together to forecast the value of an output variable. (…) with machine-learning results, causal relationships between inputs and outputs may simply not exist, no matter how intuitive such relationships might look on the surface. (…) The user of an algorithm cannot really discern which particular relationships between variables factor into the algorithm’s classification, or at which point in the algorithm they do, nor can the user determine how exactly the algorithm puts together various relationships to yield its classifications. For this reason, machine-learning algorithms are often described as transforming inputs to outputs through a black box. An analyst cannot look inside the black box to understand how that transformation occurs or describe the relationships with the same intuitive and causal language often applied to traditional statistical modelling" (COGLIANESE – LEHR, 2017).

This process includes learning (acquisition of information and rules for using information[6]), reasoning (using rules to reach approximate or definite conclusions), and interpretation – to propose one decision or different decisions as result of machine learning and not based on human reasoning.

Artificial intelligence (AI) as a technology that — when used thoughtfully — can augment human capabilities, rather than replacing them. AI is regarded as a replication of human analytical and/or decision-making capabilities. AI mechanisms can perform various functions of human intelligence: reasoning, problem solving, pattern recognition, perception, cognition, understanding, and learning. AI platforms and artifacts may support human decision making especially by probabilistic reasoning and discerning patterns in data (MITROU 2017, 10). Presently used AI solutions are narrow AI-s (ANI). A narrow AI is AI that is programmed to perform a single task — whether it's checking the weather or being able to play the board game "Go", or to analyze data to write simple texts or poems. These systems are not able to perform outside of the single task that they are designed to perform. AGI, or artificial general intelligence that is self-reflective (consciousness), multifunctional, independent and is capable of interacting with the environment does not exist. In theory, AGI would be capable of successfully performing any intellectual task that a human being can.

Consequently, it can be concluded that today's AI solutions are backed by ADM software. A further significant element shall be included in the definition section: today's robots are mostly controlled by an ADM solution, but an important common feature robots share is that they *have a physical body*, so this *differentia specifica* must be considered when embedding ADM solutions into legal frameworks. While algorithms (through machine learning) can extend the human cognitive abilities, robots are able to extend the physical skills of the humans.

ADM is concerned with decisions based on automated data management, that is, when correlations and causations discovered by machine analysis of large amounts of data are based on a decision that gives rights or obligations to or for individuals or groups of individuals. The *appearance of human involvement* in this decision-making process depends on how we incorporate it into the system: whether and to what extent it is retained in operations on the required data, and whether or not the result calculated by the algorithm is taken into account in the human decision. This spectrum can thus extend to solutions run by fully automated decision-making systems.

Certainly, trust is presently vested in such systems due to the facilitation of human decisions, the need for more complete information processing and *the need to dispel doubts* (as prerequisites for well-founded decision making), and the *pursuit of efficiency* (convenience) factors. Existing systems and solutions have evolved organically – they are

[6] This paper does not cover the analysis of machine learning as deep learning or supervised – unsupervised – reinforcement learning because the research approach does not need the differentiation based on the type of learning.

market driven and tend to operate unobtrusively, that is, without problems. Systems or software solutions employing such technology have penetrated the financial, health, governmental, public administration, welfare services sectors significantly and have had a profound impact on the societies of most developed countries ('data driven governance' BACKER 2018b). Their sporadic use is now a thing of the past; we are now witnessing an overall societal impact. Thus, even if formerly left unanswered, in the present *it is imperative* to address the emerging moral and legal questions with a general regulatory need. States and international organizations have endeavored to find the suitable regulations in this area, but on the road to finding such interventions or instruments there is still a long way ahead.

Discussing analytical tools like ADM solutions, a further distinction shall be made following HANNAH-MOFFAT's clear delineation. There is a need to differentiate between "psychologically informed actuarial tools" and "big-data informed actuarial tools" because the latter produce new types of risk for applications also within the criminal justice pipeline. HANNAH-MOFFAT stated that meanwhile psychologically informed "instruments are predicated on a discipline-based causal knowledge of criminality, and are designed in such a way as to systematically produce and organize a diverse range of information (…), big data technologies are not constrained by preconceived theoretical or methodological disciplinary norms or necessarily administered and interpreted by certified assessors" (HANNAH-MOFFAT 2019). And this conceptual difference makes both the legal environment and the application of them entirely different.

