Artificial Intelligence and Law
https://doi.org/10.1007/s10506-023-09347-w

ORIGINAL RESEARCH

Going beyond the "common suspects": to be presumed innocent in the era of algorithms, big data and artificial intelligence

Athina Sachoulidou[1]

Accepted: 24 January 2023

Abstract

This article explores the trend of increasing automation in law enforcement and criminal justice settings through three use cases: predictive policing, machine evidence and recidivism algorithms. The focus lies on artificial-intelligence-driven tools and technologies employed, whether at pre-investigation stages or within criminal proceedings, in order to decode human behaviour and facilitate decision-making as to whom to investigate, arrest, prosecute, and eventually punish. In this context, this article first underlines the existence of a persistent dilemma between the goal of increasing the operational efficiency of police and judicial authorities and that of safeguarding fundamental rights of the affected individuals. Subsequently, it shifts the focus onto key principles of criminal procedure and the presumption of innocence in particular. Using Article 6 ECHR and the Directive (EU) 2016/343 as a starting point, it discusses challenges relating to the protective scope of presumption of innocence, the burden of proof rule and the *in dubio pro reo* principle as core elements of it. Given the transformations law enforcement and criminal proceedings go through in the era of algorithms, big data and artificial intelligence, this article advocates the adoption of specific procedural safeguards that will uphold rule of law requirements, and particularly transparency, fairness and explainability. In doing so, it also takes into account EU legislative initiatives, including the reform of the EU data protection *acquis*, the E-evidence Proposal, and the Proposal for an EU AI Act. Additionally, it argues in favour of revisiting the protective scope of key fundamental rights, considering, *inter alia*, the new dimensions suspicion has acquired.

Keywords Artificial intelligence (AI) · Law enforcement · Criminal justice · (Pre-) suspect · Criminal procedural rights · Presumption of innocence

✉ Athina Sachoulidou
athina.sachoulidou@novalaw.unl.pt

[1] NOVA School of Law, CEDIS, NOVA University of Lisbon, Campus de Campolide, 1099-032 Lisbon, Portugal

Published online: 22 February 2023

1 Introduction

1.1 Background and structure of the article

Artificial intelligence (AI), a term first introduced in the 1950s, has recently experienced a revival with the availability and power of big data (Duan et al. 2019, p. 63). The term 'big data' stands for 'large sets of data from mixed sources' (European Commission for the Efficiency of Justice (CEPEJ) 2018, p. 70) combined with the analytical capabilities made possible by faster computer processors and greater data storage capacity (European Commission (EC) 2014, p. 4; Broeders et al. 2017, p. 310). Besides volume, variety and velocity, the 5 V's of big data include veracity, i.e., the accuracy of the data itself *and* the trustworthiness of data source, type and processing of it, and value, i.e., new business models and opportunities for diverse value creation (Hoffmann-Riem 2018, pp.19–20).

The use of AI to either assist or replace human decision-makers has already been seen as one of the most important applications in its history—with AI systems undertaking significantly complex tasks, ranging from individual risk assessments to emotion detection, which previously seemed unthinkable (Duan et al. 2019, p. 67). Law enforcement and criminal justice have become arenas for testing and applying AI technologies with the aim of facilitating the decision as to whom to investigate, arrest, prosecute, punish and eventually release on parole. Interestingly, the Proposal for a Regulation laying down harmonised rules on AI drafted by the European Commission (2021a; 2021b; hereinafter referred to as the EU AIA) classified AI systems intended to be used in these two areas[1] as *high risk* considering the adverse impact such systems may have on fundamental rights (Art. 6(2) EU AIA). However, as regards the area of criminal justice, the use of AI technology is not limited to the systems listed in Annex III to the EU AIA (points 6 and 8) inasmuch as, besides predictive policing software, risk assessment algorithms or similar applications, other AI-driven systems capable of observing and evaluating human behaviour with the aim of predicting future behaviour, such as safety enhancing driving systems, may serve as new investigative and evidentiary tools that challenge not only the definition of the criminal act itself, but also the organisation of the criminal procedure (Gless 2020) *and*, thus, the exercise of criminal procedural rights.

Against this backdrop, this article focuses on three use cases: predictive policing; machine evidence; and recidivism algorithms. Each of them is presented below (Sect. 2) with the help of concrete examples that showcase not only the tremendous opportunities, but also the considerable drawbacks associated with the trend of increasing automation in law enforcement and criminal justice settings. Subsequently, the use cases are examined through the lens of fundamental rights and criminal procedural rights in particular – using the example of the *presumption of innocence* (PoI) (Sect. 3). In doing so, this article sets forth three challenges: the narrow protective scope of PoI compared to the trend of broadening the contexts

[1] The EU AIA generally refers to the use of AI in judicial settings without introducing any explicit distinction relating to criminal justice. For a further analysis of the EU AIA see Sect. 4.3.

within which one may be considered as suspect; the increased 'innocence threshold' and the defence barriers posed by evidence surrounded by scientific objectivity and a sense of security; and the new meaning ascribed to the notion of doubt in the context of the criminal trial. Lastly, it provides an overview of EU legislator's initiatives in the area of law and technology, in order to determine where there is still space for improvement, and where a radical re-orientation of the regulatory efforts, including a paradigm shift in the perception of fundamental rights, may still be required in the era of algorithms, big data and AI (Sect. 4).

1.2 Scope of analysis and methodology

This article interjects in the debates taking place at the crossroads of AI, law enforcement and criminal justice. For the purposes of the following analysis, the term 'law enforcement' stands for activities carried out by public authorities 'for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security' (Art. 3(41) EU AIA; cf. Recital 12, Art. 3(7) Directive (EU) 2016/680). This article focuses on the *work of law enforcement authorities* (LEAs) as subjected to the European Convention on Human Rights (ECHR), the Charter of Fundamental Rights of the EU (CFR) and national laws. It does *not* address the work of intelligence services in the area of national security (cf. Recital 14 Directive (EU) 2016/680). Regarding this limitation of scope, the following clarification is deemed necessary:

The so-called national security clause (Art. 4(2) Treaty on European Union (TEU)) has been interpreted as meaning that the surveillance activities of national intelligence services fall within the scope of EU Member States' (MSs) competence—with the EU itself being unable of intervening and regulating them (Lachmayer 2009, p. 104; European Union Agency for Fundamental Rights (FRA) 2015, p. 10; Eskens 2021). Nevertheless, national security as a concept is neither understood in a uniform way across the EU nor defined expressly in the EU legislation or the case law of the Court of Justice of the EU (CJEU). Lack of clarity is coupled 'with the varied and seldom clearly drawn line between the areas of law enforcement and national security in individual Member States' (FRA 2015, p. 10; 2017, p. 21). This problem became evident particularly in the case of the exchange of existing intelligence among EU MSs for counter-terrorism purposes and access to such data by LEAs, and led the EC to state that solutions to the lack of clarity in the relationship between the community of law enforcement and that of intelligence are to be identified urgently (FRA 2017, p. 21). In addition, it has progressively become clear that the national security exemption is not synonymous with an absolute exclusion of the applicability of EU law—as recently confirmed by the CJEU[2]– and, in any case, it does not waive the applicability of Council of Europe (CoE) Conventions,

[2] This has been the case with the *La Quadrature and Others* and the *Privacy International* cases (both of them adjudicated by the CJEU on 6 October 2020). For a comprehensive analysis of this case law see Tracol (2021) and Eskens (2021).

including the ECHR (FRA 2015, p. 11; 2017, pp. 22–23). Against this backdrop, this article may focus on the area of law enforcement as shaped by, *inter alia*, EU law, but the following analysis is also informed by the recent CJEU case law—particularly inasmuch as it addresses the phenomenon of mass surveillance and its implications for fundamental rights in the area of intelligence *and beyond* (see Sect. 4.3).

Regarding the use of AI in the area of criminal justice in terms of an area where the EU has shared competences (Art. 3(2), 67(3) TEU), this article sheds light on the common minimum standards for criminal proceedings,[3] and particularly the safeguards enshrined in the Directive (EU) 2016/343 on strengthening of certain aspects of the presumption of innocence and of the right to be present at the trial in criminal proceedings. Choosing to focus on PoI, this article goes beyond the "common suspects", i.e., privacy, data protection and non-discrimination that have already played a leading part in the discourses on the intersection of new technologies, law enforcement and justice (cf. FRA 2020, p. 7). Without being categorically different compared to an analysis undertaken to assess the implications for privacy, data protection and non-discrimination, an analysis focused on the right to be presumed innocent rather seeks to encompass the challenges the understanding of the notions of suspicion, guilt and innocence, *being of central importance for criminal law*, is presented with in the era of algorithms, big data and AI.

Additionally, this article does not deal (at least not directly) with the admissibility of evidence generated or assessed by means of AI from a national law perspective—acknowledging its limits in terms of providing a comprehensive comparative analysis of the respective rules included in national codes of criminal procedure. This should not imply that national law is not concerned by the "intrusion" of AI in the realm of law enforcement and criminal justice. On the contrary, national legislators have to deal with an important contradiction: Criminal law is a traditional field of regulation that is significantly confined to national borders, whereas AI technologies 'embrace notions such as internationalization and globalization' (Greenstein 2022).

Lastly, with regard to the methodology used to design this article, socio-legal research methodology is employed to present the use cases (predictive policing; machine evidence; algorithmic risk assessment) with the aim of bringing together the ethical, social and legal considerations associated with them (Sect. 2). Subsequently, doctrinal legal research is employed to address the question of what safeguards are already in place when deploying AI technologies in law enforcement and criminal justice settings (Sect. 3), as well as the question of whether new safeguards should be designed to address the pressing challenges PoI is faced with in this context (Sect. 4). For this purpose, this article refers to three law & tech initiatives of the EU legislator: (1) the overall reform of the EU data protection *acquis*, (2) the *planned* adoption of rules to govern cross-border access to electronic evidence (E-evidence Proposal); and (3) the *proposed* rules for AI (EU AIA). In doing so, it does not aim

[3] These are included in Directive 2010/64/EU, Directive 2012/13/EU, Directive 2013/48/EU, Directive (EU) 2016/343, Directive (EU) 2016/80, and Directive (EU) 2016/1919.

to present these set of rules in an exhaustive way, but rather focuses on gaps related to the use of AI by the police and judicial authorities in criminal matters.

2 AI at the service of law enforcement and criminal justice

This article does not intend to engage with the history of AI (Duan et al. 2019) or the different definitions suggested to describe AI and the technologies pertaining to it (Greenstein 2022). Nevertheless, a working definition is deemed necessary for elaborating on the use of such technologies in the realm of law enforcement and criminal justice. In the EU ecosystem, there have recently been (at least) three attempts to define AI that can be summarised as follows:

As proposed in the Communication on AI for Europe (EC 2018a, p. 1), AI 'refers to systems that display intelligent behaviour by analysing their environment and taking actions—with some degree of autonomy—to achieve specific goals' and AI-based systems may be purely software-based (e.g., face recognition systems) or AI may be embedded in hardware devices (e.g., autonomous cars). The High-Level Expert Group on AI (AI HLEG), set up by the EC in 2018, has expanded this definition with the aim of, *inter alia*, facilitating further discussions on AI ethics guidelines and policy recommendations (AI HLEG 2019a, p. 1). As a result, the Group presented a twofold definition providing for the special traits of AI systems and representative examples of approaches and techniques AI includes as a scientific discipline. According to this definition, AI systems are designed by humans and 'act in the physical or digital world by *perceiving their environment*, *interpreting* the collected structured or unstructured *data*, *reasoning* on the knowledge derived from this data *and deciding the best action(s) to take* (according to pre-defined parameters)' to achieve a complex goal given to them (*ibid*, p. 7, emphasis added). These systems may also be capable of learning to adapt their behaviour by analysing the impact of their previous actions on the environment (*idem*). Machine learning (including, for instance, deep learning and reinforcement learning), machine reasoning (including, for instance, knowledge representation and reasoning) and robotics (including, for instance, control, perception, sensors and actuators) are representative approaches and techniques that pertain to AI (*idem*).

Up to now, there has not been any *legal definition* of AI except for the one included in the EU AIA that built on the work accomplished by AI HLEG during the past two years (EC 2021a, p. 8). Under Art. 3 point 1 EU AIA, an AI system means 'software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with'. The techniques and approaches mentioned above include machine learning, logic—and knowledge-based approaches, statistical approaches, Bayesian estimation, search and optimization methods (Annex I to EU AIA).

With autonomy and adaptivity characterising the AI, the latter is experiencing a revival thanks to the proliferation of available data (combined with the

demand for data from public *and* private sources; Data mining, dog sniffs and the Fourth Amendment 2014, p. 694), increased processing power and advanced mathematical algorithms (Greenstein 2022). This particularly applies to machine learning (ML) that enables the construction of 'a mathematical model from data, incorporating a large number of variables that are not known in advance' (CEPEJ 2018, p. 72).[4] Models of this kind may be employed to decode human and animal cognition with the aim of building smart machines and applications, or serve as the core technical component of a decision-making system (Greenstein 2022). They may also be deployed to predict human behaviour on the basis of the links detected and classified using training data (*idem*).

The use cases presented below are selected to discuss the use of AI to predict future human behaviour either for law enforcement purposes (*predictive policing*) or to assist decision-making at different stages of criminal proceedings (*machine evidence* and *recidivism algorithms*). In addition, they are selected to explore the implications of the use of AI for the right to be presumed innocent (Art. 6(2) ECHR; Art. 48(1) CFR). The choice to examine so different tools,[5] which are employed to deliver a significantly different output, dictates a clear distinction between: 1) information gathering at pre-investigation stages, including collecting, retaining and process data without any *individualised* cause or suspicion—context within which the applicability of PoI is questioned, if not rejected, as *suggestion for suspicion* falls outside its protective scope (Sect. 3.1) and 2) information gathering within formal criminal proceedings, including collecting and producing evidence on the basis of a *pre-established* suspicion (e.g., at the stage of issuing a search or arrest warrant or bringing charges) and, subsequently, assessing it to reach a decision on liability and impose a sentence at the trial stage or to re-evaluate the sanction imposed at a later point—context within which PoI may be fairly well anchored in criminal procedural rules, but different components of it, including, for instance, the burden of proof rule or the *in dubio pro reo* principle are confronted with new challenges (Sect. 3.2).

2.1 Predictive policing

To define predictive policing, one should take the large amount of available data as a starting point. Governments enjoy access to data collected by various public authorities, such as police, tax authorities and public energy suppliers, while the private sector accesses and expands the scope of consumers' digital footprint, including, for instance, online monetary transactions and electronic communication. *Data mining* is the term that stands for pulling together and analysing data, in terms of a task to be taken on by humans or algorithms, with the aim of detecting useful

[4] The learning phase uses the so-called *training data* to detect and classify links. ML is usually divided into three categories: (human) supervised learning, unsupervised learning and reinforcement learning: CEPEJ (2018), p. 72. For the distinction between supervised and unsupervised learning see also Greenstein (2022).

[5] The respective algorithms are different in terms of both design and the goal(s) they seek to achieve.

patterns (Data mining, dog sniffs and the Fourth Amendment 2014, p. 694).[6] When employed for law enforcement purposes, data mining can be divided into two categories: (1) *subject-based data mining*, where the focus lies on previously identified individuals; and (2) *pattern-based data mining*, where the focus shifts onto non-suspect individuals 'to identify patterns of transactions or behaviours that correlate with suspect activity' (*ibid*, p. 695). In both cases, the goal is to turn voluminous data—including, but not limited to criminal data– into useful data that assist LEAs in decision-making (Perry et al. 2013, p. 2; Meijer and Wessels 2019, p. 1033).

The term 'predictive policing' puts together police officers, software, algorithms and data sets (Egbert and Leese 2021). There may not be a unanimous definition, but some *key features* of it can be singled out on the basis of the existing scholarship: According to Perry et al. (2013, pp. 1–2), predictive policing is 'the application of analytical techniques—particularly quantitative techniques– to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions'—with the possibility of drawing on large data sets distinguishing it from old-school statistical crime analysis. The target is either the crime location and time, in the sense of areas and times of heightened criminal activity, or individuals that are likely to be involved in the criminal enterprise either as perpetrators or as victims or that match with a specific crime already committed (Meijer and Wessels 2019, pp. 1033–34; Fuster 2020, pp. 22–24; Oswald 2020, p. 22; Završnik 2020, p. 570; Strikwerda 2021, p. 423). Out of the different objectives a predictive policing tool may have, this article focuses on the *identification of likely troublemakers* as a case that has recently concerned a European court.[7]

This is the case of SyRI (Systeem Risico Indicatie), which was adjudicated by the Hague District Court on 5 February 2020.[8] SyRI can be classified as a predictive policing tool for identifying future offenders (Strikwerda 2021, p. 423), as it links citizen data stemming from various agencies, such as tax authorities and authorities administering social and unemployment policies,[9] and produces a list of individuals *with an alleged higher fraud risk*.[10] The respective risk report is handed over to the

[6] (Big) Data analytics is employed to gain insight into the past (descriptive analytics), understand and forecast the future (predictive analytics) and/or advise on possible outcomes (prescriptive analytics): Shun and Huo (2021).

[7] This choice is aligned with the other use cases inasmuch as these shift the focus onto individuals, whether suspected of having committed a crime or being about to commit one or already accused of having acted unlawfully. It is true, however, that, in the case of predictive policing, programs identifying the so-called hot spots are more popular in European countries: Singelnstein (2018). These programs use *historical data* on the time, place and type of crimes committed in the past often coupled with environmental parameters, such as population density – without necessarily deploying personally identifiable information (FRA 2020, pp. 34–35; Meijers and Wessels 2019, pp. 1033–34).

[8] Judgment available in English at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878. Accessed 24 January 2022.

[9] The input data consisted of, for instance, employment data, civic integration data, debt data, health insurance data and personal data (e.g., name, address, date of birth): Strikwerda (2021), p. 424. The Hague District Court found that a 'total of 17 data categories of various types qualify' with each of them potentially encompassing a large amount of data: *Nederlands Juristen Comité voor de Mensenrechten and Others v. The State of the Netherlands*, Hague District Court, 5 February (2020), par.6.50.

[10] For a summary of the facts and the judgment see Meuwese (2020).

authority that has requested the use of SyRI in the first place,[11] and the latter uses its discretion in starting an investigation (Meuwese (2020); Strikwerda (2021), p. 424). Having to deal with contradictory claims regarding the nature of SyRI, ranging from deep learning with big data-like features (*Nederlands Juristen Comité voor de Mensenrechten and Others v. The State of the Netherlands*, Hague District Court, 5 February (2020), par. 6.45–6.46) to data comparison with the aid of a simple decision tree (*ibid*, par. 6.47), the Court declared itself.

> 'unable to assess the correctness of the position of the State of the precise nature of SyRI because the State *has not disclosed the risk model and the indicators* of which the risk model is composed or may be composed' (*ibid*, par.6.49, emphasis added).

In addition, the Court abstained from attaching any clear technical label to SyRI, taking into consideration the fact that the respective legislation leaves the option of using predictive analyses, deep learning and data mining open, as well as that there is no clear-cut definition of the term 'big data' (*ibid*, par.6.51–6.52). Of central importance has also been the fact that the SyRI legislation provides neither 'for a duty of disclosure to those whose data are processed in SyRI' nor 'for an obligation to notify the data subjects individually, as appropriate, that a risk report has been submitted [despite its significant effect on them]'; that is, the affected individual gets informed only 'if there is a control and investigation in response to a risk report' (*ibid*, par. 6.54, 6.82).

Against this backdrop, the court reached the conclusion that the SyRI legislation fails to strike the 'fair balance' required to justify an interference with private life within the meaning of Art. 8(2) ECHR, considering, *inter alia*, the lack of transparency and verifiability coupled with insufficient attention to the principles of purpose limitation and data minimisation (*ibid*, par. 6.82, 6.86–90, 6.95–98, 6.106).[12]

While examining the use of SyRI through the lens of necessity in a democratic society in relation to the legitimate aim of achieving the economic wellbeing of the country (proportionality), the Court acknowledged the difficulties inherent in comprehending how a data subject may defend him-/herself against the submission of a risk report about him/her (*ibid*, par. 6.90), as well as the potential discriminatory and stigmatizing effects SyRI may have considering its use in the so-called 'problem districts' (*ibid*, par. 6.91–93). Nevertheless, it assessed these parameters in the light of its principal argumentation regarding the violation of Art. 8 ECHR (Meuwese 2020, p. 210). It did not address the claim that SyRI legislation was in violation of Arts. 6 and 13 ECHR (*Nederlands Juristen Comité voor de Mensenrechten and Others v. The State of the Netherlands*, Hague District Court, 5 February 2020, par. 6.107).

