LAURA NOTARO*

PREDICTIVE ALGORITHMS AND CRIMINAL JUSTICE: A SYNTHETIC OVERVIEW FROM AN ITALIAN AND EUROPEAN PERSPECTIVE**

ABSTRACT. *The paper seeks to provide a synthetic overview of the current and potential uses of "predictive algorithms" in the context of criminal justice systems, as well as of the critical issues arising from such tools.*
In the first part (§§ 2-3), the main uses of Artificial Intelligence (AI) tools within the legal systems which have implemented them will be presented. Then a synthetic framework of the possible fields of application of AI in the context of criminal justice will be provided.
In the second part (§§ 4-5), the problems arising from the introduction of algorithmic tools in criminal justice will be analysed, with special attention to the fundamental rights guaranteed at a European level, the general principles of criminal law and procedure and the further juridical limits, provided by supranational and national law.
In the conclusions (§ 6), an attempt will be made to define the acceptable scopes of application of algorithmic tools in the criminal justice system and to identify the necessary precautions and conditions for the use of new technologies to be consistent with the national and European legal framework.

* PhD Student in Criminal Law, University of Pisa.

** The paper is the result of the research work carried out by the author within the project "*Intelligenza artificiale e giustizia penale*", funded by the European Parliament and directed by Prof. Antonella Massaro.

observations – 4.2. Tools for predicting the outcome of judicial decision-making – 5. Legal limits to the introduction of AI into criminal justice systems – 5.1. The Ethical Charter of the CEPEJ of the Council of Europe – 5.2. European Union law: limits to «decisions based solely on automated processing» – 5.3. Italian law: limits to criminological expert evidence – 6. Final remarks. Scope of application and precautions for AI in criminal justice systems

1. Introduction

The use of Artificial Intelligence (AI) tools in the context of criminal justice is becoming a crucial issue also in Europe.

Even if the most significant introduction of new technologies in criminal courts can be observed in the United States' system, a discussion at a European level could be useful in order to anticipate the advances in technology and the development of new practices: it is necessary for setting up the proper cultural context[1] and a clear legal framework in order to be prepared for the challenges of modernity.

The paper is divided into five sections.

Paragraphs 2 and 3 aim at better clarifying the object of the research. First of all, the paper will examine the main uses of AI tools in those criminal justice systems which have implemented them. Then a synthetic framework of the possible fields of application of AI in the context of criminal justice will be provided.

Paragraphs 4 and 5 aim at carrying out an analysis of the problems arising from the introduction of algorithmic tools in criminal justice, with specific reference to the fundamental rights guaranteed at a European level, the general principles of criminal law and procedure and further juridical limits, provided by supranational and national law.

[1] The need that the implementation of AI tools is prepared by an interdisciplinary debate, in order to provide a framework for developing algorithms in compliance with fundamental rights, is underlined by X. Ronsin, V. Lampos, V. Maîtrepierre, *In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data*, in *Appendix I*, *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, at <www.coe.int>.

Paragraph 6 will provide some conclusions, as an attempt to define the fields of the criminal justice system in which AI tools could be admitted and also to identify the necessary precautions and conditions for the use of new technologies to be consistent with the national and European legal framework.

2. AI in criminal justice systems. Experiences and recent trends

2.1. United States

The United States experience is the first and most advanced example of introducing AI in criminal justice systems.

The fields in which predictive tools are employed are basically two: *a)* the decisions on pretrial release; *b)* the sentencing stage.

Algorithmic tools are very numerous, but the most used are, on the one hand, "PSA" (Public Safety Assessment), which was created by the Arnold Foundation, and, on the other hand, "COMPAS" (Correctional Offender Management Profiling for Alternative Sanctions) and "LSI-R" (Level of Service Inventory Revised), both developed by commercial companies.[2] Some States have also designed their own predictive algorithms.[3]

[2] For an overview of the predictive tools used in each State, see the paper published by EPIC (Electronic Privacy Information Centre), *Algorithms in the Criminal Justice System: Risk Assessment Tools*, available at the following link: <https://epic.org/algorithmic-transparency/crim-justice>. See also D. Kehl, P. Guo, S. Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing* (Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School), July 2017, available at the following link: <https://cyber.harvard.edu/publications/2017/07/Algorithms>.

[3] For example, the Ohio Department of Rehabilitation and Correction has developed the algorithm "ORAS" (Ohio Risk Assessment System) in collaboration with the University of Cincinnati (in this regard, see Kehl, Guo, Kessler, *Algorithms in the Criminal Justice System*, cit., p. 16). About the algorithm recently adopted in Pennsylvania, after the work carried out since 2010 by the Pennsylvania Commission on Sentencing, see R. Hester, *Risk Assessment at Sentencing. The Pennsylvania Experience*, in J.W. de Keijser, J.V. Roberts, J. Ryberg (Eds), *Predictive Sentencing: Normative and Empirical Perspectives*, Bloomsbury Publishing, 2019, p. 213; A. Diener, *Pennsylvania's Proposed, Questionably Constitutional, Risk Assessment Instrument*, in *Harvard Civil Rights – Civil Liberties Law Review*, 17th October 2019; A. Bashir, *Pennsylvania's Misguided Sentencing Risk-Assessment Reform*, in *The Regulatory Review*, 5th November 2019.

Actually, the spread of computational risk-assessment instruments within the United States criminal justice system is only the last step of a more general tendency of using predictive tools based on statistical-actuarial theories and methods.[4] As highlighted by some legal scholars,[5] the scientific theory on which the algorithm is developed and the "computational" nature of the risk-assessment tool should actually be distinguished,[6] at a conceptual level, in order to properly analyse the different problems arising from each of the two elements.[7]

2.1.1. Decisions on pretrial release

As regards the decisions about pretrial release, in the United States courts there are more than 20 different risk assessment tools (RATs) in use.[8] In seven States[9] the judges are required to take into account the outcome of these tools, at least in certain

[4] About the spread of evidence-based practices (EBP), see KEHL, GUO, KESSLER, *Algorithms in the Criminal Justice System*, cit., p. 7. In this regard, some criticism has been expressed by N. SCURICH, *The case against categorical risk estimates*, in *Behavioral Science Law*, 2018, p. 1 et seq.

[5] S. QUATTROCOLO, *Equo processo penale e sfide della società algoritmica*, in *BioLaw Journal*, 1/2019, p. 135 et seq.; EAD., *Questioni nuove e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cass. pen.*, 4/2019, p. 1478.

[6] See QUATTROCOLO, *Equo processo penale e sfide della società algoritmica*, cit., p. 144. For example, the predictive system "SAVRY" (*Structured Assessment of Violence Risk in Youth*), used in juvenile justice by at least nine of the United States, is a tool with an algorithmic structure, but without a digital/computational nature: it consists of a user manual accompanied by assessment forms. About SAVRY, see also G.M. VINCENT, J. CHAPMAN, N.E. COOK, *Risk-Needs Assessment in Juvenile Justice: Predictive Validity of the SAVRY, Racial Differences, and the Contribution of Needs Factors*, in *Criminal Justice & Behavior*, 2011, p. 47; QUATTROCOLO, *Questioni nuove e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, cit., p. 1478.
For a broad definition of algorithm, see T. GILLESPIE, *The relevance of Algorithms*, in T. GILLESPIE, P. BOCZKOWSKI, K. FOOT, *Media Technologies*, MIT Press, 2014, 1: «*Algorithms need not be software: in the broadest sense, they are encoded procedures for transforming input data into a desired output, based on specified calculations*». See also the study carried out by the Council of Europe, *Algorithms and Human Rights*, DGI (2017) 12, 5, in <www.coe.it>, 22nd March 2018, 5.

[7] See *infra* § 4.1.8.

[8] A. WIDGERY, *The Statutory Framework of Pretrial Release*, in *www.ncsl.org*, 8th November 2020, 7; A.Z. HUQ, *Racial Equity in Algorithmic Criminal Justice*, in *Duke Law Journal*, 2019, 1043. In the Italian literature, see M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, in *Dir. pen. cont.*, 29th May 2019, 4.