Challenges for the Legal Framework

The *basic dilemma* of the overall regulatory trials is finding the means to take advantage of technological development in a way that further innovation continues whilst remaining unhindered by the legal framework and continuing to protect against abuse and infringement. There are essentially two factors that make ADM systems more risky than "conventional" computerization solutions: first, *the new context of personal data*, that is, the fact that much more detailed personal information may be available to "virtually anyone". Second, these systems necessarily and *per definitionem* operate with the reduction or complete absence of the involvement of human factors. These two features also determine the areas through which the path toward general regulation is directed: on the one hand, data protection and data management requires intelligent and careful regulation, while on the other hand, and the framework for the enforcement of human rights requirements must be built on the consequences of the *decline in the human factor*. Furthermore, as BACKER correctly noted, "adding meaning to data may be difficult without confronting society's moral flames, however. Politics tended to refine and contain those fires in politically useful ways. It is not clear how data driven analytics-based governance will construct similarly effective constraints—or if it constituted to engage in such building. Beyond that the underlying ideologies that drive data driven governance can create paradox. One has been suggested already—the drive toward transparency and data harvesting at the heart of ideologies of accountability, against the ideology of human autonomy. Both, ironically sit at the core of the great ideological project of human rights" (BACKER 2018a, 44).

Dividing Societies in the Age of Global Technology

Disruptive technologies generate the proliferation of organic, or untargeted, regulation and as a consequence, we are witnessing a rise in significantly different regulatory frameworks that are rooted in different cultural-philosophical narrative environments, which are difficult to resolve; meanwhile, further development driven by the global market and society at large would call for a seamless solution. It should be noted here that there are essentially three major paradigms or regulatory systems in each of the two necessary areas of regulation.

The two areas are distinct based on one hand on whether the actors in the ADMS sector private companies or government agencies are and, on the other hand, based on fundamentally different approaches to human rights and data protection.

The *European Union and its Member States* have developed and applied stand-alone European standards, both in the field of data protection and the protection of human rights. Disruptive technological developments are mostly ordered by the private sector and to a lesser extent by the government, but strict adherence to and compliance with human rights and data protection requirements is a priority for all national and EU authorities. The protection of business secrets (e.g. the coding or programming of the algorithm) may be limited in relation to human rights requirements and privacy requirements. And finally – concerning the application of algorithms within the justice pipeline – there has been a regulatory and cultural aversion to systems being fully automated and void of a 'human in the loop' (JONES 2017, 47).

In the *U.S.A.*, data protection is primarily strengthened with respect to privacy, and US constitutional standards for the protection of human rights. There is no state regime for data protection, and the constitution prohibits the government from making violations. Usually, the private sector (market enterprises) is the main actor in developing and applying disruptive technologies, and such entities are the providers of software solutions for government actors (COGLIANESE – LEHR, 2017). As such, the protection of business secrets is strong and dominant. Similar cultural aversion – such as in Europe – cannot be identified in the general approach of the US. Where algorithms are deployed by private sector organizations directly, freedom of information law has limited current applicability (VEALE et al., 2019, 23). Similarly questionable – from a European point of view – are value-laden decisions concerning whether or not the design of systems in criminal justice will be outsourced.

The third global player is *China*, which has a more functional approach to data protection (DE HERT– PAPAKONSTANTINOU 2015)[7] and human rights protection. The development and application of technologies is mainly an activity of state or of semi-private

[7] The international data protection fundamentals that may be derived from all relevant regulatory instruments in force today, namely the personal data processing principles and the individual rights to information, access and rectification, are not unequivocally granted under Chinese law.

companies implemented on state's request and accordingly, the prevalence of governmental and state interests is strong, as opposed to the protection of private business secrets. The experimental use of disruptive technologies, even in an administrative (or judicial) environment, is not impeded by the sensitive human rights regime or restricted by the specific rules on data protection.

From a horizontal approach it can be stated that "in the West, *social credit* takes on a variety of forms, sometimes, but not always, driven by the private sector and supported by the growth of markets in data. Outside of China, the drivers of social credit tend to be private enterprises--for everything from rating credit to the rating of the corporate social responsibility effectiveness of enterprises. (…) Most of the elements of social credit have already been developed in the West. But the unification of the various elements, and their seamless operation would be a great innovation. For Chinese theory, that innovation would complement the move toward transformation in politics, economics, and social organization" (BACKER 2018a, 48).