[11] According to the law in place (SUWI Act and SUWI Decree), a limited number of authorities, municipalities and authorities responsible for social benefits, taxes and immigration were entitled to request the use of SyRI. In reality, the use of SyRI was anything but widespread: Meuwese (2020), p. 210.

[12] In this context, the Hague District Court cited the ECtHR judgment in the matter of *S. and Marper v. UK*: *Nederlands Juristen Comité voor de Mensenrechten and Others v. The State of the Netherlands*, Hague District Court, 5 February 2020, par. 6.84.

The Dutch government did not appeal against the court's decision, which may have been straightforward regarding the then current status of SyRI and its privacy implications, but, at the same time, provided a leeway for developing "better" data-based surveillance tools (Meuwese 2020, p. 211). Irrespective of its success in practical terms, the SyRI judgment can be considered as a core element of the energetic debate about the benefits and risks associated with the use of predictive tools.[13]

In the conceptualization of predictive policing, LEAs should be able to 'deploy their resources more efficiently and effectively' in terms of distributing them wisely in place and time (by determining, for instance, optimal patrol routes) and targeting the "real" suspects, namely those presenting a higher probability of offending in the future (Meijer and Wessels 2019, p. 1033). Both these options have been seen in a new light in times of austerity, considering the reduction in police officer numbers associated with it (Oswald 2020, p. 222). In the case of the tools signalling future wrongdoers, predictive policing is also expected to release law-abiding citizens from unnecessary encounters with the police and the respective violations of their fundamental rights, including privacy (Ferguson 2015; Data mining, dog sniffs and the Fourth Amendment 2014, p. 695). Besides, it should lead to a 'more equitable and non-discriminatory policing' by objectifying the decisions reached by police officers and reducing reliance on their subjective judgment (Joh 2016, p. 28; FRA 2020, p. 69).

To seize the benefits of predictive policing as a *data-driven* application, one needs to ensure that the algorithm is supplied with accurate data, which has been collected previously in an appropriate context (cf. Greenstein 2022), and links it properly, i.e., without leading to false positives or negatives (cf. Recital 50 EU AIA). This is particularly important inasmuch as the 'suggestion of suspicion' a predictive policing leads to, despite based on *correlations instead of causality* (Andrejevic 2017), may result in coercive measures against the affected individual. The risk of a false positive *lato* sensu becomes considerably high, when the amount of data increases, interoperability of datasets becomes the rule, and information is collected and assessed at transnational level and in very short periods of time (Fuster 2020, p. 47; Giannakoula et al 2020, p. 69, 87).

Besides, as stressed in the SyRI case-law, predictive policing tools and the laws providing for their use often do not safeguard insight into the risk model and the risk indicators the latter deploys. This implies not only insurmountable burdens for the individual who wants to challenge the risk report, but also a considerable understanding gap leading to accountability problems where police officers rely fully on the algorithmic output (Meijer and Wessels 2019, p. 1036). Lack of transparency may not allow the user of the predictive policing tool to deduce biases. The existence of biases in general cannot be excluded, as in societies where discriminatory practices are widespread, the data collected to train the algorithm may reflect these practices and, thus, reproduce and entrench bias (Mittelstadt et al. 2016, p. 5; Ryan

[13] For a review and an empirical assessment of benefits and drawbacks see Meijer and Wessels (2019).

et al. 2019, pp. 31–33; FRA 2020, p. 69; CEPEJ 2018, p. 55).[14] Relying on historical crime data is no panacea, considering that the low reporting rates in certain crime categories, such as the white collar-crime, coupled with greater access to data of more reported crimes, such as those against property, may turn the spotlight onto individuals belonging to certain professional groups or areas presenting certain demographics and, thus, expose the members/residents of them to the stigma associated with being singled-out as high-risk to commit a/any crime (cf. FRA 2020, pp. 69–73).[15]

Next, the lack of clear boundaries regarding the mass access to, *inter alia*, the digital identity of citizens within the context of predictive policing projects poses not only privacy implications (Data mining, dog sniffs and the Fourth Amendment 2014, p. 696, Strikwerda 2021, p. 427), but also results in a multi-faceted power imbalance and mistrust in the relationship between citizens and the State (Ferguson 2015, pp. 403–404; cf. Recital 38 EU AIA). This may impact on fundamental rights tightly linked to democracy, such as the freedom of expression and association (EC 2020a; Fuster 2020, p. 40)—for instance, with citizens abstaining from certain social interactions out of fear to find themselves in the line of fire of suspicious correlations. In this context, the citizens also produce data on an almost permanent basis without enjoying equal access to big data sets and without always having a reasonable alternative, i.e., a tracking-free service (Mantelero and Vaciago 2013, p. 167). The knowledge asymmetry—coupled with lack of transparency—can be turned into arms asymmetry, which, as will be shown below (Sect. 3.2.2), may even amount to an indirect shift in the burden of proof (Reidenberg 2014, p. 605; Fuster 2020, p. 23), when the affected individual seeks to contest the risk report. The safeguards (s)he may enjoy in this process are rather blurry –at least regarding criminal procedural rights—*if the risk report does not lead to an investigation*. This is the case with the right to be presumed innocent, the scope of which does not entail, as will be explained below (Sect. 3.2.2), the suggestion of suspicion for a crime not yet committed—despite the adverse impact the latter may have on the individual (e.g., refusal of permit to work).

The cases, where the risk report actually leads to deployment of investigatory powers, may fall into the protective scope of criminal procedural rights, but are not less problematic inasmuch as coercive measures are in theory based on suspicion of having committed a crime and not being a high risk to do so in the future. A risk not yet materialised or an abstract intention is categorically different from an attempted crime or even *neutral* preparatory acts. Nevertheless, the identification of such a risk by means of predictive policing may lead police officers to interpret each action of the risk-carrier as suspicious, and give them reasons to consider the interference

[14] Cf. R [Bridges] v. CC South Wales, UK Court of Appeal, EWCA Civ 1058, 11 August 2020, where the Court held, among other things, that the South Wales Police did not investigate if the 'AFR Locate', a facial recognition system used for law enforcement purposes, exhibited race or gender bias.

[15] The inevitable focus on certain areas and the increase of monitoring associated with it, when compared to the opposite trend of providing investment incentives in other areas of the same city, may lead to spatial inequality across racial and social classes: Brannon (2017), cited by Meijer and Wessels (2019), p. 1036.

with the rights of this person, for instance, by means of a search or investigative detention as legitimate or at least well-founded (Strikwerda 2021, p. 429).

2.2 Machine evidence

In her 2007 article entitled 'The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence', Murphy introduced a taxonomy of forensic evidence distinguishing the first generation of forensic evidence (e.g., handwriting, ballistics, hair and fiber analysis) from the second one stemming from the—then—new sciences and techniques, including, for instance, DNA testing, data mining, biometric scanning and electronic location scanning. In this context, she also provided a comprehensive analysis of the distinctive characteristics of each generation:

Forensic evidence of the first generation (1) is usually linked to concrete offences (for instance, the use of ballistics presupposes the use of a firearm) and deployed as supportive material in addition to other forms of evidence (e.g., eyewitnesses, documents), (2) is rather experiential and observational, less sophisticated in technical terms, and, thus, more comprehensible by laypeople, (3) rarely poses intellectual property issues and, thus, is more accessible to defence for testing purposes, (4) is predominantly reactive in the sense of serving to confirm the involvement of an already identified person into a specific crime, and (5) being, first, contained in its investigative scope and, second, capable of only providing a narrow piece of information, does not implicate (at least typically) greater questions of personal privacy (Murphy 2007, pp. 726–728).

On the contrary, *forensic evidence of the second generation* (1) applies to a wide variety of offences (considering, for instance, the breadth of citizens' digital footprint or the overall relevance of DNA samples), (2) is scientifically robust and requires specialised knowledge, fact that renders it less accessible to laypeople, but, at the same time, gives raise to claims of high probative value, (3) presents a high level of technical sophistication which often entails a significant capital expenditure to submit it to an independent analysis, (4) raises considerable concerns about the protection of the proprietary interests of those developing the technologies that underpin it, and (5) relying heavily on *big* data sets, may be deployed proactively to identify a suspect *ab initio* and impact considerably on privacy interests of both suspects and innocent third parties (*ibid*, pp. 728–730).

Meanwhile, forensic evidence pertaining to the second generation became mainstream—with digital/electronic evidence being a core element of it. Taking data mining as an example, the sources and the tools to gather information have multiplied: smartphones enabling, *inter alia*, location tracking; automatic license plate readers, electronic toll collection systems, speed cameras and car GPS devices recording travel patterns; social media websites tracking communications; e-shops facilitating the creation of consumer profiles; and digital databases storing financial data to name a few (Ferguson 2015). In this context, it is anything but coincidence that the term 'digital/electronic evidence' has started being used widely in both

political statements and legal texts.[16] Yet, there is still no common understanding on its meaning (Buono 2019, p. 308; Vazquez Maymir 2020, p. 3; Gless 2020, p. 209). Nonetheless, this does not seem to delay the emergence of a *new generation of evidence* consisting of AI-generated evidence (Gless 2020, p. 211).[17]

This new evidence category refers not only to tools designed to meet investigatory needs and to produce tangible evidence, such as the image and video comparison software employed by Interpol to connect victims, perpetrators and places in cases of child sexual abuse (Završnik 2020, p. 570),[18] but also extends to tools designed to meet commercial needs in the first place (Gless 2020, p. 198). Automated driving technology serves as a representative example[19]—with systems monitoring the vehicle's position in the lane and warning the driver when the vehicle is drifting out of it, or monitoring the driver's steering pattern, body temperature and facial movements and warning him/her to stop and take a break in cases where anomalies are detected (*ibid*, pp. 202–203).[20] Driving assistants of the second kind or similar tools go beyond conveying a message to be evaluated by a human user by rendering an "opinion" regarding, for instance, the driver's level of alertness and his/her ability to drive (*ibid*, p. 204). This "opinion" is stored in the system and can be retrieved to serve as, *inter alia*, (the single) evidence at trial, should the driver ignore the alert and cause a fatal car accident.

Like big data and data mining technologies, AI and machine learning are presented as a chance for accessing more and *more accurate* information and, thus, as

[16] For instance, the EU Ministers stressed in their Joint Statement of 13 November 2020 the great importance of the availability of and access to digital evidence, particularly for counter-terrorism purposes: Council of the EU (2020a). The EC (2020b, p. 20) highlighted in the 2020 Counter-Terrorism Agenda the need for 'a clear and robust framework for timely cross-border access to electronic evidence and investigative leads', given that 'digital evidence is needed in about 85% of all criminal investigations'. Similar, if not identical, statements can also be found in the 2020 Security Union Strategy (EC 2020c, p. 12). Additionally, the EC released in April 2018 a Proposal for a Regulation on European Production and Preservation Orders for e-evidence in criminal matters and a Proposal for a Directive laying down harmonized rules on the appointment of legal representatives for the purpose of gathering evidence in criminal proceedings (for the chronicle of the e-evidence proposals see Sachoulidou 2021a, pp. 779–780; further analysis undertaken in Sect. 4.2). Meanwhile, in January 2023, and after particularly lengthy trilogue negotiations, the Council confirmed agreement with the European Parliament on the new rules to improve cross-border access to e-evidence: https://www.consilium.europa.eu/en/press/press-releases/2023/01/25/electronic-evidence-council-confirms-agreement-with-the-european-parliament-on-new-rules-to-improve-cross-border-access-to-e-evidence/. Next, practical initiatives, such as the establishment of the e-evidence digital exchange system (eEDES) and the creation of the e-CODEX system and the JITs Collaboration Platform, are also under development: Hamran (2020), p. 2.

[17] Cf. the distinction between computer-derived and computer-generated evidence: Palmiotto (2020), pp. 7–8.

[18] More information about the International Child Sexual Exploitation (ICSE) image and video database available at: https://www.interpol.int/en/Crimes/Crimes-against-children/International-Child-Sexual-Exploitation-database. Accessed 24 January 2022.

[19] Digital assistants such as Amazon's Alexa and Google Home have also been brought to the forefront as devices recording information that may be deemed useful for criminal investigation purposes. See, for instance, Sauer (2017).

[20] Art. 6 (1) Regulation (EU) 2019/2144 stipulates that all new motor vehicles, which are placed on market, registered or entered into service, shall be equipped with, *inter alia*, driver drowsiness detection and attention warning systems (lit. c). The respective obligation shall apply from 6 July 2022.

tools a criminal justice system may leverage not only to convict but also to prevent wrongful convictions (Fairfield and Luna 2014; Hadjimatheou 2017). This possibility is not, nor should be considered, a "free pass" to criminal proceedings inasmuch as a decision on *trustworthiness and reliability* is still pending—with forensic evidence of the third (?) generation having something more in common with their predecessors: opacity and lack of accessibility to laypeople (Murphy 2007, p. 729; Gless 2020, p. 207). Opacity prevents the end-user (in the case discussed here: the judiciary) not only from comprehending how the machine "thinks" *in abstracto*, but also hampers the detection of flaws, whether inherent in the input, the procedure or the output,[21] including, *inter alia*, inherent biases.

As Burrell (2016, pp. 1–2) suggests, opacity may appear in three forms: (1) intentional corporate or institutional self-protection and concealment, namely state secrecy; (2) technical illiteracy, where writing and reading a code is a specialist's skill; and (3) 'mismatch between mathematical optimization in high-dimensionality characteristic of machine learning and the demands of human-scale reasoning and styles of semantic interpretation'. *Opacity of the first kind* has already been taken into consideration and delimited—to a greater or lesser extent—in the context of the EU data protection legislation (cf. Section 4.1).[22] Should proprietary interests be overcome, there is, however, no guarantee that the algorithm, the function of a *consumer product* is based on, is not trained to point the finger at the user to protect corporate self-interests (Gless 2020, p. 217). *Opacity of the second kind*, namely the need for special skills to scrutinise the soundness of forensic tools and techniques and the reliability of their output, may be compensated by means of expert assistance and special scientific tests to be carried out, whether upon the prosecutor's request or as part of the defence right to undertake investigations on its own (Gless 2020, pp. 211–212, 239–241).[23] Nonetheless, this possibility is usually subject to limits arising from the respective procedural framework, ranging from the reasonableness of the request itself to the trial economy (*ibid*, pp. 242–243; Murphy 2007, p. 771).[24]

[21] For instance, in the case of Intoxilyzer 5000EN, a device that was deployed in the US to measure breath alcohol content and the reliability of which was challenged before the Minnesota Supreme Court, the disclosure of its code brought to the forefront various undetected failures, "including erroneous results based on power surges, interference from cell phones, and defects in the process of self-testing and reporting errors": Chessman (2017), p. 197; see also Palmiotto (2020), p. 14.

[22] Representative examples may be found in Recital 63 Regulation (EU) 2016/679 (GDPR) and Recital 44–45 Directive (EU) 2016/680.

[23] The respective criminal procedural rules vary considerably even among the EU MSs, depending on the extent to which they follow an adversarial or an inquisitorial system or a mixture thereof. See Sellier and Weyembergh (2018), pp. 67–69. For a further analysis see Sect. 3.

[24] Taking the defence right to request that the court appoints an additional expert, as enshrined in Art.244 German Code of Criminal Procedure, as an example, the bench retains the right to reject this request, if the expert opinion is deemed 'superfluous because the matter is common knowledge, the fact to be proved is common knowledge, the fact to be proved is irrelevant to the decision or has already been proved, the evidence is wholly inappropriate or unobtainable, the application is made to protract the proceedings, or an important allegation which is intended to offer proof in exoneration of the defendant may be treated as if the alleged fact were true': Art.244(3) German Code of Criminal Procedure, as translated by Gless (2020), p. 242. The same applies, where the court holds that it 'possesses the necessary specialized knowledge' or 'if the opposite of the alleged fact has already been proved by the first expert opinion': Art.244(4) German Code of Criminal Procedure, as translated by Gless (2020), pp. 242–243.

Additionally, *opacity of the third kind*, namely understandability of the procedure and its output, jeopardises the added value of expert assistance. The need for investing greater intellectual and material sources may not be new –compared to, for instance, the difficulties inherent in scrutinising DNA evidence for the first time or analysing and presenting the model behind traditional data mining software—but AI systems distinguish themselves by operating in a way that is at least for the time being partly known and not always predictable even by their designers (Palmiotto 2020, p. 10). More specifically, complexity increases when attempting to *understand* and *explain* the interplay between extremely large datasets and the code (written with clarity in the first place) in the mechanism of the algorithm (Burrell 2016, p. 5). This is more than often beyond the capabilities of human brains (Yavar 2018).[25]

Technical complexity may impact on how the respective evidence is dealt with in the courtroom in multiple ways. Should AI-generated evidence be admitted in the first place, judges, who struggle to understand complicated mathematical formulas behind the intelligent device and the "opinion" it renders, may be at ease with abstaining from challenging it and place unyielding trust in its infallibility (Murphy 2007, pp. 768–769; Gless 2020, p. 214). This implies that the very "scientific" nature of such evidence may pose it at the same level with the eyewitness, the statements of whom are often to be trusted almost unconditionally, even though the latter can usually make him-/herself clear without any major difficulties, take the responsibility of his/her own sayings by means of oath and be prosecuted for perjury if (s)he violates this oath (Gless 2020, p. 214). Reluctance to challenge may equally concern defence attorneys, who are presented with a considerable effort-reward imbalance when attempting to contest evidence without necessarily comprehending the science behind it (Murphy 2007, pp. 765–66, 770; Palmiotto 2020, pp. 13–14). Even if the defence is "courageous" enough to accept this challenge, the assessment of such evidence presupposes not only overcoming mechanical sophistication, but also relying on databases that are usually in the control of either the State or the industry (cf. Murphy 2007, p. 749). Should proprietary interests be addressed sufficiently by means of specific rules adapted to the specificities and needs of criminal proceedings, privacy concerns remain significant, particularly where the evidence is generated through a *consumer product* and, in order to prove its fault rate, access to mass data of other consumers may be required.

Against this background, one may argue that, at the end of the day, AI-generated evidence presents *almost* the same challenges DNA tests presented at the outset of their use in judicial settings. This claim is partly true inasmuch as AI distinguishes itself by the 'ability to monitor the surrounding environment, evaluate human behaviour and act autonomously', namely a higher level of agency (Gless 2020, p. 211). Additionally, the message AI-supported devices convey –in their capacity as consumer products– presents extra layers of complexity to the extent

[25] AI systems that incorporate deep neural networks (i.e., a machine learning algorithmic structure inspired by the structure and mechanics of human brains) serve as a representative example inasmuch as they do not follow a defined logical method, but are rather based on experience and learning: Fair Trials.org (2021); Yavar (2018).

that the input data originates from a variety of sources (and is not limited to biological samples) and is assessed autonomously, that is, without humans being necessarily in the loop (*ibid*, p. 198). This inevitably raises the question of whether the current criminal justice models, which have already been adapted to face the challenges forensic evidence of previous generations presented them with, even when operating at optimal levels, may safeguard adequately the use of AI-generated evidence or shall reckon anew how to uphold criminal procedural rights in this new context.[26] This article delves into this question by turning the spotlight onto the burden of proof rule and the right to contest evidence and judgments as core elements of PoI (Sect. 3.2). In this context, the focus will lie on, *inter alia*, the need for new evidentiary rules *to the extent* that AI-generated evidence presents the defence with often insurmountable burdens by increasing significantly the 'innocence threshold' (Galetta 2013; Milaj and Misfud Bonnici 2014, p. 425). Such questions will become more pressing, once AI-generated evidence stops being too new to be reliable (Gless 2020, pp. 215–216) or even becomes the primary accuser in a criminal trial (Roth 2017).

