10

Artificial Intelligence in Criminal Justice: Strengthening or Challenging the Rule of Law?

EMMANOUIL BILLIS*

I. INTRODUCTION

In the 'global risk society',[1] crime is becoming more sophisticated, complex and transnational, while enforcement and judicial systems are becoming ponderous and overloaded. As a result, the practical significance of mechanisms and institutions aimed at enhancing (national and transnational) law enforcement and improving justice administration has grown. A key aspect in this regard is the revolutionary importance of artificial intelligence (AI) for many policy sectors. Numerous legal orders are currently resorting to this technology with the goal of strengthening the efficiency and effectiveness of crime control and criminal justice systems and of optimising the decision-making processes. In an era of multiple novel challenges in the fight against crime, a plethora of AI applications has emerged in parallel with traditional enforcement and judicial practices set to serve a variety of purposes: from predictive policing, crime prevention and crime detection to risk and recidivism assessment, the processing of evidence and the determination of criminal punishment.[2]

* Dr Emmanouil Billis, LLM is Research Group Leader at the Max Planck Institute for the Study of Crime, Security and Law (Freiburg/Germany).

[1] See U Sieber, 'The New Architecture of Security Law – Crime Control in the Global Risk Society' in U Sieber, V Mitsilegas, C Mylonopoulos, E Billis and N Knust (eds), *Alternative Systems of Crime Control. National, Transnational, and International Dimensions* (Berlin, Duncker & Humblot, 2018) 3ff.

[2] See, eg, in this volume the contributions by G Petri, Y Razmetaeva, K Ligeti, M Caianiello, E Bampasika, C Salvi, R Fortson and T Quintel and D Cole. See also E Billis, N Knust and JP Rui, 'Künstliche Intelligenz und der Grundsatz der Verhältnismäßigkeit' in M Engelhart, H Kudlich and B Vogel (eds), *Digitalisierung, Globalisierung und Risikoprävention – Festschrift für Ulrich Sieber zum 70. Geburtstag, Teilband II* (Berlin, Duncker & Humblot, 2021) 693, 698ff. More generally on the various possible uses of AI, *inter alia*, Organisation for Economic Cooperation and Development (OECD), 'Artificial Intelligence in Society' (Paris, OECD Publishing, 2019) 47–80; F Pereira,

As the studies in this volume so clearly demonstrate, fundamental research and legal policy are called to address not only the opportunities but also the considerable risks for the peaceful coexistence of humans that come with such an evolution mainly in two ways. On the one hand, compared to prior (conventional) technological advancements, employing new AI technology to realise ambitious anti-crime plans might result in broader, more direct and multi-layered threats to established rights and freedoms. In view of this, extensive *a priori* bans of AI uses identified as particularly dangerous for individuals or societies may be deemed necessary. On the other (hand), the debate over the 'inevitability' of the expansion of AI should not only be about getting the most out of this technology in terms of effective crime fighting. The first priority must be to develop the algorithms ('step-by-step procedure for calculation, data processing, evaluation and automated reasoning and decision-making')[3] and to program the machines in accordance with the overriding objectives of substantially protecting and securing respect for the most basic human and social values. In this context, research and policy are called to enable continuous cooperation and knowledge exchange between legal scholars, justice authorities and computer scientists from the AI field. The corresponding objectives of such a collaboration could be to promote mutual understanding about the actual operation of new technologies in the legal world and to develop strategies on how to successfully and dynamically 'translate' diachronic legal notions and protective principles into programming language.[4] Exploring particularly sensitive matters connected to broader socio-legal and legal-ethical questions about justice, legitimacy and democracy should be at the heart of any such research. In this context, issues of privacy, data protection, security, reliability, transparency and objectivity, bias and discrimination as well as the explicability and accountability parameters of the AI applications constitute significant individual subjects of discourse.[5]

P Machado, E Costa and A Cardoso (eds), *Progress in Artificial Intelligence* (Cham, Springer, 2015); M Kment and S Borchert, *Künstliche Intelligenz und Algorithmen in der Rechtsanwendung* (Munich, Beck, 2022); J Wagner, *Legal Tech und Legal Robots. Der Wandel im Rechtsmarkt durch neue Technologien und künstliche Intelligenz* (Wiesbaden, Springer, 2018).