We can make further comparisons based on the role of the human in the picture. JONES pointed out that "the US and EU regions negotiate the interplay between computer automation and personhood in shared ways, but Europeans[8] have drawn important lines regarding what it means to be a human, based on how computational technologies must adhere to certain restrictions on fully automated processing – injecting a *human in the loop*. American policymakers, on the other hand, have been less eager to draw such lines, rejecting the idea that individuals need to be protected from being treated as data subjects to be automatically processed, instead embracing the notion of the neutral platform and portraying the computable and the computational as aspirational" (JONES 217). The perception of the role of humans in Chinese society also differs; therefore, the definition of the basic criteria for automation processes will be distinctive in terms of human participation or rather, it will be more similar to the American approach. "Protecting American personhood has meant subjecting individuals and groups to as much accuracy, fairness and objectivity – computational neutrality – as possible. (…) The EU and member countries' consistent insistence on the categorization of data protection and a human in the loop as a fundamental right" (JONES 230).

Algorithms in the Criminal Justice Pipeline

The conceptualization of the broad infiltration and deep intrusion of algorithms into the everyday life of our societies on a global scale discusses "governing algorithms" (BAROCAS –HOOD –ZIEWITZ 2013) and about "algorithmic governmentality" (ROUVROY –BERNS 2013) and asks for ethical and legal possibilities of application of ADM solutions within the justice sector of the societies. As ZAVSRNIK summarized correctly "big data, coupled with

[8] See the main policy instruments: Declaration of cooperation on Artificial Intelligence (10 April 2018) of the EU Member States; Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe (24 April 2018) COM(2018) 237 final; European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics and the Ethics Guidelines for Trustworthy Artificial Intelligence (High-Level Expert Group on AI, 8 April 2019)

algorithms and machine learning, has become a central theme of intelligence, security, defence, anti-terrorist and crime policy efforts, as computers help the military find its targets and intelligence agencies justify carrying out massive pre-emptive surveillance of public telecommunications networks" (Zavsrnik 2019, 2). Moreover "we are witnessing a gradual movement away from the traditional, retrospective, individualized model of criminal justice, which prioritizes a deliberated and personalized approach to pursuing justice and truth, towards a prospective, aggregated model, which involves a more ostensibly efficient, yet impersonal and distanced, approach. 'Actuarial justice' is based on a 'risk management' or 'actuarial' approach to the regulation of crime and the administration of justice" (Marks –Bowling –Keenan 2015).

There is a peculiar paradox in this area: although there is no comprehensive policy on the use of algorithms and algorithmic decision-making in within the justice process, the use of tools that utilize such technology is almost universal. The reason for this is that the appearance of these devices was (or is) considered technological advancement, but we fail to take notice of systemic differences and specifics, and perhaps we did not even realize them. The novelty of the problems and the exclusion of analogue (not algorithmic decision making) solutions has led to the situation of the present, confrontation with their presence and massive penetration worldwide.

An important frame circumstance is that the situations presented here do not yet affect all justice systems and are not homogeneous: thus, while in one country algorithmic tools are only relevant at a certain procedural stage, in another given state they are related to another stage. This does not mean, however, that it is not *possible to hypothetically bind* the tools known and used today to provide an overview that could describe the future.

The application of ADM solutions can be fundamentally *different depending on the stage* of the criminal justice pipeline at which they are implemented. ADM solutions can be deployed for prevention, detection, investigation, prosecution of crimes, in court proceedings and also in the course of the penal execution. These solutions are designed for the specific stage of the criminal justice pipeline and therefore use different datasets and machine learning concepts, and thus each shall comply with the different legal requirements depending on the given procedural stage. For instance, the rights of the concerned individual to be informed are fundamentally different at the stage of prevention compared to the investigation phase or in the courts. Where public authorities use ADM solutions that obviously have an effect on a given person, the content of the information obligation depends on the legal status of the person concerned, as well as the law applicable to him or her. Therefore, the information obligation may vary in different phases. The classification assigned to the different phases of the criminal justice process is dynamic and serves for understanding the situation. However, *the classification shall be deepened* by answering the following questions: a) Are potential offenders, victims, witnesses, or members of the criminal justice system affected? b) To what extent is human intervention related to using ADMS? Next, it is necessary to examine and identify which individual standards and requirements of human rights and data protection are currently associated with each solution, what can be covered by analogy and what are the risks of infringement. Further consideration is needed in terms of what changes would be legally possible if for example a tool of one stage of the criminal procedure were to be applied in another stage of the criminal justice pipeline: if tools for prevention were implemented for investigational purposes (e.g. person-based predictive policing, etc.).