2.3 Recidivism algorithms

The use of algorithms, big data and AI in criminal justice settings usually refers to two distinct "moments": (1) *before the criminal trial*, where the focus lies on predictive policing instruments employed, as shown above (Sect. 2.1), before the judicial process or before a court referral, or AI-supported investigatory tools applied in the prosecution of crime; and (2) *during the criminal trial*, where the focus lies, besides AI-generated evidence (Sect. 2.2), on predictive tools designed to assess the risk of recidivism and facilitate decision-making in the courtroom (CEPEJ 2018, pp. 49–53).[27] If an algorithm is a 'sequence of computational steps that transform the input into the output' (Cormen et al. 2009, p. 5), then a recidivism algorithm transforms sets of data gathered from samples of a *relevant* population into an assessment of the risk of re-offending relating to a *concrete* individual that has already been accused or even found guilty of having committed a crime (Quattrocolo 2020, p. 131).[28] There are different procedural stages at which such a risk-assessment may be deployed, including (Quattrocolo 2020, pp. 132–135):

[26] It is worthy to stress that criminal procedural rights are linked tightly to the democratic character of a State considering the functions criminal procedure serves: protecting legal interests against crime on the one side and safeguarding personal freedoms on the other: Giannakoula et al. (2020), p. 49.

[27] These tools are usually referred to as 'algorithmic justice' or 'preventive justice'.

[28] This scheme already reveals a contradiction inasmuch as recidivism algorithms seem to be employed to assess the defendant's future behaviour in the context of a procedure designed to 'reconstruct a fact that occurred in the past [and] to establish the culpability for it': Quattrocolo (2020), p. 131. It is true that the purposes of criminal punishment go beyond repression and extend to prevention of future crime – with the doctrinal approach to criminal sanctions varying considerably even among legal orders that follow the same legal tradition: *ibid*, pp. 137–146. This article neither delves into this debate nor addresses the different approaches to sentencing at national level.

1) *pre-trial coercive measures*, such as pre-trial detention and bail, namely measures imposed—before the facts and the defendant's liability are determined definitively—to address, for instance, the risk of re-offending, escaping or tampering with evidence;
2) *sentencing* following a fact-finding-based conviction—with criminal punishment usually consisting of retribution-oriented penalties, which are based on proportionality between the criminal act, individual guilt and the sanction, and preventive measures seeking to address individual dangerousness and to protect the society from further harm;
3) *re-evaluation of the sentence already imposed*; that is, for instance, the case of early release on parole, where the focus lies on the risk of re-offending.

In the European Ethical Charter on the use of AI in judicial systems and their environment, CEPEJ identifitied only one predicitive tool that is employed before *European* criminal courts for the purposes outlined above. This is the case of HART (Harm Assessment Risk Tool), a software based on ML and trained with the use of Durham Police archives dating from 2008 to 2012, in order to assess the risk of re-offending on the basis of 34 risk predictors (CEPEJ 2018, p. 51; Oswald et al. 2018, pp. 227–229). Offernders are classified in three risk groups: high risk to commit a new *serious* offence[29]; moderate risk to commit a *non-serious* offence; and low risk to commit *any* offence. The critical time framework is the same in all three categories: two years after the commission of the first offence (Oswald et al. 2018, p.227; Palmiotto 2021, p.62).[30]

Two characterists of HART have attracted scholarly attention, if not concern: the risk predictors used to build the model and the built-in error classification. The behavioural predictors (29 in total), which are related to the individual offending history, are coupled with age, gender, two forms of residential postcode, out of which the primary one is limited to the first four digits and, thus, encompasses a rather large geographic area, and the police intelligence reports relating to the specific offender (Oswald et al. 2018, p.228). This combination raises at least three major concerns. *First*, this model takes into account parameters that fall outside the scope of the offender's control, namely age and gender, and, thus, poses not only discrimination risks, but also contradicts moral liability as a cornerstone of criminal law. *Second*, the use of the residential postcode as a variable may result in a 'feedback loop that may perpetuarte or amplify existing patterns of offending' (*idem*). This is the case when increased police focus on the highest-risk postcode areas leads to increased arrests of people residing there and, subsequently, these arrests serve as input data to generate a risk assessment in the context of the same model.[31]

[29] In Durham, the following offences are classified as serious: murder, attempted murder, aggravated violent offences (e.g., grievous bodily harm), robbery, sexual crimes and firearm offences: Oswald et al. (2018), p.227.

[30] The offenders classified as 'moderate risk' are entitled to attend the Checkpoint programme, a culture-changing initiative within Durham Constabulary: Oswald et al. (2018), p.227.

[31] The counterargument is that advanced algorithms differ from earlier methods of forecasting inasmuch as they 'are based upon millions of nested and conditionally-dependant decision points, spread across many hundreds of unique trees': Oswald et al. (2018), p.228.

This may ultimately lead to discriminatory bias towards geographical areas usually marked by community deprivation (Palmiotto 2021, p.63). *Third*, the legality of the chosen parameters should not be taken for granted. As Oswald et al (2018, p.239) explain, this may be the case with the inclusion of *spent* convictions as a risk variable, considering that Section 4 of the Rehabilitation of Offenders Act 1975 stipulates that a rehabilitated person is to be treated—for all purposes in law– as a person who has not committed the partiuclar offence. This paves the way for contesting a decision to reject, for instance, a bail request, which was reached on the basis of, *inter alia*, an algorithmic risk assessment, 'on the grounds that this was an *ultra vires* decision' (*idem*).[32]

Regarding the built-in error classification, the HART algorithm has been trained to distinguish between cautious errors, i.e., the over-estimation of a risk, and dangerous ones, i.e., the under-estimation of a risk. The former is linked to a lower cost compared to the latter and, thus, may occur more frequently (Oswald et al. 2018, p.228). This choice is translated into the following reality: a low risk individual is more likely to be subjected to coercive measures than a high risk individual to be released (Palmiotto 2021, p.64). In other words, false positives compensate for false negatives. This contradicts the error classification in criminal justice settings, where '[i]t is better that ten guilty persons escape than that one innocent suffers' (Blackstone 1893, p.358). Additionally, error assessment of this kind raises the question of who decides about what. The prioritisation of false negatives against false positives or the decision about the risk thresholds should lie with democratically elected decision makers and be open to scrutiny (cf. Završnik 2021, p.633). Or, as Green (2018) argues, this is a 'political exercise that involves making normative judgments about the tradeoffs between reducing incarcenation and reducing crime'. That said, a change in the algorithmic error classification or risk thresholds should not be dealt with as a mere technical tweak, but rather as a public policy change to be scrutinized at political level (*idem;* Palmiotto 2021, p.64) in the context of an *informed* democratic discourse.

Similar concerns have been expressed in the light of the US-American experience with COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)[33] and the landmark decision in *Loomis v. Wisconsin* case. Eric Loomis, who was allegedly involved in a drive-by shooting, was charged with first-degree recklessly endangering safety, attempting to flee or elude a traffic officer, operating a motor vehicle without the owner's consent, possession of a firearm by a felon, and possession of a short-barreled shortgun or rifle—in all cases as a repeat offender (*Loomis*, 881 N.W.2d, par. 754). The court accepted the defendant's plea, which was limited to two of the lesser charges, and ordered a pre-sentence investigation—with the respective report including a COMPAS risk assessment. According to the latter, Loomis was classified as a high risk of recidivism in all three categories COMPAS

[32] See also The Law Society of England and Wales (2019) and Palmiotto (2021), p.64, who turn the spotlight onto the lawfulness of the origin of the data and its re-use on behalf of the police.

[33] COMPAS is owned by a company named Erthpoint (previously called Northpointe): Wisser (2019), p.1814.

produces an assessment (*ibid*, par. 755), namely pre-trial release risk, general recidivism and violent recidivism (Wisser 2019, p.1814), and was denied parole on this basis.

Loomis challenged the algorithm as a violation of his right to be sentenced using accurate information, to have an individualised sentence as well as to non-discrimination, considering that gender has been one of the risk predictors (*Loomis*, 881 N.W.2d, par. 757). The court's final decision was in favour of the State, but stressed some limitations and cautions courts shall observe to avoid due process violations in the event of using a COMPAS risk assessment (*idem*). More specifically, (1) risk assessments may not be used as the determinative factor in sentencing; (2) in addition to the risk assessment, the choice of a specific sentence has to be justified; (3) risk assessments are not to be used to determine whether someone will be incarcerated or not or how severe his/her sentence will be; and (4) courts deploying COMPAS shall be provided with written advisement including limitations of and cautions about it (summarised by Wisser 2019, p.1815).

This case law has also triggered an energetic debate on the possibility of using AI-supported algorithms to assist decision-making in criminal matters, not necessarily limited to sentencing, before European courts.[34] Pro-arguments focus on the right to a hearing within a reasonable time and the right to such a hearing by an independent and impartial tribunal as core elements of the fair trial principle (Art. 6(1) ECHR). Algorithms are presented as a solution to the load of often undermanned criminal courts, lengthy and costly procedures and human mistakes (Melzer 2020, p.148), and as a source of new knowledge. Besides this, their output is expected to be "free" from subjective criteria or sympathies with the one or the other side and, as such, to decrease arbitrary judgments (*ibid*; Dreyer and Schmees 2019, p.759), as well as to grant the defendants equal opportunities before all criminal courts (Ebersbach 2020, pp.27, 32–33).

However, fair decision-making presupposes not only access to complete and bias-free data algorithms[35] (as already shown in the case of predictive policing; Sect. 2.1), but also the possibility to translate different decision-making components into input data for the algorithm, including leniency, human values and social perceptions. Additionally, equal *criminal* justice does not necessarily mean reaching a decision by matching predetermined clues and results as part of a mathematical formula. It is rather based on *individualised* culpability—shaped by factors within the control of the accused person—for an action, which has already been materialised in the empirical world by harming or endangering a legally protected interest, and leads to an *individualised* punishment that is shaped on the basis of the criteria the law provided for prior to the crime commission (Papadimitrakis 2019). Behind algorithmic risk assessment, however, there is a model that is not trained to evaluate the

[34] Currently, criminal cases fall outside of the scope of application of software-based tools employed in the judicial sector: CEPEJ (2018), pp.14, 17–18.

[35] For instance, in the case of COMPAS, ProPublica found that black defendants were more likely to be wrongly classified as high risk to re-offend than white ones, while the latter were more likely to be wrongly predicted as being a low risk: Angwin and Larson (2016).

individual as a unique personality. It is rather the individual that has to fit into predetermined categories and sets of factors (Greenstein 2022).

Additionally, and as already mentioned in the case of HART, certain risk variables or decision parameters in general may be inappropriate to bring into criminal sentencing or equivalent (Kehl et al. 2017, p.23) inasmuch as the laws in force do not provide any guidelines for this purpose or even include an express prohibition as to their inclusion into a certain judgment. In that sense, the algorithm that takes into account these parameters may, for instance, embrace an 'other-wise condemned discrimination, sanitized by scientific language' (Starr 2014, p.806). This does not necessarily signal malicious intentions, but rather that (criminal) law and computer science may approach these variables differently—with a programmer considering the inclusion of gender as risk parameter as a means to promote the accuracy of the risk assessment and not as a discrimination enabler (Wisser 2019, p.1818).

Next, the use of the recidivism algorithm needs to be substantiated, like in the case of algorithms employed for predictive policing or evidentiary purposes, in the sense of being subject to review as such. This entails a distinction that the court failed to make in the *Loomis v. Wisconsin* case, namely the distinction between the individual pieces of data the algorithm is supplied with and the scoring method as review subjects (Kehl et al. 2017, p.23; Wisser 2019, p.1822). Accessing this information and having it translated into a language one understands is a prerequisite for safeguarding transparency in the realm of criminal justice and protecting the right to participate effectively in a criminal trial and to defend oneself (Art. 6(3) ECHR). That said, Quattrocolo's (2020, p.93) request for either establishing technical standards that enable *ex ante* reverse engineering or setting review models that offer *ex post* validation of the algorithmic output is a legitimate one. In this context, it is not only the defendant as a layperson that needs an intelligible explanation, but also the judge (Greenstein 2022), so that (s)he can reach an informed (non-arbitrary) decision about how much to rely on the algorithmic output and how to interpret the latter—the same way that (s)he is expected to assess critically the expert opinions presented to him/her. Otherwise, it is highly unlikely that a judge will reach the "courageous" decision to distance him/herself from the algorithmic output (*idem*; CEPEJ 2018, p.56).

This article discusses algorithmic risk assessments, including the use of recidivism algorithms as outlined above, while turning the spotlight onto the current, past-oriented reading of PoI and examining the need for expanding it beyond the stages of criminal proceedings that exclusively focus on past crimes (Sommerer 2018) alongside the need for revising the notion of doubt in criminal justice settings (Sect. 3.2.3). These research questions are aligned with the observed shift of suspicion's focus from past onto future. This can be considered as part of a general trend of over-focusing on risk management among the *multiple* purposes of criminal law and sentencing in particular (Kehl et al. 2017, pp.26–27). This trend is not based on a stable empirical basis; that is, there is, for instance, no clear indication whether longer imprisonment decreases the risk of re-offending (Završnik 2021, p.627). Besides this, it is not limited to the area of sentencing. Pre-emption-oriented decision-making is also observed at law-making level—with a gradual (but consistent) transition from a post-crime to a pre-crime logic revealing a transfer of criteria and

ideologies from other fields of regulation, such as police law, into criminal law. This transition (also referred to as 'precognitive paradigm of criminal law') is already reflected in the adoption of pre-preventive measures, ranging from criminalisation itself to broad surveillance and data processing for the purposes of combatting crime (Giannakoula et al. 2020, p.57, 87), as part of various criminal policies with the EU counter-terrorism measures[36] offering maybe the most representative example (*ibid*, pp.51–52, 56–63; Kaiafa-Gbandi 2019).

3 To be presumed innocent in the era of algorithms, big data and artificial intelligence

FRA published in 2020 a report entitled 'Getting the Future Right—Artificial Intelligence and Fundamental Rights', which was based on 91 interviews with public officials and employees in the private sector in selected EU MSs. The survey included questions about 'their use of AI, their awareness of fundamental rights issues involved, and practices in terms of assessing and mitigating risks linked to the use of AI' (FRA 2020, p.6). The majority of the interviewees acknowledged that the use of AI-driven applications may have an impact on fundamental rights—with their responses varying depending on how they use AI (e.g., ML-based pension forecasts, social benefit algorithms, targeted advertising) and what they understand under fundamental rights (*ibid*, pp.58–59). Privacy, data protection and non-discrimination appeared to be the "protagonists". In addition, there were brief references to human dignity, the right to a fair trial and the right to effective remedy (*ibid*, p. 59). The right to be presumed innocent *as such*, namely as a standalone right, has been only referred by public sector representatives (*idem*). This "outsider" is chosen below to feature challenges often remaining under the radar of practitioners and policymakers working on AI. This choice is aligned with the first opinion FRA supports in the report mentioned above:

> 'When introducing new policies and adopting new legislation on AI, the EU legislator and the Member States, acting within the scope of EU law, must ensure that respect for the *full spectrum of fundamental rights*, as enshrined in the Charter and the EU Treaties, is taken into account' (*ibid*, p.7, emphasis added).

Both ECHR and CFR endorse PoI using an almost identical formulation: Everyone charged with a criminal offence shall be presumed innocent until proven guilty

[36] The issue of forwarding data to a public authority or retaining it in a general or indiscriminate way as part of the fight against terrorism arose in all four cases recently adjudicated by the CJEU (*Privacy International v. Secretary for Foreign and Commonwealth Affairs*, Case C-623/17, CJEU, 6 October 2020; *Quadrature du Net and Others v. Premier ministre and Others*, Joined Cases 511/18, C-512/18 and 520/18, CJEU, 6 October 2020; cf. Opinion of Advocate General Campos Sánchez-Bordona in Case C-623/17, 2020, and in Joined Cases 511/18 and C-512/18 and Case C-520–18; see Tracol 2021).

according to law (Art. 6(2) ECHR).[37] In the case law of the European Court of Human Rights (ECtHR), PoI has also been associated with various *meta-rules*, such as the principle of objectivity requiring that the judiciary 'should not start with the preconceived idea that the accused has committed the offence charged' (*Barberà, Messegue and Jabardo v. Spain*, ECtHR, 6 December 1988, par. 77), or the obligation to refrain from judicial pronouncements of guilt prior to a court finding of it (Commission of the European Communities 2006, p.5).

The EU legislator laid down common minimum rules concerning certain aspects of PoI in the Directive (EU) 2016/343,[38] which was adopted as part of the Roadmap for strengthening the procedural rights of suspected or accused persons in criminal proceedings (also referred to as 'the Roadmap').[39] This Directive applies to natural persons *suspected or accused in criminal proceedings*, *for the duration of the proceedings* (Villamarín López 2017, p.339). It follows the same pattern the ECtHR followed in the past by taking into consideration the twofold effects PoI has during criminal proceedings: first, it requires that the suspect or accused person is treated in a way that corresponds to his/her legal situation of not having yet been found guilty; and, second, it stipulates that the conviction of the accused person presupposes that the prosecution has presented sufficient incriminating evidence for the court to undermine his/her PoI (*ibid*, p.343). Using the general rule of PoI (Art. 3) as a starting point, the Directive deals with it, first, as a *rule of treatment* throughout the criminal proceedings with regard to public references to guilt (Art. 4) and the presentation of suspects and accused persons (Art. 5), and, second, as a *rule of judgment* by reference to the burden of proof (Art. 6) and the right to remain silent and not to incriminate oneself (Art. 7) (*idem*).

The following analysis first examines the EU legislator's choices relating to the scope of PoI in the light of the use of AI in law enforcement and criminal justice settings (Sect. 3.1). Next, it delves into PoI safeguards—with a special emphasis on the equality of arms principle, the burden of rule principle and the *in dubio pro reo* principle (Sect. 3.2).

3.1 The scope of PoI and the challenges posed to innocence outside of it

Under the terms of Art. 2 Directive (EU) 2016/343, PoI can be invoked by 'natural persons who are suspects or accused persons in criminal proceedings' and 'at all stages of criminal proceedings'. Regarding the first requirement, the formulation chosen follows the case law of the ECtHR[40] inasmuch as it implies that those who are not subject to any criminal investigation fall outside the scope of PoI (Villamarín

[37] Art. 48(1) CFR stipulates that *everyone who has been charged* shall be presumed innocent until proved guilty according to law.

[38] Besides PoI, the Directive regulates the right to be present at one's trial.

[39] The Roadmap was adopted by the Council on 30 November 2009 and, subsequently, was welcomed by the European Council and made part of the Stockholm programme – An open and secure Europe serving and protecting citizens on 11 December 2009.

[40] E.g., *Zollman v. UK*, ECtHR 27 November 2003.

Lopes 2017, p.340). The same applies to those that are called to testify before the police without being suspects or accused, but obtain this procedural "label" during the course of the interrogation—situation which continues being covered by the ECtHR case law and the Directive 2013/48/EU on legal assistance (*idem*). Regarding the second requirement, the Directive (EU) 2016/343 goes beyond the ECtHR case law, according to which the application of PoI depends on the existence of a criminal charge referring, thus, to persons subject to an advanced stage of criminal proceedings (ECtHR 2020, pp.9–11). According to Art. 2 Directive (EU) 2016/343, PoI applies.

> 'at all stages of the criminal proceedings, from the moment when a person is suspected or accused of having committed a criminal offence, or an alleged criminal offence, until the decision on the final determination of whether that person has committed the criminal offence concerned has become definitive'.

Furthermore, Recital 12 Directive (EU) 2016/343 stipulates that PoI may be invoked even before the affected natural person 'is made aware by the competent authorities of a Member State, by official notification or otherwise, that he or she is a suspect or accused person'. The EU legislator ascribes particular importance to the pre-trial phase, during which *most* evidence collection and analysis actions take place with a significant impact on suspect's future defence (FRA 2021, p.27). In other words, being under suspicion in criminal proceedings, not yet concluded by a definite resolution, suffices for being entitled to PoI safeguards.