[9] Alaska, Delaware, Hawaii, Indiana, Kentucky, New Jersey and Vermont.

cases, whilst in eight other States[10] their adoption is authorised and encouraged.[11] Some legislative interventions at a state level[12] have introduced regulations on the use of predictive tools in pretrial release decisions: requirements of impartiality and non-discrimination, guarantees of transparency and accessibility of the systems and data processed by the algorithms, as well as the need of periodic validation review of the predictive instrument.[13]

When examining the United States experience, a warning is needed: in that legal tradition, which provides money bail as a condition for the release of the arrested, the introduction of algorithmic risk assessment tools is strictly connected to the debate on the abolition of bail.[14] The case of California could be an example of what just mentioned. In 2018, the State of California abolished money bail and replaced it with the obligation of making use of risk assessment tools to decide whether to release the arrested person and on what conditions. Nevertheless, the reform was repealed in 2020, after a referendum, which was supported by a diverse group of stakeholders: on the one hand, the insurance companies interested in keeping bail money (the so-called "bail industry") and, on the other, some human rights organisations which, though supporting the abolition of bail, have always pointed out the risk of machine bias and the lack of transparency of predictive algorithms.

That said, among the algorithms employed in the context of pretrial decisions, the most used is "PSA" (Public Safety Assessment).[15] It is developed by a non-profit

[10] Colorado, Illinois, Montana, New York, Pennsylvania, Rhode Island, Virginia and West Virginia.

[11] Widgery, *The Statutory Framework of Pretrial Release*, cit., p. 7.

[12] For example in Idaho, New York and California.

[13] See, for example, House Bill 118 in Idaho, Senate Bill 1509 in the State of New York and Senate Bill 36 in California.

[14] Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 8; J.L. Koepke, D.G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, in *Washington Law Review*, 4/2018, p. 1725. On money bail, in the Italian literature, see V. Tondi, *Il* Bail. *La libertà su cauzione negli ordinamenti anglosassoni*, Cedam-Wolters Kluwer, 2016.

[15] In this regard, in the Italian literature, see Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 7; F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, 29th September 2019, p. 18.

organisation ("Laura and John Arnold Foundation") and is used in four States[16] and in numerous important jurisdictions in the United States.[17] The input of the algorithm consists of nine risk factors for the individual,[18] which are compared with a database of about 750,000 cases from about 300 jurisdictions. The output produced by the system consists of a score of 1 to 6 for each of the three "risks" to assess: Failure To Appear (FTA), New Criminal Arrest (NCA) and New Violent Criminal Arrest (NVCA). The criteria for allocating the scores are published online.[19]

2.1.2. Sentencing

For the sentencing stage, more than 60 different risk assessment tools are employed[20] and the use of predictive instruments is mandatory in some States.[21] Also the well-known "Loomis case,"[22] which put the "COMPAS" algorithm at the centre of the discussion all over the world, was connected to the sentencing context.

COMPAS is a computational predictive tool developed in 1998 by the company Northpointe (now Equivant). The system works on data sampled from over 30,000 COMPAS assessments conducted between January 2004 and November 2005

[16] Arizona, Kentucky, New Jersey and Utah.

[17] Among those, Allegheny County (Pittsburgh, Pennsylvania), Cook County (Chicago, Illinois), Harris County (Houston, Texas), Mecklenburg County (Charlotte, North Carolina), Milwaukee County (Wisconsin) and San Francisco County (California). See "*Where is PSA currently used?*" at *advancingpretrial.org*.

[18] The elements considered are: 1) age at current arrest; 2) current violent offence; 3) pending charge at the time of the arrest; 4) prior misdemeanour conviction; 5) prior felony conviction; 6) prior violent conviction; 7) prior failure to appear in the past 2 years; 8) prior failure to appear older than 2 years; 9) prior sentence to incarceration. See *"How It Works*", at <advancingpretrial.org>.

[19] See *"How It Works*", at <advancingpretrial.org>.

[20] A.Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, cit., p. 1075; A.M. Barry-Jester, B. Casselman, D. Goldstein, *Should Prison Sentences Be Based on Crimes That Haven't Been Committed Yet?*, in <fivethirtyeight.com>, 4th August 2015. More generally, on the use of algorithms in sentencing, see Kehl, Guo, Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, cit., p. 13.

[21] Kehl, Guo, Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, cit., 16; Huq, *Racial Equity in Algorithmic Criminal Justice*, cit., 1075. For example, the legislations of Arizona, Kentucky, Ohio and Pennsylvania require the use of predictive tools in the sentencing stage. Ohio and Pennsylvania have also developed their own risk assessment algorithm.

[22] *State* v. *Loomis*, 881 N.W.2d 749, 753 (Wis. 2016).

across the United States.[23]

The input of the algorithm is made of the information collected from the file and the answers given to a series of questions by the person concerned.[24] The output consists of a *risk* assessment and a *needs* assessment. Within the *risk* assessment, a score on a scale of 1 to 10 is assigned for each of the three risks of recidivism considered: pretrial recidivism risk, general recidivism risk and violent recidivism risk.

The framework of the elements considered and the results returned by COMPAS is therefore very complex and shows that the algorithm was not initially developed for the sentencing stage[25] but as an aid for judges and other competent authorities in the field of pretrial release decisions and in the sentence enforcement (e.g. for the admission to parole).

2.2. European tendencies and experiences

2.2.1. HART (United Kingdom)

In the European experience, the algorithm that can be compared to the predictive tools employed in the United States is called "HART" (Harm Assessment Risk Tool)[26] and is being tested in the United Kingdom. It was developed in a partnership between the University of Cambridge and the Constabulary of Durham, with the purpose of assessing the risk of recidivism two years after the arrest. The

[23] *Practitioner's Guide to COMPAS Core* (2019), at <www.equivant.com>, 11.

[24] *State* v. *Loomis*, cit., § 13. The COMPAS algorithm can consider up to 137 factors (see Ronsin, Lampos, Maîtrepierre, *In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data*, in *Appendix I*, *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, at <www.coe.int>, § 129; J. Dressel, H. Farid, *The accuracy, fairness, and limits of predicting recidivism*, in *Science Advances*, 4/2018, 1; J. Nieva Fenoll, *Intelligenza artificiale e processo*, translated by P. Comoglio, Giappichelli, 2019, p. 59).

[25] Kehl, Guo, Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, cit., p. 11.

[26] In this regard, M. Oswald, J. Grace, S. Urwin, G.C. Barnes, *Algorithmic risk assessment policing models: lessons from the Durham HART model and "Experimental" proportionality*, in *Information and Communications Technology Law*, 2/2018, pp. 223 et seq. In the Italian literature, Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 10.

algorithm was designed for a better implementation of "Checkpoint,"[27] a diversion programme aimed at avoiding criminal prosecution and conviction for the offenders whose risk of recidivism is lower or related to less-serious offences.

In this context, HART has the function of assisting the police in the selection of those eligible for the programme. It places the offenders into three risk categories regarding the risk of committing an offence in the following two years: a) high risk (risk of committing serious offences); b) moderate risk (risk of committing less-serious offences); c) low risk (no risk of recidivism).[28] The Checkpoint programme is for the offenders placed in the second category.

The input of the algorithm consists of 34 predictive factors: 29 are connected with the individual's criminal history; the others concern age, gender, two different post codes and the number of "police intelligence reports" on the offender.[29] HART is built upon a database of 104,000 cases occurred in Durham between 2008 and 2012.[30]

A significant aspect is the different relevance given to "dangerous errors" (false negatives) and "cautious errors" (false positives) at the stage of programming the algorithm: the choice in favour of greater accuracy of the low risk assessments entails an overestimation of high-risk individuals.[31] False negatives are thus very rare, whilst the possibility of obtaining a significant percentage of false positives should be taken into consideration.

One of the most widely discussed issues within the civil society debate – also in the light of some studies carried out by non-profit organisations for the protection of fundamental rights – concerns one of the two post codes included in the predictive elements which are inserted in HART's input: the *Mosaic code.* It is a geo-demographic

[27] Information on the "Checkpoint" project, implemented since 2015, is available on the Durham Constabulary website <www.durham.police.uk>. In this regard, see Oswald, Grace, Urwin, Barnes, *Algorithmic risk assessment policing models*, cit., 227. On the use of HART, since 2017, see S. Carlo, *Big Brother Watch's written evidence on algorithms in the justice system for the Law Society's Technology and the Law Policy Commission*, in *bigbrotherwatch.org.uk*, 2.

[28] Oswald, Grace, Urwin, Barnes, *Algorithmic risk assessment policing models*, cit., p. 227.

[29] Oswald, Grace, Urwin, Barnes, *Algorithmic risk assessment policing models*, cit., p. 228.