Dangers of the Application of Algorithms in the Criminal Justice Pipeline

In order to map the possible risks and dangers associated with the application of algorithms in the criminal justice pipeline, I used the expert report of the British Law Society (2019) as a grid and have integrated the results of other groups and researchers into that structure. Four main categories can be differentiated, as follows: as such instrumental, dignitary, justificatory, and systematic concerns.

Instrumental concerns

a) *Bias and discrimination*: The way in which data that serves as input into systems is labeled, measured, and classified is subjective and can be a source of bias. "If, as is commonly known, the justice system does under-serve certain populations or over-police others, these biases will be reflected in the data, meaning it will be a biased measurement of the phenomena of interest, such as criminal activity" (VEALE et al., 2019 19). Algorithms do not differentiate between information (data) they use all kind of data as possible base of judgement (decision), but we, members of the society do. For instance, ethnicity will be avoided to assign as a sole basis of judgement. However, studies proved also that protecting informational privacy, excluding information about protected attributes from a decision (especially the race or ethnicity), and equalizing classification rates across protected groups do not generally increase algorithmic fairness. Further, under some conditions, these interventions may reduce fairness (ALTMAN –WOOD - VAYENA 2018, 34). In general, "algorithmic systems trained on past biased data without careful consideration are inherently likely to recreate or even exacerbate discrimination seen in past decision-making. … Machine learning systems are designed to discriminate—that is, to discern—but some forms of discrimination seem socially unacceptable" (EDWARDS –VEALE, 2017, 11).

b) *Oversimplification of complex issues:* algorithms work with datasets and rules, but both data and rules should be formalized (quantified) in a mathematically workable format. "A core question for algorithms in the justice systems is what insight and information is lost in this process. (…) Relying on algorithmic systems might result in some decisions being made on a shallow view of evidence and without a deep, contextual consideration of the facts" (VEALE et al., 2018, 20).

Dignitary concerns

a) *Individuals not treated as such*: while the violation of rules (e.g. commission of an offense) presupposes individuals who breach the law often willingly, meaning that more often than not, these behaviors are unique (even in cases that are comparable), such human behaviors are nothing more than items of the datasets. As such, individuality or special circumstances can be taken into consideration if they can be mathematically manipulated or described. VEALE and others stated that "machine learning systems are similarity engines, seeking to find cases with traits that are similar to cases that were present in the training data and classify them similarly. (…) Membership of a group or similarity to other cases in a dataset do not cause criminality, victimisation, or other focuses of algorithms in the justice system – but a heightened emphasis on correlation, simply because it is computationally possible, may cause the conflation of the two and place dignity at risk. (…) Data protection as a regime has highly individual foundations, and this has been reflected in its provisions. Many problems in algorithmic bias can best be understood as disadvantaging a group or a community, rather than an individual. As a result, it becomes important to consider whether a use of an algorithm will affect a group rather than an individual" (ALTMAN et al., 31). Other researchers, ALTMAN et al. underline that "a group may be harmed disproportionately to the social benefit, and this is especially problematic if the group is a historically disadvantaged group. In addition, if the harms are not connected to choices made by the individuals, it may be clear that the absolute cost to those individuals is too high. (…)." Viewed through the lens of the potential-outcomes framework, intuitively, an algorithmic decision may cause harm to an individual when the expected outcome for that person given the decision is worse than the expected outcome for that person absent the decision (ALTMAN et al., 6).

b) *Loss of autonomy*: according to the expert group "algorithmic systems might manipulate people into situations they would not have been in otherwise. (…) The idea that one is being constantly technologically surveilled with behavior predicted may moderate an individual into a bland, constrained course of action, in fear of triggering systems designed to detect anomalies or deviation" (VEALE et al., 2018, 21).

c) *Privacy*: if private or sensitive data are the basis of machine learning, the outputs or results of the usage of algorithms can lead to violations and misuse of protected information.