A contrario, neither the Directive nor PoI as such, including the meta-rules mentioned above, apply when being under suspicion *outside the context of criminal proceedings*. This is, *inter alia*, the case of those for whom risk reports are generated by means of predictive policing in the kind of pattern-based data mining (Sect. 2.1) and who are classified as an alleged high risk to commit a/any crime in the future—taking into consideration that the generation of a risk report does not always, nor automatically, result in the initiation of criminal proceedings. Pre-suspicion of this kind is more than often coupled with gathering and analysing information on behalf of LEAs, in order to be prepared for the "fulfilment of the prophecy", namely when the pre-suspect commits that/any crime. The pre-suspect is often called a 'person of interest', namely a person who may have not been investigated for having committed a concrete crime, but is singled out by the algorithm and becomes a target of surveillance (Galetta 2013) with the aim of accumulating knowledge for future reference (Giannakoula et al. 2020, pp.58, 60–62, 68–69).

This reality, which is supported and enhanced by technological advancements such as AI (Greenstein 2022), has the potential for sabotaging the trust relationship between citizens and the State (cf. EC 2014, p.3; Fuster 2020, p.11). This relationship entails the value choice that citizens do not live under constant surveillance (Giannakoula et al. 2020, p.72) and manifests itself not only in privacy, but also in PoI (Milaj and Misfud Bonnici 2014, p.423; Campbell 2013). This signals a major turning point as far as the perception of PoI is concerned, and, particularly, the intersection of its procedural, political and philosophical elements (Sachoulidou 2021b).

One may argue that it is not the first time PoI in the sense of a broader State-citizen trust relation is compromised to accommodate police discretion when investigating suspects. Traditionally, police officers choose to focus on specific individuals through observation,

which may also include personal hunches, questioning and information conveyed by witnesses, victims or other third parties, the identity of whom may even remain secret (Joh 2016, p.15)—with their experience often leading them to target individuals with particular types of characteristics (Koss 2015, pp.302–304; cf. Gless 2018). New technologies, including the 'algorithms, big data, and AI' cluster, call out for re-examining that compromise by broadening significantly the scope of investigation and, thus, expanding law enforcement powers as well as by shifting the focus from past crime onto future threats.

This expansion places not only the need for new tools of police accountability at the forefront (Joh 2016, p.16), but also the need for revisiting the protective scope of fundamental rights, including the right to be presumed innocent. This includes answering, by means of legal regulation, the question of whether the police have to comply with the standards of individualised suspicion *before* supplying the algorithm with the data needed to generate the risk report, namely *before any intervention pertaining to criminal proceedings* takes place (cf. *ibid*, pp.18–19). Such *ex-ante* individualised suspicion is, for the moment, not required—with regard to both past *and future* wrongdoers (Data mining, dog sniffs and the Fourth Amendment 2014, p.695). To answer this question in an informed way, one needs to comprehend the specificities of AI-supported police discretion, namely *scale* and *future-driven decision-making*, compared to traditional surveillance means and investigative leads.

Predictive policing may be targeted at *everyone*. On the contrary, employing, for instance, sniffer drug dogs or relying upon instinct to stop and investigate a driver crossing the county's borders does not permit the police to 'surveil round the clock and track down every piece of information without at least some whiff of wrongdoing' (*ibid*, p. 696). The scale of data collection and analysis by means of predictive policing implies that the *average citizen* cannot trust that (s)he will not incur such close scrutiny (*idem*) before becoming a suspect in the procedural sense of the term, as well as a subject of PoI.

Predictive policing is part of a larger shift from a post-crime onto a pre-crime society (cf. Sect. 2.3) and from post hoc onto pre-emptive ordering practices (Strikwerda 2021, p.426). The focus lies progressively on 'anticipat[ing] and forestall[ing] that which has not yet occurred *and may never do so*' and this is a context where crime becomes a risk instead of a fact (Zedner 2007, p.262, emphasis added; Strikwerda 2021, pp.426–427) and the individual becomes a risk to commit a crime instead of a suspect for having already done so. This is a new context where innocence is challenged without PoI or the principle of personal accountability and guilt applying, and where lack of transparency and explainability become considerably problematic in the light of the decisions that may be based on the risk assessment, such as the initiation of a (this time) targeted –not necessarily neutral– police investigation (Strikwerda 2021, p.428) or even pre-trial coercive measures in the case of using recidivism algorithms (Sommerer 2018). In such a context, the question of whether PoI should apply goes hand in hand with that of whether a claim can be brought to an independent court against *an allegedly wrongful risk report* and—should this be the case—what is the evidentiary threshold the affected individual will have to meet (cf. Eckes

2021).[41] Next, a legal order that employs future-driven predictive policing tools should be prepared for enabling the defence to 'contest a conviction for biased predictive policing' (Gless 2018). This already raises the question of how predictive policing evidence as another kind of machine evidence (Sect. 2.2) may be employed in compliance with other PoI safeguards, such as the equality of arms principle and the burden of proof rule (see below).

Against this backdrop, the EU legislator should, *inter alia*, revisit the decision not to refer expressly to new-age police intuition and predictive policing in the Directive (EU) 2016/343. (S)he may do so in the future by revisiting the ECtHR case law and the guiding principle mentioned above that dictates that criminal proceedings may not be initiated with the preconceived notion that an individual has committed the offence in question (*Barberà, Messegue and Jabardo v. Spain*, ECtHR, 6 December 1988, par. 77). Otherwise, PoI may 'lose its place as a guiding principle' in the era of algorithms, big data, AI and the ubiquitous surveillance those (may) facilitate (Gless 2018).

3.2 PoI safeguards, new (?) evidentiary thresholds and the space left for rebutting and benefiting from doubt

The burden of proof rule –despite not being stated expressly in the main international human rights instruments, including the ECHR and the CFR—is recognised as an integral aspect of PoI by international human rights law (FRA 2021, p.65). Following EU law and CoE standards, Art. 6(1) Directive (EU) 2016/343 stipulates that.

> 'Member States shall ensure that the burden of proof for establishing the guilt of suspects and accused persons is on the prosecution […] without prejudice to any obligation on the judge or the competent court to seek both inculpatory and exculpatory evidence, and to the right of the defence to submit evidence in accordance with the applicable national law.'

Next, Art. 6(2) Directive (EU) 2016/343 incorporates the so-called *in dubio pro reo* principle into the burden of proof rule dictating that any doubt as to the question of guilt shall benefit the suspect or accused person.

[41] This should not imply that law enforcement and criminal proceedings are the sole areas where risk assessment may pose challenges to fundamental rights, including PoI. This may also be the case with, for instance, restrictive measures in the kind of freezing of assets and travel bans imposed by the EU against private persons in order to pursue its foreign policy objectives – following listing decisions: see the example of the EU global human rights sanctions regime at Eckes (2021). Choosing to focus on what occurs in the area of law enforcement and criminal proceedings, this article does not address this context or similar ones where PoI may even be challenged not by States themselves but in private settings.

Recital 22 Directive (EU) 2016/343 underlines that shifting the burden of proof from the prosecution to the defence would infringe PoI without prejudice to, *inter alia*, legal and factual presumptions of criminal liability.[42] The defence rights shall be maintained in this context; that is, the aforementioned presumptions should be 'reasonably proportionate to the legitimate aim pursued […] *rebuttable* and in any event, should be used only where the rights of the defence are respected' (emphasis added). Furthermore, Recital 23 Directive (EU) 2016/343 acknowledges that, in the case of legal orders following an inquisitorial system, it is upon the judges and competent courts to seek both inculpatory and exculpatory evidence. Such legal orders 'should be able to maintain their current system provided that it complies with this Directive and with other relevant provisions of Union and international law'. That implies that the burden of proof rule applies in both adversarial and non-adversarial systems (FRA 2021, p.65). In this context, and despite being governed by different procedural rules, fact-finding, whether in inquisitorial or adversarial systems, aims at establishing the truth (Gless 2020, p.218). Evidence generated by means of AI, whether in the context of predictive policing or in the larger context of human-technology interaction, appears to have the possibility to enhance fact-finding (*ibid*, p. 219). It remains, however, questionable to what extent it fits into the current procedural framework. This is examined in the following subsections in the light of the equality of arms principle as a field of interplay between PoI and the fair trial principle, the rule of the burden of proof *stricto* sensu, and the *in dubio pro reo* principle.

3.2.1 Equality of arms and evidence admissibility rules

To admit machine evidence—no matter how this may be classified (e.g., witness, documentary evidence)—before criminal courts, one should ensure that there are tools in place that allow judges, prosecutors and the defence to examine it adequately, whether by means of an expert opinion or in another way to be defined, should evidence of this kind become prevalent (Gless 2020, p.219). As particularly regards the defence, it is not only knowledge of all evidence adduced or observations filed that is required, but also the *possibility to comment on them with a view to influencing the court's decision* (e.g., *Brandstetter v. Austria*, ECtHR, 28 August 1991, par. 67) as a core element of the equality of arms principle that applies to both civil and criminal cases (ECtHR 2020, p.33). This is where PoI and the burden of proof in particular "interact" with the fair trial principle,[43] as enshrined in Art. 6(1)

[42] The regulation of the reversal of the burden of proof in the Directive has been subject to considerable controversies and disagreements – with the LIBE Committee arguing that such a reversal in criminal proceedings is unacceptable and the rule of the burden of proof should be respected. Against this backdrop, the reversal of the burden of proof has been included in a Recital (and not in an Article as originally planned). Recital 22 may almost adopt the protective content of the ECtHR judgment in the case *Salabiaku*, but it does so without expressly referring to this case law: Villamarín López (2017), p.352.

[43] The fair trial as a key principle enshrined in Art. 6 ECHR provides only a procedural and not a substantive guarantee; that said, errors of fact or law are to be taken into consideration only inasmuch as they amount to an infringement of the rights and freedoms enshrined in ECHR – with the ECtHR being primarily concerned about the *overall fairness* of criminal proceedings: ECtHR (2020); Palmiotto (2021), pp.58–59.

ECHR (and Art. 47(2) CFR). Additionally, the principle of equality of arms overlaps partly with the specific guarantees of Art. 6(3) ECHR (*idem*), including the defence right to examine an incriminating witness (Art. 6(3) lit. d ECHR).[44]

If national criminal courts go down the road of admitting machine evidence, equality of arms will presuppose:

1) awareness that machine evidence has been deployed in a specific case (cf. Greenstein 2022);
2) awareness of how and where to request explanations as well as what kind of explanations to request (e.g., calculation method, training data);
3) enough means to challenge the underlying decision (cf. FRA 2020, p.76).[45]

Such an approach to equality of arms, which is *prima facie* adapted to adversarial systems, should be placed within the context of the ECtHR's attempt to go beyond the dichotomy between adversarial and inquisitorial systems (considering the level of cross-pollination observed at national level) and create a cross-jurisdictional notion of procedural fairness (cf. Sellier and Weyembergh 2018, p.67; Gless 2020, p.221). However, the enforcement of such a system may be challenged by: the lack of specific procedural rules; competing interests, including secrecy surrounding LEAs' investigatory practices; and increased complexity of machine evidence that impacts adversely on intelligibility (cf. Ashworth and Zedner 2008; Quattrocolo 2020).

Should AI-generated evidence become mainstream in the future, the failure to lay down specific criminal procedural rules concerning relevance and reliability tests of this evidence as well as the means to contest the message it conveys would breach equality of arms. The primary purpose of criminal procedural rules in general is to protect the defendant against power abuses. Therefore, it would be the defence that would most likely suffer from omissions and lack of clarity in such rules (cf. *Coëme and Others v. Belgium*, ECtHR, 22 June 2000, par. 102).

Regarding competing interests as a source of opacity (cf. Sect. 2.2), measures restricting the defence rights may actually be put in place, but those should remain strictly necessary to meet the requirements of Art. 6(1) ECHR (cf. *Van Mechelen and Others v. The Netherlands*, ECtHR, 23 April 1997, par. 58; *Paci v. Belgium*, ECtHR, 17 April 2018, par. 85). Next, in order to ensure a fair trial, the difficulties the defence has to deal with in the light of limitations on its rights have to be counterbalanced by the procedures the judicial authorities follow (*Rowe and Davis v. UK*, ECtHR, 16 February 2000, par. 61; *Doorson v. The Netherlands*, ECtHR, 26 March 1996, par. 72—where the Court dealt with the case of anonymous witnesses; ECtHR

[44] Following the autonomous and broad interpretation of the term 'witness', Art. 6(3) lit. d ECHR may apply to algorithm-based evidence, when the latter is employed at trial and appeal proceedings, and can serve as a basis for the defendant's conviction: Palmiotto (2021), p.60.

[45] Case law from German Higher Regional Courts (e.g., Oberlandesgericht Bamberg, 13 June 2018, 3 Ss Owi 626/18) suggests that national judges may consider the idea of granting access to raw measure data in order to enable examining machine evidence adequately: Gless (2020), p.221.

2020, p.36). As will be explained below, promoting the use of open-source data and codes may be a fair (even just partial) solution to this problem.

Lastly, safeguarding equality of arms may call out for denying admissibility in the case of evidence the access to which for testing and contesting purposes is constrained due to high-level complexity. Civil-law jurisdictions usually do not provide for express rules on evidence admissibility—with all *relevant* evidence being admissible as a natural part of the courts' truth-seeking mission (Gless 2020, p.222). Interestingly, the US legal system, which also opts for the criterion of relevance (Rules 401–402 US Federal Rules of Evidence), provides for the exclusion of relevant evidence, in cases where:

> 'its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence' (Rule 403 US Federal Rules of Evidence).

This rule is interpreted narrowly with the focus laying on the judge's reasoning, particularly with respect to the potential of the evidence at stake for confusing or misleading the jury (Gless 2020, p.223). The *black box and lacking explainability* related issues that often arise in the case of machine evidence may be seen as another field of application for this rule or a similar one beyond the US legal order (*idem*).

Next, admissibility may be denied on the grounds of lack of reliability. In this regard, the fate of early lie detectors in criminal courtrooms could inform the future of AI-generated evidence. Starting with the *Frye* decision, the use of evidence generated with the help of Marston's lie detector was denied. Among other things, it was argued that, due to many variables, the research behind the machine was based on probabilities and the machine itself was not infallible (denial grounds listed and summarised by Oswald 2020, p.218). In the subsequent judgment on the appeal, Associate Justice Van Orsdel is quoted with stressing 'the difficulty of defining when a scientific principle or discovery crosses the line between "experimental and demonstrable"' (*Frye v. United States*, 293 F. 1013 (D.C. Cir 1923), cited by Oswald 2020, p.219), as well as with suggesting the general acceptance test by reference to the particular field to which a means of generating evidence pertains (*idem*). This test was only challenged 70 years after the Frye judgment in the case *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, where the so-called Daubert's four-factor test was suggested to determine admissibility.

According to the Daubert's test, it should be examined, whether a scientific technique: (1) withstands testing successfully; (2) has been subjected to peer review and publication; (3) has a known error rate and standards to be subjected to operation control; and (4) is generally accepted in a scientific community (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), par. 593–594, cited by Murphy 2007, p.576). This test introduces a burden of evidentiary admissibility that is, however, invalidated when scientific methodologies are treated in a law-like way. This is the case, as Murphy (2007, p.764) explains by means of reference to the US-American experience, when a scientific methodology is presumed admissible 'unless a party demonstrates by some unascertainable standard that other courts erred in admitting it, or that the science has undergone a significant change that warrants

revisiting a prior court's findings'—adopting an efficiency- and consistency-driven approach.

Compared to the leeway US-American courts granted to lie detectors (Oswald 2020, pp.219–220), the German Federal Court of Justice held that polygraph evidence is 'a completely unsuitable means of proof' that lacks probative value (Bundesgerichtshof, 1 StR 1998, 156/98, cited by Gless 2020, p.224). Moreover, the polygraph's measurements of bodily functions were deemed lacking in sufficient scientific methodology to be admitted as reliable evidence (*ibid*). To what extent AI-generated evidence will have a similar fate, inevitably depends on the level of explainability and intelligibility to be achieved in the near future, while the lack thereof may call out for adopting an exclusionary rule.

This is also the case with evidence generated by means of a biased algorithm, including the ones employed for predictive policing purposes—should the existence of bias be proved and despite the partly true positives. The decision to exclude pieces of evidence may downside efficiency of criminal justice in the sense of leaving aside potentially incriminating evidence, but is aligned with rule-of-law requirements. Respect for rule of law presupposes that both prosecution and criminal punishment are organised in an honest and transparent way and are based on sound, non-arbitrary judgments (Gless 2018). This is equally important as regards PoI *stricto* sensu: '*Everyone* charged with a criminal offence shall be presumed innocent until proven guilty according to law (Art. 6(2) ECHR, emphasis added). If bias intrudes in legal proceedings, PoI stops applying *equally* to *everyone*. This particularly regards the perceived level of dangerousness to the extent that this may lead first to one-sided criminal investigations and then to "pre-determined" criminal convictions.

3.2.2 The burden of proof, rebuttable (?) presumptions and available means to rebut

Substantial concerns may arise relating to the (even just indirect) impact on the burden of proof rule. As explained above, the burden of proof may lie on the prosecutor/judge (Art. 6, Recital 22 Directive (EU) 2016/343), but this does not automatically exclude *rebuttable* presumptions, whether of fact or of law concerning the criminal liability of a suspect or accused person, which respect proportionality and the use of which complies with the defence rights. This becomes relevant, for instance, in the case of persons caught in possession of drugs (of a certain quantity), weapons or other illicit goods that are expected to prove *otherwise* (FRA 2021, p.69). In the era of algorithms, big data and AI, the same person may be arrested and prosecuted after the predictive policing software raises a red flag, or trialled with the driver drowsiness detection system speaking against him/her in terms of responding to fatigue related warnings, or have his/her sanction determined on the grounds of an algorithmic assessment of his/her risk of re-offending. This may not lead to a direct reversal of the burden of proof, but raises questions relating to the actual *possibility to rebut the respective presumptions of fact that are not necessarily limited*

to past actions[46] or presumptions of fact in general that are consolidated through AI-supported tools.

The main challenge is to prove that such presumptions are inaccurate and, thus, should not be taken into consideration for decision-making purposes (cf. Galetta 2013; Milaj and Misfud Bonnici 2014, p.425; Marx 2005, p.357). To have a real chance to rebut them as well as machine evidence in general as part of both PoI and the right to defence (Art. 6(3) ECHR), one has to overcome the resources asymmetry and question the algorithmic output with a view to influencing the decisions of criminal justice authorities (Sect. 3.2.1). The difficulties inherent in challenging the algorithmic output become considerably higher considering the fact that machine evidence and the presumptions based on it are surrounded by increased objectivity and a sense of security. The judge may find it difficult to leave the algorithmic "comfort zone",[47] when the suspect or accused person is not able to present equally "strong" evidence. And this will be a problem to be solved in a context, where decision-making should take place within a reasonable time framework (Art. 6(1) ECHR) and it cannot be excluded that the evidence gathered by the police will be given, from the very beginning, stronger weight than that collected by the defence (cf. FRA 2021, p.67).

Challenges of this kind first call out for scrutinising once again the origin and the type of data that supports the generation of machine evidence as well as the way the latter takes place. Defence rights cannot be invalidated in the name of State secrecy, let alone in the name of proprietary interests of third private parties. In that sense, exclusive reliance on open-source materials, including codes, may be a fair compromise solution inasmuch as materials of this kind do not qualify legally as confidential ones (cf. Eckes 2021; Gless 2020, p.252; EP 2021, par. 17). Nevertheless, open-source data is neither automatically reliable nor free of bias, and open-source codes do not automatically protect the suspect or the accused person from false positives.

Equally important is the way machine evidence will reach criminal courts in the future inasmuch as this may impact on the defendant's ability to rebut it as well as the presumptions of fact based on it. This may occur by means of a *written report to be drafted by a court-appointed expert*, which, being part of the case file, should be accessible to the defence (Gless 2020, p.225). Reports of this kind usually include the tests administered by the expert and their results; that is, provided the respective device (for instance, a software embedded in vehicles) has been certified as evidentiary tool, there will be no reference to the raw data the device was supplied with or any detailed information about its design (*idem*). Such information could present, however, the defence with a realistic opportunity to challenge the evidence generated with the help of this device efficiently.

