Maastricht Law

Faculty of Law
Working Paper series

2020/01

Regulating Algorithmic Opacity in Criminal Proceedings: an opportunity for the EU Legislator?[*]

by Francesca Palmiotto

[*] This collection of working papers stems from the first Jean Monnet NOVA-EU Workshop on "Digitalisation, Ethics and EU Fundamental Rights" which was organised at Maastricht University on 9-10 January 2020. The project is co-founded by the Erasmus+ Programme of the European Union.

Working Paper submitted by
Francesca Palmiotto
Francesca.Palmiotto@eui.eu
PhD Researcher at European University Institute (EUI) – Florence

Regulating Algorithmic Opacity in Criminal Proceedings

An opportunity for the EU Legislator?

Francesca Palmiotto[*]

Abstract Algorithms are increasingly used in criminal proceedings worldwide for evidentiary purposes and for supporting decision-making. In a worrying trend, these tools are still concealed in secrecy and opacity preventing us from understanding how their specific output has been generated. Many cases demonstrated that software can be biased and that they should not be presumed reliable. The main concern relates to their opacity, as the defence cannot challenge the use of software in criminal proceedings without further insight into their inner workings. This requires an appropriate legal framework creating procedural rules to balance transparency with other legitimate purposes of the criminal process. The paper questions whether the EU legal response on data protection is up to the challenge and which role, if any, the EU legislator should play in regulating algorithmic opacity in criminal proceedings.

Keywords criminal justice; algorithmic transparency; algorithmic opacity; right to confrontation; electronic evidence

Table of content

I. Introduction

"I call to the stand: Alexa". In Florida, the smart device might be a key 'witness' in a murder case.[1] The police believe that this device might have heard and recorded relevant information for the criminal investigation in the case.[2] SyRi is a software that was used

[*] Francesca Palmiotto, PhD Researcher
European University Institute, Florence, IT
Francesca.Palmiotto@EUI.eu

[1] Alexa is a virtual assistant AI technology developed by Amazon. It functions as a companion speaker that can answer to questions, research on Internet, command other devices or stream music when you call its name.

[2] Gerald Sauer, 'A Murder Case Tests Alexa's Devotion to Your Privacy' *Wired* (28 February 2017) <https://www.wired.com/2017/02/murder-case-tests-alexas-devotion-privacy/> accessed 5 March 2020.

to identify individuals at risk of committing frauds in the Netherlands.[3] Citizens that received a high-risk label could be further investigated by the police. Uber's drivers are managed by an algorithmic faceless boss;[4] Amazon uses Artificial Intelligence (AI) engines for recruiting.[5] Nowadays, algorithms are used for firing, hiring, profiling, targeting, ranking, cheating and even taking decisions with significant effects on individuals' lives.

The pervasive use of algorithms has caused many concerns, particularly as regard to their reliability. To provide few examples, research has shown that Amazon's AI recruiting tool was biased against women;[6] that facial recognition applications may be consistently less accurate when identifying black people;[7] that SyRi could have discriminatory effects against people with low income or with minority background[8] (recently the use of this system was declared to be in violation with human rights by the Court of the Hague).[9] In a worrying trend, however, these tools are still concealed in secrecy and opacity preventing individuals from understanding how their specific output has been generated. In the light of these concerns, the literature advocates for transparency and explanation against the adverse effects of the "*Black Box Society*".[10]

In the specific context of criminal justice, when algorithms are used for evidentiary purposes, it is essential to ensure the rights of the defence enshrined in Article 6 of the

[3] van Schendel S, 'The Challenges of Risk Profiling Used by Law Enforcement: Examining the Cases of COMPAS and SyRI' in Leonie Reins (ed), *Regulating New Technologies in Uncertain Times* (TMC Asser Press 2019); Anton Ekker, 'Landslide Victory in SyRI case: Dutch Court bans risk profiling' (*Ekker Advocatuur*, 2 February 2020) <https://ekker.legal/2020/02/02/syri/> accessed 6 February 2020.

[4] Alex Rosenblat, *Uberland: How Algorithms Are Rewriting the Rules of Work* (University of California Press 2018).

[5] Julien Lauret, 'Amazon's Sexist AI Recruiting Tool: How Did It Go so Wrong?' *Medium* (16 August 2019) <https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e> accessed 29 February 2020.

[6] ibid.

[7] Natasha Singer and Cade Metz, 'Many Facial-Recognition Systems Are Biased, Says U.S. Study' *The New York Times* (19 December 2019) <https://www.nytimes.com/2019/12/19/technology/facial-recognition-bias.html> accessed 5 March 2020.

[8] Brief by the United Nations Special Rapporteur on extreme poverty and human rights as *Amicus Curiae* in the case of *NJCM v De Staat der Nederlanden (SyRi)* before the District Court of The Hague, case number C/09/550982/HA ZA 18/388.

[9] Judgment of the 5 February 2020, *NJCM c.s./De Staat der Nederlanden (Syri),* C/09/550982/HA ZA 18/38, NL:RBDHA:2020:865.

[10] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015).

European Convention of Human Rights (ECHR). In particular, according to the case-law of the European Court of Human Rights (ECtHR) on Article 6§3(d), the right to confrontation not only requires that defendants should be in a position to challenge the probity and credibility of evidence against them, but that they should also be able to test their truthfulness and reliability.[11] Nevertheless, in order to do so, it is necessary to comprehend how the software works and how it is used. This requires an appropriate legal framework creating procedural rules to balance transparency with other legitimate purposes of the criminal process. This paper will firstly show the challenges brought by the pervasive use of algorithms in criminal proceedings and the threats that their opacity poses to the right to confrontation. Secondly, it will focus on the EU Data Protection Regulation and will investigate if this legal response can be considered adequate. Lastly, through the example of the Proposals on the European Production and Preservation Order for electronic evidence in criminal matters, this paper will show that the EU legislator can play a role in this matter by enhancing a transparent framework for the acquisition of such evidence.

II. Algorithmic Opacity and the Rights of the Defence

1. The use of algorithms in criminal proceedings

Nowadays, almost every daily action leaves a trail of digital information. As an example, take a single day in a life.

Tea is resident in Maastricht; on Friday night before going to sleep, she asks Alexa to set an alarm for the following day at 9 AM. She is going to visit her friend Linda and she does not want to be late. In the morning, while eating breakfast, she checks the news on her smartphone. Meanwhile, Linda sends her position through WhatsApp. Tea then asks Siri on her iPhone what is the best way to reach that location by car. On her way, she stops at a gas station to buy some food and other items. She then realizes that her wallet is empty, but fortunately the store accepts payment with Apple Pay. She finally arrives at destination. Linda and Tea spend together a very nice Saturday, and round 7 PM Tea decides to leave. In the car, she sets the heating system on via app, so that the house is warm when she returns, and asks Siri what is the best way back to Maastricht without payment of tolls. On the way back, her attention is captured by the sight of a large landfill. A few days later, Linda's body is found in that site. After having seized the phone of the victim, Tea is investigated for the murder of her friend Linda.

[11] *Al-Khawaja and Tahery v UK* ECHR 2011–VI 191.

In that single day, Tea has left sufficient digital traces to allow a detailed reconstruction of her activities and locations. Belgium law enforcement would surely be willing to use this digital information to prove or disprove her involvement in the crime.

For these purposes, different algorithms may be used for acquiring digital information, such as, in this case, forensic software for collecting data stored in her phone, the digital images of the cameras in the gas station, the information in WhatsApp or Apple Pay. Due to the gravity of the crime, judicial authorities might even use a malware for taping her phone or for acquiring encrypted data. Additionally, algorithms may serve as forensic tools for automated analysis of the DNA found at the crime scene and a facial recognition software might be useful for the identification of the person captured by the security cameras in the landfill.

As this example shows, due to the pervasiveness of technology surrounding individuals, digital information has become a cornerstone of criminal investigations for its rich probative value. Nowadays, proceedings not relying on algorithms represent an exception in European countries.

In the broadest sense, an algorithm is a procedure for solving a problem, akin to cooking recipes, instructions on how to play a game or the assembly instructions of IKEA. As a longstanding concept, algorithms can be defined as "any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output. An algorithm is thus a sequence of computational steps that transform the input into the output. We can also view an algorithm as a tool for solving a well-specified computational problem".[12] Based on the algorithm, an input can be transformed into different sets of outputs. When this set of instructions is implemented as a computer program, the machine will execute it in a speed up, automated and efficient way.