Justificatory concerns

a) *Opacity preventing scrutiny of justification and of rulemaking*: in the judicial process, all decisions must be open to scrutiny in order to assess whether the judiciary was fair, legal, justified, and legitimate. I think that similar expectations must be met by machine decisions. This means that the programming source, the training datasets, and the eventual encoded rules shall be transparent and understandable at least for forensic experts. In the context of transparency there are two main approaches to ADM systems, and for software developed for justice: they can follow the black box or the white box approach. According to the general understanding, the black box approach analyzes the behavior of the ADS without any knowledge of its code. Explanations are constructed from observations of the relationships between the inputs and outputs of the system. Under this approach, the operator or provider of the ADS is uncollaborative, i.e. does not agree to disclose the code in order to protect business secrets. The white box approach assumes that analysis of the ADS code is possible which means that the solution is explainable (CASTELLUCCIA - LE MÉTAYER 8).

b) *Power and function creep* from information infrastructures, meaning that even "where algorithmic systems and their associated informational infrastructures are deployed proportionally today, the tools deployed may not have appropriate safeguards to prevent them from being misused in the future. CCTV systems, for example, were deployed in an era where technology did not allow their contents to be analysed at scale automatically" (VEALE et al, 2018, 23).

Systemic concerns

a) *Human rights*: ADM solutions applied in the justice pipeline can intrude human rights. Human rights concerns may depend on the stage in which the ADM system is applied within the criminal justice pipeline, so that, for example, violation of the right to a fair trial does not occur during the crime prevention phase. Online judicial decision-making may raise the need for a fair trial, but the reality in criminal matters is not long-term. It is also true, however, that in the course of investigation, the gathering of evidence may already affect fundamental rights, and in this case the evidence (such as face recognition, mobile data, etc.) collected or provided by algorithms can play a significant role.

b) *Changing nature of law*: the British commission brought an important – or in my view the most important – point to the discussion, namely "algorithmic systems in the justice sector which look at past data to predict the future run the risk of stagnation, holding the evolution of justice anchored in the past rather than free to evolve." (VEALE et al, 2018, 25).

Changing the reality of social coexistence also brings with it changes in law, coding legislative changes into algorithms seems easy to implement, but judicial practice or jurisprudence will not change as algorithms *will not seek a new direction based on the "old" pattern*. So, if there were still a human factor in the decision-making process, he would not receive a new or different impulse, and if this were the case, the "algorithmic case law" would consequently remain unchanged. It is also important to see that if previously decided cases constitute the algorithm's teaching database, we also obstinately assume that each case was adjudicated correctly by human judges, since there is no moral or legal basis for questioning the correctness of (final) decisions in past procedures.

Limitations under the Current Data Protection Regime (EU)

Article 16 of the Treaty on the Functioning of the European Union provides legal basis and entitlement for the establishment of the EU data protection regime. The General Data Protection Regulation (GDPR)[9] issued in 2016, as already mentioned, contains strong and uniform rules on the use of personal data. It is important to note, however, that as a rule, the use of personal data in the criminal justice pipeline is not yet addressed here, but legal relationship is regulated by the Directive on data protection in the area of police and justice[10] (LED).

[9] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. Official Journal L 119, 4.5.2016 1-88

[10] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of

ADM solutions used in criminal justice may fall within the scope of the GDPR and LED, depending on who is processing the personal data and for what purposes. For example, while an ADM solution has been developed by a private company – even if its creation was carried out based on a mandate of a state authority or as an outsourced service – and the algorithm needs personal dataset for learning, the GDPR would apply for to data processing by the company, but later LED would cover the data processing by the authorities if they were to use the result of the ADM software for criminal justice purposes (prevention, investigation, detection or prosecution of criminal offenses etc.).

General Data Protection Regulation (GDPR)

A restricted concept of applying ADM solutions under Article 22 of the GDPR was introduced, which contains dedicated norms to ADM systems under the title of "Automated individual decision-making, including profiling". According to this Article:

"1. The data subject shall have the *right not to be subject to a decision* based *solely* on automated processing, including profiling, which produces *legal effects* concerning him or her or *similarly significantly affects* him or her. 2. Paragraph 1 shall not apply if the decision: (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) is based on the data subject's explicit consent. 3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least *the right to obtain human intervention* on the part of the controller, to express his or her point of view and to contest the decision. 4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1)[11], unless point (a) or (g) of Article 9(2) applies[12] and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place."

Prima facie, the GDPR contains a *general ban* of ADM; if all of the requirements listed below are fulfilled, the given ADM solution cannot be deployed:

- it concerns a fully automated decision-making solution and

criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA Official Journal L 119, 4.5.2016 89-131

[11] Article 9 (1) of GDPR: „Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited."

[12] Article 9 (2) of GDPR: Paragraph 1 shall not apply if one of the following applies: a) the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject; (…) g) processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

- the basis for such a decision is a personal dataset, and
- the result has legal effect, or it significantly affects the position of the data entity.