[46] The case of risk reports generated by means of recidivism algorithms are relevant inasmuch as sentencing is guilt-based.

[47] Additionally, one may also take into consideration the high amount of trust placed into judiciary regarding its impartiality *and* its ability to detect unreliable evidence that is coupled with the lack of clear evidentiary rules, and the (often over-) reliance on review of the established facts by appellate courts: Gless (2020), pp.226–227.

One may claim that, in any case, the defence is entitled to conduct its own, parallel investigations, including scientific tests, to prove otherwise. This possibility is associated with the way equality of arms (Sect. 3.2.1) is perceived in the light of national criminal procedural rules, and this varies considerably across European legal orders. For instance, in Italy and Ireland, the defence is expressly entitled to undertake investigations on its own and present the respective evidence at the trial alongside the prosecutorial material (Sellier and Weyembergh 2018, pp.67–68). In other countries, such as Germany, the Netherlands, Hungary, Spain and France, the defence is limited to a rather "reactive" role inasmuch as it has to rely on the prosecution or the judge, to whom it has to request the authorisation of further investigative acts (*ibid*, pp.68–69). In some legal orders, like the German one, the defendant may offer expert evidence *informally* to support his/her claims during the investigation phase. Next, the respective report will be incorporated into the case file and the defence will contribute the opinion of its own expert witness (Gless 2020, p.226). To profit from this possibility requires that one can afford hiring an external expert, the latter will have enough resources and access to the information needed to contest the machine evidence the prosecution intends to employ or the court assesses (*idem*), as well as that his/her findings will surpass the great deal of trust placed on public, State-led laboratories, the staff of which is usually appointed by the court to provide expert testimonies.

Alternatively, machine evidence may be introduced by means of written report and be subsumed under testimonial evidence in the sense of *a testimony of a witness that is not available to testify before the court*. The ECtHR has clarified the principles that apply in such a case in several decisions (e.g., *Al-Khawaja and Tahery v. UK* (GC), ECtHR, 15 December 2011, par. 119–147; *Seton v. UK*, ECtHR, 12 September 2016, par. 58–59; *Dimović v. Serbia*, ECtHR, 28 September 2016, par. 36–40; *T.K. v. Lithuania*, ECtHR, 3 December 2018, par. 95–96) as follows:

First, there must be a good reason for admitting the evidence of an absent witness considering that, as a general rule, witnesses provide evidence during the trial and all reasonable efforts should be made to safeguard their presence. Second, the admission of testimonial evidence in the place of live evidence at trial is to be treated as a last resort, when a witness has not been examined at any prior procedural stage. Third, this may place the defence at a disadvantage. Fourth, defence rights are unduly restricted in cases where the conviction is solely or mainly based on evidence of this kind. Fifth, Art. 6(3) ECHR is, however, to be interpreted in a holistic way. Against this backdrop, the decision to admit testimonial evidence as the sole or decisive evidence may not automatically result in a breach of Art. 6(1) ECHR, but the Court has to scrutinise the respective proceedings. In this context, it is required that sufficient *counterbalancing factors* exist, including strong procedural safeguards, 'to permit a fair trial and proper assessment of the reliability of that evidence to take place' (summarised at ECtHR 2020, pp.89–90).

In *Schatschaschwili v. Germany* (par. 111–131), the ECtHR confirmed, *inter alia*, that the absence of a good reason for the non-attendance of a witness remains an important factor when assessing the overall fairness of the proceedings (even if it is not necessarily conclusive of its lack) that 'might tip the balance in favour of finding a breach of Article 6 §§ 1 and 3(d)' (ECtHR 2020, p.90). Besides this, it explained

that these principles are also applicable in cases where it was unclear whether the testimony of the absent witness was the sole or the decisive evidence for convicting the defendant, but it played an important role and its use by the court had an adverse impact on defence (*idem*).

The ECtHR case law on Art. 6 ECtHR may have enhanced defence rights, particularly in inquisitorial systems, where cross-examination of witnesses is less formalised or there is no express prohibition of admitting hearsay compared to adversarial systems,[48] and as regards knowledge parity among the trial parties, but the test described above is relatively vague (Gless 2020, p.232). Irrespective of this, the machine itself cannot attend the trial, nor be replaced by a human being (manufacturer, designer, trainer or equivalent). The production/design of an AI-driven software or device is rarely a one-man business. Even where this is the case, the replacement of the machine by a human as a witness, whether expert or not, would only make sense if the latter would be able to decode the operational process behind reaching a certain conclusion, the judiciary would be able to ask the "right" questions as part of their truth-finding mission, and the defendant would have access to adequate resources to rebut (Palmiotto 2021, p.60). In other words, explainability of AI has to be ensured (Gless 2020, pp.233–234, 239–240) in order to provide the defence with the right means to contest AI findings. This is a goal already set by the scientific community, but at the same time a project *under construction* (e.g., Adadi and Berrada 2018). Until further progress is achieved, algorithmic opacity has an adverse impact not only on the burden of proof, as a core element of PoI, the equality of arms, and the fair trial principle as a whole, but also hampers reasoning of judicial decisions (Art. 6(1) ECHR) as well as the defence right to appeal that presupposes a comprehensible judgment reasoning (Hildebrandt 2018; Palmiotto 2021, pp.60–61).

3.2.3 In dubio pro reo

Starting with the pre-trial stages, criminal justice authorities must adapt their attitude towards the suspect or accused person in accordance with the burden of proof rule outlined above (Sect. 3.2.2). Next, they should charge him/her, only if –following the fact-finding procedure and weighing up the evidence presented in this context, and *respecting* defence rights– are convinced of his/her guilt (Art. 6(1) Directive (EU) 2016/343; Villamarín López 2017, p.351). Following this rule, any doubt as to the question of guilt shall benefit the suspect or accused person. Art. 6(2) Directive (EU) 2016/343 also adopts this rule –commonly referred to as the *in dubio pro reo* principle. It does so, however, without introducing any common standard of proof, like in the case of the Anglo-Saxon rule of 'beyond any reasonable doubt' (*ibid*, p.352).

[48] Instead, inquisitorial systems, such as the German one, opt for the principle of immediacy dictating that the judgment shall be based on what has been said and done at the public trial: Gless (2020), p.234. This does not negate the fact that machine evidence will be scrutinised 'behind the "closed doors" of the device', nor means that the national law does not provide for exceptions: *ibid*, pp.236–237.

The *in dubio pro reo* principle has already been observed in the ECtHR case law as a specific expression of PoI (*Barberà, Messegue and Jabardo v. Spain*, ECtHR, 6 December 1988, par. 77; *Tsalkitzis v. Greece (no. 2)*, ECtHR, 19 October 2017, par. 60). This principle may be violated in cases where national courts' decisions finding a suspect or accused person guilty are *not sufficiently reasoned* (*Melich and Beck v. the Czech Republic*, ECtHR, 24 July 2008, par. 49–55; *Ajdarić v. Croatia*, ECtHR, 13 December 2011, par. 51), or where an extreme and *unattainable burden of proof* is placed on the defence removing even the slightest prospect of success (*Nemtsov v. Russia*, ECtHR, 31 July 2014, par. 92; *Topić v. Croatia*, ECtHR, 10 October 2013, par. 45; *Frumkin v. Russia*, ECtHR, 5 January 2016, par. 166; summarised at ECtHR 2020, p.69).

Both scenarios become relevant when employing: (1) the output of predictive policing algorithms before criminal justice authorities to support the charge and, eventually, the guilt on the basis of a risk assessment; (2) machine evidence generated by means of AI; and (3) recidivism algorithms, particularly at the stage of guilt-based sentencing. Regarding the use of recidivism algorithms in general, one may argue that the decisions based on their output is not (always) about the guilt, but rather about coercive measures that may not even be guilt-based, such as pre-trial detention. However, the use of recidivism algorithms—even if not raising the question of whether the accused is deemed guilty of a *past* crime—does raise the question of whether the accused is deemed guilty of a *potential* crime (Greenstein 2022). In such a context, where one is scored and classified as low or high risk, the importance of disposition seems to be neglected. The latter is, however, of key importance to the extent that, in the case of *humans* as subjects of risk prediction, 'accuracy is likely to be subject to some fundamental limit due to the importance of extrinsic external factors relevant to the *specific individual*' (Oswald 2020, p. 224, emphasis added) or because the person concerned simply decides to act *otherwise*.

In all the contexts outlined above, the opacity of the algorithmic output may have an adverse impact on the reliability test and, thus, on the quality of the judicial reasoning rendering it insufficient, should the judiciary be unable to decode the information presented to it (Palmiotto 2021, pp.60–61). The same applies regarding the burden of proof when employing evidence that is hardly contestable and without granting access to proper resources, including time, money, and access to codes and databases, to contest it. Additionally, one should reconsider whether, at the end of the day, there is any space left for doubt in a context where the algorithmic output is surrounded by objectivity and a scientific language (EP 2021, par. 15) and, thus, "beats" those that are apt to lie. Should the algorithm leave space for doubt, one should also decide "how much of this doubt" would actually benefit the suspect or accused person. In other words, one should decide whether 1% false positive rate would be enough to exonerate the defendant.

Lastly, the power equilibrium should also be taken into consideration, particularly if it is almost about the technology to determine the amount of doubt. This transfer of governance to the algorithms is associated with a significant monopoly inasmuch as only those governing can access the resources required to produce or purchase the algorithm that is employed to make decisions about citizens (Greenstein 2022), including those regarding their own innocence.

4 The missed (?) and the eagerly awaited opportunities for the EU legislator

4.1 The reform of the EU data protection *acquis*

In 2016, the EU experienced an overall reform of its data protection rules with the adoption of the Regulation (EU) 2016/679 (GDPR) and the Directive (EU) 2016/680 (Law Enforcement Directive (LED))—with LED applying to personal data processing, whether AI-supported or not, in law enforcement and criminal justice settings (Art. 1(1), 2, Recital 11 LED).[49] Similar to the GDPR, the LED[50] was adopted to enhance public trust between the State (and the police in particular) and the society, facilitate the cooperation and data exchange among the EU MSs, and reinforce human rights in an era of rapid technological developments (Art. 1(2) LED; Marquenie 2017, p.328; Sajfert and Quintel 2019, p.3). It does so without addressing *expressly* technological developments, such as big data, AI and ML. Instead, like the GDPR (Recital 15), it opts for technological neutrality (Recital 18 LED; Zarsky 2017, p.1002; Pagallo 2017, p.37, 43; Gonçalves 2017, p.105).[51] Nevertheless, it entails provisions that are related to the data processing reality as shaped by new and emerging technologies, as well as automated decision-making as such and the rights of data subjects in this context.

Starting with *Art. 6 LED*, the data controller has to distinguish between (a) those for whom there are serious grounds to believe that they have committed or *are about to* commit a criminal offence (b) those convicted of a criminal offence; (c) victims of a criminal offence or those with regard to whom certain facts give rise to reasons for believing that they could be victims of a criminal offence; and (d) other parties to a criminal offence (e.g., witnesses, informants, contacts or associates of the persons referred to in previous categories). This distinction demonstrates the need to classify and treat data differently depending on the kind and the degree of one's involvement in the criminal enterprise (Marquenie 2017, p.330)—reflecting, thus, an—at least at first sight– reasonable decision of the EU legislator from a human rights perspective. This conclusion can be overturned, once one shifts the focus onto the (apparently) equal treatment of actual and potential suspects (Art. 6 lit. a LED) without any guidance as to how to distinguish between these two categories (Sachoulidou 2021b). To do so, one has to examine what is the meaning ascribed to the term 'potential suspect' by the national legislator. Should this entail those classified as a high-risk to commit a/any crime in the future, Art. 6 lit. a LED paves the way not only to indiscriminate data processing (cf. Marquenie 2017, p.331), but also contradicts the nexus between actual (not potential) wrongdoing and a certain kind

[49] That said, the GDPR provisions and their analysis falls outside the scope of this article.

[50] Regulating data protection in the field of judicial cooperation in criminal matters and police cooperation became possible through the abolition of the pillar structure by the Lisbon Treaty (Marquenie 2017, p.325; De Hert and Papakonstantinou 2009, p.410).

[51] Nonetheless, EU data protection laws are seen as a 'first attempt to enhance human interpretability in algorithmic design': Palmiotto (2020), p.18.

and level of suspicion that is required in order to mobilise the mechanisms of law enforcement and criminal justice (Sachoulidou 2021b). Recital 31 LED suggests that the (blurry) distinction introduced in Art. 6 LED:

> 'should not prevent the application of the right of presumption of innocence as guaranteed by the Charter and by the ECHR, as interpreted in the case law of the Court of Justice and by the European Court of Human Rights respectively.'

This may be a well-desired guidance, but it rather belongs to the legally binding provisions of LED and not to one of its Recitals.

Next, *Art. 11 LED* provides for a prohibition on decision-making *solely* based on automated means, including profiling, provided this affects the data subject adversely or significantly. Profiling is defined in Art. 3(4) LED as 'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to *a natural person*' (emphasis added). Solely automated decision-making is further described as 'the ability to make decisions by technological means without human involvement (Article 29 Data Protection Working Party 2018, p.8). The combination of these provisions and definitions suggests the following: First, profiling-based decision-making that is not *solely* based on automated means does not fall in the scope of Art. 11 LED (Sajfert and Quintel 2019, pp.8–9). Second, Art. 11 LED is not applicable to group or collective profiling (*idem*). Third, automated data processing for purposes other than decision-making fall outside the scope of Art. 11 LED. This may include, for instance, data processing by means of AI for enabling the function of a consumer product, which, however, may serve evidentiary purposes without making decisions *stricto* sensu (Palmiotto 2020, p.19). Finally, LED does not define the term 'adverse legal effect'. The latter shall presumably refer to results that affect the legal status of the data subject by altering his/her rights negatively (Sajfert and Quintel 2019, p.9). It remains questionable whether a classification as high risk to commit a/any crime pertains to this context. Similarly, it is questionable whether it is the machine evidence that affects the data subject adversely or significantly or the decision reached, *inter alia*, on the basis thereof (Palmiotto 2020, p.19).[52] The European Parliament has recently answered these questions in a positive way stressing that the relationship between fundamental rights protection and effective policing is to be seen as a core element in the discussion on whether and how AI can be used in law enforcement settings 'where decisions may have long-lasting consequences on the life and freedom of individuals' (EP 2021, point L and par. 3, 16). The same should apply to criminal justice settings, considering that AI may become a permanent part of them providing investigative analysis and assistance (*idem*).

The prohibition entailed in Art. 11 LED is not absolute. It can be removed in cases where automated decision-making is authorized by Union or Member State law, which appropriately *safeguards* the rights and freedoms of the data subject providing him/her with at least the right to obtain human intervention on the part of the controller (Art. 11(1) LED). Suitable measures to *safeguard* the data subjects' rights and freedoms are also required, when the decisions of this kind are based on special

[52] Palmiotto (2020, p.19) argues that risk-assessment software falls in the scope of Art. 11 LED.

categories of personal data (Art. 11(2) LED), such as data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs (Art. 10 LED). There is no such "flexibility" in the case of individual profiling that results in discrimination against natural persons (Art. 11(3) LED)—with the LED promoting impartial, data-driven police investigations (Sajfert and Quintel 2019, p.11). This prohibition should be coupled—even without this being expressly required by Art. 11 LED—with the duty to conduct a Data Protection Impact Assessment, *whenever the data processing is likely to result in a high risk for individuals* according to Art. 27 LED (*ibid*, p.12).

Regarding the rights of data subjects in the context shaped by Art. 11 LED[53] and beyond, Arts. 13 and 14 LED provide for the rights to receive and to access information. The right of access is subject to the limitations enshrined in Art. 15(1) LED, which, among other things, include the need to 'avoid prejudicing the prevention, detection, investigation or prosecution of criminal offences or the execution of criminal penalties' (lit. b), as well as the need to 'protect the rights and freedoms of others (lit. e). The combination of the respective provisions and the recitals accompanying those suggests the following: First, LED does not address the implications for criminal procedural rights, but rather "limits itself" in declaring their respect (e.g., Recital 38, 104). Second, Art. 11 LED does not specify the kind of human intervention that would be deemed sufficient to protect the rights of the affected individuals (Sajfert and Quintel 2019, p.10). Third, regarding the rights to receive and to access information, the LED does not refer expressly to the 'logic involved, as well as the significance and the envisaged consequences of such processing for the data subject' (Arts. 14, 15 GDPR) (Palmiotto 2020, p.20). Lastly, the aforementioned limitations included in Art. 15 LED call out for striking a balance between the data subject rights and the transparency requirements, on the one side, and the purposes of law enforcement and, particularly, criminal justice, and the involved interests of third parties (e.g., State, private entities) on the other. It is true that the data protection context may be optimal for doing so with regard to the right to privacy (*idem*), but this is not the case with criminal procedural rights, including the right to be presumed innocent. In other words, the data protection rules in place do not compensate for the lack of specific criminal procedural safeguards.

4.2 The E-evidence proposal(s)[54]

With information and communication technologies being employed throughout the whole spectrum of committing crimes, including *but not limited to* cybercrimes,

[53] According to Recital 38 LED, the safeguards mentioned in Art. 11 LED shall include, besides the right to human intervention, 'the provision of specific information to the data subject and the right to […] express his or her point of view, to obtain an explanation of the decision reached after such assessment or to challenge the decision'.

[54] This Section only discusses the E-evidence Proposal the EC released in April 2018 as well as the respective compromise proposal voted by the LIBE Committee in December 2020. In the meantime, in early 2023, the Council confirmed agreement with the EP on new rules to improve cross-border access to e-evidence:https://www.consilium.europa.eu/en/press/press-releases/2023/01/25/electronic-evidence-council-confirms-agreement-with-the-european-parliament-on-new-rules-to-improve-cross-borderaccess-to-e-evidence/. This agreement is based on the final compromise texts that were made publicly available on 20 January 2023: https://data.consilium.europa.eu/doc/document/ST-5448-2023-INIT/en/pdf and https://data.consilium.europa.eu/doc/document/ST-5449-2023-INIT/en/pdf.

evidence in electronic form[55] may assist in incriminating or exonerating the suspect or accused person (cf. Europol 2020a; 2020b; 2021). Against this backdrop and in a context shaped by multiple security concerns and increasing digitalisation (Sachoulidou 2021a, pp.779–780), as well as acknowledging the fact that investigative leads in electronic form usually have a short lifecycle and may be located in foreign or multiple jurisdictions (Kleijssen and Perri 2016), the European Commission released in April 2018 the so-called E-evidence Proposal. The latter included two sets of rules, one Regulation (EC 2018b; hereinafter referred to as Draft EPO-R) and one Directive (EC 2018c),[56] with the aim of governing cross-border access to e-evidence in criminal matters.[57]

These draft rules will apply to data, whether content or non-content (Art. 2(7–10) Draft EPO-R),[58] held by private service providers (Art. 2(3), 3(1) Draft EPO-R), that enables the identification of individuals or entities involved in or victimised by criminal activities, and, thus, may be of importance for criminal proceedings (Recital 18 Draft EPO-R; EC 2018b, p.14). This data may be either produced by means of a European Production Order (EPO; Art. 2(1) Draft EPO-R) or preserved through a European Preservation Order (EPO-PR; Art. 2(2) Draft EPO-R)—with the requirements for issuing an EPO varying depending on the data category at stake (Art. 5 Draft EPO-R, given the different levels of interference with fundamental rights (EC 2018b, p.14)).

The E-evidence Proposal may not refer expressly to evidence produced by means of algorithms, whether AI-supported or not, but its future applicability to transnational exchange of evidence of this kind cannot be excluded. That said, the challenges the defence is already presented with due to algorithmic opacity, particularly regarding the reliability test, may be exacerbated in the transnational context, where more actors are involved, *lex loci*, *lex fori* and European law interplay with each other, and the respective information, documents and forms are not always available in the language one understands (Palmiotto 2020, pp.21–22). This reality calls out for adopting rules at EU level that will empower the defence (*ibid*, p.22) no matter if the latter addresses machine evidence that is collected across the borders or not. And

[55] The term 'e-evidence' may stand for 'any data resulting from the output of an analogue device and/or a digital device of potential probative value that are generated by, processed by, stored on or transmitted by any electronic device': Biasiotti (2017), p.4.