Algorithms can be of two types, depending on their purpose: a) decision-support algorithms and b) algorithmic evidence. The first type is used to inform human decision-making (*e.g.* crime data mining, risk-assessment software for bail hearings, sentencing, probation and parole). The most famous example is COMPAS, a software used in the US

[12] Thomas H. Cormen and others, *Introduction to Algorithms* (3rd edn, The MIT Press 2009) 5.

for evaluating the risk of recidivism of an individual to aid the judge in sentencing.[13] At the moment, the use of algorithms for supporting decision-making in criminal trials is very rare in Europe. To date, the only predictive software identified in Europe is HART (Harm Assessment Risk Tool) developed in partnership with Cambridge University and used in the United Kingdom by the Durham police. This technology, based on machine learning, is expected to be able to assess the risk - low, medium or high - of reoffending of suspects.[14] For the purpose of this analysis, attention will be on the second category, namely algorithmic evidence.[15]

Although tautological, *algorithmic evidence* can be simply defined as evidence procured through algorithmic processes.[16] In this case the algorithm takes some data as input to produce evidence as output. Such evidence can be of two types: computer-derived evidence (*e.g.* GPS positions and data stored in the smartphone) and computer-generated evidence (*e.g.* breathalysers or DNA analysers).[17]

[13] Tim Brennan, William Dieterich and Beate Ehret, 'Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System' (2009) 36 Criminal Justice and Behavior 21; Katherine Freeman, 'Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis' (2016) 18 North Carolina Journal of Law and Technology 75.

[14] For more on HART see Francesca Palmiotto, 'The Impact of Algorithmic Opacity on Fair Trial Rights in Criminal Proceedings' in Martin Ebers and Marta Cantero Gamito (eds), *Algorithmic Governance and Governance of Algorithms* (Springer Forthcoming); Marion Oswald and others, 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and "Experimental" Proportionality' (2018) 27 Information & Communications Technology Law 223.

[15] Nevertheless, in Section III.1, important reference is made to this distinction.

[16] Evidence can also be in *contained in* the algorithm, when its inner working are directly relevant to prove accountability, to adjudicate who is answerable and blameworthy and for the traceability of fault and causation. One clear example is the *Google Shopping case*, where the European Commission fined Google 2.42 billion Euros for breaching EU antitrust rules. Google has abused its market dominance as a search engine by giving an illegal advantage to another Google product, its comparison shopping service. In this case, the Commission had to gather evidence on the functioning of the algorithm to demonstrate that it was intentionally violating antitrust law. For more see Patterson MR, 'Algorithmic Opacity and Exclusion in Antitrust Law' (2018) 5 Rivista Italiana di Antitrust / Italian Antitrust Review <http://doi.org/10.12870/iar-12870> accessed 6 February 2020. When evidence is contained in the algorithm, the party bearing the burden of proof needs to have access to information regarding the algorithm, *i.e.* in criminal cases the public prosecution. Such evidence, however, is in most of the cases in the hands of defendants, causing problems of compatibility with the right to silence. Consequently, this type of evidence will not be the focus of this work.

[17] Martha M Jenkins, 'Computer-Generated Evidence Specially Prepared for Use at Trial Law and Technology Symposium' (1975) 52 Chicago-Kent Law Review 600.

Computer-derived evidence refers to 'digital evidence' or 'electronic evidence'.[18] In criminal proceedings any digital data can constitute evidence (*e.g.* images, Word documents, metadata, conversations on Skype, messages on WhatsApp). In order to be presented at trial, digital data needs to be collected and acquired from its source (hard-disk, computer, smartphone, Internet, apps). The process of acquisition, collection and preservation of digital evidence requires the use of specific forensic tools algorithms. In this case algorithms are used as a mean of obtaining digital evidence. Consequently, the quality of the evidence presented is strictly dependent on the quality of the software and techniques used to acquire it.

Computer-generated evidence refers to automated forensic techniques, such as DNA analysis, identification with facial recognition software, drones for mapping crime scenes, breath analysers for determining alcohol content, license plate readers. In these cases, the input (*e.g.* biological samples) is processed by the algorithm to 'generate' evidence (e.g. DNA results). To some extent, these algorithms offer an 'expert evidence'.[19]

2. *Algorithmic opacity and what is hiding behind*

As far as computers are concerned, there is a common expectation of objectivity, precision and impartiality by non-specialists.[20] If you watch a video, for instance, you would expect that what you are seeing on the screen perfectly corresponds to what happened in the reality. However, sometimes, this may not be the case.[21] Even computer programs can go wrong. As in algorithmic models there is a direct correlation between

[18] In this paper, electronic evidence and digital evidence will be used as synonyms. The definition adopted is the one proposed by the Evidence Project (European Data Informatics Exchange Framework for Courts and Evidence) and states as follow: "Electronic evidence is any data resulting from the output of an analogue device and/or a digital device of potential probative value that are generated, processed, stored or transmitted using any electronic device. Digital evidence is that electronic evidence that is generated or converted to a numerical format". See Maria Angela Biasiotti and others (eds), *Handling and Exchanging Electronic Evidence Across Europe* (Springer International Publishing 2018).

[19] Christian Chessman, 'A Source of Error: Computer Code, Criminal Defendants, and the Constitution' (2017) 105 Cal. L. Rev. 179.

[20] Sergey Bratus, Ashlyn Lembree and Anna Shubina, 'Software on the Witness Stand: What Should It Take for Us to Trust It?' in Alessandro Acquisti, Sean W Smith and Ahmad-Reza Sadeghi (eds), *Trust and Trustworthy Computing* (Springer Berlin Heidelberg 2010).

[21] For instance, in a recent case in Ontario, the video of the crime scene may not be admitted at trial due to its poor quality, leaving key elements, such as the speed and force of police officers, open to misinterpretation. See Laura Osman, 'Key Abdi Video Unreliable, Expert Tells Officer's Manslaughter Trial' (*CBC News*, 20 June 2019) <https://www.cbc.ca/news/canada/ottawa/video-expert-montsion-1.5181755> accessed 29 February 2020.

the input, the procedure and the output. Any flaw or error present in this interplay can alter the quality of the resulting output.

In of Linda's murder, I mention at least three different types of algorithmic evidence: a) a malware, b) a forensic software for acquiring data in the smartphone and c) a DNA analyser. If this case is fictitious, the following examples are not. Recently, a group of researchers[22] discovered that Exodus, a malware used for hacking by Italian law enforcement, was designed to conduct activities that are not permitted under Italian law, thus resulting in an illegal evidence.[23] EnCase is a popular forensic tool used by forensic experts worldwide to collect computer-derived evidence, including timestamps.[24] Although the reliability of EnCase has rarely been questioned, relevant research has shown that by using an undocumented Windows system call, instead of another system call, all timestamps could be undetectably modified.[25] STRmix DNA is an automated forensic software that uses complex mathematical formulae to examine the statistical likelihood that a certain genotype can be attributed to one specific individual over another. This instrument is used in several jurisdictions in the US, Canada and Europe. In 2015, a partial inspection of the source code revealed different technical errors which reduced the probability of a DNA sample matching the defendant.[26]

[22] Security without Borders, 'Exodus: New Android Spyware Made in Italy' (*Security without Borders,* 29 March 2019) <https://securitywithoutborders.org/blog/2019/03/29/exodus.html> accessed 1 April 2019

[23] A malware, simply put, is a malicious software that is installed secretively on a device, where it can then conduct a wide range of invasive activities. *Exodus* was disguised in other apps, seemingly harmless, available on Google Play and on the Apple store. When the user installed one of these apps, *Exodus* could hack the phone and steal data from it. After download, it was able to (for the Android version) retrieve browsing history, media exchanged through WhatsApp and SMS, record surroundings using the microphone and phone calls, take pictures with the camera, extract the GPS position, events from the Calendar, the call logs, contact list from Facebook etc. *Exodus* relied on highly invasive techniques that significantly reduced the overall security of the device, making it vulnerable to other hacking attacks. Furthermore, the software was never remotely disinfected by the operator. For more on Exodus and the illegal activities carried out with it see Palmiotto F, 'Algorithmic Opacity as a Challenge to the Rights of the Defense' (*RAILS*, 6 September 2019) <https://ai-laws.org/en/2019/09/algorithmic-opacity-challenge-to-rights-of-the-defense/> accessed 1 March 2020.

[24] A timestamp is an information about the time of creation, access and modification of a certain file. For more on the software see the website https://www.guidancesoftware.com/encase-forensic.

[25] Eric Van Buskirk and Vincent T Liu, 'Digital Evidence: Challenging the Presumption of Reliability' (2006) 1 Journal of Digital Forensic Practice 19.

[26] Chessman (n 19) 188; see also Health Support Queensland and Forensic and Scientific Services, 'STRmix Miscoding Error'<https://www.health.qld.gov.au/__data/assets/pdf_file/0029/633368/dohdl1617012.pdf> accessed 13 February 2019.