[30] *Ibidem.*

[31] Oswald, Grace, Urwin, Barnes, *Algorithmic risk assessment policing models*, cit., p. 230.

tool, developed and sold by a marketing company called "Experian,"[32] which profiles adult residents in the United Kingdom by placing them into 66 categories and works on a base of 850 million data (amongst which there are family composition, ethnicity, online data, occupation, health data, gas and electricity consumption and school performance).[33]

2.2.2. Predictive justice algorithms in civil justice and "judicial analytics" (France)

Even if still limited to the context of civil justice, some predictive tools designed in France should be taken into consideration for their potential impact on the justice system in general.

Since 2016, when *Loi* No. 2016-1321 «*sur la République numérique*» made all the judicial decisions by French courts available to the public, some start-ups have used that huge open access database to develop predictive algorithms aimed at supporting lawyers in their work:[34] the main service offered is the indication of the probability of success for a lawsuit, which could be useful for deciding whether undertaking a legal action and for choosing the best strategy.[35]

The algorithms which have raised the greatest perplexity are those giving relevance to the identity of the individual judge in order to analyse his/her work. Actually, the issue did not emerge from the use of commercial applications designed by LegalTech start-ups, but from the publication of a study on the website "SupraLegem" which provokingly highlighted the significant variance in the percentages of rejection of the appeals by the different judges of the *Conseil d'Etat* who were competent for asylum law cases.[36]

[32] For a description of the product see the company's website at the following link: <www.experian.co.uk/assets/marketing-services/brochures/mosaic-ps-brochure.pdf>.

[33] See Carlo, *Big Brother Watch's written evidence on algorithms in the justice system for the Law Society's Technology and the Law Policy Commission*, cit., 1; Big Brother Watch Team, *A closer look at Experian Big Data and Artificial Intelligence in Durham Police*, at <bigbrotherwatch.org.uk>, 6th April 2018; Big Brother Watch Team, *Police uses Experian marketing data for AI custody decisions*, at <bigbrotherwatch.org.uk>, 6th April 2018.

[34] They are developed by *Predictice*, *Case Law Analytics*, *Doctrine.fr*, *Tyr Legal*. See C. Szwarc, *La justice predictive: une autre justice?*, in *Le Mag des Avocats*, 34, 9/2017, p. 5.

[35] Tests of one of these algorithms (*Predictice*) have been promoted by the French Ministry of Justice at the Courts of Appeal of Rennes and Douai. See C. Castelli, D. Piana, *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Quest. giust.*, 4/2018, p. 156.

[36] The study was conducted by the lawyer Michaël Benesty and the IT engineer Anthony Sypniewski. On this issue,

In 2019, also in the light of the protests promoted by the judges against predictive tools, the French legislature provided that the judges' names and the data that could identify them must be removed from the material made available to the public: the main reason for this intervention was the foreseeable risk that profiling each individual judge could affect the judiciary's independence.[37] The Parliament also criminalised the use of the judges' data having the purpose or the effect of assessing, analysing or comparing them or to predict the judges' future decisions.[38]

3. Possible fields of application for AI in criminal justice

After examining the current experiences of the use of AI systems in the context of criminal justice, it is possible to identify the fields of relevance of those tools and place them within the categories of criminal law and procedure. The perspective will be mainly that of the Italian system, even if some reference to the comparative framework will be made.

3.1. Assessment of the re-offending risk

As seen above, the majority of AI systems experienced in criminal justice are employed for assessing the re-offending risk in its different spheres of relevance, such as the decisions on pretrial release or the sentencing.

In the Italian system, a prognostic assessment of the re-offending risk can be relevant, first of all, in the context of decisions about pretrial precautionary measures.

The Code of Criminal Procedure requires the existence of one of the three so-called *pericula libertatis* listed in Article 274 of the Code of Criminal Procedure, as well

see M. Langford, M.R. Madsen, *France Criminalises Research on Judges*, in *verfassungsblog.de,* 22nd June 2019; M. Benesty, *The Judge Statistical Data Ban. My Story*, at <www.artificiallawyer.com>, 7th June 2019.

[37] B. Galgani, *Considerazioni sui "precedenti" dell'imputato e del giudice al cospetto dell'IA nel processo penale*, in *Sist. pen.*, 4/2020, p. 87.

[38] See article 33, al. 3, Loi n. 2019-222: «*Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées*».

as «serious evidence of guilt» (Article 273 of the Code of Criminal Procedure), for the application of a pretrial precautionary measure. The measures grounded on the first two *pericula* – regarding the risk for evidence and the flight risk – are clearly connected with procedural purposes. The measure ordered because of the third one (Article 274, letter c, of the Code of Criminal Procedure) aims at dealing with the risk of committing «serious crimes»[39] or offences «of the same type of the one under prosecution» and punished by law with a maximum sentence of at least four years (at least five years for ordering prison custody), thus performing a special-preventive function that is not always strictly connected with the proceeding within which the measure is ordered.[40]

Moreover, such examples of pretrial coercive measures – without a strictly procedural function but aimed at special-preventive needs – are not an Italian peculiarity: as a matter of fact, measures with a similar aim can be found in all the main European criminal procedure systems.[41]

A second area of relevance for the risk of committing new offences is the stage of sentencing.

In the Italian system, the decision on the measure of the penalty to be applied is given by the same judge who establishes whether the accused is guilty. The provision about the sentence's measure (Article 133 of the Criminal Code) lists a series of elements which the judge must take into account, divided into aspects regarding the seriousness of the offence and those regarding the so-called "*capacità a delinquere*"[42] (capacity to

[39] More precisely, Article 274, letter *c*, of the Italian Code of Criminal Procedure refers to serious offences with the use of weapons or any other tool against persons or the constitutional system, or offences connected to organised crime.

[40] For some criticisms about the third *periculum* and its special-preventive (and thus non-procedural) function, see F. Callari, *Il* periculum libertatis *costituito dal rischio di realizzazione di determinati reati e le misure cautelari: il fine giustifica i mezzi?*, in *Dir. pen. cont.*, 12th November 2012, 1.

[41] See, for example, Article 144 of the French Code of Criminal Procedure, § 112a of the German Code of Criminal Procedure and Article 503 of the Spanish *Ley de Enjuiciamiento Criminal*. On the issue, also for further references, see R. Vogler, S. Fouladvand, *Standards for making factual determinations in arrest and pre-trial detention: a comparative analysis of law and practice*, in J.E. Ross, S.C. Thaman (Eds), *Comparative Criminal Procedure*, Edward Elgar Publishing, 2016, p. 191.

[42] On this notion, see, for all, G. Fiandaca, E. Musco, *Diritto penale. Parte generale*, VIII ed., Zanichelli, 2019, p. 796.

offend). This latter must be deducted from *i)* the reasons for offending and the nature of the crime; *ii)* criminal records; *iii)* the conduct at the time of and after the crime; *iv)* the person's individual, family and social conditions. That said, it is easy to find a similarity between the elements from which the so-called capacity to offend is deducted and those included in the input of the predictive algorithms used in the United States.

The indicators for determining the sentence are not different, on a general level, to those provided in the main European systems: although a concept similar to the Italian "*capacità a delinquere*" cannot be found in the comparative framework, the French, Spanish and German Criminal Codes do mention personality, family and socio-economic situation of the offender among the elements to be taken into account for establishing the sentence.[43]

Also the identity of the judge competent for the decision about the accused's guilt and for establishing the sentence is common to the main civil law systems. That differentiates them from the common law ones, in which the judgement is structured in two stages and the sentencing hearing – for establishing the sentence to be applied – takes place after the conviction, before a different judge.

A third context, in which the recidivism risk is relevant, is the decision on the applicability of some provisions, for which the system requires, explicitly or implicitly, the exclusion of re-offending risk. Besides the so-called "*sospensione condizionale*" (Articles 163 of the Italian Criminal Code), for which this requirement is clearly stated, several other measures provided by Italian criminal law should be mentioned, which entail a lower level of special-preventive needs: the "*sanzioni sostitutive*" (substitute penalties) provided by Law No. 689 of 1981, the "alternative measures" provided by the Prison Law (*i.e.* Law No. 354 of 1975) and other measures such as release on temporary licence (Article 30-*ter* of the Prison Law) and conditional release (Article 176 of the Italian Criminal Code).