[3] On this description of an algorithm, see the European Union Agency for Fundamental Rights (FRA) and Council of Europe, 'Handbook on European Data Protection Law' (April 2018) 351.

[4] On the problem of 'the programmability of law (i. e. whether it can be transposed in computer code instructions which a machine can follow and execute)' and 'the algorithmization of the law', see E Hilgendorf, 'Introduction: Digitization and the Law – a European Perspective' in E Hilgendorf and J Feldle (eds), *Digitization and the Law* (Baden-Baden, Nomos, 2018) 13ff. See further the studies in S Deakin and C Markou (eds), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence* (Oxford, Hart Publishing, 2020).

[5] See in general the Council of Europe, European Commission for the Efficiency of Justice (CEPEJ), 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment', adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3–4 December 2018); OECD, 'Recommendation of the Council on Artificial Intelligence' (22 May 2019, OECD/LEGAL/0449). See further Artificial Intelligence in Society (n 2) 81–120. According to the EU Commission, High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (8 April 2019) 2: '(…) Trustworthy AI has three components, which should be met throughout the system's entire life cycle: (1) it should be lawful, complying with all applicable laws and regulations (2) it should be

Equally important is the general discussion about whether algorithms and machines can perceive and employ fundamental legal concepts for the delivery of criminal justice, such as fairness and proportionality, and about the challenges of incorporating AI into criminal justice in a way that is consistent with core human rights standards and rule of law principles.

The design and implementation of effective and transparent policy plans and regulations in these fields require, on the one hand, a better understanding of the inner workings of the AI applications on the part of legal theory and practice. On the other (hand), AI developers need to have a firm grasp of the main ideas behind key legal concepts and their potential differentiations in different legal traditions. The joint study of the many (attempts at) definitions of AI, especially of their scope, practical significance and limitations (for example, regarding the controversial issue of autonomous or automated AI) is just one essential element.[6] In liberal democratic orders employing AI technology, the primary focus must be on the constitutive requirements of the rule of law: the principle of legality, including the consistent and impartial application of foreseeable, clear and transparent norms and institutions; the principles of equality and proportionality; the nonarbitrary use of power and the respect for fundamental rights and procedural guarantees, and the separation of state powers and the control of their exercise by independent and impartial judicial organs.[7]

ethical, ensuring adherence to ethical principles and values and (3) it should be robust, both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm. (…) [It must be ensured] that the development, deployment and use of AI systems meets the seven key requirements for Trustworthy AI: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability.' See further the Council of Europe, Committee of Experts on Internet Intermediaries (MSI-NET), 'Algorithms and Human Rights. Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications' DGI(2017)12. On important ethical considerations in the creation of AI systems, see The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 'Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems' (2019). See also T Bynum, 'Computer and Information Ethics' in E Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), available at https://plato.stanford.edu/archives/sum2018/entries/ethics-computer/.

[6] See, eg, the definitions of AI systems and techniques in: Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Art 3; Ethics Guidelines for Trustworthy AI (n 5) 36; EU Commission, High-Level Expert Group on Artificial Intelligence, 'A Definition of AI: Main Capabilities and Scientific Disciplines' (8 April 2019) 3, 7; Artificial Intelligence in Society (n 2) 15; European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment (n 5) 69–70; Council of Europe, Commissioner for Human Rights, 'Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights', Recommendation, May 2019, 5; Communication from the Commission, 'Artificial Intelligence for Europe', 25.4.2018, COM(2018) 237 final, 1; Commission Staff Working Document, 'Liability for Emerging Digital Technologies', 25.4.2018, SWD(2018) 137 final; Communication from the Commission, 'Coordinated Plan on Artificial Intelligence', 7.12.2018, COM(2018) 795 final, 1.