These examples show that any flaw affecting the algorithm will consequently impact the quality of the resulting evidence in terms of legality, validity and accuracy.[27] Detecting these flaws, however, is extremely difficult when algorithms are concealed in opacity and secrecy that prevent insight into their functioning. Akin to the "Wanderer above the sea of fog",[28] our view is also blurred.

According to Burrell, algorithmic opacity can be of different kinds. It can stem from intentional corporate or state secrecy, technical illiteracy or from the lack of interpretability.[29] Pasquale, on the other hand, distinguishes between three types of secrecy: legal, real and obfuscation.[30] Arguably, the notion of opacity not only refers to barriers between access and information, such as the protection of trade secret, but it also concerns the issue of understandability of such information.

This second aspect has become particularly crucial after the rise of AI systems.[31] Traditional software is in fact based on a fixed and known model, whereas AI systems operate in environments that are at best partially known, even by the system designer.[32] In this case, understanding how the output has been generated is more of a challenge.[33] This lack of insight on the system represents an urgent legal concern as it prevents testing the reliability of computer systems.

[27] In a previous work, I refer to these flaws with the term 'miscode' and I offer a threefold categorization. See Palmiotto (n 14).

[28] Caspar David Friedrich, *Wanderer above the Sea of Fog* (1818, oil on canvas). In this painting, a man stands upon a rocky precipice in front of a landscape covered in thick fog.

[29] Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society, 1.

[30] According to the author, real secrecy establishes a barrier between hidden content and unauthorized access to it. Legal secrecy obliges those privy to certain information to keep it secret. Obfuscation involves deliberate attempt at concealment when secrecy has been compromised. See Pasquale (n 10) 12.

[31] With regard to the General Data Protection Regulation (GDPR), Malgieri et al. refer to the concept of 'legibility' that is the capability of individuals to autonomously understand the logic, the significance and the envisaged consequences of an algorithmic decision-making. See Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 243.

[32] While datasets may be extremely large but possible to comprehend and code may be written with clarity, the interplay between the two in the mechanism of the algorithm is what yields the complexity (and thus opacity). See Burrel (n 29) 5.

[33] Marc Beillevaire, 'Inside the Black Box: How to Explain Individual Predictions of a Machine Learning Model' (Master thesis, KTH Royal Institute of Technology School of Electrical Engineering and Computer Science 2017).

3. Calling algorithms to the stand? The right to confront algorithmic evidence

Algorithms are a set of instructions to solve a problem, and yet they may create one. Insofar as individuals are not able to understand, even with the help of legal counsel, complex algorithmic systems used to process evidence against them, there is a significant threat to due process rights.[34] Algorithmic evidence challenges, *inter alia*, the right to confrontation when such evidence cannot be examined by the defence due to its opacity.[35] In this regard, the notion of 'fair trial' provided by Article 6 of the European Convention of Human Rights (hereafter ECHR) will be adopted.

In the criminal limb, the guarantees of Article 6 ECHR apply when a 'criminal charge' is being determined. This concept has an autonomous meaning in the Convention. In the determination of a 'criminal offence', the starting-point for the assessment is based on the criteria outlined in *Engel and Others v. the Netherlands*,[36] that are as follow: 1) classification in domestic law 2) nature of the offence and 3) severity of the penalty that the person concerned risks incurring. According to the following Court's case-law, a 'charge' exists from the moment that an individual is officially notified by the competent authority of an allegation that he has committed a criminal offence, or from the point at which his situation has been substantially affected by actions taken by the authorities as a result of a suspicion against him.[37] As a result, the guarantees of the fair trial are applicable throughout the entire proceeding, including the pre-trial stage,[38] the sentencing process,[39] appeal proceedings against convictions or sentences that are in fact provided

[34] Veale M, 'Algorithm Use in the Criminal Justice System Report' (The Law Society of England and Wales 2019) 57.

[35] On the right to confrontation and computer systems see Joseph Clarke Celentino, 'Face-to-Face With Facial Recognition Evidence: Admissibility Under the Post-Crawford Confrontation Clause' (2016) 114 Michigan Law Review 1317; Karen Neville, 'Programmers and Forensic Analyses: Accusers Under the Confrontation Clause' (2011) 10 Duke Law & Technology Review 1; Andrea Roth, 'Machine Testimony' (2017) 126 Yale Law Journal 1972; Brian Sites, 'Rise of the Machines: Machine-Generated Data and the Confrontation Clause' (2014) 16 The Columbia Science & Technology Law Review 36.

[36] *Engel and Others v. the Netherlands* (1976) Series A no 22, paras 82-83.

[37] *Deweer v. Belgium*, (1980) Series A no 35, paras 42-46; *Eckle v. Germany* (1982) Series A no 51, para 73; *McFarlane v. Ireland* App no 31333/06 (ECtHR, 10 September 2010), para 143; *Ibrahim and Others v. the United Kingdom* App no 50541/08, 50571/08, 50573/08 and 40351/09 (ECtHR, 13 September 2016), para 249.

[38] *Dvorski v. Croatia* ECHR 2015-VI 405, para 76.

[39] *Phillips v. the United Kingdom* ECHR 2001-VII 20, para 39.

for[40] and, similarly, the execution of final judgments in criminal cases.[41] Consequently, within this timeline, the use of algorithms for evidentiary purposes must be compatible with the guarantees of a 'fair trial' set in Article 6 ECHR.

With regard to algorithmic evidence, the right to confrontation is particularly relevant. For the purpose of a fair trial, the defendant[42] must have the possibility of confronting evidence against him.[43] Specifically, Article 6§3(d) ECHR enshrines the principle that, before an accused can be convicted, "all evidence against him must normally be produced in his presence at a public hearing with a view to adversarial argument".[44] In the light of the autonomous and broad interpretation of the term 'witness' adopted by the Court, this provision applies to algorithmic evidence when such evidence is 1) used at trial and appeal proceedings and 2) when it can serve to a material degree as the basis for a conviction.[45]

Recalling the categories illustrated above, all types of algorithmic evidence can potentially fulfil these requirements. Algorithmic evidence conveys information that is necessary to establish facts in criminal proceedings and its statements "have become so probative and powerful that an algorithm like STRmix[46] can become the primary accuser in a criminal trial".[47] While a computer-derived evidence, such as the GPS position, conveys information about where an individual was and at what time, a computer-generated evidence could convey information about, for instance, the probability that the DNA found on the victim matches the DNA of the defendant.

[40] *Eckle v. Germany* (1982) Series A no 51, para 73.

[41] David John Harris and others, *Law of the European Convention on Human Rights* (3rd edn, Oxford University Press 2014).

[42] In this paper, the words 'defendant', 'accused' and 'defence' will be used as synonyms in line with the terminology of the ECtHR on Article 6§3. Article 6§3 ECHR, in fact, encompasses a list of rights granted to "everyone charged with a criminal offence", that are commonly referred to as the 'rights of the defence'.

[43] Mehmet Arslan, 'The Right to Examination of Prosecution Witnesses' (2018) 6 ZIS 218.

[44] *Al-Khawaja and Tahery* (no 11), para 118.

[45] For instance, in the case *Georgios Papageorgiou v. Greece* ECHR 2003-VI 229 the Court examined under Article 6§3(d) ECHR the issue of access to the original documents and computer files relevant to the criminal accusations against the applicant.

[46] For more information about this software see above Algorithmic opacity and what is hiding behind.

[47] Roth (n 35) 2044.

To some extent, algorithms do what witnesses do: “they make claims relied upon by factfinders for their truth”.[48] However, these claims are not reliable *a priori* as they may be flawed. According to the leading case *Al-Khawaja and Tahery v UK*, the right to confrontation not only requires that defendants should be in a position to challenge the *probity* and *credibility* of their evidence, but that they should also be able to test their *truthfulness* and *reliability.*[49]

Algorithmic opacity is an obstacle to this test. In order to achieve the purpose of the confrontation defendants need to have insights on the system. Oftentimes, this cannot be easily achieved as they might lack access to information and/or resources to understand it. With regard to this second aspect, in an expert interview conducted for the study “*Legal Framework for Hacking by Law Enforcement*”, Giovanni Ziccardi raised concerns also on the *de facto* inability to challenge the evidence collected through technological tools by legal professionals, due to their lack of adequate knowledge.[50] Some examples provide further clarification on this matter.

In criminal investigations, malware is used as a mean to obtain digital evidence (i.e. algorithmic evidence). Consequently, the quality of the evidence presented is strictly dependent on the quality of the software and techniques used to acquire it. When evidence is gathered through technological means, the defence has the right to check that the investigative operations have been carried out in compliance with the relevant provisions and that no further manipulation of the collected data has occurred. In the case of Exodus, the defence should have been in the position to confront the evidence collected through it and to show that it was (potentially) illegally obtained.[51] Nevertheless, very little was

[48] Roth (n 35) 2001.