[43] For example, Article 132-1 of the French Criminal Code refers to the personality of the offender and his/her economic, family and social situation ("*la personnalité de son auteur (…) sa situation matérielle, familiale et sociale*"), § 46 of the German Criminal Code includes the personal and economic conditions of the offender ("*seine persönlichen und wirtschaftlichen Verhältnisse*") among the elements considered for establishing the sentence, Article 66 ap. 6 of the Spanish Criminal Code provides that the personal circumstances of the convicted person ("*las circunstancias personales del delincuente*") must be taken into account.

Also the comparative framework is full of similar measures, provided with the same function and whose applicability is therefore grounded on the exclusion of recidivism risk.[44]

Further spheres of relevance for the re-offending risk can be found in the regulation of the security measures (*misure di sicurezza*)[45] and the *ante delictum* preventive measures (*misure di prevenzione*),[46] for which a judgement of "social dangerousness"[47] of the person concerned is required.

Also in the German and Spanish systems[48] a dual-track (*doppio binario*) of penalties and "security measures" can be found. In France and the United Kingdom, instead, no system of "security measures" is provided within the criminal law framework: the penalty is the only criminal measure and the health system is competent for taking charge of people who have been acquitted because of mental disorder but are still considered dangerous.[49] The absence of criminal measures in these cases, however, does not exclude completely a component of coercion, since, for example, strict procedural requirements could be provided for the release of patients placed in a psychiatric hospital.[50] Moreover, the tendency of the last decades is to introduce security measures for imputable persons, which therefore come in addition to the penalty: that is happening also in those systems, like the German[51] and Spanish[52] ones, which, unlike the Italian, were characterised by a "*pure* dual-track" (*doppio binario puro*), with a clear

[44] See, for example, the measures provided by the French (Articles 132-29 and 132-40), German (§ 56) and Spanish (Article 80) criminal codes.

[45] About the personal security measures, also in a comparative perspective, see for all M. Pelissero, *Pericolosità sociale e doppio binario. Vecchi e nuovi modelli di incapacitazione*, Giappichelli, 2008, *passim*.

[46] On social dangerousness in the regulation about personal *ante delictum* preventive measures, see A. Martini, *Essere pericolosi. Giudizi soggettivi e misure personali*, Giappichelli, 2017, spec. p. 81 et seq., p. 121 et seq.

[47] For an overview of the fields of relevance of the "social dangerousness", see Martini, *Essere pericolosi. Giudizi soggettivi e misure personali*, Giappichelli, 2017, *passim*.

[48] See §§ 66 et seq. of the German Criminal Code and Article 96 et seq. of the Spanish Criminal Code.

[49] Pelissero, *Pericolosità sociale e doppio binario*, cit., p. 151 et seq.

[50] Pelissero, *Pericolosità sociale e doppio binario*, cit., p. 163 et seq., p. 171 et seq.

[51] Pelissero, *Pericolosità sociale e doppio binario*, cit., spec. p. 161 et seq., p. 255 et seq.

[52] J. Lascuraín Sánchez (dir.), *Manual de Introduccion al Derecho Penal*, Agencía Estatal BOE, 2019, 280; X. Etxebarria Zarrabeitia, *Medidas de seguridad: presupuestos de su aplicación*, in *Cuadernos penales José María Lindón*, 10/2014, p. 125 et seq.

distinction between those concerned with punishment (imputable persons) and those with security measures (non-imputable persons). The area of relevance of the so-called "social dangerousness" – that is basically the risk of re-offending – appears ever wider.

With regard to the *ante delictum* preventive measures regulated by Legislative Decree No. 159 of 2011, it is difficult to find similar provisions in the comparative framework. Those measures, as a matter of fact, though already widely enforced during Fascism and even before, have survived also in the current constitutional context, since they are strictly connected to the contrast to organised crime.

3.2. Predicting the outcome of judicial decision-making

Besides the possibility of using tools that assess the re-offending risk, the potential of the systems aimed at predicting the outcome of judicial decision-making requires to be framed within the categories of criminal justice.

Although the main experience at the moment is the French one, which consists in a defence tool used by lawyers and is limited to the civil justice field, it is not difficult to imagine the predictive instruments' potential impact on the activity of prosecutors and judges, whether employed in the criminal justice system.

In the Italian debate, for example, some have taken into consideration the possible uses of predictive justice tools for the decision by the Prosecutor's office about going on with criminal prosecution.[53]

In a system like the Italian one, which is informed by the principle of mandatory criminal prosecution,[54] the choice between prosecuting the suspect and dropping the charges is made according to the parameter of the «groundlessness of the *notitia criminis*»[55] (*infondatezza della notizia di reato*), which is defined as the insufficiency of

[53] C. Parodi, V. Sellaroli, *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in *Dir. pen. cont.*, 6/2019, p. 64.

[54] The mandatory criminal prosecution is provided also in Germany (§ 152, II, StPO) but it has been progressively eroded by the introduction of an area of discretionary power for the Prosecutor with reference to minor offences (*Bagatelldelikte*). In this regard, see M. Delmas-Marty, M. Chiavario (Eds), *Procedure penali d'Europa*, Cedam, 2001, 193. About the area of discretion which emerges, as a matter of fact, within the activity of the prosecutors and about priority criteria for criminal prosecution in Italy, see N. Galantini, *Il principio di obbligatorietà dell'azione penale tra interesse alla persecuzione penale e interesse all'efficienza giudiziaria*, in *Dir. pen. cont.*, 23rd September 2019, p. 1 et seq.

[55] For this translation, see Gialuz, L. Lupària, F. Scarpa, *The Italian Code of Criminal Procedure. Critical Essays and*

the evidence gathered during preliminary investigations in order to uphold the accusation before the court at the trial stage.[56] In this context, an algorithm indicating the probability of conviction could assist the Prosecutor in their assessments.[57]

In those legal systems which instead provide a significant discretionary power within the decision-making about criminal prosecutions, like the French one and the common law ones,[58] a predictive tool like the one imagined could help in programming the activities of the prosecution offices, guiding their choices according to predetermined criteria and guidelines.

A further scenario, which has been examined in the debate over the last few years,[59] is the use of predictive algorithms on the part of the judge, even if, at the moment, it does not find significant confirmation in the practice of criminal justice. Algorithms like those developed in France for lawyers in the civil justice field could be a support for the judge in the preliminary case-law research which is required for the decision-making, even in systems that do not provide a rule of *stare decisis*.

4. Risks of the use of AI technologies in the criminal justice system

The paragraphs above have tried to provide an overview of the current – and some potential – uses of AI in criminal justice, in an attempt to frame its possible uses in the context of the main European systems. The problematic issues arising in the last few years with reference to predictive algorithms shall now be examined.

English Translation, Wolters Kluwer, 2014, p. 324.

[56] See Article 125 of the provisions for the implementation of the Italian Code of Criminal Procedure.

[57] In any case, even if a predictive tool indicating the likelihood of conviction when certain evidence is available could be developed, the possibility that an algorithmic tool could assess the evidence or establish that a person is guilty "beyond any reasonable doubt" is currently to be excluded.

[58] Delmas-Marty, Chiavario (Eds), *Procedure penali d'Europa*, cit., pp. 119, 244.

[59] See O. Di Giovine, *Il* judge-bot *e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale)*, in *Cass. pen.*, 3/2020, p. 951 et seq.; A. Massaro, *Determinatezza della norma penale e calcolabilità giuridica*, Editoriale Scientifica, 2020, 501; Basile, *Intelligenza artificiale e diritto penale*, cit., p. 14.

4.1. Risk-assessment tools

As mentioned above, the systems for assessing the risk of recidivism, already in use in the United States and experienced also in Europe, have a primary position in the discussion about the use of AI tools in criminal justice.

A lot of critical issues have emerged, on both sides of the Atlantic, in the discussion among legal scholars and in the civil society debate.

Each of the arising problems shall be synthetically examined and put in relation to the general principles of criminal law and procedure which could be affected by the introduction of "intelligent" risk assessment systems. At the end of this overview, a distinction between the issues connected with the computational nature of the predictive tools and those linked to statistical/actuarial theories at the basis of the algorithms should be considered.[60]

4.1.1. De-individualisation of decisions

One of the main problems emerging from the use of risk-assessment algorithms is that the "predictive" output is obtained by extending to the individual case the outcome of statistical processing carried out on a sample of cases.[61] A decision grounded (exclusively) on the outcome of the algorithmic tool would necessarily be the product of a generalisation and could not be considered as an individualised assessment of the case at stake.

In the context of sentencing, the criticism just mentioned could end up affecting the fundamental principle of culpability, which requires an individual assessment not only for deciding about conviction but also for establishing the sentence.