[7] See Communication from the Commission to the European Parliament, the European Council and the Council, 'Further Strengthening the Rule of Law within the Union', 3.4.2019, COM(2019) 163 final, 1: 'The rule of law is one of the founding values of the European Union, as well as a reflec-

In terms of the relationship between AI and the rule of law specifically, the challenge is twofold: to proactively program AI tools in a way that excludes arbitrariness in decision-making processes involving such tools and to optimise the operation and learning processes of AI with the overall purpose of complementing the traditional justice sector in producing more accurate, objective and fair results. The present chapter discusses characteristic questions and problems of contemporary importance for legal theory, policy and practice associated with the key notions of human dignity, legality, proportionality, privacy, equality and procedural justice. It focuses on the meaning and significance of rule of law and human rights considerations in designing and employing AI tools for crime control and criminal justice purposes.

II. HUMAN DIGNITY, HUMAN RIGHTS AND THE RULE OF LAW: THE DEMAND FOR HUMAN-CENTRIC AND LAWFUL AI

> (…) The notion of the Rule of Law requires a system of certain and foreseeable law, where everyone has the right to be treated by all decision-makers with dignity, equality and rationality and in accordance with the laws, and to have the opportunity to challenge decisions before independent and impartial courts through fair procedures. (…) The Rule of Law and human rights are interlinked (…) The Rule of Law would just be an empty shell without permitting access to human rights. Vice-versa, the protection and promotion of human rights are realised only through respect for the Rule of Law: a strong regime of Rule of Law is vital to the protection of human rights. In addition, the Rule of Law and several human rights (such as fair trial and freedom of expression) overlap. (…) The Rule of Law is linked not only to human rights but

tion of our common identity and common constitutional traditions. It is the basis of the democratic system in all Member States, necessary to ensure the protection of fundamental rights. (…) The rule of law ensures that Member States and their citizens can work together in a spirit of mutual trust; trust in public institutions, including in the justice system, is crucial for the smooth functioning of democratic societies. The rule of law is also one of the principles guiding the EU's external action. (…) The rule of law is enshrined in Article 2 of the Treaty on European Union as one of the founding values of the Union. Under the rule of law, all public powers always act within the constraints set out by law, in accordance with the values of democracy and fundamental rights, and under the control of independent and impartial courts. The rule of law includes, among others, principles such as legality, implying a transparent, accountable, democratic and pluralistic process for enacting laws; legal certainty; prohibiting the arbitrary exercise of executive power; effective judicial protection by independent and impartial courts, effective judicial review including respect for fundamental rights; separation of powers; and equality before the law. These principles have been recognised by the European Court of Justice and the European Court of Human Rights.' See also Communication from the Commission to the European Parliament and the Council, 'A New EU Framework to Strengthen the Rule of Law', 11.3.2014, COM(2014) 158 final, 4: 'The precise content of the principles and standards stemming from the rule of law may vary at national level, depending on each Member State's constitutional system.' See, further, Rule of Law Checklist, adopted by the Council of Europe Venice Commission at its 106th Plenary Session (Venice, 11–12 March 2016) paras 9ff, 15–18, 31ff; A Perego, 'The European Commission and the EU Rule of Law Policy' in AB Engelbrekt, A Moberg and J Nergelius (eds), *Rule of Law in the EU* (Oxford, Hart Publishing, 2021) 291. See also Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Council of Europe Treaty Series – No [225], 5.9.2024, Art 5.

> also to democracy. (…) Democracy relates to the involvement of the people in the decision-making process in a society; human rights seek to protect individuals from arbitrary and excessive interferences with their freedoms and liberties and to secure human dignity; the Rule of Law focuses on limiting and independently reviewing the exercise of public powers. The Rule of Law promotes democracy by establishing accountability of those wielding public power and by safeguarding human rights, which protect minorities against arbitrary majority rules.[8]

Liberal legal orders are principally bound by the same (sometimes conflicting) obligations: to effectively prevent and suppress unlawful conduct, to aim at truthful and fair dispute resolutions, to preserve procedural economy and social peace as well as to respect the basic principles of human dignity, proportionality and leniency. At its creation, the European Union (EU) system was conceived on the basis of the ideas and fundamental values of human dignity, freedom, democracy, equality, the rule of law and human rights.[9] The notions of human dignity, human rights and the rule of law are interconnected. Mindful of this, it is first and foremost human dignity, 'the real basis of fundamental rights',[10] as enshrined in Article 1 of the EU Charter of Fundamental Rights (EU Charter) and in various national Constitutions, along with associated specific dignity rights (right to life, right to physical and mental integrity, prohibition of torture and of inhuman or degrading treatment, prohibition of slavery and forced labour), that position the individual at the centre of the states' and the EU's actions.