[49] *Al-Khawaja and Tahery* (no 11) para127.

[50] It is reported that the use of hacking techniques and tools, and the evidence gathered through these means, is not challenged in court as many legal professionals do not have the required knowledge. This inability to challenge the evidence collected limits the ability of targeted persons to gain effective remedy. See Mirja Gutheil and Quentin Liger, ‘Legal Frameworks for Hacking by Law Enforcement: Identification, Evaluation and Comparison of Practices’ (Study requested by the European Parliament’s Committee on Civil Liberties, Justice and Home Affairs and was commissioned, overseen and published by the Policy Department for Citizens’ Rights and Constitutional Affairs, 2017) 56.

[51] For more on Exodus see Francesca Palmiotto, ‘Algorithmic Opacity as a Challenge to the Rights of the Defense’ (*RAILS*, 6 September 2019) <https://ai-laws.org/en/2019/09/algorithmic-opacity-challenge-to-rights-of-the-defense/> accessed 1 March 2020.

known on this software before research drew more attention to the problem.[52] To date, all the public information on Exodus comes from extensive research and investigation work by media. Could the defence have had the same information? Probably no. Paradoxically, the numerous barriers to information would not even have allowed the defence to know what kind of software was used, the name of the company that developed it or its technical characteristics.

As an example of computer-generated evidence, a fascinating case is Intoxilyzer 5000EN, a device used to measure breath alcohol content in the US. In a series of cases in Minnesota, defendants asked for the discovery of source code underlying the software to be able to challenge its results.[53] The main argument of the defendants was that the report produced by the machine effectively acts as a witness, as it makes statements about them.[54] Consequently, the defence needed access to information regarding the machine's inner workings, its design and programming. In this regard, even a thorough analysis of the operator who administered the test could not be considered sufficient to satisfy this demand. The analysis of the code finally revealed that "the machine was susceptible to a variety of undetected failures, including erroneous results based on power surges, interference from cell phones, and defects in the process of self-testing and reporting errors".[55] Furthermore, whilst breath alcohol content can be used to estimate blood alcohol content (BAC), the conversion factor for breath to blood can vary from one person to another. This particular device operated on the basis of Henri's Law and used a partition ratio of 2100:1. However, not every person has this ratio. A wrong partition ratio can cause significantly erroneous results, as people with a lower ratio may have their BAC overestimated, and those with a higher ratio may have their BAC underestimated.[56]

[52] Security Without Borders (no 23); Emanuel Maiberg and others, 'Researchers Find Google Play Store Apps Were Actually Government Malware' (*Motherboard*, 2019 2019) <https://motherboard.vice.com/en_us/article/43z93g/hackers-hid-android-malware-in-google-play-store-exodus-esurv> accessed 1 April 2019; 'Più di mille italiani intercettati sul cellulare, per errore, da un hacker di Stato' *La Repubblica* (30 March 2019) <https://www.repubblica.it/tecnologia/sicurezza/2019/03/30/news/molte_centinaia_di_italiani_intercettati_su_cellulare_per_errore_da_hacker_di_stato-222865990/> accessed 2 April 2019.

[53] *In re Source Code Evidentiary Hearings in Implied Consent Matters*, 816 N.W.2d 525 (Minn. 2012). For more on the case see David Liebow, 'DWI Source Code Motions after Underdahl' (2010) 11 Minn. JL Sci. & Tech. 853 and Chessman (n 19) 196 and *State v Underdahl* (*Underdahl III)* 767 N.W.2d 677 (Minn. 2009).

[54] *State v Underdahl* (*Underdahl III)* 767 N.W.2d 677 (Minn. 2009).

[55] Chessman (n 19) 197.

[56] Liebow (n 53) 856.

After disclosure of the source, the defence was then able to test the adequacy of the scientific method embedded in the software.

Although the probative value of evidence will be ultimately evaluated by the judge, the defence must have the opportunity to challenge algorithmic evidence. Beyond the admissibility stage, the defence should "be allowed to impeach machines at trial, just as the opponent can impeach human witnesses and declarants even when a judge deems their assertions reliable".[57] In order to achieve the purpose of the confrontation, the defence needs to comprehend how the software used for acquiring or generating the evidence works, clearing the view from the fog of opacity. The procedural question remains how the defence can obtain information and expert knowledge to determine the significance of such information.

4. *Transparency and Explanation as proposed solutions*

A growing number of scholars from several disciplines are currently studying the issue of algorithmic opacity from different angles.[58] The tools proposed in the literature are mainly twofold: transparency and explanation.

A first strong strand of research advocates for algorithmic transparency as instrumental to understand how ubiquitous algorithms work. In the common sense, transparency is the opposite of opacity and refers to accessing the mechanisms by which the model works.[59] The major trend is to consider 'algorithmic transparency' (as opposed to opacity) a key-instrument for enhancing accountability of these systems,[60] for providing explanations to

[57] Roth (n 35) 1982.

[58] For an excellent literature review see Brent Daniel Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society.

[59] Zachary C Lipton, 'The Mythos of Model Interpretability' [2017] arXiv:1606.03490 [cs, stat] <http://arxiv.org/abs/1606.03490> accessed 11 December 2019.

[60] Joshua A Kroll and others, 'Accountable Algorithms' (2016) 165 U. Pa. L. Rev. 633; Mike Ananny and Kate Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 New Media & Society 973; Paul B de Laat, 'Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?' [2017] Philosophy & Technology; Nicholas Diakopoulos, 'Algorithmic Accountability Reporting: On the Investigation of Black Boxes' [2014] Tow Center for Digital Journalism, Columbia University <https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2> accessed 5 March 2020.

individuals affected by their use[61] and for limiting the adverse effects of the 'Black Box' on society.[62]

Nevertheless, the concept of algorithmic transparency has also been criticized for its shortcomings. Particularly, Ananny and Crawford argue that transparency, in its literal significance as 'being able to see a system', does not equate to *understand* algorithms.[63] We then "see without knowing".[64] Furthermore, transparency can even have undesirable consequences, such as allowing strategic gaming of the system[65] and harm competitiveness.[66] In light of these limitations, a second strand in the literature refers to the concepts of explanation and/or interpretability.[67] Explanation has been defined as a "human-interpretable description of the process by which a decision-maker took a particular set of inputs and reached a particular conclusion".[68] This term is usually referred to in order to overcome the shortcomings of transparency and mitigate concerns about privacy and trade secrets.

Literature on algorithmic transparency and explainable AI can contribute to understanding the upstream technical problem of algorithmic opacity and provide technical solutions. The right to confrontation, however, requires more, particularly a specific framework creating procedural rules. It is the role of procedural justice to ensure

[61] Tae Wan Kim and Bryan R Routledge, 'Informational Privacy, A Right to Explanation, and Interpretable AI' (*2018 IEEE Symposium on Privacy-Aware Computing (PAC)* 2018) <https://ieeexplore.ieee.org/document/8511831/> accessed 5 March 2020; Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2018) 16 IEEE Security & Privacy 46; Finale Doshi-Velez and others, 'Accountability of AI Under the Law: The Role of Explanation' (2017) arXiv:1711.01134 [cs, stat] <http://arxiv.org/abs/1711.01134> accessed 23 December 2019; Burrell (n 35).

[62] Pasquale (n 10); Danielle Keats Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions Essay' (2014) 89 Washington Law Review 1; Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (First edition, Crown 2016); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (First Edition, St Martin's Press 2017).

[63] Ananny and Crawford (n 60).

[64] Ibid 977.

[65] Bruno Lepri and others, 'Fair, Transparent, and Accountable Algorithmic Decision-Making Processes' (2018) 31 Philosophy & Technology 611.

[66] de Laat (n 60).

[67] Lipton (n 59); Lepri and others (n 65); Mittelstadt and others (n 58); Brent Mittelstadt, Chris Russell and Sandra Wachter, 'Explaining Explanations in AI', (*Proceedings of the Conference on Fairness, Accountability, and Transparency* (FAT) *2019* ACM Press); Kim and Routledge (n 61); Doshi-Velez and others (n 61); Andrew D. Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233.

[68] Doshi-Velez and others (n 61).

challengeability of evidence and relieve the burden that opacity places on the defence asking for disclosure of information and seeking for expert knowledge to determine the significance of such information.[69] In this sense, many procedural questions should be carefully addressed to examine the practicability of potential solutions[70] and how they can be hinged in the criminal justice system. The goal of overcoming the barriers posed by algorithmic opacity must be balanced with the other legitimate purposes of the criminal process, such as the protection of other individual rights,[71] the reasonable time of the proceeding and the need of legal certainty. The following Section will investigate whether the EU legal response can be considered adequate.