[56] For a comprehensive, critical assessment of the Commission's proposal see: Böse (2018); Mitsilegas (2018); Tinoco-Pastrana (2020); Tosza (2020); Vasquez Maymir (2020).

[57] At transnational level, this matter is currently regulated by means of: 1) numerous agreements of mutual legal assistance (MLA) at the level of the Council of Europe, including the Budapest Convention on Cybercrime and its newly adopted Second Additional Protocol on enhanced co-operation and disclosure of electronic evidence, which has been open for signature since May 2022; 2) the EU MLA regime and the Directive 2014/41/EU (European Investigation Order Directive) which takes precedence over the former (Art. 35); 3) bilateral agreements singed by the EU MSs or the EU itself and third countries: Sachoulidou (2021a), p.778. At national level, the ways of handling (e-)evidence in criminal proceedings varies considerably: Smuha (2018), pp.93–95. For the impact of cross-border gathering of evidence on defence rights see: Bachmaier Winter (2013).

[58] Regarding non-content data, the Draft EPO-R distinguishes between subscriber, access and transactional data (Art. 2(7–9)).

this need becomes more pressing in the light of the declared interest of the Council of the EU (2018) to facilitate the use of new technologies, including AI-based ones, for law enforcement purposes in compliance with fundamental rights.[59]

Indeed, fundamental rights and criminal procedural rights are to be respected in a procedural context that is predominantly shaped by *speed* (see, for instance, the strict deadlines the draft EPO-R provides for executing and enforcing EPOs and EPOs-PR in Arts. 9, 10, 14) and *efficiency* (cf. Mitsilegas 2018). To this end, the E-evidence Proposal entails several declaratory references to criminal procedural rights (e.g. EC 2018b, p.10; Recital 14 EPO-R) and *Art. 17 Draft EPO-R* provides for a right to an effective remedy against the EPO during the criminal proceedings, for which the order was issued.[60] The *ratione personae* scope of this right includes suspects and accused persons whose data was obtained with an EPO (Art. 17(1) Draft EPO-R), as well as *those whose data was obtained without them being a suspect or accused person in criminal proceedings* (Art. 17(2) Draft EPO-R). At first sight, the inclusion of the second category is to be welcomed, as it increases the level of protection for those potentially affected by an EPO. Nonetheless, it is problematic inasmuch as it does not comply with the application scope of the E-evidence Proposal itself, which only covers *concrete* criminal investigations or proceedings (Art. 3(2), Recital 24 Draft EPO-R; EC 2018b, p.6, 15); that is, preventive (not strictly investigative) measures and pre-suspects should fall outside its scope (cf. EDRi 2020). In addition, the Proposal does not explain how non-suspects might be exposed to such (intrusive) investigatory means in the first place, if the principle of proportionality is actually respected (Recital 12 Draft EPO-R) and EPOs and EPOs-PR are not (to be) deployed for mass surveillance purposes (Sachoulidou 2021b).

Next, *Art. 17(6) Draft EPO-R* stipulates that the rights of the defence and the fairness of the proceedings are to be respected in criminal proceedings in the State issuing an EPO, when assessing evidence obtained through that order. Similarly, according to Recital 54 Draft EPO-R, suspects and defendants affected by an EPO shall benefit from all procedural guarantees applicable to them, such as the right to information (Recital 54). When attempting to assess these safeguards in the light of the algorithmic reality outlined in the previous Sections, the E-evidence Proposal does not seem to provide for any 'common standards and transparency requirements with regard to the process of data acquisition, access and search activities conducted by service providers and authorities', no matter if the accuracy of the data itself cannot always be taken for granted (Palmiotto 2020, p.25). In addition, the drafters of the Proposal do not take any position as regards the potential conflicts between the rights of the affected individuals and proprietary interests (Sachoulidou 2021b). Both a transparent framework and a clear balance of interests, however, could have had a positive impact on

[59] The Council (2018) underlined, *inter alia*, that 'the transparency and correctness of algorithms used in all applications of artificial intelligence as well as other appropriate safeguards need to be looked at in order to maintain the ability to verify the credibility of the results proposed and to ensure the overall accountability and lawfulness of such algorithms': cited by Palmiotto (2020), p.23.

[60] This right includes the possibility to challenge the legality of the measure, including its necessity and proportionality (Art. 17(3) EPO-R).

the exercise of the defence right to contest pieces of evidence of this kind and, thus, on PoI and the fair trial principle in general – instead of leading to a chain of (blind) trust in the quality of the evidence requested by LEAs and collected by private service providers (Palmiotto 2020, p.25).

Following its release, the E-Evidence Proposal has been criticised, *inter alia*, on the grounds of its compatibility with fundamental rights protection (LIBE Committee 2019, pp.144–145). Birgit Sippel, the LIBE Committee's Rapporteur for the Proposal, sought to address these concerns in her draft report, which was published in October 2019 (*idem*; Christakis 2020) and on the basis of which the LIBE Committee voted in favour of a *compromise* proposal introducing substantial changes to the Commission's proposal in December 2020 (EP 2020a). These two proposals have been subjected to inter-institutional negotiations since February 2021 (Wahl 2021). The added value of the compromise proposal has been considerable from a fundamental rights perspective inasmuch as it, *inter alia* (Sachoulidou 2021a, pp.785–787): (1) chooses the neutral term 'e-information' and explains that the information to be preserved and produced does not automatically count as admissible evidence (Recital 15 Draft EPO-R LIBE Committee Version)—respecting, thus, the right to a fair trial; (2) prioritises data integrity by means of suggesting the creation of a Common European Exchange System (Art. 7a Draft EPO-R LIBE Committee Version); (3) opts for informing *by default* the individual affected by an EPO or EPO-PR, unless there is a duly justified judicial order specifying the duration of the confidentiality duty that is subject to periodical review (Art. 11(1a) Draft EPO-R LIBE Committee Version); (4) provides for data protection safeguards in the form of a purpose limitation (Art. 11a Draft EPO-R LIBE Committee Version) and erasure of information that is illicitly obtained or no longer necessary (Art. 11b Draft EPO-R LIBE Committee Version); (5) excludes expressly the admissibility of information obtained in breach of the Regulation (Art. 11c Draft EPO-R LIBE Committee Version); and (6) erases the references to the *unknown* suspect/perpetrator, as well as the Art.17(2) Draft EPO-R. Nonetheless, when erasing Art. 17(6) Draft EPO-R and replacing it with express rules governing the admissibility of illicitly obtained e-information, the compromise proposal does not make any step towards addressing admissibility related concerns that may arise in the case of e-information generated by means of AI, or suggesting concrete means to re-enforce criminal procedural rights in the era of algorithms, big data and AI (Sachoulidou 2021a, p.789). This could have been achieved by stressing, for instance, that the examination of e-evidence shall take place pursuant to Art. 6 ECHR in the sense of promoting the right to cross-examination and limiting the bench-dominated approach to evidence examination, particularly regarding inquisitorial systems (cf. Gless 2020, p.249). The absence of proposals of this kind—both in the Commission's proposal and in the compromise proposal—is a missed (should the final legal text to be adopted move towards the same direction) opportunity for the EU legislator to set the tone with regard to the protection of the right to contest machine evidence as a core element of PoI, defence rights and the fair trial principle in general (similarly Palmiotto 2020, p.22, 25).

4.3 The AI proposal and next steps

In April 2018, the EU released its AI Strategy with a twofold aim: to make the EU a world-class hub for AI and to ensure the human-centric and trustworthy character of AI (EC 2018a). Next, AI HLEG presented in 2019 the 'Ethics Guidelines for Trustworthy Artificial Intelligence', according to which (AI HLEG 2019b, p.5): Trustworthy AI systems are—throughout their entire life cycle– *lawful* (that is, they comply with all applicable laws and regulations), *ethical* (that is, they ensure adherence to ethical principles and values), and *robust* from a technical and social perspective. To achieve this goal, they should meet the following key requirements: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; environmental and societal well-being; and accountability (*ibid*, p.2).[61]

In February 2020, the EU proceeded with releasing the White Paper on AI, where it stressed that the use of AI can affect EU values and lead to breach of fundamental rights, including the right to non-discrimination, an effective judicial remedy and fair trial (EC 2020a, p.11). Risks may arise from flaws in the overall design of AI systems or the use of data without correcting possible bias (*idem*). Similarly, the Council of the EU (2020b, p.5) put 'opacity, complexity, bias […] unpredictability and partially autonomous behaviour' at the centre of the attempts to ensure the compatibility of automated systems with fundamental rights. At the end of the same year, the European Parliament proposed legislative action with the aim of harnessing the opportunities and benefits associated with the use of AI *and* safeguarding the protection of ethical principles in this context—with an emphasis on, *inter alia*: human-centric, human-made and human-controlled AI; mandatory compliance assessment of high-risk AI; safety, transparency and accountability; safeguards and remedies against bias and discrimination; respect for privacy; and good governance relating to AI, including the data used or produced by it (EP 2020b, Annex to the Resolution, point A.III.).

It was against this backdrop that the EC (2021a, 2021b, 2021c) drafted and released the Proposal for a Regulation laying down harmonised rules on AI in April 2021, in order to govern the development of AI systems intended to be employed,

[61] Aiming to facilitate compliance with the seven key requirements, the AI HLEG (2020) also presented a detailed assessment list (Assessment List for Trustworthy AI (ALTAI)) and developed a prototype web based tool for practical guidance purposes: https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence (accessed 24 January 2022). In addition, further soft-law tools have been suggested with the aim of enabling the legal and ethical use of AI technologies, including the OECD (2019) principles on AI, the SHERPA Guidelines for the ethical use of AI and big data systems (Brey et al. 2019), and the EU's Guidance Note on 'Ethics by design and ethics of use approaches for AI' (EC 2021d). The CoE's Ad Hoc Committee on AI (CAHAI) also adopted the recommendation on the 'Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rules of law' in December 2021. These elements are designed to be included in legally binding or non-legally binding instruments that will make up the CoE legal framework on AI, and are intended to be submitted to the Committee of Ministers for further consideration: https://ai-regulation.com/council-of-europe-cahai-ai-recommendation/ (accessed 24 January 2022).

among other things, in law enforcement and judicial settings. More specifically, the EU AIA dictates that AI systems intended to be used by LEAs for:

- 'making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for potential victims of criminal offences';
- '[serving] as polygraphs and similar tools or to detect the emotional state of a natural person';
- 'detect[ing] deep fakes'
- 'evaluat[ing] the reliability of evidence in the course of investigation or prosecution of criminal offences'
- 'predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons as referred to in [Art. 3(4) LED] or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups';
- 'profiling of natural persons as referred to in [Art 3(4) LED] in the course of detection, investigation or prosecution of criminal offences'; and
- 'allowing [them] to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data' (Annex 3, point 6 EU AIA)

as well as those 'intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts' (Annex 3, point 8 EU AIA) are to be classified as *high risk* (Art 6(2) EU AIA), considering the implications of their use for fundamental rights, including the right to an effective remedy and to a fair trial, as well as the right of defence and PoI. These rights may be hampered where AI systems are not sufficiently transparent, explainable and documented (Recital 38 EU AIA). In the light of these risks, the EU AIA provides for a series of legal obligations on AI designers, programmers and developers relating to: implementation, documentation and maintenance of a risk management system (Art 9); appropriate data governance and management practices (Art 10); technical documentation before the system is placed on the market or put into service (Art 11); record-keeping (Art 12); transparency and provision of information to users (Art 13); human oversight (Art 14); accuracy, robustness and cybersecurity (Art. 15).

Being already an important step towards ensuring compatibility of AI systems with fundamental rights, the EU AIA should be informed by the subsequent Resolution the European Parliament adopted on 6 October 2021, which pays attention to the specificities of the *use of AI by the police and judicial authorities in criminal matters* and, thus, the specificities of *criminal law itself* – seeking to address risks for the protection of fundamental rights, including criminal procedural rights (EP 2021, point O). In this context, the EP particularly stressed that the use of AI is anything but a mere technical feasibility. On the contrary, it is 'rather a political decision concerning the design and the objectives of law enforcement and of criminal justice systems' (*idem*, point Q). Against this background, it suggested, among other things, the following:

The use of AI applications shall be prohibited when incompatible with fundamental rights (par. 2). Safety, robustness, security, fit-for-purpose operation, respect for the principles of fairness, data minimization, accountability, transparency, non-discrimination and explainability are the minimum safeguards when AI tools are developed or employed by LEAs or the judiciary; and such tools should be subject to risk assessment and strict necessity and proportionality testing (par. 4). AI systems employed in this area, and, particularly, those that may be repurposed for mass surveillance or mass profiling should be subject to strict democratic control and independent oversight (par. 6). LEAs and the judiciary should only use AI applications that comply with privacy and data protection by design (par. 11). To ensure the effectiveness of the exercise of defence rights and transparency of national criminal justice systems, EU MSs should adopt *specific, clear and precise rules on the conditions, modalities and consequences* of the use of AI for law enforcement and criminal justice purposes—with a focus on the rights of targeted individuals, complaint and redress procedures, including judicial redress, and the right to information relating to data collection process, the related assessments and the use of AI applications in a specific case (par. 14). Besides high legal standards and human intervention, the sovereign discretion of judges and decision-making on a case-by-case basis should be preserved; that is, AI is not to be employed for proposing judicial decisions (par. 16). Next, the EU MSs should only allow LEAs and judicial authorities to purchase tools and systems, the algorithms and logic of which are 'auditable and accessible at least to [them and] independent auditors, to allow for their evaluation, auditing and vetting' and which will not be closed or labeled as proprietary by the vendors (par. 17). In that sense, the use of open-source software is recommended where possible (cf. Sects. 3.2.1 and 3.2.2). Additionally, appropriate public procurement processes should be adopted, in order to ensure compliance with fundamental rights and applicable laws, and public–private partnerships, contracts and acquisitions and the purposes for which AI systems are procured should be disclosed to the public (par. 18). A compulsory fundamental rights impact assessment is to be conducted prior to employing any AI system in law enforcement and judicial settings (par. 20). The latter should also be subject to periodic mandatory auditing in the context of a clear institutional framework that will guarantee, *inter alia*, an informed democratic debate on the necessity and proportionality of AI in these areas (par. 21). In this context, of particular importance is interdisciplinary research and input (par. 22), as well as the specialised training of involved professionals relating to the ethical provisions, potential dangers, limitations and proper use of AI technology (par. 23). Lastly, the EU MSs should inform about the tools deployed by their LEAs and judiciary, their type and purpose, the types of crime they are applied to, their developers, as well as disclose through their authorities false positive and false negative rates of the technology in question (par. 33).

In other words, the authorities employing AI and the affected individuals should be placed on equal footing relating to access to information about, *inter alia*, the training data, the calculations and the assessment methods. In addition, AI's unique status in the realm of criminal law should be recognised and the respective safeguards should be adapted to the specificities of the interests at stake, including (but not limited to) personal freedom. In this context, of key importance is the collaboration with

experts, in order to understand not only the technology, but also to explain the underlying legal concepts (Gless 2020, p.252). Finally, the cost of *both* false positives and false negatives should be brought to the attention of policy-makers *and* the public.

When shaping EU *and* national law on the use of AI in law enforcement and criminal justice settings, next steps should also be informed by the recent CJEU case law on *neighbouring* scenarios, namely the preventive retention of traffic and location data or IP addresses and data relating to civil identity to combat crime and safeguard public security, the expedited retention, automated analysis, and real-time collection of such data, or the general and indiscriminate retention of data by access providers to online public communication services and hosting service providers for such purposes (Tracol 2021, p.10).[62] More specifically, the following statements of the Court may serve as guidance, particularly when employing AI or similar technologies for crime prevention purposes, that is, in law enforcement settings[63]:

First, collecting and processing data in a general and indiscriminate way –a possibility that may be enhanced by means of AI—constitutes a particularly serious interference with fundamental rights *where there is no link between the conduct of the affected individuals and the objective pursued by the legislation at issue* (*Quadrature du Net and Others v. Premier ministre and Others*, Joined Cases 511/18, C-512/18 and 520/18*,* CJEU, 6 October 2020, par. 143, 145). Targeted measures of this kind should be limited to what is absolutely necessary with relation to the categories of data to be collected and processed, the means of communication affected, the persons concerned, and their temporal and geographical scope (*ibid*, par. 147, 150, 178). This presupposes defining, for instance, *which crimes can be considered serious* enough to justify such intrusive preventive measures,[64] that is, the criterion of being a high risk to commit any crime is to be deemed invalid, as well as *which exactly persons can be the target of counter-terrorism preventive measures* (*ibid*, par. 188). Besides this, concrete geographical criteria should be determined in a non-discriminatory way.

Second, should a clear and precise legal basis for exceptions to the rule of trust between citizens and the State be detected, a *proportionality test* remains indispensable (Art. 52(1) CFR; *Digital Rights Ireland and Seitlinger and others*, Joined Cases C-293/12 and C-594/12, CJEU, 8 April 2014; *Privacy International v. Secretary for Foreign and Commonwealth Affairs*, Case C-623/17, CJEU, 6 October 2020, par. 63–64, 68, 76–78; *Quadrature du Net and Others v. Premier ministre and Others*, Joined Cases 511/18, C-512/18 and 520/18*,* CJEU, 6 October 2020, par. 113, 132; cf. Gaughran v. UK, ECtHR, 13 February 2020, par. 89). Besides

[62] In the case of these scenarios, the CJEU examined primarily the applicability of the Directive 2002/58/EC.

[63] Some of the CJEU positions have already been adopted by the European Data Protection Board (2020) in its Recommendations on the European Essential Guarantees for surveillance measures: Tracol (2021), pp.11–12.

[64] Art. 83(1) of the Treaty on the Functioning of the EU could serve as guidance: Tracol (2021), p.10. Even in this context, it remains problematic that newly established criminal offences, such as the ones included in the EU terrorism criminal legislation (see Directive (EU) 2017/541), often involve neutral acts (e.g., receiving training, travelling): Kaiafa (2019); Giannakoula et al. (2020), p.52, 59.

balancing the satisfaction of some rights and the damage to others, respect for proportionality also means that the exceptions to the rule of trust mentioned above shall not render EU law and fundamental rights protection standards in particular inapplicable (*Privacy International v. Secretary for Foreign and Commonwealth Affairs*, Case C-623/17, CJEU, 6 October 2020, par. 44). And once adopted, such exceptions should not be turned into the rule by the national legislator (*ibid*, par. 59).

Third, the decisions that may enable mass data collection and processing should be based not only on a clear and precise legal basis *in terms of providing for substantive and procedural conditions*, but also be subject to 'effective review, either by a court or by an independent administrative body whose decision is binding'. By means of this review it should specifically be verified that 'a situation justifying that measure exists and that the conditions and safeguards that must be laid down are observed' (*Quadrature du Net and Others v. Premier ministre and Others*, Joined Cases 511/18, C-512/18 and 520/18, CJEU, 6 October 2020, par. 179, 189, 192). In case of automated data analysis, the CJEU also highlighted that the algorithm must be based on specific and reliable pre-established models and criteria – to be regularly re-examined – and not on sensitive data in isolation (*ibid*, par. 180–182). Additionally, such analyses must be subject to human re-examination before a measure that may adversely affect the concerned individual is adopted (*ibid*, par. 182).

Lastly, when authorising derogations from the rule of trust between citizens and the State, *defining procedural safeguards* should also include the adaptation of their scope to the specific circumstances. Should the use of AI in law enforcement settings and for risk assessment purposes be the use case at hand, this calls out for extending the scope of the procedural safeguards beyond the "traditional" pre-trial and trial proceedings. In the case of PoI, this would mean extending its protective scope to individuals that have not yet acquired the "procedural label" of suspect or accused person, and to stages at which the individual may be "only" classified as a high risk to commit a/any crime in the future (cf. Campbell 2013, p.689). This suggests shifting the focus from the epistemic (rule of the burden of proof) onto the non-epistemic dimension of PoI (Sachoulidou 2021b), which gives effect to the moral and political claim of citizens to be treated by the State as law-abiding until it proves otherwise (Dennis 2011, p.354); that is, a general and prospective claim of civic trust that protects individuals from becoming defendants (Duff 2013, pp.180–181). Such an extensive reading of PoI would consolidate the logical argument that, in democratic States governed by rule-of-law principles, everyone who is not suspected or accused of having committed a specific crime cannot, nor should de facto, be presumed guilty or a risk that justifies surveillance or other coercive measures.