To acknowledge dignity as inviolable and absolute means to protect human beings from being disregarded or arbitrarily treated as mere objects by the state.[11] How to express this normative limit as an algorithmic one can be a challenge. The negative duty of rule of law systems refers not only to singling out and banning the use of applications posing direct threats to human dignity but also to programming algorithms applied in enforcement and judicial proceedings in a manner that secures respect for it. With the respect, protection and advancement of core human values being the overriding objectives of any technological evolution comes also the positive duty to direct machines towards safeguarding the individual against arbitrarily and cruelly coercive, non-lenient or disproportional actions of state organs.

Thus the necessity for a human-centric approach regarding all issues related to the operation and use of AI follows from the fundamental values of human dignity, autonomy, freedom and the rule of law. The European Commission's

[8] Rule of Law Checklist, ibid 15, 31, 33.

[9] Treaty on European Union (TEU) Preamble and Art 2.

[10] EU Charter explanations.

[11] See on this, eg, the findings of the German Constitutional Court in the cases BVerfGE 27, 1 (6) [1969] (*Mikrozensus*) and BVerfGE 30, 1 (25-26) [1970] (*Abhörurteil*); for an overview regarding the different conceptions of human dignity in German literature and jurisprudence, see R Poscher, '§ 17 Menschenwürde' in M Herdegen et al (eds), *Handbuch des Verfassungsrechts* (Munich, Beck, 2021) at 55–63 and 79–104.

Ethics Guidelines for Trustworthy AI of April 2019, stating that 'trustworthy AI' should be lawful, ethical and robust, already refer to this approach.[12] Furthermore, the AI Regulation (AI Act), which lays down harmonised rules for the EU, accepts the need to continue developing and evolving AI technology in the modern world. At the same time, it also adopts the human-first approach, *inter alia* by requiring extra checks and guarantees for specific uses of AI, especially applications with the highest potential for harming people, including with respect to systems assisting decision-making and law enforcement systems. The Regulation also identifies specific AI uses as unacceptable and prohibited, such as in the case of AI systems that provide social scoring of natural persons for the purpose of evaluating or classifying their trustworthiness. Further categorisations include applications considered 'high risk' and rather dangerous, intrusive and problematic, *inter alia* due to concerns of discrimination, bias or lack of transparency. Pertinent examples are facial recognition in public places and predictive policing systems. In these cases, the AI Act considers it necessary to restrict the use of AI based on proportionality assessments.[13]

The European Parliament's resolution on AI in criminal law of October 2021 is pointing in a similar direction.[14] According to the European Parliament, AI systems deployed for law enforcement and criminal justice purposes need to fully respect the principles of human dignity and non-discrimination, the privacy rights, the freedom of movement, the presumption of innocence and the rights of the defence (right to silence, freedom of expression and information, equality before the law, equality of arms, rights to an effective remedy and a fair trial) in accordance with the EU Charter and the European Convention on Human Rights (ECHR). The Parliament called for

[12] Ethics Guidelines for Trustworthy AI (n 5) 37: 'The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union, all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoy a unique and inalienable moral status. This also entails consideration of the natural environment and of other living beings that are part of the human ecosystem, as well as a sustainable approach enabling the flourishing of future generations to come.' See also Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 'Building Trust in Human-Centric Artificial Intelligence', 8.4.2019, COM(2019) 168 final.

[13] Artificial Intelligence Act (n 6) Arts 5ff.

[14] European Parliament's resolution on 'artificial intelligence (AI) in criminal law and its use by the police and judicial authorities in criminal matters' of 6 October 2021 (2020/2016(INI)). With respect, for example, to the scoring or persons, the resolution notes at 32: 'Supports the recommendations (…) for a ban on AI-enabled mass scale scoring of individuals; considers that any form of normative citizen scoring on a large scale by public authorities, in particular within the field of law enforcement and the judiciary, leads to the loss of autonomy, endangers the principle of non-discrimination and cannot be considered in line with fundamental rights, in particular human dignity, as codified in EU law.'