III. A ROLE FOR THE EU LEGISLATOR?

1. The inadequacy of the EU Data Protection Legislation

Recently, the issue of algorithmic opacity of AI or, more generally, automated systems has become the focus of many soft-law instruments and governance strategies.[72] Furthermore, it has also seized the attention of the EU legislator in the Data Protection legislation, composed of the General Data Protection Regulation 2016/679 (hereafter GDPR) and the Directive EU/2016/680 (hereafter LED). Both instruments contain different provisions regarding automated decision-making and the rights of data subjects to receive information on such processing.[73]

As to the GDPR, Article 13(2)(f), Article 14(2)(g) and Article 15(1)(h) respectively enshrine the right to receive and to access meaningful information regarding the existence

[69] See Ananny and Crawford (n 60) 979 with regard to the concept of transparency as burden.

[70] Such as the source code disclosure, independent validation, certificates, reverse engineering etc..

[71] Secrecy in algorithmic design can be instrumental to the protection of the *ius excludendi alios*, such as proprietary protection, state or corporate secrecy or privacy rights.

[72] See *inter alia* the Independent High-Level Expert Group on Artificial Intelligence setup by the European Commission, that authored the 'Ethics Guidelines for Trustworthy AI' (December 2019) available at https://ec.europa.eu/futurium/en/ai-alliance-consultation.

[73] In general, the principle of 'transparency of data processing', epitomized in Article 5(1)(a) GDPR, requires the information to be "*concise, easily accessible and easy to understand*" (emphasis added) and that the data subject is informed "*of the existence of the processing operation and its purposes*". Focus thus is not only on mere access to information, but also on the need to make such information comprehensible to the data subject. With particular regard to automated decision-making, next to the general prohibition of decision-making based solely on automated processing (Article 22 GDPR and Article 11 LED), different provisions enshrine the right of the data subject to receive or have access to meaningful information on such processing.

of automated decision-making, the logic involved and the significance and envisaged consequences of such processing. The LED refers to automated decision-making in Article 11, where it states a general prohibition of automated decisions which produce an *adverse*[74] legal effect concerning the data subject or significantly affects him or her. The rights to information in the LED are more generic and do not specifically refer to automated decision-making. In the Directive, Article 13 and 14 enshrine the right to receive or to access to information, whereas the following Article 15 lists the limitations to these rights.

Undoubtedly, the EU Data Protection legislation can be considered as a first attempt to enhance human interpretability in algorithmic design.[75] However, can this legal response be considered adequate for criminal proceedings? Firstly, due to the differences in terms of provisions and legal protection of the data subject, it is important to investigate whether the GDPR or the LED apply to algorithms used in criminal proceedings.

In criminal proceedings only the provisions of the Directive apply. In fact, according to Article 1(1), Article 2 and Recital 11, the LED applies to the processing of personal data by competent authorities[76] for the purposes of 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. Where competent authorities process personal data for purposes other than for the purposes of the Directive, the GDPR applies. Consequently, only the provisions of the LED are relevant to the scope of this paper.

[74] Note that in the GDPR the adjective "adverse" is not included in Article 22.

[75] For more on the subject see Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond' (2019) 27 International Journal of Law and Information Technology 91; Juraj Sajfert and Teresa Quintel, 'Data Protection Directive (EU) 2016/680 for Police and Criminal Justice Authorities', *Cole/Boehm GDPR Commentary* (Edward Elgar Publishing, Forthcoming) <https://papers.ssrn.com/abstract=3285873> accessed 23 December 2019; Bryce Goodman and Seth Flaxman, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"' (2017) 38 AI Magazine 50; Gianclaudio Malgieri, 'Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations' (2019) 35 Computer Law & Security Review; Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76.

[76] According to Recital 11 LED: "Such competent authorities may include not only public authorities such as the judicial authorities, the police or other law-enforcement authorities but also any other body or entity entrusted by Member State law to exercise public authority and public powers for the purposes of this Directive."

However, the provisions regarding automated decision-making (Article 11 LED) and rights to information (Article 13 and 14 LED) cannot be considered sufficient to face the challenges brought by the use of algorithms in criminal proceedings for two reasons: a) the narrow scope of the notions of profiling and automated decision-making and b) the narrow applicability of the rights to information.

Firstly, pursuant to Article 11 LED, decisions based *solely* on automated processing, including profiling, which produce an *adverse legal effect* on the data subject or *significantly affects him or her*, should be prohibited. However, Union or Member State law can authorize such a processing if appropriate safeguards are provided, such as the right to obtain human intervention. Profiling is defined in Article 3(4) LED as a form of automated processing for the evaluation of personal aspects about a natural person, whereas solely automated decision-making is defined in the Guidelines on automated individual decision-making provided by the Article 29 Working Party as the "ability to make decisions by technological means without human involvement".[77] If one applies these definitions to the taxonomy of algorithms provided above, one should note that the applicability of this provision is limited. In fact, in criminal proceedings algorithms are mainly used for evidentiary purposes rather than for making decisions autonomously.[78]

It could be argued, however, that profiling is also not a decision *per se*. In the Guidelines of Article 29 Working Party, profiling is actually defined as an activity carried out more generally to *aid* decision-making, akin to the evidence in a trial. However, it would be difficult to argue that an algorithmic evidence can produce *adverse legal effects* or *significantly affect* an individual. What produces these effects is rather the decision that is based on such evidence.

So far, the only category that seems to be included in the notion of Article 11 LED regards algorithms used for the evaluation of personal aspects – such as risk-assessment software – whereas the much broader category of algorithmic evidence falls out of the scope of this Article. In my view, the only concrete field of application of Article 11 LED is to discourage the use of fully automated systems for evaluating personal aspects of the data subject. For instance, in the French transposition of the LED, as far as judicial decisions

[77] Working Party on the protection of individuals with regard to the processing of personal data, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, adopted on 3 October 2017.

[78] See above Section II.1 on the distinction between algorithmic evidence and decision-support algorithms.

are concerned, there is a total prohibition of semi or fully automated decision-making if such processing is intended to evaluate aspects of personality.[79]

Secondly, as to the rights to information, there is a crucial difference between the Regulation and the Directive. In the LED the right to access of the data subject has an extremely narrow application. Article 15 LED lists several limitations that can severely hamper the right to access by the defence, such as *inter alia* the interest to avoid prejudicing the prevention, detection, investigation or prosecution of criminal offences or to protect the rights and freedom of others. In this regard, one might question which kind of balance should be struck, for instance, between the request of the defendant to access information relevant for challenging a piece of evidence on the one hand, and the investigative secret or privacy rights on the other. These limitations come from the needed balancing between the rights of the data subject and other legitimate purposes of criminal justice. Furthermore, the possibility to access to meaningful information about "the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject" envisaged in Article 15 GDPR is not even mentioned in the Directive.

However, the fact that transparency requirements are sacrificed in the Directive, compared to the GDPR, further supports the vision that transparency should be balanced with other legitimate purposes of the criminal process. In my view, the Directive struck this balance because transparency is linked to the right to privacy, a right that is considerably waived in the criminal process. The challenges brought by the use of opaque algorithms on the rights of the defence thus need to be addressed from the lens of criminal procedural law, rather than privacy and data protection law. Data Protection regulation cannot be a panacea for every technology-related issue.

2. *The European Proposal on Electronic Evidence: an opportunity*

Another main characteristic of the algorithmic evidence is that it is cross-border by nature. Digital data are often not located in the same territory of the investigating authority, but

[79] Article 10(1), Loi n. 78-17 du 6 janvier 1978 as amended by Loi n. 2018-493 du 20 juin 2018. The case of UK is also interesting, since the right to contest the decision has been embraced in the right to request the controller to reconsider the decision (Section 14[4][b][i] UK Data Protection Act 2018) The data controller must consider the request, "including any information provided by the data subject that is relevant to it". For more on the national implementation see Malgieri (n 75).

in foreign or multiple jurisdictions.[80] Recalling the example of Tea's murder, the relevant evidence can be collected by different actors, including private companies, that may be located in different countries, such as data collected from WhatsApp[81] or Alexa. Additionally, freedom of movement in the EU territory fostered the transnational dimension of criminal investigations. As a result, authorities of EU Member States increasingly require access to data that might serve as evidence and that is stored outside their country and/or by service providers in other countries.The European instruments of mutual recognition serve this purpose.[82] However, there is still a lack of a unified legal framework and shared rules that make it possible to handle evidence and its possible exchange in a uniform manner across the Member States,[83] as well as a lack of attention on the rights of the defence.[84]

With regard to the latter aspect, it should be noted that the issue of algorithmic opacity is further exacerbated in transnational criminal investigations. As explained above, the lack of knowledge on how the output of a system, namely the evidence, is generated represents an obstacle to the right to confrontation, as the defence cannot test its reliability and trustfulness.[85] When evidence is collected abroad, this possibility is even more hinged by the intrinsic complexity of the system in terms of actors, legal framework (*lex loci, lex*

[80] For a detailed reconstruction on the cross-border nature of electronic evidence see Bonnic Mifsud, Tudorica Melania and Cannataci Joseph, 'The European Legal Framework on Electronic Evidence: Complex and in Need of Reform' in Maria Angela Biasiotti and others (eds), *Handling and Exchanging Electronic Evidence Across Europe* (Springer International Publishing 2018).