However, the GDPR can impede the application of an ADM solution if any of these three criteria are not met. The meaning and interpretation of the elements remains the subject of debate,[13] therefore only a narrower application can be allowed. As such, this also means that the application of ADM solutions permitted when the algorithm recommends or prepares the human decisions. Hence the GDPR's narrow definition of automated decisions and the far-reaching exceptions will help create an environment in which interaction with ADM systems *will remain an everyday occurrence* (DREYER - SCHULZ 2019, 45).

In connection with the debate concerning the *disclosure of business secrets*, Article 13 (2) f) and Article 14 (2) g) contains the *right to explanation*. The sub-points therein lay down the obligation to inform the data subject about personal data collected, i.e. "the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject." Both legal and computer science experts and academics struggle in interdisciplinary debates to define the components of the so-called right to explanation and in particular its content, and it seems to be a general consensus that this right can be triggered properly only in some of the cases, many ADM solutions with machine learning engines can be unlikely explained (EDWARDS – VEALE 64). The debate on the content and the depth of the communication continues; participants have yet to decide how the text of the GDPR shall be interpreted, and whether the source code shall be disclosed or only abstract information is needed for compliance with the GDPR or the structure and internal design of the ADM solution shall be explainable and understandable.

It is important to understand that specific applications of algorithms do not fall under the scope of the GDPR if they do not use personal data: location-centered predictive policing is the most significant example of this.

Law Enforcement Directive (LED)

In addition to the GDPR, the EU data protection reform package of 2016 also included a directive on data protection in the area of police and justice (applicable as of May 6, 2018). The LED lays down the rules relating to the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection, or prosecution of criminal offenses or the execution of criminal penalties, including safeguarding against and the prevention of threats to public security (Article 1). While in theory, setting specific purposes of the data processing offers easy delineation between the application of GDPR and LED, in reality *the differentiation could lead* to improper data processing (e.g. mistake in purpose concerning data processing in connection with migration) and misuse of rules (SAJFERT –QUINTEL 2019).

[13] See more in the Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679; Article 29 Data Protection Working Party, 2017 (17/EN WP251rev.01)

Art. 11 of the directive states the following concerning automated individual decision-making:

“1. Member States shall provide for a decision based solely on automated processing, including profiling, which produces an adverse *legal effect concerning the data subject or significantly affects him or her*, to be prohibited unless authorised by Union or Member State law to which the controller is subject and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller. 2. Decisions referred to in paragraph 1 of this Article shall not be based on special categories of personal data referred to in Article 10, unless suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests are in place. 3. Profiling that results in discrimination against natural persons on the basis of special categories of personal data referred to in Article 10 shall be prohibited, in accordance with Union law.”

The ban of applying ADM solutions for criminal justice purposes provided by the LED has limitations as well: first, if the data processing does not result in a fully automated process; second, if personal data have not been used, or third if authorization is provided by law (even automated processing), then Article 11 shall not apply. Comparison of the wording of Article 22 of the GDPR and this Article reveals a clear difference, as SAJFERT AND QUINTEL correctly pointed out: “the GDPR is not limited to adverse legal effects, but includes all outcomes legally affecting data subjects as a result of automated decision-making processes. However, it could be argued that Art. 22 GDPR kicks in only when data subjects are being affected by legal effects or similarly significantly affected by the outcome of the automated decision-making, while the language in the Directive is stronger and more straightforward, prohibiting automated decision-making when the data subject is significantly affected, without requiring a correlation with a legal effect” (SAJFERT –QUINTEL 2019).

However, it is important to underline that this ban is not applicable in cases where the ADM solution is utilized in preparing the basis for a human decision, as in this case, the eventual data processing falls outside of the scope of Article 11 – in compliance with the LED and its national implementation.

On the contrary, Article 11 Section 3 contains a definitive prohibition on discriminative profiling, which means that a profile may not consist exclusively of data relating to the race, ethnicity or religious affiliation of the data subject (etc.), if use of such characteristics runs the risk of leading to any unjustified discrimination of the data subject.

Recital No 38 provides that automated processing should be subject to suitable safeguards, including the provision of specific information to the data subject and the right to obtain human intervention, in particular to express his or her point of view, to obtain an explanation of the decision reached after such assessment, or to challenge the decision. Unlike the GDPR, the LED does not contain legal provisions on any specific obligation for the data controller to inform the data subject about the logic of the ADM solution, only the general rules on the right to be informed (see LED Chapter III) applies.