> algorithmic explainability, transparency, traceability and verification as a necessary part of oversight, in order to ensure that the development, deployment and use of AI systems for the judiciary and law enforcement comply with fundamental rights, and are trusted by citizens, as well as in order to ensure that results generated by AI algorithms can be rendered intelligible to users and to those subject to these systems, and that there is transparency on the source data and how the system arrived at a certain conclusion.[15]

The Parliament also noted the importance of human intervention with respect to all law enforcement applications of AI and stated that 'the decision giving legal or similar effect always needs to be taken by a human, who can be held accountable for the decisions made'.[16]

In this light, the human-centric approach should not be seen as a further obstacle to technological progress and especially not when technology is employed to enhance the protection of human dignity, human rights and the rule of law. However, acceptance and advancement of technological innovation do not exclude the possibility for the *a priori* imposition of oversight and intervention mechanisms, strict operating requirements, broad use restrictions and – if necessary to effectively secure dignity, autonomy and freedom – total bans on particular AI applications. At the same time, some of the protective requirements, rule of law standards and human rights guarantees prescribed in the aforementioned texts and in other contemporary soft-law and hard-law instruments may not be so easy to implement and comply with in practice, at least in the present state of technological and legal evolution. Training data sets free of errors or the risk of bias; fully explainable and transparent machine learning systems irrespective of their internal complexity or the external transparency barriers raised by intellectual property interests; objective and non-discriminating automated decisions and outputs; functioning models of accountability and responsibility: these are just some of the still unresolved issues to be urgently addressed in policy debates and interdisciplinary research.

Furthermore, the core of the human dignity concept may be subject to direct violations inflicted by (the use of) AI in the context of law enforcement and criminal justice. The normative (constitutional) importance of dignity as inviolable and absolute does not leave much room in such cases for the balancing of competing interests; nor can it be truly satisfied by the fragmentary adoption of partial restrictions and formal control measures. Hence, a more holistic approach to prohibiting certain types of algorithmic systems may be necessary, even in less obvious situations. A case in point is a voluntary social credit system that, ultimately, may also be used, directly or indirectly, for surveillance and law enforcement purposes: 'voluntariness' does not automatically equal or

[15] ibid at 17.
[16] ibid at 16.

imply respect for human dignity, autonomy and personal self-determination.[17] Similarly regarding coercive tools, which, at least for the time being, seem to have a restricted (supporting) role and more of a symbolic significance. In an otherwise human-centric world, most characteristic is the example of robots 'mimicking' human law enforcers. Judging by specific policing and investigative measures some countries imposed during the Covid pandemic, the dystopic scenario of being ordered and stopped, questioned, searched, arrested or fined with an administrative or criminal sanction by autonomous animal-shaped robots is not too distant after all.[18]

III. PROPORTIONALITY

Further drawing on control strategies and measures adopted by governments and international organisations during the outbreak of the Covid pandemic as well as on alternative and technologically advanced security,[19] prevention and enforcement practices applied in response to current global threats (international and eco terrorism, terrorist financing, cyberterrorism and other wide-scale digital attacks, novel forms of transnational organised crime and money laundering, etc.): It has become evident that the modern tools used in the war on the various kinds of 'visible' and 'invisible' enemies no longer target only traditional suspects or criminals but also more recent categories of security risks, pre-suspects and potential criminals. Enabled by the current technological progress, these tools exhibit an extensive range of intrusive and coercive characteristics potentially affecting the rights of much larger sections of populations compared to the conventional criminal justice apparatus.

[17] *cf* on Mantello's 'Ikeaveillance' and on the Chinese social credit, R Vogler, 'Big Data and Criminal Justice. Proportionality, Efficiency and Risk in a Global Context' in E Billis, N Knust and JP Rui (eds), *Proportionality in Crime Control and Criminal Justice* (Oxford, Hart Publishing, 2021) 165, 174ff.