[81] From the privacy policy it is not clear where the data are stored. Even if the data controller for EU citizens is WhatsApp Ireland, in the policy is stated that: "Information controlled by WhatsApp Ireland will be transferred or transmitted to, or stored and processed, in the United States or other countries outside of where you live for the purposes as described in this Privacy Policy". See in the website https://www.whatsapp.com/legal/#privacy-policy.

[82] The current EU legal framework consists of Union cooperation instruments such as, among others, the Directive 2014/41/EU on the European Investigation Order (EIO Directive), the Convention on Mutual Assistance in Criminal Matters between MS, the Council Framework Decision 2002/465/JHA on joint investigation team, and other agreements such as the Agreement on Mutual Legal Assistance (MLA) between US and EU. Additionally, another crucial tool for the regulation of cross-border access to electronic evidence is the Council of Europe's Budapest Convention on Cybercrime (CETS No 185), ratified by most EU Member States.

[83] Biasiotti and others (n 18).

[84] On the transnational gathering of evidence and the rights of the defense see also Lorena Bachmaier Winter, 'Transnational Criminal Proceedings, Witness Evidence and Confrontation: Lessons from the ECtHR's Case Law' (2013) 9 Utrecht L. Rev. 127.

[85] See Section 3.

fori and European law),[86] different languages used in the relevant documents and norms, availability and access to information and documents. In this sense, the system of cross-border access to the electronic evidence adds an extra layer of opacity which weighs on the defence. Reviewing the whole process that brought the evidence ultimately before a court requires a high level of expertise and resources. Henceforth, it is necessary to establish specific rules to ensure that the evidence that is transferred from one Member State to another can be successfully challenged by the defence.[87]

Even if criminal law is regulated at national level, it is almost necessary to have a transnational regulation considering globalization and modern technologies, mobilization of individuals and especially the very volatile and cross-border nature of digital evidence.[88] In this regard, the EU level can be the best suited to address a more effective and uniform protection of the rights of the defence in cross-border proceedings, ensuring mechanisms to grant the right to confront algorithmic evidence in every Member State, in compliance with European fair trial rights. This was also the main outcome of the Evidence Project[89] that underlined the urgent need of a uniform European Regulation regarding admissibility, reliability and the security of the tools used for acquiring and preserving digital evidence, through clear and specific rules setting common definitions

[86] On this aspect see Lorena Bachmaier Winter, 'Transnational Evidence: Towards the Transposition of Directive 2014/41 Regarding the European Investigation Order in Criminal Matters' [2015] Eucrim: the European Criminal Law Associations' fórum 47. As argued by the author: "It is often taken for granted that the executing State taking care that the investigative measure performed in its territory respects the lex loci, adjusted, if need be, to the formalities of the lex fori. But who controls in the main proceedings if such rules have been respected? In most countries, the defence is supposed to take care of it, but what are the real possibilities if the defence does not know how the evidence was obtained in the foreign country and what the applicable rules in that country are".

[87] For instance, with regard to digital evidence, the mere process of collection of such evidence could and should be regulated following the international standards and forensic protocols that must be applied when digital investigations are carried out. In this way, the defence will have a transparent (and translated) framework to test the taking of such evidence. This is a suggestion, although not sufficient, that could however reduce the effort of the defence in gaining and understanding information relevant to challenging the evidence obtained abroad.

[88] Mifsud and others (n 80).

[89] The Evidence Project, funded by the European Commission under the 7th Framework Programme, started in March 2013 and ended in October 2016, used a multidisciplinary approach in close collaboration with the various stakeholders to identify, define and evaluate a set of actions that should be undertaken at the EU level and national measures to enable the electronic exchange of evidence between the competent authorities in Europe. The aim of the EVIDENCE project was to create a Common European Framework for the correct and harmonised handling of electronic evidence during its entire lifecycle: the collection, preservation, use and – in particular – exchange of electronic evidence. The main goal was to draft a Roadmap including policy recommendations, guidelines, technical standards, legislation, further research, etc. for realising this Common European Framework. See for more www.evidenceproject.eu

and approximation of legal procedures.[90] Shared standards for maintaining the integrity of the electronic evidence are also necessary for enhancing a form of admissibility of evidence based on mutual trust and for giving the defence a clear and transparent framework for challenging the evidence obtained in this fashion.

The growing attention of the EU legislator towards the regulation of electronic evidence must include reflections on the how effective protection of the defence across Member States can be guaranteed. In the document of September 2019 entitled *The future of EU internal security: new technologies and internal security*, the European Council expressed a strong interest in the use of new technologies for law enforcement purposes, such as 5G and AI technologies, but also underlined the need to find a balance with the protection of fundamental rights. Additionally, it takes into account the specific issue of algorithmic opacity in stating 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".[91]

The same concerns and considerations are unfortunately not present in the draft Regulation on the European Production and Preservation Order.[92] In order to improve cross-border access to electronic evidence (that are currently under consideration), the European Commission released two proposals in April 2018: one for a Regulation[93] and

[90] In the edited book outcome of the project, the authors affirm that: "It is essential to activate a single legal framework on a European Level for the collection, processing and exchange of electronic evidence. This European framework should be a compromise between the need to ensure efficient police investigation and respect for fundamental rights of every citizen, on which the new technologies have a major impact". See Biasiotti (n 18).

[91] Preparation of the Council Debate, *The future of EU internal security: new technologies and internal security*, 18 September 2019 (OR. en) 12224/19.

[92] Nor they are expressed in the Opinion of the European Economic and Social Committee on '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) 225 final — 2018/0108(COD)). In the Opinion, concerns regarding the right of the defense are expressed only in very general terms in paragraph 3.7: "*The EESC underlines the fact that this Regulation must respect fundamental rights and observe the principles recognised in particular by the Charter of Fundamental Rights of the European Union and the Member States' constitutions. These include the right to liberty and security, the respect for private and family life, the protection of personal data, the freedom to conduct a business, the right to property, the right to an effective remedy and to a fair trial, the presumption of innocence and right of defence, the principles of legality and proportionality, and the right not to be tried or punished twice in criminal proceedings for the same criminal offence. The EESC points out that protecting these rights also depends on the conditions under which these rights may be encroached upon, and who decides on this*".

[93] Commission, 'Proposal on European Production and Preservation Orders for electronic evidence in criminal matters' COM (2018) 225 final.

one for a Directive[94]. The draft Regulation establishes a system to compel private service providers offering services (such as social networks, online marketplaces and other providers of internet infrastructure) in the EU to produce (Production Order) or preserve (Preservation Order) electronic evidence, regardless of its location. The issuing public authority will be able to demand through a direct channel with the service provider the production or the preservation of electronic evidence. The orders are directly addressed to a legal representative designated by the service provider, following the rules tabled in the draft Directive for the appointment of legal representative.[95]

Both proposals are justified by the Commission in the light of the need to speed up the process to secure and obtain electronic evidence, that is stored and/or held in another jurisdiction. It also intends to improve legal certainty for authorities, service providers and persons affected and maintain a high standard for law enforcement requests, thus ensuring protection of fundamental rights, transparency and accountability.[96] To better achieve these objectives, these measures need to be adopted at Union level.[97] On the one hand, the goal of enhancing legal certainty is surely beneficial, also from the perspective of the defence. However, the proposals do not make any reference to technical standards

[94] Commission, 'Proposal laying down harmonised rules on the appointment of legal representatives for the purpose of gathering evidence in criminal proceedings' COM/2018/226 final.

[95] For a more detailed comment on the proposals see Valsamis Mitsilegas, 'The Privatisation of Mutual Trust in Europe's Area of Criminal Justice: The Case of e-Evidence' (2018) 25 Maastricht Journal of European and Comparative Law 263; Nathalie A Smuha, 'Towards the EU Harmonization of Access to Cross-Border E-Evidence: Challenges for Fundamental Rights & Consistency' <https://papers.ssrn.com/abstract=3501421> accessed 1 March 2020; Marco Stefan and Gloria González Fuster, 'Cross-Border Access to Electronic Data Through Judicial Cooperation in Criminal Matters' (Social Science Research Network 2018) SSRN Scholarly Paper ID 3298705 <https://papers.ssrn.com/abstract=3298705> accessed 5 March 2020; Stanislaw Tosza, 'The European Commission's Proposal on Cross-Border Access to E-Evidence : Overview and Critical Remarks' [2018] eucrim - The European Criminal Law Associations' Forum <https://eucrim.eu/articles/european-commissions-proposal-cross-border-access-e-evidence/> accessed 2 March 2020; Stanislaw Tosza, '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' (Social Science Research Network 2019) SSRN Scholarly Paper ID 3517866 <https://papers.ssrn.com/abstract=3517866> accessed 1 March 2020.