5 Conclusions

Algorithms, big data and AI have recently been put at the centre of attention in the realm of criminal law not only with new criminal liability related concerns, but also with their inherent potential to revolutionise law enforcement and criminal justice, and particularly to increase speed, efficiency and accuracy. Predictive policing, machine evidence and recidivism algorithms are examples usually deployed to showcase substantial benefits. However, the latter come hand in hand with risks for the protection of fundamental rights, including but not limited to criminal procedural rights. Being a multi-layered right as enshrined in ECHR, CFR and the Directive (EU) 2016/343, PoI serves as a representative example—with its protective scope often falling behind the algorithmic reality, the burden of proof rule as a core element of it being challenged by knowledge and access disparity, and the *in dubio pro reo* principle having to be re-visited to assess the space left for benefiting from doubt in the era of AI.

At the crossroads of AI, law enforcement and criminal justice, the rights of the suspect or accused person have to be balanced with other interests, including State and police secrecy and proprietary interests of third parties, as well as with other purposes of criminal proceedings, such as the efficient prevention, detection and prosecution of crime. This calls out for delicate regulatory steps, particularly in the area of criminal procedural law – acknowledging that, for the time being, false positives and algorithmic opacity are (or should be seen as) considerable burdens in the attempt of making AI part of the daily routine of law enforcement and judicial authorities in criminal matters, as well as that the affected individuals may experience grave adverse consequences, ranging from social exclusion on the basis of a risk assessment, the non-discriminatory character of which cannot be taken for granted, to violation of their personal freedom.

The EU legislator has already addressed similar concerns in neighbouring areas of regulation. This is the case with the EU data protection reform, including the adoption of the LED. The latter may represent a significant progress towards safeguarding the protection of privacy in the area of law enforcement and criminal justice, but does not do the same for criminal procedural rights, nor has it been expected to do so. The E-evidence Proposal, being a more recent legislative initiative of the EU legislator in the area of police and judicial cooperation in criminal matters, could have been a possible forum for introducing uniform procedural safeguards when the defence has to contest e-evidence generated by means of AI. Even after the significant involvement of the LIBE Committee into the respective legislative procedure, this goal has not been achieved. Lastly, the EU legislator has already proposed an innovative set of rules to govern AI—with its use in law enforcement and judicial settings having already been classified as high-risk. This article suggests that the EU AIA should be informed by the European Parliament's Resolution, which further addresses the specificities of criminal law as a field of application of AI tools, and the recent CJEU case law on preventive measures that may enable mass data collection and processing and, thus, mass surveillance. Besides this, it argues in favour of designing

specific procedural safeguards (e.g., defence access to training data, calculations and assessment methods, prioritisation of open-source codes, toning up the cross-examination side of criminal proceedings), as well as, if and where needed, introducing new layers of fundamental rights protection. Otherwise, until the technical standards become fundamental rights compliant (Hildebrandt 2014; EP 2021) and the law provides for sufficient safeguards, including democratic control and human oversight, a moratorium on the use of AI for law enforcement and criminal justice purposes should be considered.

In this context, the legislative initiatives taken at EU level should also serve as guidance and as an inspiration platform for the national legislator who is responsible for drafting not only rules that will increase public trust without hampering innovation, but also for striking a delicate balance between efficiency and fundamental rights protection. Such a balance is of vital importance in general and in the realm of criminal law in particular.

Funding Open access funding provided by FCT|FCCN (b-on). The author did not receive financial support from any organisation for the submitted work.

Data availability No data was used for the research described in the article.

Code availability No code was used for the research described in the article.

Declarations

Conflict of interest The author has no conflict of interest to declare that is relevant to the content of this article.

Consent to participate Not applicable.

Consent for publication Not applicable.

Ethics approval Not applicable.

References

Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Andrejevic M (2017) Digital citizenship and surveillance| To Pre-empt a thief. Int J Commun 11:18

Angwin M, Larson J (2016) How we analyzed the COMPAS recidivism algorithm. ProPublica, 23 May. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. Accessed 24 Jan 2022

Article 29 Data Protection Working Party (2018) Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679. https://ec.europa.eu/newsroom/article29/items/612053. Accessed 24 Jan 2022

Ashworth A, Zedner L (2008) Defending the criminal law: reflections on the changing character of crime, procedure and sanctions. Crim Law Philos 2(1):21–51. https://doi.org/10.1007/s11572-007-9033-2

Bachmaier Winter L (2013) Transnational criminal proceedings, witness evidence and confrontation: lessons from the ECtHR's case law. Utrecht Law Rev 9(4):127–146

Biasiotti MA (2017) A proposed electronic evidence exchange across the European Union. Digit Evid Elec Signat L Rev 14:1

Blackstone W (1893) Commentaries on the laws of England. J.B. Lippincott Co., Philadelphia

Böse M (2018) An assessment of the Commission's proposals on electronic evidence. Study requested by the LIBE Committee. https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOLSTU(2018)604989. Accessed 24 Jan 2022

Brey P, Lundgren B, Macnish K, Ryan M (2019) Guidelines for the ethical use of AI and big data systems. SHERPA Project. https://www.project-sherpa.eu/wpcontent/uploads/2019/12/use-final.pdf. Accessed 24 Jan 2022

Brannon M (2017) Datafied and divided: techno-dimensions of inequality in American cities. City Commun 16(1):20–24. https://doi.org/10.1111/cico.12220

Broeders D, Schrijvers E, Van der Sloot B, Van Brakel R, De Hoog J, Ballin H (2017) Big data and security policies: towards a framework for regulating the phases of analytics and use of Big Data. CLSR 33:309–323. https://doi.org/10.1016/j.clsr.2017.03.002

Buono L (2019) The genesis of the European Union's new proposed legal instrument(s) on e-evidence. Towards the EU Production and Preservation Orders. ERA Forum 19:307–312. https://doi.org/10.1007/s12027-018-0525-4

Burrell J (2016) How the machine 'thinks': understanding opacity in machine learning algorithms. Big Data 3(1):1–12. https://doi.org/10.1177/2053951715622512

Campbell L (2013) Criminal labels, the European Convention on Human Rights and the presumption of innocence. MLR 76(4):681–707. https://doi.org/10.1111/1468-2230.12030

Chessman C (2017) A "source of error": computer code, criminal defendants and the constitution. California Law Review 105: 179–228. https://papers.ssrn.com/sol3/papers.cfm?abstractid=2707101. Accessed 24 Jan 2022

Christakis T (2020) E-Evidence in the EU Parliament: Basic features of Birgit Sippel's draft report. European Law Blog. https://europeanlawblog.eu/2020/01/21/e-evidence-in-the-eu-parliament-basic-features-of-birgit-sippels-draft-report/. Accessed 24 Jan 2022

Commission of the European Communities (2006) Green Paper. The Presumption of Innocence. COM (2006) 176 final. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52006DC0174&from=GA. Accessed 24 Jan 2022

Cormen TH, Leiderson CE, Rivest RL, Stein C (2009) Introduction to algorithms, 3rd edn. MIT Press, Cambridge

Council of the EU (2009) Resolution of the Council of 30 November 2009 on a Roadmap for strengthening procedural rights of suspected or accused persons in criminal proceedings. 2009/C 295/01. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AC%3A2009%3A295%3ATOC. Accessed 24 Jan 2022

Council of the EU (2018) The future direction of EU internal security: new technologies and internal security—Preparation of the Council debate. 12224/19. https://data.consilium.europa.eu/doc/document/ST-12224-2019-INIT/en/pdf. Accessed 24 Jan 2022

Council of the EU (2020a) Joint Statement by the EU Home Affairs Ministers on the recent terrorist attacks in Europe. https://www.consilium.europa.eu/en/press/press-releases/2020a/11/13/joint-statement-by-the-eu-home-affairs-ministers-on-the-recent-terrorist-attacks-in-europe/. Accessed 24 Jan 2022

Council of the EU (2020b) Presidency conclusions – The Charter of Fundamental Rights in the context of artificial intelligence and digital change. https://www.consilium.europa.eu/media/46496/st11481-en20.pdf. Accessed 24 Jan 2022

Data mining, dog sniffs, and the Fourth Amendment. A framework for evaluating suspicionless mass surveillance programs (2014) Harvard Law Review 128(2): 691–712. https://harvardlawreview.org/2014/12/data-mining-dog-sniffs-and-the-fourth-amendment/. Accessed 24 Jan 2022

De Hert P, Papakonstantinou V (2009) The data protection framework decision of 27 November 2008 regarding police and judicial cooperation in criminal matters—A modest achievement however not the improvement some have hoped for. CLSR 25(5):403–414. https://doi.org/10.1016/j.clsr.2009.07.008

Dennis I (2011) The Human Rights Act and the law of criminal evidence: ten years on. Sydney Law Rev 33(3):333–357

Dreyer S, Schmees J (2019) Künstliche Intelligenz als Richter? Wo keine Trainingsdaten, da kein Richter – Hindernisse, Risiken und Chancen der Automatisierung gerichtlicher Entscheidungen. Comput Und Recht 35(11):758–764. https://doi.org/10.9785/cr-2019-351120

Duan Y, Edwards JS, Dwivedi YK (2019) Artificial intelligence for decision making in the era of Big Data—evolution, challenges and research agenda. Int J Inf Manage 48:63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Duff A (2013) Who must presume whom to be innocent of what? NJLP 42(3):170–192. https://doi.org/10.5553/NJLP/221307132013042003002

Ebersbach M (2020) Big Data, Algorithmen und Bewährungsentscheidungen. In Momsen C, Schwarze M (eds) Strafrecht im Zeitalter von Digitalisierung und Datafizierung, pp.26–37. https://kripoz.de/wp-content/uploads/2020/06/Ebersbach-Big-Data-Algorithmen-und-Bew%C3%A4hrungsentscheidungen.pdf. Accessed 24 Jan 2022

Eckes C (2021) EU global human rights sanctions regime: Is the genie out of the bottle? J Contemp Eur Stud. https://doi.org/10.1080/14782804.2021.1965556

EDRi (2020) "E-evidence": Mixed results in the European Parliament. https://edri.org/our-work/e-evidence-mixed-results/. Accessed 24 Jan 2022

Egbert S, Leese M (2021) Criminal futures. Predictive policing and everyday police work. Routledge, London

Eskens S (2021) Large scale surveillance to protect national security: Under EU control? https://papers.ssrn.com/sol3/papers.cfm?abstractid=3902158. Accessed 24 Jan 2022

European Commission (2014) Towards a thriving data-driven economy. COM (2014) 442 final. https://ec.europa.eu/transparency/regdoc/rep/1/2014/EN/1-2014-442-EN-F1-1.Pdf. Accessed 24 Jan 2022

European Commission (2018a) Artificial Intelligence for Europe. COM (2018a) 237 final. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018aDC0237&from=EN. Accessed 24 Jan 2022

European Commission (2018b) Proposal for a Regulation of the European Parliament and of the Council on European Production and Preservation Orders for electronic evidence in criminal matters. COM/2018a/225 final - 2018b/0108 (COD). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018b%3A225%3AFIN. Accessed 24 Jan 2022

European Commission (2018c) Proposal for a Directive of the European Parliament and of the Council laying down harmonised rules on the appointment of legal representatives for the purpose of gathering evidence in criminal proceedings. COM/2018b/226 final—2018c/0107 (COD). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018c%3A226%3AFIN. Accessed 24 Jan 2022

European Commission (2020a) White Paper on Artificial Intelligence—A European approach to excellence and trust. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020en.pdf. Accessed 24 Jan 2022

European Commission (2020b) A Counter-Terrorism Agenda for the EU: Anticipate, Prevent, Protect, Respond. COM (2020b) 795 final. https://ec.europa.eu/home-affairs/sites/homeaffairs/files/pdf/09122020communicationcommissioneuropeanparliamentthecouncileuagendacounterterrorismpo-2020b-9031com-2020b795en.pdf. Accessed 24 Jan 2022

European Commission (2020c) Communication on the EU Security Union Strategy. COM (2020c) 605 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020cDC0605. Accessed 24 Jan 2022

European Commission (2021a) Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. COM (2021a) 206 final. 2021a/0106 (COD).

https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence. Accessed 24 Jan 2022

European Commission (2021b) Annexes to the Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. COM (2021b) 206 final. Annexes 1 to 9. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence. Accessed 24 Jan 2022

European Commission (2021c) Fostering a European Approach to Artificial Intelligence. COM (2021c) 205 final. https://ec.europa.eu/transparency/regdoc/rep/1/2021c/EN/COM-2021c-205-F1-EN-MAIN-PART-1.PDF. Accessed 24 Jan 2022

European Commission (2021d) Ethics by design and ethics of use approaches for AI. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021d-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf. Accessed 24 Jan 2022

European Commission for the Efficiency of Justice (CEPEJ) (2018) European ethical charter on the use of artificial intelligence in judicial systems and their environment. https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment. Accessed 24 Jan 2022

European Council (2010) The Stockholm Programme—An open and secure Europe serving and protecting citizens. 2010/C115/01. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AC%3A2010%3A115%3ATOC. Accessed 24 Jan 2022

European Court of Human Rights (ECtHR) (2020) Guide on Article 6 of the European Convention on Human Rights. Right to a fair trial [criminal limb]. https://www.echr.coe.int/documents/Guide_Art6criminalENG.pdf. Accessed 24 Jan 2022

European Data Protection Board (2020) Recommendations 02/2020 on the European Essential Guarantees for surveillance measures. https://edpb.europa.eu/our-work-tools/our-documents/recommendations/recommendations-022020-european-essential-guaranteesen. Accessed 24 Jan 2022

European Parliament (2020a) Report on the proposal for a Regulation of the European Parliament and of the Council on European Production and Preservation Orders for electronic evidence in criminal matters (COM(2018)0225 – C8-0155/2018–2018/0108 (COD)). https://www.europarl.europa.eu/doceo/document/A-9-2020a-0256_EN.pdf. Accessed 24 Jan 2022

European Parliament (2020b) Resolution of 20 October 2020b with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020b/2012(INL)). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020bIP0275&from=EN. Accessed 24 Jan 2022

European Parliament (2021) Resolution of 6 October 2021 on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters (2020/2021(INI)). https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.pdf. Accessed 24 Jan 2022

European Parliament, Committee on Civil Liberties, Justice and Home Affairs (LIBE Committee) (2019) Draft Report on the proposal for a regulation of the European Parliament and of the Council on European Production and Preservation Orders for electronic evidence in criminal matters (COM (2018) 0225 – C8–0155/2018 – 2018/0108(COD)). Rapporteur: Birgit Sippel. https://birgitsippel.de/wp-content/uploads/2020/10/Draft-ReportRegulation-on-European-Production-and-Preservation-Orders-for-electronic-evidence-in-criminal-matters.pdf. Accessed 24 Jan 2022

European Union Agency for Fundamental Rights (FRA) (2015) Surveillance by intelligence services: Fundamental right safeguards and remedies in the EU. Volume I: Mapping Member States' legal frameworks. https://fra.europa.eu/en/publication/2015/surveillance-intelligence-services-volume-i-member-states-legal-frameworks. Accessed 24 Jan 2022

European Union Agency for Fundamental Rights (FRA) (2017) Surveillance by intelligence services: Fundamental right safeguards and remedies in the EU. Volume II: Field perspectives and legal update. https://fra.europa.eu/en/publication/2017/surveillance-intelligence-services-fundamental-rights-safeguards-and-remedies-eu. Accessed 24 Jan 2022

European Union Agency for Fundamental Rights (FRA) (2020) Getting the future right. Artificial intelligence and fundamental rights. https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights. Accessed 24 Jan 2022

European Union Agency for Fundamental Rights (FRA) (2021) Presumption of innocence and related rights. Professional Perspectives. https://fra.europa.eu/en/publication/2021/presumption-of-innocence. Accessed 24 Jan 2022

Europol (2020a) Europol and the European Commission inaugurate new decryption platform to tackle the challenge of encrypted material for law enforcement investigations. https://www.europol.europa.eu/media-press/newsroom/news/europol-and-european-commission-inaugurate-new-decryption-platform-to-tackle-challenge-of-encrypted-material-for-law-enforcement. Accessed 24 Jan 2022

Europol (2020b) Europol and Eurojust sign new contribution agreement expanding cooperation on the Sirius project. https://www.europol.europa.eu/newsroom/news/europol-and-eurojust-sign-new-contribution-agreement-expanding-cooperation-sirius-project. Accessed 24 Jan 2022

Europol (2021) SIRIUS EU Digital Evidence Situation Report. 3rd Annual Report 2021. https://www.europol.europa.eu/publications-events/publications/sirius-eu-digital-evidence-situation-report-3rd-annual-report-2021. Accessed 24 Jan 2022

Fair Trials.org (2021) Regulating artificial intelligence for use in criminal justice systems in the EU. Policy Paper. https://www.fairtrials.org/app/uploads/2022/01/Regulating-Artificial-Intelligence-for-Use-in-Criminal-Justice-Systems-Fair-Trials.pdf. Accessed 24 Jan 2022

Fairfield JA, Luna E (2013) Digital innocence. Cornell L Rev 99:981

Ferguson AG (2015) Big data and predictive reasonable suspicion. Univ Pa L Rev 163(2):327–410

Fuster GG (2020) Artificial intelligence and law enforcement. Impact on fundamental rights. Study requested by the LIBE Committee. https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOLSTU(2020)656295. Accessed 24 Jan 2022

Galetta A (2013) The changing nature of the presumption of innocence in today's surveillance societies: rewrite human rights or regulate the use of surveillance technologies? EJLT 4(2). http://ejlt.org/index.php/ejlt/article/view/221. Accessed 24 Jan 2022

Giannakoula A, Lima D, Kaiafa Gbandi M (2020) Combating crime in the digital age: a critical review of EU information systems in the area of freedom, security and justice in the post-interoperability era. Challenges for criminal law and personal data protection. Brill, Leiden.

Gless S (2018) Predictive policing In defence of 'true positives.' In: Bayamlioglu E, Baraliuc I, Janssens L, Hildebrandt M (eds) Being profiled Cogitas ergo sum: 10 years of profiling the European citizen. Amsterdam University Press, Amsterdam, pp. 62–65.