[18] *cf* R McMorrow and G Li, 'The Robot Dogs Policing Shanghai's Strict Lockdown' *Financial Times* (14 April 2022), available at www.ft.com/content/5c437146-2d18-466b-84af-24a47b32de59. Problematic in terms of securing respect for human dignity and the rule of law may also be the 'reverse' situation where an interaction takes place between an individual and an AI system, and the individual may be led to believe that it is human or may not be able to realise that the system is non-human. To ensure *inter alia* the transparency of AI systems, the EU Commission High-Level Expert Group on Artificial Intelligence in its Policy and Investment Recommendations for Trustworthy AI of 26 June 2019 (at 12) suggested introducing the mandatory self-identification of AI systems and noted that, if there is such a likelihood, deployers of AI systems should be attributed a general responsibility to disclose that the system is actually non-human. However, especially in the context of crime control and law enforcement, where the interference with personal freedoms and rights takes its most severe form, a total ban of AI systems enabling or operating on the basis of such misleading impressions should also be considered; see, however, Artificial Intelligence Act (n 6) Art 52. On other interesting aspects of the relationship between human dignity and human-like machines and systems (from a constitutional law perspective), see C Geminn, 'Menschenwürde und menschenähnliche Maschinen und Systeme' (2020) 24 *Die Öffentliche Verwaltung* 172ff.

[19] On the 'new security architecture', see Sieber (n 1) 3–34.

Mass surveillance and the (automated) collection and analysis of personal data for the principally legitimate reasons of security and public safety have become a worldwide (national and transnational) phenomenon. The extended application of highly evolved electronic and AI technologies for the bulk registration, analytic observation and/or tracking of actual or potential hazards (and 'dangerous' or 'non-compliant' individuals) may indeed lead to broad interferences with privacy and other rights for unspecified numbers of individuals who may not even be remotely linked to criminal behaviour of any kind. Nowadays, even the most intrusive temporary and emergency legislations and the various exceptional administrative and enforcement measures restricting basic freedoms and personal rights appear to be easily justifiable as *sine qua non* for fulfilling the purpose of keeping populations safe and secure. In light of all the global, grave and imminent risks to the principal legal interests of health and life, current legal-policy agendas tend to focus more on designing the most effective security mechanisms than on the rule of law and human rights concerns surrounding their use. Further, the potential consequences in the fields of crime prevention and repression are sometimes overlooked: Exceptional monitoring measures may eventually turn into permanent and continuous (real-time) state supervision.[20] And the normalisation of pre-emptive but still massively invasive restrictions of liberty, movement and privacy rights for public security purposes may substantially transform the traditional objectives, operational methods and, most importantly, the protective limits and principles of current and future systems of crime control and criminal justice.[21]

It is therefore a positive development that – even if the absolute core of human dignity is not directly violated and the prohibition of a certain practice is not deemed *a priori* necessary – the aforementioned regulatory and soft-law instruments prescribe minimum rule of law and human rights guarantees concerning the operation of AI within the last resort sectors of law enforcement, crime control and criminal justice. It is in the framework of these particular public law branches where the most effective measures, now enhanced by current technologies, will usually turn out to be also the most intrusive and coercive, capable of widely endangering and restricting the full exercise of rights and personal freedoms. Even more so in cases where no human actor/decision-maker who can intuitively comprehend the meaning and importance of dignity and human rights is actively involved in applying the measure. In this respect, an immanent rule of law notion of crucial importance to criminal and security law in terms of its protective function against arbitrary interferences with individual rights and vested freedoms is that of proportionality.[22]

[20] *cf* YN Harari, 'The World After Coronavirus' *Financial Times* (20 March 2020), available at www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75.

[21] For further analysis with examples and references, see E Billis, N Knust and JP Rui, 'The Typology of Proportionality' in Billis, Knust and Rui (n 17) 3, 5–11.

[22] *cf* Ethics Guidelines for Trustworthy AI (n 5) 12–13; Artificial Intelligence Act (n 6) 7, 11, 21ff.; European Parliament's resolution on AI (n 14).