Further Discussion

As mentioned, the aim of this paper is to present fundamental concepts, introduce the definitions used and to discuss the criminological framework of applying ADM solutions within criminal justice. In order to provide a comprehensive overview of the issues, the main risks have been mapped. Part One also touched over the data protection challenges involved, while human rights implications will be examined in Part Two. A general approach to the use of specific algorithms within the criminal justice pipeline was also discussed, while legal analysis will be presented in Part Two.

Summarizing the general social trends, I agree with ZAVRSNIK who stated that "the introduction of AI into criminal justice settings also reinforces the social power of the new emerging digital elite, which capitalizes on the ideology of big data and algorithmic impartiality. However, delivering justice with new tools – big data, algorithms and machine learning – does not lead to a de-biased digital bonanza" (ZAVRSNIK, 15). However, in criminal justice, decisions are made, obviously and necessarily, on the basis of available data – i.e. evidence, and therefore, the *statistical approach and the probabilistic reasoning* provided by ADM solutions could not form the *sole* basis for decisions. Notwithstanding "statistical evidence and probabilistic reasoning today play an important and expanding role in criminal investigations, prosecutions and trials, not least in relation to forensic scientific evidence (including DNA) produced by expert witnesses" (AITKEN –ROBERTS – JACKSON 2010). Although criminal statistics, court records, and recidivism rates alone cannot determine decisions by authorities, in the recent decades, there has nonetheless been undeniable focus on collecting and recording such data, creating databases based on these and carrying out scientific analysis. Patterns and trends have been identified and intensive widespread knowledge has been accumulated about phenomena of criminality and criminal justice. The *game-changer* in this significant shift has been the practically unlimited access to any type of data due to the technological revolution and digitalization of the last 15 years. Solid understanding and proper integration of statistical and probability reasoning within the criminal justice process has become more important than ever, "these inductive generalizations about patterns of crime form the basis for attempts to develop the prediction capacity of AI in criminal justice settings"(BROADHURST –BROWN –MAXIM –TRIVEDI – WANG 2019) and possibilities for forensic science to use the advantageous existence of the recent data-tsunami (Orbán, 2018; MIHES, 2018, 208). "Therefore *lawyers need to understand* enough to be able to question the use made of statistics or probabilities and to probe the strengths and expose any weaknesses in the evidence presented to the court; judges need to understand enough to direct jurors clearly and effectively on the statistical or probabilistic aspects of the case; and expert witnesses need to understand enough to be able to satisfy themselves that the content and quality of their evidence is commensurate with their professional status and, no less importantly, with an expert witness's duties to the court and to justice" (AITKEN – ROBERTS – JACKSON, 4; MARKS – BOWLING – KEENAN, 5; MURPHY, 2007).

The *mission* of enhancing understanding and knowledge of law professionals is followed by this research as well. [14]

[14] This research was supported by the project nr. EFOP-3.6.2-16-2017-00007, titled Aspects on the development of intelligent, sustainable and inclusive society: social, technological, innovation networks in

References:

BRIGITTE ALFTER – RALPH MÜLLER-EISELT – MATTHIAS SPIELKAMP: Automating Society: Taking Stock of Automated Decision-Making in the EU. A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations. 1st edition, January 2019

COLIN AITKEN – PAUL ROBERTS – GRAHAM JACKSON: Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, 2010, https://www.researchgate.net/publication/259088224_Fundamentals_of_Probability_and_Statistical_Evidence_in_Criminal_Proceedings_Guidance_for_Judges_Lawyers_Forensic_Scientists_and_Expert_Witnesses

MICAH ALTMAN – ALEXANDRA WOOD - EFFY VAYENA: A Harm-Reduction Framework for Algorithmic Fairness. IEEE Security & Privacy, 2018, vol. 16, no. 3, 34-45

LARRY CATÁ BACKER: And an Algorithm to Bind them All? Social Credit, Data Driven Governance, and the Emergence of an Operating System for Global Normative Orders. Entangled Legalities Workshop; 24 & 25 May 2018, Geneva. Electronic copy available at: https://ssrn.com/abstract=3182889 (2018a)

LARRY CATÁ BACKER: Next Generation Law: Data Driven Governance and Accountability Based Regulatory Systems in the West, and Social Credit Regimes in China (July 7, 2018) http://dx.doi.org/10.2139/ssrn.3209997 (2018b)