Whether considered in the context of pretrial precautionary measures, the risk of a de-individualisation of the decision could be contrary to the guarantees aimed at protecting personal liberty: the limitation of that fundamental right can only be ordered

[60] For this distinction, see QUATTROCOLO, *Equo processo penale e sfide della società algoritmica*, cit., p. 144.

[61] See L. MALDONATO, *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale*, in *Dir. pen. cont.*, 2/2019, p. 411; L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena. A proposito dell'esperienza statunitense nel c.d.* evidence-based sentencing, in *Dir. pen. cont.*, 2/2019, p. 356; GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 21.

as *extrema ratio*, when relevant precautionary needs require coercive measures with specific reference to the person concerned.

4.1.2. Discriminatory effects

The risk of de-individualisation of decisions, as described above, and the use of statistical generalisations appear strictly connected to another critical issue concerning the discriminatory effects[62] of the use of risk-assessment tools in criminal justice (which is probably the most well-known and controversial aspect, also in the public opinion debate).

If the algorithm has been "trained" on a data-set consisting of criminal records, it could easily reproduce existing discriminations, connected to the conditions of social marginalisation of certain communities, even when ethnicity data or, more generally, the data about being part of a minority group are not included in the system's input.[63] The functioning of the most advanced computational tools is based – besides on the data provided at the time of programming the algorithm – on the connections identified by the algorithm itself by measuring the frequency of certain recurring elements. The system could therefore recognise "false" connections between data which are not linked at an etiological level[64] and thus could end up producing discriminatory outcomes.

A problem of compatibility with the principle of equality is therefore arising, both with reference to decisions on pretrial detention and to the sentencing stage.

4.1.3. Criminal determinism

Moreover, with regard to sentencing, it is possible that a predictive system affected by implicit "prejudices" against specific categories may lead to a more severe sentence as a consequence of belonging to a certain group: in other words, a person could

[62] See MALDONATO, *Algoritmi predittivi e discrezionalità del giudice*, cit., p. 407; D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, cit., p. 364; GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 21.

[63] MALDONATO, *Algoritmi predittivi e discrezionalità del giudice*, cit., p. 407.

[64] See RONSIN, LAMPOS, MAÎTREPIERRE, *In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data*, cit., § 56 et seq., spec. § 66, 71.

be punished not for what he/she *has done*, but for what he/she *is*, according to a pattern which recalls the model of "*Tätertyp*" or "*tipo d'autore*."[65] That appears to be inconsistent with the so-called materiality principle (*nullum crimen sine actione*) and the harm principle (*nullum crimen sine iniuria*), which are vital for any liberal criminal system.

The criticisms just described – which is connected to the de-individualisation of decisions, arising from the statistical origin of the predictive outcome, to its tendency to make generalisations, and to the consequent discriminatory effects – entails the risk of moving towards a *deterministic* approach by reintroducing some outdated assumptions of the Italian Positive School.

4.1.4. Lack of transparency

Besides the criticisms around the generalising and potentially discriminatory effect of the predictive outcomes, one of the most widely discussed problems concerning the use of AI tools in the criminal justice system is about the lack of transparency of the predictive algorithms' mechanisms.

The impossibility for the person concerned by the judicial decision to access the system could be in conflict with the principles connected to fair trial and the right of defence: the proceedings would be carried out with a strong limitation to the principles of cross-examination and equality of arms, since the defence cannot dispute the validity of the predictive tool and its outcome. Moreover, the right of appeal would be essentially denied, whether the defence lacks adequate tools for contesting the judicial reasoning which makes reference to the output of the predictive algorithm.

The inaccessibility of computational risk-assessment tools is normally connected to the "commercial" nature of many predictive algorithms in use in the United States,[66] whose source codes are under trade secret and thus cannot be disclosed to either the judge or the parties (or the experts that may intervene in the proceeding). This issue could probably be solved by providing a mandatory publication of the source code and

[65] See MANES, *L'oracolo algoritmico e la giustizia penale*, cit., 17; GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 21.

[66] *Supra*, § 1.1. A distinction between algorithms developed by public institutions, "commercial" algorithms and non-profit algorithms has been suggested by GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 5.

other relevant information for accessing the algorithm, or through the provision of competitive procedures for the predictive services' providers, or, more radically, by prohibiting the use of "private" tools and entrusting public administrations with the development of the algorithms to be used in criminal justice systems.

In any case, the principle of equality of arms and the right to defence would be frustrated, whether the analysis of the algorithm in order to dispute its validity requires high specialisation. The problem is not new, since similar problems arise in every proceeding within which scientific evidence is admitted, thus requiring the intervention of experts. Nevertheless, with reference to predictive algorithms, the problem could be more serious because of the potential extent of the scope of application of these tools.

Moreover, it is not completely clear whether the access to the source code is enough to allow full knowledge of the tool's mechanisms, especially whether it concerns algorithms able at developing their self-learning proceedings (so-called "machine learning").[67]

4.1.5. Risks for the system's and data's integrity

Further criticism could also arise from the computational nature of predictive algorithms: the risk of alterations of the system or intrusions aimed at influencing the decision-making should be taken into account. For example, someone might illicitly intervene on the algorithm or on the data[68] in order to guide the judge towards a specific decision (favourable to the intruder or unfavourable to a political opponent or economic rival) or might manipulate the system without a specific purpose, but still conditioning the decision-making process.

Of course, this could be avoided through technological protection measures designed to guarantee the security and integrity of the systems used in criminal justice. The harmful potential of unauthorised intrusions, however, deserves to be considered, with special reference to the independence of judges.

[67] In this regard, a reference to the «ontological inaccessibility of the functioning mechanism» is made by Maldonato, *Algoritmi predittivi e discrezionalità del giudice*, cit., p. 408. See also J.P. Davis, *Law without mind. AI, Ethics and Jurisprudence*, in *University of San Francisco Law research Paper*, 2018, p. 6.

[68] See Di Giovine, *Il* judge-bot *e le sequenze giuridiche in materia penale*, cit., p. 957.

4.1.6. Cognitive biases

The use of quantitative and automated risk-assessment tools in criminal justice systems might cause cognitive errors or distortions in the judicial decision-making process. This might entail a substantial delegation of the decision to the computational tool, even whether the algorithm's output is not considered as binding for the judge.[69]

The judge who receives the outcome of a predictive tool could run into a first logical error, connected to the cognitive bias known as "anchoring":[70] it refers to the influence of the first information obtained on all the following steps of a decision-making process.

Even assuming that this phenomenon could be avoided or reduced – for example by introducing the contribution of the predictive tool only after the judge has autonomously come to a decision[71] – there would still be the risk of further cognitive distortions, connected to so-called "automation bias":[72] when using computational risk-assessment tools, the decision-makers (such as the judges) would tend to prefer the solutions suggested by the automated system over those reached without their use.

Both the "cognitive prejudices" examined lead to an excessive and unjustified reliance on the predictive tool's outcome. This means that, even provided that algorithms are used merely as "advisors" or "assistants" to the judge, the decision-making concerning the re-offending risk might end up being delegated to the predictive tool.[73]

4.1.7. Contamination of the decision on the accused's guilt

A further problem – which is also connected to cognitive errors – is that the introduction of evidence regarding the accused's personal conditions and his/her

[69] See BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 22.

[70] In this regard, see MALDONATO, *Algoritmi predittivi e discrezionalità del giudice*, cit., 410; NIEVA FENOLL, *Intelligenza artificiale e processo*, cit., p. 38.

[71] The use of predictive tools as a "double check" instrument after the judge's assessment has been carried out is taken into consideration by MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 20.

[72] In this regard, G. UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sist. pen.*, 11th November 2020, 4; P. COMOGLIO, *Prefazione*, in NIEVA-FENOLL, *Intelligenza artificiale e processo*, italian translation by P. Comoglio, Giappichelli, 2019, X.

[73] BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 22.

personality might lead to a contamination of the decision on his/her guilt[74] and raise some concerns with reference to the principles of materiality (*nullum crimen sine actione*) and harm (*nullum crimen sine iniuria*), as well as with the presumption of innocence.