Proportionality shapes and limits the exercise of all state powers. It is a legal reasoning technique, a temperate method of controlling public authority and, at the same time, a rational factor for enhancing the social acceptance of actually necessary coercive and rights intrusive measures – one that can also contribute towards securing the functionality of legal systems in the long term.[23] Within the framework of proportionality assessments, the competing interests are weighted at the three different levels of public power, affecting measures and decisions of all three branches: the legislative, the executive and the judicial. At the core of such assessments is the examination of the following generally recognised elements: the existence of a legitimate aim justifying the adoption of a certain measure; the suitability of the measure for reaching this concrete aim; the necessity of this particular measure in view of the aim (requirement of the least intrusive measure); and the appropriateness of the measure in terms of its (rights-limiting) effects balanced against the benefits of the aim pursued (proportionality *stricto sensu*).[24]

In the broad context of crime control and criminal justice, the various types of proportionality have been relevant, *inter alia*, with respect to the diachronic substantive law questions of criminalisation and punishment as well as in terms of policing and criminal procedure, particularly regarding the definition and review of the applicability conditions and limits for investigations, searches and other coercive measures (communication interceptions, arrest warrants, pre-trial detentions and bails, etc).[25] In line with the above-described developments, the significance of the proportionality concept has further increased with respect to the use of the new technologies of (bulk) surveillance and AI for security and

[23] Billis, Knust and Rui (n 21) 11ff. *cf* L Zedner, 'Ends and Means: Why Effective Counter-Terrorism Requires Respect for Proportionality and Rights' in Billis, Knust and Rui (n 17) 125ff.

[24] Billis, Knust and Rui (n 21) 24. This proportionality concept is incorporated in most liberal rule of law orders, national and international, see M Kremnitzer, T Steiner and A Lang (eds), *Proportionality in Action: Comparative and Empirical Perspectives on the Judicial Practice* (Cambridge, Cambridge University Press, 2020); M Bothe and E-C Gillard, 'The Proportionality Principle in Comparative Public, European Union and International Law – "Reflections on the Proportionality Equation"' in Billis, Knust and Rui (n 17) 277ff. Within the ECHR system, the decisions and judgments of the European Court of Human Rights (ECtHR) appear to accept the applicability of all four proportionality components, see, eg, J McBride, 'Proportionality and the European Convention on Human Rights' in E Ellis (ed), *The Principle of Proportionality in the Laws of Europe* (London, Hart Publishing, 1999) 23ff; A Barak, *Proportionality. Constitutional Rights and Their Limitations* (Cambridge, Cambridge University Press, 2012) 183–84; J Gerards, 'How to Improve the Necessity Test of the European Court of Human Rights?' (2013) 11 *International Journal of Constitutional Law* 466ff; E Brems and L Lavrysen, '"Don't Use a Sledgehammer to Crack a Nut": Less Restrictive Means in the Case Law of the European Court of Human Rights' (2015) 15 *Human Rights Law Review* 139ff. The EU proportionality test, as applied by the Court of Justice of the European Union, is focusing, at least for the time being, on the elements of suitability and necessity, see V Mitsilegas and E Billis, 'Article 49 – Principles of Legality and Proportionality of Criminal Offences and Penalties' in S Peers et al (eds), *The EU Charter of Fundamental Rights. A Commentary*, 2nd edn (Oxford, Hart Publishing, 2021) 1473, 1485ff, 1498ff with further references.

[25] *cf* RA Duff, 'Proportionality and the Criminal Law: Proportionality of What to What?' in Billis, Knust and Rui (n 17) 29ff.

crime-fighting purposes.[26] Again, the problem is twofold. First, regarding the proportionate employment of AI: Legislators must be able to define clear and non-arbitrary criteria and the administrative and judicial bodies must be able to exercise self-restraint and control for ensuring the legitimate, suitable, least intrusive and appropriate utilisation of AI in the various constellations. Second, regarding the proportionate output of AI and machine learning: The main challenge here is to develop the algorithms and 'train' the machines such that they can reach decisions and take actions in conformity with the basic elements of proportionality.