SOLON BAROCAS – SOPHIE HOOD – MALTE ZIEWITZ, Governing Algorithms: A Provocation Piece (March 29, 2013) http://dx.doi.org/10.2139/ssrn.2245322

RODERIC BROADHURST – PAIGE BROWN – DONALD MAXIM – HARSHIT TRIVEDI – JOY WANG: *Artificial Intelligence and Crime*, Research Paper, Korean Institute of Criminology and Australian National University Cybercrime Observatory, College of Asia and the Pacific, Canberra, June 2019. https://ssrn.com/abstract=3407779

CLAUDE CASTELLUCCIA – DANIEL LE MÉTAYER: Understanding algorithmic decision-making: Opportunities and challenges. Scientific Foresight Unit within the Directorate-General for Parliamentary, 2019.

CARY COGLIANESE – DAVID LEHR: Regulating by Robot: Administrative Decision Making in the Machine-Learning Era. The Georgetown Law Journal (2017) Vol 105. http://scholarship.law.upenn.edu/faculty_scholarship/1734

KELLY HANNAH-MOFFAT: Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. Theoretical Criminology 2019, Vol. 23(4) 453–470; https://doi.org/10.1177/1362480618763582

PAUL DE HERT – VAGELIS PAPAKONSTANTINOU: The Data Protection Regime in China. In-Depth Analysis (November 19, 2015). Brussels Privacy Hub Working Paper, Volume 1, Number 4 https://ssrn.com/abstract=2773577

employment and digital economy. The project has been supported by the European Union, co-financed by the European Social JURAJ SAJFERT – TERESA QUINTEL: Data Protection Directive (EU) 2016/680 for Police and Criminal Justice Authorities. In MARK COLE – FRANZISKA BOEHM (eds.), Commentary on the General Data Protection Regulation, Edward Elgar Publishing, 2019, Forthcoming. Available at SSRN: https://ssrn.com/abstract=3285873

Fund and the budget of Hungary.

STEPHAN DREYER - WOLFGANG SCHULZ: The General Data Protection Regulation and Automated Decision-making: Will it deliver? Potentials and limitations in ensuring the rights and freedoms of individuals, groups and society as a whole. Bertelsmann Stiftung, January 2019, DOI 10.11586/2018018, 45

LILIAN EDWARDS – MICHAEL VEALE: Slave to the Algorithm: Why a „right to an explanation" is probably not the remedy you are looking for. Duke Law & Technology Review 2017, Vol 16 No 1

MEG LETA JONES: The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood. Social Studies of Science Volume 47 Issue 2 2017

Konrad LISCHKA – ANITA KLINGEL: Wenn Maschinen Menschen bewerten. Internationale Fallbeispiele für Prozesse algorithmischer Entscheidungsfindung. Arbeitspapier, May 2017, Bertelsmann Stiftung. DOI 10.11586/2017025

AMBER MARKS – BEN BOWLING – COLMAN KEENAN: Automatic Justice? Technology, Crime, and Social Control. (October 19, 2015). R. Brownsword, E. Scotford and K. Yeung (eds), The Oxford Handbook of the Law and Regulation of Technology, OUP https://ssrn.com/abstract=2676154

LILIAN MITROU: Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'? (December 31, 2018). http://dx.doi.org/10.2139/ssrn.3386914

ERIN MURPHY: The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence. California Law Review 2007 Vol 95, Issue 3, DOI: 10.15779/Z38R404

JÓZSEF ORBÁN: Bayes-hálók a bűnügyekben. Doktori értekezés, Pécs, 2018.

CRISTIAN MIHEŞ: New Dares in Criminal Investigation, Annales Universitatis Apulensis, Series JURISPRUDENTIA, 21/2018, ProUniversitaria, Bucureşti, 2018,

ANTOINETTE ROUVROY – THOMAS BERNS: Gouvernementalité algorithmique et perspectives d'émancipation », Réseaux 2013/1 (No 177), p. 163-196 DOI 10.3917/res.177.0163. Translated by Elizabeth Libbrecht [Algorithmic Governmentality and Prospects of Emancipation]

VALENTIN MIRIŞAN – LAVINIA ONICA-CHIPEA, A Chance for Roma Children: "Children of Promise" Foundation, Revista de Cercetare si Intervenție Socială, vol.45/2014, https://www.rcis.ro/images/documente/rcis45_16.pdf

ALEŠ ZAVRŠNIK: Algorithmic justice: Algorithms and big data in criminal justice settings. European Journal of Criminology 2019, DOI: 10.1177/1477370819876762