Such a contamination is what the Italian legislature intended to prevent through Article 220, paragraph 2, of the Code of Criminal Procedure, which forbids any assessment aimed at establishing the character or personality of the accused or his/her «psychic qualities independent of pathological causes.» In the perspective of the legislator of the time, a similar measure was necessary to avoid undue contaminations and alterations of the decision on the accused's responsibility, within a system like the Italian one, which – as well as other European legislations – does not provide a separation of the judgement on the defendant's guilt and the sentencing stage.[75]

4.1.8. Some observations

After the synthetic overview of the emerging criticisms with reference to risk-assessment algorithms in the criminal field, it is possible to connect some of the problems to the tools' computational nature and others to the statistical-actuarial character of the scientific theories according to which the algorithm has been designed.[76]

For example, at the origin of the risk of de-individualising the decision there is the use of generalisations based on outputs of statistical nature. Also the discriminatory outcomes seem to be linked, first and foremost, to the inevitable generalising effect connected to the use of statistical data and methods.[77]

Even the perplexities of a wider range – regarding the possible tendencies towards criminal determinism – seem to come from the use of generalising methods, which lead to reasoning by categories (or "types") of offenders.

[74] In this regard, see Di Giovine, *Il* judge-bot *e le sequenze giuridiche in materia penale*, cit., p. 959; Maldonato, *Algoritmi predittivi e discrezionalità del giudice*, cit., p. 411; D'Agostino, *Gli algoritmi predittivi per la commisurazione della pena*, cit., p. 367; A. Garapon, J. Lassègue, *Justice digitale. Révolution graphique et rupture anthropologique*, Puf, 2018, p. 279.

[75] About the two-stage structure of criminal proceedings in common law systems and the autonomy of the sentencing stage, see D'Agostino, *Gli algoritmi predittivi per la commisurazione della pena*, cit., p. 367.

[76] As suggested by Quattrocolo, *Equo processo penale e sfide della società algoritmica*, cit., p. 144.

[77] This however does not exclude that it is the tool's computational nature that allows to quickly process the database and to identify a statistical correlation between some recurring elements and the recidivism prediction.

Otherwise, both the risks arising from the opacity of the algorithms' functioning – the danger of manipulations of the system and the risks concerning the right of defence – seem to be directly connected to the computational nature of these systems.

As for the problematic issues connected with the risk of cognitive biases, on the one hand, there is the so-called "automation bias" which is determined by the "computational nature" of the tool, whilst in other cases (e.g. the "anchoring") the cognitive errors are independent from that aspect.

The observations above cannot (and are not meant to) lead to a minimisation of the problems of AI tools for criminal justice systems: they are only an attempt to take into account the origins of the reported problems, in order to contextualise each of them within broader and well-known issues – one for all, the introduction of scientific evidence in the criminal proceeding – and hence to better identify the appropriate precautions.

4.2. Tools for predicting the outcome of judicial decision-making

Some criticisms can also be highlighted with reference to the use of tools which aim at predicting the outcome of judicial decision-making.

As regards the use of them by private parties, the main problem is the one emerging from the French debate: in that context, someone has underlined the risk of an undue control on the judges' activities, which might be against the judiciary's independence.

Actually, as said above, the problem arose after the publication of a study which analysed the tendencies of the individual judges in the field of asylum law and not from the use of predictive algorithms by law firms. In any case, the issue is not easy to tackle, considering the (equal) constitutional value of all the interests at stake: if, on the one hand, there is the judiciary's autonomy, on the other hand, there are the guarantees regarding the accessibility of judicial decisions as a democratic control of the judicial power, as well as the right to information and the freedom of expression.[78] The solution

[78] In this regard, see RONSIN, LAMPOS, MAÎTREPIERRE, *In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data*, cit., § 43.

provided by the French legislator by prescribing to remove any data referring to the individual judge – and by criminalising the analysis and comparison of those data – still raises some perplexities.[79]

With regard to the use of predictive tools on the part of the judges, the main risk is that of generating a tendency to conform with the dominant direction (the so-called "herd effect"[80]), which might result in a standardisation of law interpretation and judicial decisions[81] and might lead the judges to excessively rely on the algorithmic tool's output.[82]

Actually, a similar effect occurs, to a certain extent, also regardless of the use of predictive algorithms: judges tend to make ever wider reference to case-law precedents within the grounds of their judgements, even in legal systems like the Italian one, which do not recognise the *stare decisis* rule.

Such a "decisional conformism" may be seen partly as a physiological effect of the Supreme Courts' activity, aimed at ensuring homogeneous interpretation and application of the law, and partly as an unorthodox remedy to the need to deal with complex legal issues and to cope, at the same time, with an increasingly heavy workload.

In any case, what should be stressed is not so much the compliance with the dominant case-law in itself, but rather *how* the "prevailing" opinion is identified. As already highlighted with reference to risk-assessment tools, the algorithm – by working on the data-set provided during the programming stage – might identify false correlations between elements that are actually not connected at all in the legal reasoning which grounds the judicial decisions examined.

One last hypothetical application of predicting tools which is worth to be taken into account is that of an algorithm used as an aid for the Prosecutor at the time of choosing between prosecuting the suspect and dropping the charges.[83] In this regard, the principle of mandatory criminal prosecution – in those systems, like the Italian one,

[79] Langford, Madsen, *France Criminalises Research on Judges*, cit.

[80] Di Giovine, *Il* judge-bot *e le sequenze giuridiche in materia penale*, cit., p. 959.

[81] See Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, cit., p. 13, who foresees the risk of an "ossification" of case law.

[82] Basile, *Intelligenza artificiale e diritto penale*, cit., p. 22.

[83] Parodi, Sellaroli, *Sistema penale e intelligenza artificiale*, cit., p. 64.

in which it is provided – does not appear to be at stake, since the decision of dropping the charges would in any case fall within the jurisdiction of a judicial authority (called "*Giudice per le Indagini Preliminari*", i.e. Judge for the Preliminary Investigations).

5. Legal limits to the introduction of AI into criminal justice systems

The introduction of AI systems in criminal justice, as seen above, might conflict with some fundamental principles of criminal law and procedure.

As regards the risk-assessment tools for predicting recidivism, some concerns have been raised with reference to the principle of culpability, to the principle of equality and to the principles of materiality (*nullum crimen sine actione*) and harm (*nullum crimen sine iniuria*); at a procedural level, the problems highlighted concern the right of defence and fair trial, which might be frustrated by the lack of transparency, especially when the algorithm is private-owned and its functioning mechanism is covered by trade secret.

As regards the algorithms that offer a prediction of the outcome of judicial decision-making, and especially the use of them on the part of the judges, the main risk is that of a standardisation of decisions, which, without full clarity on the functioning of the algorithmic tool, might end up eroding the principle of legality and the judiciary's independence.

At this stage, it is necessary to examine the limits that currently exist at an international level, as well as in the context of European Union and national law, for the possible introduction of AI in criminal justice systems.

The "European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment", adopted by the European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe, even if not binding, has certainly a primary relevance at an interpretative level.

In the context of European Union law, the regulations on the protection of personal data are of special importance within the debate on the limits of AI in criminal justice systems.

Finally, with reference to the Italian system, Article 220, paragraph 2, of the Code of Criminal Procedure should be taken into consideration, since it might be seen

as an obstacle to the use of predictive tools for assessing the risk of recidivism.

5.1. The Ethical Charter of the CEPEJ of the Council of Europe

The "European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment"[84] was adopted in 2018 by the European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe.[85]

The document affirms five principles: 1) the «principle of respect for fundamental rights» (ensuring that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights); 2) the «principle of non-discrimination» (specifically preventing the development and intensification of any discrimination between individuals or groups of individuals); 3) the «principle of quality and security» of the data (with regard to the processing of judicial decisions and data, using certified sources and intangible data with models elaborated in a multi-disciplinary manner, in a secure technological environment); 4) the «principle of transparency, impartiality and fairness» (making data processing accessible and comprehensible and authorising external audits); 5) the «principle "under user control"» (precluding a prescriptive approach and ensuring that the users are informed actors and in control over the choices made).

5.2. European Union law: limits to «decisions based solely on automated processing»

In the context of European Union law, the acts adopted in 2016 for the protection of personal data (EU Regulation 679/2016 and Directive 2016/680) include some provisions of great interest for the debate on the introduction of AI systems in justice systems.

For the criminal law field, the main reference is Article 11 of Directive 2016/680, which specifically concerns the «processing of personal data by competent

[84] CEPEJ (2018)14.

[85] The contents of the Charter, with a focus on the main questions for the criminal sector, are analysed by QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della carta etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *Leg. pen.*, 18th December 2018.

authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties.»