[96] Commission, Proposal COM (2018) 225 final, 2.

[97] See Proposal COM (2018) 225 final, 6 where it states that: "Given the cross-border dimension of the problems addressed, the measures included in the proposal need to be adopted at Union level in order to achieve the objectives. […]Given the diversity of legal approaches, the number of policy areas concerned (security, fundamental rights including procedural rights and protection of personal data, economic issues), and the large range of stakeholders, Union-level legislation is the most appropriate means to address the identified problems".

and shared practices for the gathering of data requested, as it was suggested by the Evidence Project.[98]

When electronic evidence is sourced from private service providers, the quality of the data should not be taken for granted, as computer scientists explain.[99] In these procedures, law enforcement should check the quality of the data obtained from private parties, the integrity of the technician to execute the search in a trustworthy manner and the tools used to collect the data and the infrastructure to retrieve, reassemble, and report the data.[100] The absence of common standards and transparency requirements with regard to the process of data acquisition, access and search activities conducted by service providers and authorities represents a concern in this regard.[101] The result is a chain of (blind) trust where law enforcement firstly and the defence secondly (blindly) trust in the quality of the evidence collected by service providers.

Additionally, the draft texts do no enshrine specific safeguards for the defence, apart from Article 17 (1-3) that allows the suspect, the accused or the persons whose data are obtained through the order to challenge the legality, necessity, proportionality of the *order*. With regard to the *evidence* so obtained, Member States shall ensure the respect of the rights of the defence when assessing evidence obtained through the European Production Order, pursuant to Article 17 (6). Consequently, the defence can challenge the legality, proportionality and necessity of the Order, but it is unsure how the defence can effectively challenge the evidence so obtained, especially in the absence of specific standards for its acquisition. In this regard, the standardization of the procedures for handling the evidence could have had a beneficial effect on the right to confrontation – by offering to the defence a transparent framework – and ultimately enhance trust in the reliability of such evidence.

[98] Moreover, at the EU level there are guidelines and best practices that provide practical guidance for handling digital evidence set for instance by the European Union Agency for Network and Information Security (ENISA) and by the Council of Europe as the Electronic Evidence Guide (EEG). However in the proposal no common standards for the acquisition, collection, custody and exchange of electronic evidence are set or referred to.

[99] Josiah Dykstra and Alan T Sherman, 'Acquiring Forensic Evidence from Infrastructure-as-a-Service Cloud Computing: Exploring and Evaluating Tools, Trust, and Techniques' (2012) 9 Digital Investigation S90.

[100] ibid.

[101] On the transparency of the procedure to acquire the evidence see also Smuha (n 95) 20.

As a conclusive remark, the proposals on the Production and Preservation orders cannot be considered as the appropriate instrument to realize the desired harmonization of rules in handling digital evidence in Europe, as they intervene in a very specific aspect of e-evidence exchange procedures. At the same time, however, it might also represent a missed opportunity for enforcing the right of the defence to challenge algorithmic evidence obtained abroad.

IV. CONCLUSION

The use of new technologies in our everyday life has created new forms of evidence that derive from computer systems[102]. At the same time, the shift to automation in forensic analyses, previously conducted by humans with a higher cost/savings effort, has also radically changed the nature of such evidence.[103] In the light of the pervasiveness and the efficiency of technology, algorithmic evidence will be the main type of criminal evidence in the very near future.

In criminal proceedings, defendants must have the possibility to challenge evidence against them.[104] Specifically, the right to confrontation requires that defendants should be in a position to test its probity, credibility, truthfulness and reliability. This principle, a cornerstone of the right to a fair trial, applies also to algorithmic evidence. However, the opacity surrounding algorithms is an obstacle to this test, when the defendant lacks access to information and/or resources to understand how the system works.[105] This is particularly worrisome considering not only the growing reliance on this type of evidence, but also the general presumption of reliability of algorithmic outputs in courtrooms.

Algorithmic opacity has been studied from different points of view in the literature.[106] However, the challenges arising in the specific context of criminal proceedings require more than transparency or explanation, namely a set of procedural rules. A legal framework is also necessary to balance the rights of the defence with other legitimate purposes of the criminal process, including the reasonable time of the proceeding.

[102] See Section II.1. on computer-derived evidence.

[103] See Section II.1. on computer-generated evidence.

[104] See Section II.3. on the right to confrontation in Article 6§3(d) ECHR.

[105] See Section II.2. on algorithmic opacity.

[106] See Section II.4. for a short literature overview on transparency and explanation.

At the EU level, an example for a legal response to the issues of algorithmic opacity is the Data Protection Legislation.[107] However, in the LED, that applies to personal data processed in criminal proceedings, the provisions regarding automated decision-making (Article 11 LED) and rights to information (Article 13 and 14 LED) cannot be considered adequate for the challenges. This conclusion is not surprising considering that the primary objective of the Directive is the protection of privacy of data subjects involved in criminal proceedings, rather than the protection of the rights of the defence.

The issues arising from the use of opaque algorithmic evidence need to be addressed through the lens of criminal procedural law. Even though this field has always been considered as a national prerogative, the role of the EU legislator should not be disregarded. Due to the cross-border nature of digital evidence and the freedom of movements of individuals, the EU level can be considered as best suited for granting a uniform and effective protection of the rights of the defence in cross-border proceedings.[108]

As far as algorithmic evidence is concerned, specific rules for the collection of evidence – setting standards, shared definitions and common procedures – need to be included for safeguarding the rights of the defence. The recent proposals by the Commission on the European Production and Preservation Orders for electronic evidence, however, have not met this expectation. In the absence of common standards regarding the gathering of data by service providers, the defence can only 'have faith' in the quality of the evidence collected abroad. Trusting the reliability of evidence, without having possibility to effectively test its quality, collides with the essence of the right to confrontation.

Conclusively, notwithstanding the growing attention on the ethical and legal implications of the use of new technologies in law enforcement, the role of the defence is dangerously overlooked. The right to confrontation requires defendants (and their attorneys) to monitor and test the quality of the evidence produced against them. In this sense, the defence plays a crucial role in algorithmic monitoring and, to some extent, contributes to the public oversight of these systems. The importance of detecting flaws and errors in algorithms is beneficial not only for the single case, but it has a broader impact when it

[107] See Section III.1. on the inadequacy of Data Protection Legislation.

[108] See Section III.2. on the evidence gathering in transnational proceedings.

ultimately contributes to improving the quality of these tools and, consequently, enhance trust in their use.

References

[1] Ananny M and Crawford K, ‘Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability’ (2018) 20 New Media & Society 973

[2] Arslan M, ‘The Right to Examination of Prosecution Witnesses’ (2018) 6 ZIS 218

[3] Biasiotti MA and others (eds), *Handling and Exchanging Electronic Evidence Across Europe* (Springer International Publishing 2018)

[4] Bratus S, Lembree A and Shubina A, ‘Software on the Witness Stand: What Should It Take for Us to Trust It?’ in Alessandro Acquisti, Sean W Smith and Ahmad-Reza Sadeghi (eds), *Trust and Trustworthy Computing* (Springer Berlin Heidelberg 2010)

[5] Brennan T, Dieterich W and Ehret B, ‘Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System’ (2009) 36 Criminal Justice and Behavior 21

[6] Brkan M, ‘Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond’ (2019) 27 International Journal of Law and Information Technology 91

[7] Burrell J, ‘How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms’ (2016) 3 Big Data & Society <https://doi.org/10.1177/2053951715622512> accessed 25 October 2018

[8] Celentino JC, ‘Face-to-Face With Facial Recognition Evidence: Admissibility Under the Post-Crawford Confrontation Clause’ (2016) 114 Michigan Law Review 1317

[9] Chessman C, ‘A Source of Error: Computer Code, Criminal Defendants, and the Constitution’ (2017) 105 Cal. L. Rev. 179

[10] Chung H, Park J and Lee S, ‘Digital Forensic Approaches for Amazon Alexa Ecosystem’ (2017) 22 Digital Investigation S15

[11] Citron DK and Pasquale F, ‘The Scored Society: Due Process for Automated Predictions Essay’ (2014) 89 Washington Law Review 1

[12] Cormen TH and others, *Introduction to Algorithms* (3rd edn, The MIT Press 2009)

[13] de Laat PB, ‘Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?’ [2017] Philosophy & Technology