Gless S (2020) AI in the courtroom: a comparative analysis of machine evidence in criminal trials. Georgetown J Int Law 51(2):195–253

Gonçalves ME (2017) The EU data protection reform and the challenges of big data: remaining uncertainties and ways forward. Inform Comm Tech Law 26(2):90–115. https://doi.org/10.1080/13600834.2017.1295838

Green B (2018) "Fair" risk assessments: a precarious approach for criminal justice reform. Presented at the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2018), Stockholm, Sweden. https://scholar.harvard.edu/files/bgreen/files/18-fatml.pdf. Accessed 24 Jan 2022

Greenstein S (2022) Preserving the rule of law in the era of artificial intelligence (AI). Artif Intell Law. 30: 291–323. https://doi.org/10.1007/s10506-021-09294-4

Hadjimatheou K (2017) Surveillance technologies, wrongful criminalisation, and the presumption of innocence. Philos Technol 30:39–54. https://doi.org/10.1007/s13347-016-0218-2

Hamran L (2020) The admissibility of e-evidence in criminal proceedings within the EU. ELI Newsletter November-December 2020: 2–3. https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Newsletter/2020/Flip_Nov_2020/PDF.pdf. Accessed 24 Jan 2022

High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019a) A definition of AI: Main capabilities and disciplines. Definition developed for the purpose of AI HLEG's deliverables. https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines. Accessed 24 Jan 2022

High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019b) Ethics guidelines for trustworthy AI. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top. Accessed 24 Jan 2022

High-Level Expert Group on Artificial Intelligence (AI HLEG) (2020) The assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment. https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence. Accessed 24 Jan 2022

Hildebrandt M (2014) Criminal law and technology in a data-driven society. In: Dubber M, Hörnle T (eds) The Oxford Handbook of Criminal Law. Oxford University Press, Oxford, pp 175–198

Hildebrandt M (2018) Algorithmic regulation and the rule of law. Phil Trans R Soc A 376:20170355. https://doi.org/10.1098/rsta.2017.0355

Hoffmann-Riem W (2018) Rechtliche Rahmenbedingungen für und regulative Herausforderungen durch Big Data. In: Hoffmann-Riem M (ed) Big Data—Regulative Herausforderungen. Nomos Verlagsgesellschaft, Baden-Baden, pp 11–78

Joh E (2016) The new surveillance discretion: automated suspicion, big data, and policing. HarvLRev 10:15–42

Kaiafa-Gbandi M (2019) Information exchange for the purpose of crime control: the EU paradigm for controlling terrorism—challenges of an 'open' system for collecting and exchanging personal data. EuCLR 2:141–174. https://doi.org/10.5771/2193-5505-2019-2-141

Kehl D, Guo P, Kessler S (2017) 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. https://dash.harvard.edu/handle/1/33746041. Accessed 24 Jan 2022

Kleijssen J, Perri P (2016) Cybercrime, evidence and territoriality: issues and options. In: Kuijer M, Werner W (eds) Netherlands yearbook of international law. T.M.C. Asser Press, The Hague, pp 147–173

Koss KK (2015) Leveraging predictive policing algorithms to restore Fourth Amendment protections in high-crime areas in a post-Wardlow world. Chicago-Kent Law Rev 90(1):301–334

Lachmayer K (2009) European policy cooperation and its limits: From intelligence-led to coercive measures? In: Barnard C, Odudu O (eds) The outer limits of European Union law. Hart Publishing, Oxford, pp 89–118

Mantelero A, Vaciago G (2013) The 'dark side' of big data: Private and public interaction in social surveillance. Cri 14(6):161–169. https://doi.org/10.9785/ovs-cri-2013-161

Marquenie T (2017) The police and criminal justice authorities directive: data protection standards and impact on the legal framework. CLSR 33(3):324–340. https://doi.org/10.1016/j.clsr.2017.03.009

Marx G (2005) Seeing hazily, but not darkly, through the lens: some recent empirical studies of surveillance technologies. LSI 30(2):339–400. https://doi.org/10.1111/j.1747-4469

Meijer A, Wessels M (2019) Predictive policing: review of benefits and drawbacks. Int J Public Adm 42(12):1031–1039. https://doi.org/10.1080/01900692.2019.1575664

Melzer J (2020) Auswirkungen von künstlicher Intelligenz auf Völkerrecht, insbesondere die Gewährleistung der Garantien des Art. 6 EMRK. Zeitschrift Zum Innovations- Und Technikrecht 3:145–150

Meuwese A (2020) Regulating algorithmic decision-making one case at the time: a note on the Dutch 'SyRI' judgment. Eur Rev Digit Adm Law 1(1–2):209–211. https://doi.org/10.4399/978882553896019

Milaj J, Mifsud Bonnici JP (2014) Unwitting subjects of surveillance and the presumption of innocence. CLSR 30(4):419–428. https://doi.org/10.1016/j.clsr.2014.05.009

Mitsilegas V (2018) The privatisation of mutual trust in Europe's area of criminal justice: the case of e-evidence. MJ 25(3):263–265

Mittelstadt BD, Allo P, Taddeo D, Wachter S, Floridi L (2016) The ethics of algorithms: mapping the debate. Big Data Soc. https://doi.org/10.1177/2053951716679679

Murphy E (2007) The new forensics: Criminal justice, false certainty, and the second generation of scientific evidence. Calif Law Rev 95(3):721–797. https://doi.org/10.2307/20439109

Organisation for Economic Co-operation and Development (OECD) (2019) Recommendation of the Council on Artificial Intelligence. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Accessed 24 Jan 2022

Oswald M, Grace J, Urwin S, Barnes GC (2018) Algorithmic risk assessment policing models: lessons from the Durham HART model and 'experimental' proportionality. Inform Commun Technol Law 27(2):223–250. https://doi.org/10.1080/13600834.2018.1458455

Oswald M (2020) Technologies in the twilight zone: early lie detectors, machine learning and reformist legal realism. Int Rev Law Comput Technol 34(2):214–231. https://doi.org/10.1080/13600869.2020.1733758

Pagallo U (2017) The legal challenges of big data: putting secondary rules first in the field of EU data protection. EDPL 36(1):36–45. https://doi.org/10.21552/edpl/2017/1/7

Palmiotto F (2020) Regulating algorithmic opacity in criminal proceedings: an opportunity for the EU legislator? Maastricht Law 1:1–32

Palmiotto F (2021) The black box on trial: The impact of algorithmic opacity on fair trial rights in criminal proceedings. In: Ebers M, Cantero Gamito M (eds) Algorithmic governance and governance of algorithms. Springer, Cham, pp 49–70

Papadimitrakis G (2019) Big data and algorithmic risk studies. New challenges in the area of penology [in Greek]. Poiniki Dikaiosyni 10:1045–1054

Perry WL, Mclnnis B, Price CC, Smith SC, Hollywood JS (2013) Predictive policing. The role of crime forecasting in law enforcement operations. RAND Corporation. https://www.rand.org/content/dam/rand/pubs/researchreports/RR200/RR233/RANDRR233.pdf. Accessed 24 Jan 2022

Quattrocolo S (2020) Artificial intelligence, computational modelling and criminal proceedings. A framework for a European legal discussion. Springer, Cham

Reidenberg J (2014) The data surveillance state in Europe and the United States. Wake Forest L Rev 49:583–608

Roth A (2017) Machine testimony. Yale Law J 126:1972–2053

Ryan M, Brey P, Machnish K, Hatzakis T, King O, Maas J, Haasjes R, Fernandez A, Martorana S, Oluoch I, Eren S, Van Der Puil R (2019) Report on ethical tensions and social impacts. WP 1. D1.4 SHERPA. https://doi.org/10.21253/DMU.8397134

Sachoulidou A (2021a) The key elements of the LIBE Committee's compromise proposal on e-evidence: a critical overview through a fundamental rights lens. Global Affairs 7(5):777–793. https://doi.org/10.1080/23340460.2021.1999173

Sachoulidou A (2021b) OK Google: is (s)he guilty? J Contemp Eur Stud. https://doi.org/10.1080/14782804.2021.1987863

Sajfert J, Quintel T (2019) Data Protection Directive (EU) 2016/680 for police and criminal justice authorities (December 1, 2017). In: Cole M, Boehm F (eds) GDPR Commentary. Edward Elgar Publishing, Cheltenham, pp 1–22

Sauer G (2017) A murder case tests Alexa's devotion to your privacy. https://www.wired.com/2017/02/murder-case-tests-alexas-devotion-privacy/. Accessed 24 Jan 2022

Sellier E, Weyembergh A (2018) Criminal procedural laws across the European Union – A comparative analysis selected main differences and the impact they have over the development of EU legislation. Study requested by the LIBE Committee. https://op.europa.eu/en/publication-detail/-/publication/70e80c5d-b64e-11e8-99ee-01aa75ed71a1. Accessed 24 Jan 2022

Shun Z, Huo Y (2021) The spectrum of big data analytics. J Comput Inform Syst 61(2):154–162. https://doi.org/10.1080/08874417.2019.1571456

Singelnstein T (2018) Predictive Policing: Algorithmenbasierte Straftatprognosen zur vorausschauenden Kriminalintervention. Neue Zeitschrift für Strafrecht: 1–9. https://scholar.google.com/scholar_lookup?title=Predictive%20policing%3A%20algorithmenbasierte%20Straftatprognosen%20zur%20vorausschauenden%20Kriminalintervention&journal=Neue%20Zeitschrift%20f%C3%BCr%20Strafrecht%20%28NStZ%9&publication_year=2018&author=Singelnstein%2CT#d=gs_cit&t=1676456410000&u=%2Fscholar%3Fq%3Dinfo%3Az15xpVDSMowJ%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den

Smuha N (2018) Towards the EU harmonization of access to cross-border e-evidence: challenges for fundamental rights & consistency. EuCLR 8(1):83–115. https://doi.org/10.5771/2193-5505-2018-1-83

Sommerer L (2018) The presumption of innocence's Janus head in data-driven government. In: Bayamlioglu E, Baraliuc I, Janssens L, Hildebrandt M (eds) Being profiled Cogitas ergo sum: 10 years of profiling the European citizen. Amsterdam University Press, Amsterdam, pp 58–61

Starr S (2014) Evidence-based sentencing and the scientific rationalization of discrimination. Stanford Law Rev 66(4):803–872

Strikwerda L (2021) Predictive policing: The risks associated with risk assessment. Police J Theory Pract Princ 94(3):422–436. https://doi.org/10.1177/0032258X20947749

The Law Society of England and Wales (2019) Algorithm use in the criminal justice system report. https://www.lawsociety.org.uk/en/topics/research/algorithm-use-in-the-criminal-justice-system-report. Accessed 24 Jan 2022

Tinoco-Pastrana A (2020) The Proposal on Electronic Evidence in the European Union. Eucrim 1:46–50

Tosza S (2020) All evidence is equal, but electronic evidence is more equal than any other: the relationship between the European Investigation Order and the European Production Order. New J Eur Crim Law 11(2):161–183. https://doi.org/10.1177/2032284420919802

Tracol X (2021) The two judgments of the European Court of Justice in the four cases of Privacy International, La Quadrature du Net and Others, French Data Network and Others and Ordre des Barreaux francophones et germanophone and Others: the Grand Chamber is trying hard to square the circle of data retention. CLSR 41:1–13. https://doi.org/10.1016/j.clsr.2021.105540

Vasquez Maymir S (2020) Anchoring the need to revise cross-border access to e-evidence. Internet Policy Rev 9(3):1–24. https://doi.org/10.14763/2020.3.1495

Villamarín López ML (2017) The presumption of innocence in Directive 2016/343/EU of 9 March 2016. ERA Forum 18:335–353. https://doi.org/10.1007/s12027-017-0480-5

Wahl T (2021) New E-evidence legislation: Trilogue started – Criticism on EP stance. Eucrim. https://eucrim.eu/news/new-e-evidence-legislation-trilogue-started-critics-on-ep-stance/. Accessed 24 Jan 2022
Wisser L (2019) Pandora's algorithmic black box: the challenges of using algorithmic risk assessments in sentencing. ACLR 56:1811–1832
Yavar B (2018) The artificial intelligence black box and the failure of intent and causation. Harvard J Law Technol 31(2):890–938
Zarsky T (2017) Incompatible: the GDPR in the age of big sata. Seton Hall L Rev 47:995–1020
Završnik A (2020) Criminal justice, artificial intelligence systems, and human rights. ERA Forum 20(4):567–583. https://doi.org/10.1007/s12027-020-00602-0
Završnik A (2021) Algorithmic justice: algorithms and big data in criminal justice settings. Eur J Criminol 18(5):623–642. https://doi.org/10.1177/1477370819876762
Zedner L (2007) Pre-crime and post-criminology. Theor Criminol 11(2):261–281. https://doi.org/10.1177/1362480607075851

Cases cited

Advocate General Sánchez-Bordona C (2020a) Opinion in Case C-623/17. https://curia.europa.eu/juris/document/document.jsf?text=&docid=222262&pageIndex=0&doclang=en&mode=lst&dir=&occ=first&part=1&cid=5343495. Accessed 24 Jan 2022
Advocate General Sánchez-Bordona C (2020b) Opinion in Joined Cases 511/18 and C-512/18 and Case C-520/18. https://curia.europa.eu/juris/document/document.jsf?text=&docid=222263&pageIndex=0&doclang=en&mode=lst&dir=&occ=first&part=1&cid=5347732. Accessed 24 Jan 2022
Bundesgerichtshof, 1 StR 1998, 156/98. https://dejure.org/dienste/vernetzung/rechtsprechung?Gericht=BGH&Datum=17.12.1998&Aktenzeichen=1%20StR%20156/98. Accessed 24 Jan 2022
Court of Justice of the European Union (CJEU) Digital Rights Ireland and Seitlinger and others, Joined Cases C-293/12 and C-594/12, ECLI:EU:C:2014:238, 8 April 2014. https://curia.europa.eu/juris/documents.jsf?language=EN&critereEcli=ECLI:EU:C:2014:238. Accessed 24 Jan 2022
Court of Justice of the European Union (CJEU) Privacy International v. Secretary for Foreign and Commonwealth Affairs, Case C-623/17, ECLI:EU:C:2020:790, 6 October 2020. https://curia.europa.eu/juris/document/document.jsf?text=&docid=232083&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=183052. Accessed 24 Jan 2022
Court of Justice of the European Union (CJEU) Quadrature du Net and Others v. Premier ministre and Others, Joined Cases 511/18, C-512/18 and 520/18, ECLI:EU:C:2020:791, 6 October 2020. https://curia.europa.eu/juris/document/document.jsf?text=&docid=232084&pageIndex=0&doclang=en&mode=req&dir=&occ=first&part=1&cid=183346. Accessed 24 Jan 2022
Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993). https://supreme.justia.com/cases/federal/us/509/579/. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Salabiaku v. France (Application No. 10519/83), 7 October 1988. https://hudoc.echr.coe.int/eng?i=001-57570. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Barberà, Messegue and Jabardo v. Spain (Application No. 10590/83), 6 December 1988. https://hudoc.echr.coe.int/eng?i=001-57429. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Brandstetter v. Austria (Application Nos. 11170/84, 12876/87, 13468/87), 28 August 1991. https://www.legal-tools.org/doc/deb99c/pdf/. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Doorson v. The Netherlands (Application No. 20524/92), 26 March 1996. https://hudoc.echr.coe.int/eng?i=001-57972. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Van Mechelen and Others v. The Netherlands (Application No. 55/1996/674/861–864), 23 April 1997. https://www.refworld.org/cases,ECHR,3ae6b6778.html. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Rowe and Davis v. The United Kingdom (Application No. 28901/95), 16 February 2000. https://hudoc.echr.coe.int/eng?i=001-58496. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Coëme and Others v. Belgium (Applications Nos. 32492/96, 32547/96, 32548/96, 33209/96, 33210/96), 22 June 2000. https://www.legislationline.org/download/id/4076/file/ECHR_Coeme_Belgium_2000.pdf. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Melich and Beck v. The Czech Republic (Application No. 35450/04) 24 July 2008. https://hudoc.echr.coe.int/eng?i=001-161180. Accessed 24 Jan 2022

European Court of Human Rights (ECtHR) Adjarić v. Croatia (Application No. 20883/09) 13 December 2011. https://hudoc.echr.coe.int/eng?i=001-107989. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Al-Khawaja and Tahery v. UK (Application Nos. 26766/05 and 22228/06) 15 December 2011. https://hudoc.echr.coe.int/eng?i=001-108072. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Topić v. Croatia (Application no. 51355/10) 10 October 2013. https://hudoc.echr.coe.int/eng?i=001-126638. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Schatschaschwili v. Germany (Application No. 9154/10) 15 December 2015. https://hudoc.echr.coe.int/eng?i=001-159566. Accessed 24 January 2022.
European Court of Human Rights (ECtHR) Seton v. UK (Application No. 55287/10) 12 September 2016. https://hudoc.echr.coe.int/eng?i=001-161738. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Dimović v. Serbia (Application No. 24463/11) 28 September 2016. https://hudoc.echr.coe.int/eng?i=001-164314. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) T.K. v. Lithuania (Application No. 14000/12) 3 December 2018. https://hudoc.echr.coe.int/eng?i=001-183542. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Zollman v. UK (Application No. 62902/00), 27 November 2003. https://www.stradalex.com/en/slsrcpubljurint/document/echr62902-00. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) S. and Marper v. UK (Applications Nos. 30562/04 and 30566/04), 4 December 2008. https://rm.coe.int/168067d216. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Nemtsov v. Russia (Application No. 1774/11), 31 July 2014. https://hudoc.echr.coe.int/eng?i=001-145784. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Frumkin v. Russia (Application No. 74568/12), 5 January 2016. https://hudoc.echr.coe.int/eng?i=001-159762. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Tsalkitzis v. Greece (No. 2) (Application No. 72624/10), 19 October 2017. https://hudoc.echr.coe.int/eng?i=001-177691. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Paci v. Belgium (Application No. 45597/09), 17 April 2018. https://hudoc.echr.coe.int/eng?i=001-182232. Accessed 24 Jan 2022
European Court of Human Rights (ECtHR) Gaughran v. UK (Application No. 45245/15), 13 February 2020. https://hudoc.echr.coe.int/eng?i=001-200817. Accessed 24 Jan 2022
Frye v. United States, 293 F. 1013 (D.C. Cir 1923). https://casetext.com/case/frye-v-united-states-7. Accessed 24 Jan 2022
Loomis v. Wisconsin, 881 N.W.2d. https://www.courts.ca.gov/documents/BTB24-2L-3.pdf. Accessed 24 Jan 2022
Oberlandesgericht Bamberg, 13 June 2018, 3 Ss Owi 626/18. https://dejure.org/dienste/vernetzung/rechtsprechung?Gericht=OLG%20Bamberg&Datum=13.06.2018&Aktenzeichen=3%20Ss%20OWi%20626/18. Accessed 24 Jan 2022
The Hague District Court, Nederlands Juristen Comité voor de Mensenrechten and Others v. The State of the Netherlands, C/09/550982/HAZA18–388, ECLI:NL:RBDHA:2020:865. https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878. Accessed 24 Jan 2022
UK Court of Appeal, R [Bridges] v. CC South Wales, EWCA Civ 1058, 11 August 2020. https://www.judiciary.uk/wp-content/uploads/2020/08/R-Bridges-v-CC-South-Wales-ors-Judgment.pdf. Accessed 24 Jan 2022

Statutes cited

Charter of Fundamental Rights of the European Union (CFR). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT. Accessed 24 Jan 2022
Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32002L0058. Accessed 24 Jan 2022
Directive 2010/64/EU of the European Parliament and of the Council of 20 October 2010 on the right to interpretation and translation in criminal proceedings. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32010L0064. Accessed 24 Jan 2022

Directive 2012/13/EU of the European Parliament and of the Council of 22 May 2012 on the right to information in criminal proceedings. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32012L0013. Accessed 24 Jan 2022

Directive 2013/48/EU of the European Parliament and of the Council of 22 October 2013 on the right of access to a lawyer in criminal proceedings and in European arrest warrant proceedings, and on the right to have a third party informed upon deprivation of liberty and to communicate with third persons and with consular authorities while deprived of liberty. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32013L0048. Accessed 24 Jan 2022

Directive 2014/41/EU of the European Parliament and of the Council of 3 April 2014 regarding the European Investigation Order in criminal matters. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014L0041. Accessed 24 Jan 2022

Directive (EU) 2016/343 of the European Parliament and of the Council on strengthening of certain aspects of the presumption of innocence and of the right to be present at the trial in criminal proceedings. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L0343. Accessed 24 Jan 2022

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 criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016L0680. Accessed 24 Jan 2022

Directive (EU) 2016/800 of the European Parliament and of the Council of 11 May 2016 on procedural safeguards for children who are suspects or accused persons in criminal proceedings. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L0800. Accessed 24 Jan 2022

Directive (EU) 2016/1919 of the European Parliament and of the Council of 26 October 2016 on legal aid for suspects and accused persons in criminal proceedings and for requested persons in European arrest warrant proceedings. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L1919. Accessed 24 Jan 2022

European Convention on Human Rights (ECHR). https://www.echr.coe.int/documents/convention_eng.pdf. Accessed 24 Jan 2022

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 (General Data Protection Regulation). https://eur-lex.europa.eu/eli/reg/2016/679/oj. Accessed 24 Jan 2022

Regulation (EU) 2019/2144 of the European Parliament and of the Council of 27 November 2019 on type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019R2144&from=EL. Accessed 24 Jan 2022

Treaty on European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012M%2FTXT. Accessed 24 Jan 2022

US Federal Rules of Evidence. 2022 Edition. https://www.rulesofevidence.org/table-of-contents/. Accessed 24 Jan 2022