The first paragraph of the provision requires the Member States to prohibit any «decision based solely on automated processing which produces an adverse legal effect» on the persons concerned or has otherwise a significant impact on them.[86] Such decisions are admitted only whether authorised by the European Union law or by the law of the State to which the "controller" is subject, and provided that appropriate safeguards are introduced for the rights and freedoms of the "data subject" and, among these, «at least the right to obtain human intervention on the part of the controller.»

Moreover, the second paragraph provides that decisions on criminal convictions and offences must not be based on automated processing of so-called "sensitive data," listed in Article 10. An automated processing of those data is only allowed when suitable measures for the protection of the data subject's rights are provided.

Finally, the third paragraph introduces an absolute prohibition of profiling activities which produce discriminatory effects on «natural persons.»

The interpretation of the notion of «decision based solely on automated processing» is crucial for identifying the scope of application of the prohibition in Article 11 of the Directive (as also in Article 22 of the Regulation)[87].

A restrictive interpretation would limit the prohibition to the case of a decision entirely entrusted to a machine. In this perspective, the use of AI tools would be permitted when it consists of a mere aid to the decision-making process and the human intervention is not limited to endorsing the machine's decision but consists of a substantial assessment of the case.[88]

[86] A similar prohibition is provided by Article 22 of EU Regulation 2016/679 ("Automated individual decision-making, including profiling"). A prohibition of decisions based solely on automated processing had already been established by Article 15 of Directive 95/46/EC.

[87] Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 16.

[88] See M. Brkan, *Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond*, in *Electronic Journal*, 2017, p. 10.
In a similar meaning, see the *Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679* (17/IT; WP 251 rev.01), p. 23, in the amended version adopted on 6th February 2018 by the data protection Working Group, set up by Article 29 of Directive 95/46/EC.

The wording of Article 11, however, also allows to be interpreted in a broader meaning: according to some legal scholars,[89] the prohibition of automated processing would entail a specific rule for evidence assessment, which requires the algorithmic output to be corroborated by other elements.

5.3. Italian law: limits to criminological expert evidence

In the Italian system, there is a significant limit to the use of algorithmic tools for the assessment of the re-offending risk, like those widespread in the United States and tested in the United Kingdom.

The second paragraph of Article 220 of the Italian Code of Criminal Procedure, as a matter of fact, establishes that «expert evidence shall not be admitted to determine whether the accused is a habitual or professional offender or he/she has a tendency to commit offences or to establish the character or personality of the accused and, in general, his/her qualities independent of pathological causes,[90]» unless it is provided for the purposes of executing the sentence or security measure.[91]

At the basis of the choice made by the legislator of 1988 to confirm the regulation provided under the former Code of Criminal Procedure,[92] there were two

[89] GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., 16; G. MALGIERI, G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law,* vol. 7, 1st November 2017, p. 14; MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 20.

[90] For this translation, see GIALUZ, LUPÀRIA, SCARPA, *The Italian Code of Criminal Procedure*, cit., p. 193.

[91] On the prohibition of psychological and criminological expert evidence in the Italian system, see F. ERAMO, *Il divieto di perizie psicologiche nel processo penale: una nuova conferma dalla Cassazione*, in *Dir. pen. proc.*, 7/2007, p. 931 et seq.; P. MOSCARINI, *La perizia psicologica e il 'giusto processo'*, in *Dir. pen proc.*, 8/2006, p. 929 et seq.; P. MARTUCCI, *Il contributo del criminologo nel processo penale: un problema ancora aperto*, in *Dir. pen. proc.*, 6/2004, p. 744 et seq.; I. GIANNINI, *Il dibattito sull'ammissibilità della perizia e della consulenza criminologica nel processo penale*, in *Rass. penit. crim.*, 3/2003, p. 87 et seq.; A. SAPONARO, *L'esame della personalità del reo nel processo penale: evoluzione, involuzione, modelli alternativi, prospettive,* Cacucci, 2000, spec. p. 104 et seq.; P. RIVELLO, voce *Perito e perizia*, *Dig. disc. pen.*, IX, p. 474 et seq.; D. BIELLI, *Periti e consulenti nel nuovo processo penale*, in *Giust. pen.*, 2/1991, p. 65 et seq.; E. AMODIO, *Perizia e consulenza tecnica nel quadro probatorio del nuovo processo penale*, in *Cass. pen.*, 1989, p. 170 et seq.; G. TRANCHINA, *Il divieto di perizia psicologica sull'imputato: una limitazione sicuramente anticostituzionale*, in *Riv. it. dir. proc. pen.*, 1971, p. 1325 et seq.

[92] The wording of Article 314, paragraph 2, of the Italian Code of Criminal Procedure of 1930 was almost identical to the one in force today: «Expert evidence is not permitted to determine whether the accused is a habitual or professional offender, or he/she has a tendency to commit offences, or to establish the character or personality of the ac-

fundamental concerns, related to as many procedural guarantees. First and foremost, the prohibition of criminological and psychological expert evidence was informed by the presumption of innocence, which could be affected by expert evidence on the personality of the accused being introduced when his/her guilt is still to be established. Secondly, the provision examined is grounded on the same concerns for the accused's moral freedom that are at the basis of Article 188 of the Code of Criminal Procedure, which prohibits the use of any «methods or techniques which may influence the freedom of self-determination or alter the capacity to recall and evaluate facts.»[93]

However, two provisions in the Italian procedural system might open to the introduction of a judgement on the personality of the accused within the criminal trial. Article 194 of the Code of Criminal Procedure, while prohibiting testimony on the morality of the accused, admits it when «it concerns specific facts that may be suitable for qualifying his/her personality in connection with the offence and his/her to social danger». Article 236 allows the acquisition of criminal record certificates and final judgements, but also the documents held at social services offices and at the office of the judge entrusted with the supervision of the sentence's enforcement (*Ufficio di Sorveglianza*): all these documents might contain criminological or psychological opinions on the personality of the accused. The case law, however, tends to limit the use of the documentation mentioned in Article 236 only for deciding on the application of the so-called "general mitigating circumstances" (*circostanze attenuanti generiche*) and for the decisions about sentencing and probation (*sospensione condizionale*).[94]

Moreover, an exception to the prohibition of criminological expert evidence is provided within the juvenile criminal procedure.[95] The juvenile justice system, as a matter of fact, allows assessments on the minor's personality (Article 9 of D.P.R. No.

cused and, in general, his/her psychic qualities independent of pathological causes».

[93] For this translation, see Gialuz, Lupària, Scarpa, *The Italian Code of Criminal Procedure. Critical Essays and English Translation*, cit., p. 177.

[94] Cass., sez. VI pen., 24th September 2013, no. 42823. In this regard, see Quattrocolo, *Questioni nuove e soluzioni antiche?*, cit., p. 1763.

[95] In this regard, see C. De Luca, *Gli accertamenti sulla personalità dell'autore del reato minorenne e il divieto di perizia psicologica nel rito ordinario: riflessioni e nuove prospettive*, in *Cass. pen.*, 2018, p. 2146 et seq.; L. Carceni, *Processo penale minorile*, in *Enc. Dir.*, *Agg.*, vol. IV, Giuffrè, 2000, p. 1018 et seq.

448 of 1988)[96] during criminal proceedings. Even if their function is mainly that of taking into consideration the access to a diversion programme as an alternative to criminal trial and sentence, the use of the outcomes of those assessments in the sentencing stage cannot be excluded.

Finally, the use of psychological and criminological assessments is allowed at the stage of sentence enforcement, as provided by Article 220 of the Code of Criminal Procedure. As a matter of fact, in the system designed by the Prison Law (*i.e.* Law No. 354 of 1975) the "scientific observation" of the convict's personality is given primary importance, since it grounds the individualised rehabilitation programme.

Looking at the comparative framework, a prohibition of criminological expert evidence such as the one established by Italian law cannot be found in any of the main European legal systems. On the contrary, the introduction of information and assessments on the accused's personal conditions is explicitly provided. For example, the French system provides for "quick social enquiries" (*enquêtes sociales rapides*) on the economic, family and social situation of the accused, which are aimed at verifying the practicability of certain sentencing options and at identifying the appropriate measures for social rehabilitation.[97] More detailed investigations on the personality of the accused and his/her economic, family and social situation can then be ordered by the investigation judge.[98] The results of those investigations are gathered into a "personality dossier," which has the function of providing the judicial authority with information on the person's past and present life and «does not have the purpose of searching for evidence of his/her guilt.»[99] The German system allows the Prosecutor to request the

[96] See paragraph 2 of Article 9.