[14] Diakopoulos N, ‘Algorithmic Accountability Reporting: On the Investigation of Black Boxes’ [2014] Tow Center for Digital Journalism, Columbia University <https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2> accessed 5 March 2020

[15] Dilek S, Çakır H and Aydın M, ‘Applications of Artificial Intelligence Techniques to Combating Cyber Crimes: A Review’ (2015) 6 International Journal of Artificial Intelligence & Applications 21

[16] Doshi-Velez F and others, ‘Accountability of AI Under the Law: The Role of Explanation’ [2017] arXiv:1711.01134 [cs, stat] <http://arxiv.org/abs/1711.01134> accessed 23 December 2019

[17] Edwards L and Veale M, ‘Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?’ (2018) 16 IEEE Security & Privacy 46

[18] Eubanks V, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (First Edition, St Martin's Press 2017)

[19] Freeman K, 'Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis' (2016) 18 North Carolina Journal of Law and Technology 75

[20] Goodman B and Flaxman S, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"' (2017) 38 AI Magazine 50

[21] Green B, 'Fair Risk Assessments: A Precarious Approach for Criminal Justice Reform' (2018)

[22] Gutheil M and Liger Q, 'Legal Frameworks for Hacking by Law Enforcement: Identification, Evaluation and Comparison of Practices' 142

[23] Harris DJ and others, *Law of the European Convention on Human Rights* (3rd edn, Oxford University Press 2014)

[24] Hayes P, van de Poel I and Steen M, 'Algorithms and Values in Justice and Security' [2020] AI & SOCIETY <http://link.springer.com/10.1007/s00146-019-00932-9> accessed 13 January 2020

[25] Hepenstal S and others, 'Algorithmic Transparency of Conversational Agents', *IUI Workshops* (2019)

[26] Horsman G, 'Tool Testing and Reliability Issues in the Field of Digital Forensics' (2019) 28 Digital Investigation 163

[27] Jenkins MM, 'Computer-Generated Evidence Specially Prepared for Use at Trial Law and Technology Symposium' (1975) 52 Chicage-Kent Law Review 600

[28] Kim TW and Routledge BR, 'Informational Privacy, A Right to Explanation, and Interpretable AI', *2018 IEEE Symposium on Privacy-Aware Computing (PAC)* (IEEE 2018) <https://ieeexplore.ieee.org/document/8511831/> accessed 5 March 2020

[29] Kroll JA and others, 'Accountable Algorithms' (2016) 165 U. Pa. L. Rev. 633

[30] Lauret J, 'Amazon's Sexist AI Recruiting Tool: How Did It Go so Wrong?' *Medium* (16 August 2019) <https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e> accessed 29 February 2020

[31] Lepri B and others, 'Fair, Transparent, and Accountable Algorithmic Decision-Making Processes' (2018) 31 Philosophy & Technology 611

[32] Liebow D, 'DWI Source Code Motions after Underdahl' (2010) 11 Minn. JL Sci. & Tech. 853

[33] Lipton ZC, 'The Mythos of Model Interpretability' [2017] arXiv:1606.03490 [cs, stat] <http://arxiv.org/abs/1606.03490> accessed 11 December 2019

[34] Malgieri G, 'Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations' (2019) 35 Computer Law & Security Review <http://www.sciencedirect.com/science/article/pii/S0267364918303753> accessed 13 January 2020

[35] Malgieri G and Comandé G, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 243

[36] Mifsud B, Melania T and Joseph C, 'The European Legal Framework on Electronic Evidence: Complex and in Need of Reform' in Maria Angela Biasiotti and others (eds), *Handling and Exchanging Electronic Evidence Across Europe* (Springer International Publishing 2018) <https://www.springer.com/gp/book/9783319748719> accessed 28 February 2020

[37] Mitsilegas V, 'The Privatisation of Mutual Trust in Europe's Area of Criminal Justice: The Case of e-Evidence' (2018) 25 Maastricht Journal of European and Comparative Law 263

[38] Mittelstadt B, Russell C and Wachter S, 'Explaining Explanations in AI', *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19* (ACM Press 2019)

[39] Mittelstadt BD and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society <http://journals.sagepub.com/doi/10.1177/2053951716679679> accessed 21 December 2019

[40] Neville K, 'Programmers and Forensic Analyses: Accusers Under the Confrontation Clause' (2011) 10 Duke Law & Technology Review 1

[41] O'Neil C, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (First edition, Crown 2016)

[42] Osman L, 'Key Abdi Video Unreliable, Expert Tells Officer's Manslaughter Trial | CBC News' *CBC* (20 June 2019) <https://www.cbc.ca/news/canada/ottawa/video-expert-montsion-1.5181755> accessed 29 February 2020

[43] Oswald M and others, 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and "Experimental" Proportionality' (2018) 27 Information & Communications Technology Law 223

[44] Palmiotto F, 'The Impact of Algorithmic Opacity on Fair Trial Rights in Criminal Proceedings' in Martin Ebers and Marta Cantero Gamito (eds), *Algorithmic Governance and Governance of Algorithms* (Springer Forthcoming)

[45] ——, 'Algorithmic Opacity as a Challenge to the Rights of the Defense' (*RAILS*, 6 September 2019) <https://ai-laws.org/en/2019/09/algorithmic-opacity-challenge-to-rights-of-the-defense/> accessed 1 March 2020

[46] Pasquale F, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015)

[47] Rosenblat A, *Uberland: How Algorithms Are Rewriting the Rules of Work* (University of California Press 2018)

[48] Roth A, 'Machine Testimony' (2017) 126 Yale Law Journal 1972

[49] Sajfert J and Quintel T, 'Data Protection Directive (EU) 2016/680 for Police and Criminal Justice Authorities', *Cole/Boehm GDPR Commentary* (Edward Elgar Publishing, Forthcoming) <https://papers.ssrn.com/abstract=3285873> accessed 23 December 2019

[50] Sauer G, 'A Murder Case Tests Alexa's Devotion to Your Privacy' *Wired* (28 February 2017) <https://www.wired.com/2017/02/murder-case-tests-alexas-devotion-privacy/> accessed 19 March 2019

[51] Security without Borders, 'Exodus: New Android Spyware Made in Italy' (March 2019) <https://securitywithoutborders.org/blog/2019/03/29/exodus.html> accessed 1 April 2019

[52] Selbst AD and Powles J, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233

[53] Singer N and Metz C, 'Many Facial-Recognition Systems Are Biased, Says U.S. Study' *The New York Times* (19 December 2019) <https://www.nytimes.com/2019/12/19/technology/facial-recognition-bias.html> accessed 5 March 2020

[54] Sites B, 'Rise of the Machines: Machine-Generated Data and the Confrontation Clause' (2014) 16 The Columbia Science & Technology Law Review 36

[55] Smuha NA, 'Towards the EU Harmonization of Access to Cross-Border E-Evidence: Challenges for Fundamental Rights & Consistency' <https://papers.ssrn.com/abstract=3501421> accessed 1 March 2020

[56] Stefan M and González Fuster G, 'Cross-Border Access to Electronic Data Through Judicial Cooperation in Criminal Matters' (Social Science Research Network 2018) SSRN Scholarly Paper ID 3298705 <https://papers.ssrn.com/abstract=3298705> accessed 5 March 2020

[57] Tosza S, 'The European Commission's Proposal on Cross-Border Access to E-Evidence : Overview and Critical Remarks' [2018] eucrim - The European Criminal Law Associations' Forum <https://eucrim.eu/articles/european-commissions-proposal-cross-border-access-e-evidence/> accessed 2 March 2020

[58] ——, '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' (Social Science Research Network 2019) SSRN Scholarly Paper ID 3517866 <https://papers.ssrn.com/abstract=3517866> accessed 1 March 2020

[59] Van Buskirk E and Liu VT, 'Digital Evidence: Challenging the Presumption of Reliability' (2006) 1 Journal of Digital Forensic Practice 19

[60] Veale M, 'Algorithm Use in the Criminal Justice System Report' (The Law Society of England and Wales 2019)

[61] Vowles E and Report JSC for TL, 'Your Smart Device Could Be Used against You in Court' (*ABC News*, 9 March 2018) <https://www.abc.net.au/news/2018-03-09/your-google-home-or-fit-bit-could-be-used-against-you-in-court/9510368> accessed 19 March 2019

[62] Wachter S, Mittelstadt B and Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76

[63] Winter LB, 'Transnational Criminal Proceedings, Witness Evidence and Confrontation: Lessons from the ECtHR's Case Law' (2013) 9 Utrecht L. Rev. 127

[64] ——, 'Transnational Evidence: Towards the Transposition of Directive 2014/41 Regarding the European Investigation Order in Criminal Matters' [2015] Eucrim: the European Criminal Law Associations' fórum 47