[97] Delmas-Marty, Chiavario (Eds), *Procedure penali d'Europa*, cit., p. 147. The "*enquêtes rapides*" can be ordered by the Public Prosecutor (Article 41, paragraph 8, of the French Code of Criminal Procedure), the investigating judge (Article 81, paragraph 7, of the French Code of Criminal Procedure) or the correctional court (Article 396, paragraph 2, of the French Code of Criminal Procedure).

[98] Delmas-Marty, Chiavario (Eds), *Procedure penali d'Europa*, cit., p. 147. See Article 81, paragraph 6 of the French Code of Criminal Procedure.

[99] See Article D16 of the French Code of Criminal Procedure. «*L'enquête sur la personnalité des personnes mises en examen ainsi que sur leur situation matérielle, familiale ou sociale prévue à l'article 81, alinéa 6, du code de procédure pénale et les examens, notamment médical et médico-psychologique, mentionnés à l'alinéa 7 du dit article, constituent le dossier de personnalité de la personne mise en examen. Ce dossier a pour objet de fournir à l'autorité judiciaire, sous une*

social services to prepare a dossier (*Gerichtshilfe*), by gathering together all the elements regarding the environmental, social, family and psychological situation of the accused.[100] Such an investigation is mandatory for persons aged 18 up to 21.[101]

In order to establish whether the use of algorithmic tools for predicting recidivism comes under Article 220 of the Italian Code of Criminal Procedure, it is necessary to clarify whether the use of such tools falls within the scope of the notion of "expert evidence".

Some legal scholars have supported that conclusion: according to their opinion, risk-assessment tools like those widespread in the United States should be considered as a technical-scientific assessment, similar to expert evidence, since the algorithm is programmed on the grounds of psychological and criminological models and theories.[102]

6. Final remarks. Scope of application and precautions for AI in criminal justice

The overview of the algorithmic tools used in the context of criminal justice and the further tendencies emerging in civil justice have allowed to identify two main uses of AI systems worth of attention.

On the one hand, the United States experience – which at the moment has not been replicated in Europe, except for one example in the United Kingdom – has brought to light the use of tools aimed at assessing the re-offending risk, in the context of decisions about pretrial custody and those regarding the sentencing. Moreover, the

forme objective et sans en tirer de conclusion touchant à l'affaire en cours, des éléments d'appréciation sur le mode de vie passé et présent de la personne mise en examen. Il ne saurait avoir pour but la recherche des preuves de la culpabilité».

[100] Delmas-Marty, Chiavario (Eds), *Procedure penali d'Europa*, cit., 209. See § 160 of the German Code of Criminal Procedure.

[101] Delmas-Marty, Chiavario (Eds), *Procedure penali d'Europa*, cit., 209. See § 38 of the *Jungengerichtsgesetz* (JGG).

[102] In this meaning, Quattrocolo, *Questioni nuove e soluzioni antiche?*, cit., 1762; Maldonato, *Algoritmi predittivi e discrezionalità del giudice*, cit., p. 412. However, the similarity between the use of risk assessment tools and the criminological expert evidence is not self-evident, according to Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 20.

recidivism risk-assessment might gain a much wider area of relevance, considering provisions such as "alternative measures" and other prison law provisions, as well as the so-called security measures and the *ante delictum* preventive measures.

Risk-assessment tools aimed at predicting recidivism, however, are at the centre of a wide debate among legal scholars and within the civil society, because of the numerous problems that their use in the context of criminal justice seem to arise. Some of those issues can be connected to the validity of scientific theories and statistical models which ground the algorithmic tools' functioning: making reference to an output generated by processing statistical data, according to a generalising approach, might lead to a de-individualisation of the judgement and, potentially, to discriminatory effects against minority groups. On the other hand, the computational nature of the tools considered is linked to the lack of accessibility to the systems and to the codes governing their functioning, with the risk of affecting fair trial and the right of defence, especially in case of private-owned algorithms covered by trade secret.

The second category of tools which has emerged from the analysis of the current experiences is that of the predictive algorithms – developed mainly in France in the civil law field – which offer a prediction of the outcome of judicial decision-making. Although they are mainly used by law firms and have been tested by very few courts, these tools need to be considered because of their potential impact and for the risks of a wider use of them, especially on the part of the judges.

The risk highlighted with reference to these algorithmic systems concerns, on the one hand, the independence of judges, in case of undue attempts at profiling each individual judge by analysing and comparing his/her previous decisions. The main problem is however the one connected to the risk of standardising decisions, not because of the effect of conformity but because of the method for identifying the "prevalent" case law, which might be based on merely quantitative data or statistical correlations erroneously detected by the machine.

In the light of the criticism considered, the European law framework, also of soft law, suggests an extremely cautious approach with regard to the introduction of AI tools into criminal justice systems. The fundamental principles set out by the CEPEJ Ethical Charter of the Council of Europe (respect of fundamental rights, non-discrimination, quality and security of the data, transparency, impartiality and fairness,

guarantee of human control) and the provisions established by the European Union, which prohibit decisions based solely on an automated processing of personal data, are an important starting point for the European Union and domestic legislators in order to further regulate the use of such tools. Moreover, as regards the algorithms for predicting the risk of re-offending, the Italian procedural system, unlike others, has always excluded the possibility of introducing criminological or psychological expert evidence in criminal trials, since Article 220 of the Code of Criminal Procedure – at least according to the interpretation given by some legal scholars – might be an obstacle to the use of risk-assessment tools.

In the light of the framework thus outlined, some final remarks can be expressed about possible precautions for avoiding the highlighted risks.

With reference to the instruments which assess the probability of re-offending, the risk of a de-individualisation of the decision and the potential discriminatory effects seem to need two fundamental precautions: a serious multi-disciplinary debate on the quality of data and the validity of the models generated by their processing, as well as an adequate training of judges who may use the algorithms' outcomes. Legal professionals, most of all, should be warned about the generalising effects of extending the outcome of the algorithmic process to the individual case at stake: deeper awareness of the limits of the machine's "reasoning" may avoid the risk of automation biases and acritical acceptance of the "judgement" suggested by the algorithmic system.

The transparency problem – which is tied to the need to guarantee the effectiveness of the right of defence and fair trial principles – requires at least a clear public law regulation which, besides providing strict requirements for the selection of the service provider, should also establish obligations of disclosure in order to give the defence the widest possible access to the functioning mechanisms of private-owned systems. The solution to be preferred, however, would be entrusting public institutions with the development of the algorithms to be used in the justice field. On the other hand, the risk of the system being manipulated must be addressed through specific measures in order to guarantee the security of the environment where the software in use in the criminal justice system is kept and used.

Finally, in order to prevent possible violations of the presumption of innocence, as a consequence of introducing information about the accused's personality, a

prohibition of the use of those data for deciding on guilt should be clearly stated. Even if that would appear not sufficient to remove the risk of contaminations of the decision-making process concerning the accused's responsibility, the idea of introducing a separation between the decision on responsibility and the time of sentencing, also in systems like the Italian one, could be at least taken into consideration. In any case, such an intervention would be of extremely relevant impact for the legal system and thus needs to be very cautiously considered, bearing in mind the overall balance of the procedural law in force.

As regards the AI systems offering a prediction of the outcome of judicial decision-making, the fundamental precaution to be taken into account would be training legal professionals: judges, above all, must be adequately informed on how algorithmic tools work so that they can be aware, while carrying out their activities, of the fundamental difference between human reasoning and the "reasoning" of the algorithm. The algorithmic output is grounded on correlations established on the basis of the statistical frequency of certain groups of words and not on a real evaluation of the argumentations at the origin of the case law precedents. As regards the use of algorithms for profiling the judges, the main precaution would be prohibiting the publication of any data regarding the identity of the judges and any use of them for the analysis and comparison of the judges' decisions.

In any case, besides all the precautions aimed at addressing the issues arising the use of AI tools criminal justice systems, the most important need is that of wider awareness on how algorithms work: only justice professionals' greater knowledge of the algorithmic process and the consequent limits of its "product" may avoid the risks of de-individualisation of decisions on re-offending risk and the «ossification»[103] of case law. What is needed is a multi-disciplinary debate aimed at clarifying whether the risk of false correlations and the problem of technical inaccessibility to the algorithms with "self-learning" functions can be reduced, at least to a measure considered acceptable for their use in the context of criminal justice.

[103] Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, cit., p. 13.