In the current state of technological progress, this is easier said than done – and it may prove to be more of a challenge than a support for the entire rule of law and proportionality endeavour. Not to disregard also the opinion that calculations and decisions about the proportionality or excessiveness of a certain action cannot really result from applying a mechanical formula but (must) remain 'a value judgement' to be made by the responsible human actor/operator 'in good faith and in a reasonable manner'.[27] The field of international humanitarian (and international criminal) law, specifically concerning the use of autonomous weapon systems (AWS) in armed conflict situations, provides a characteristic example.[28] AWS processors and the algorithms set to assess the information gathered for carrying out target selections and identifying incidental harm expected from a possible military strike are not susceptible to stress and other human emotions. Free from inherent emotional weaknesses and the risk of human error that has led to criminal incidents and unjustified war escalations in the past, AWS are therefore (or, at some point, will be) able to reach more objective and proportionate decisions. However, in contrast to the *de facto* complexity of armed conflicts and the unpredicted and often spontaneous or reactive nature of military conduct, the mathematic processes and proportionality assessments conducted by AWS are rather linear. They are mostly based on technical and prefixed parameters for the collection of information by using sensors and on the rigid interpretation and application of previously defined legal rules by processors and algorithms. What AWS may still lack is the human ability to understand and adapt to constantly changing realities in a variety of ways using learned inferences that enable individuals to perceive their own actions also from the perspective of others. If military decision-making involving the choice of suitable, least intrusive and appropriate means and targets requires contextual intelligence operating, where necessary, beyond predetermined algorithms, it remains questionable whether contemporary AWS are actually capable of 'deciding', 'acting' and 'reacting' in a temperate and proportionate manner.[29]

[26] Billis, Knust and Rui (n 2).

[27] Bothe and Gillard (n 24) 295.

[28] See the analysis in Billis, Knust and Rui (n 2) 715ff.

[29] See E Billis and N Knust, 'Proportionality (Principle of)' in P Caeiro et al (eds), *Elgar Encyclopedia of Crime and Criminal Justice* (Cheltenham, Edward Elgar Publishing, 2023) with references to R Geiss, *Die völkerrechtliche Dimension autonomer Waffensysteme* (Friedrich-Ebert-Stiftung,

IV. PRIVACY

Similar considerations generally apply to the rapid technological advancements enabling the automated collection, selection and 'real-time' processing of large volumes of personal data and the highly sophisticated electronic systems assigned to (assist in) the relevant decision-making processes. The ethical and pragmatic challenges in achieving proportionate AI outputs in the fields of security, crime control and criminal justice in no way change the fact that proportionality remains an important limitation and control tool against any arbitrarily coercive and highly intrusive exercise of state powers – especially when it comes to establishing the extent of permissible use of AI. Guaranteeing legitimacy and proportionality in exercising powers has been an ever-present prerequisite, *inter alia*, for the effective protection of personal data and the safeguarding of privacy.[30] Nonetheless, in light of today's unprecedented technological dominance in (mass) surveillance and other areas of privacy intrusions,[31] proportionality is just one among many rule of law related concerns that theory, practice and policy are called to address in this context.

The personal freedoms and rights of self-determination, which are directly connected with the 'mother right'[32] to human dignity as 'the freedom to shape one's life',[33] are undisputedly crucially important in modern liberal and democratic systems. Yet, the privacy rights, which make up a significant part of them, are not absolute. However, permitted restrictions are usually subject to strict limitations: Interferences with the exercise of these rights must be provided for by law, respect their core essence and be necessary in a democratic society for protecting national security, public safety and other important public interests or the rights and freedoms of others.[34] Relevant examples in the present context are the definitions of privacy rights under Articles 7 and 8 EU Charter (respect for private life and protection of personal data) as well as the right to respect for private life under Article 8 ECHR.[35]

2015), available at https://library.fes.de/pdf-files/id/ipa/11444-20150619.pdf; M Hildebrandt, 'The Artificial Intelligence of European Union Law' (2020) 21 *German Law Journal* 74; P Scharre, 'Why Unmanned' (2011) 61 *Joint Force Quarterly* 89; and NE Sharkey, 'The Evitability of Autonomous Robot Warfare' (2012) 94 *International Review of the Red Cross* 787.

[30] See, eg, L Bachmaier Winter, 'Proportionality, Mass Surveillance and Criminal Investigation: The Strasbourg Court Facing Big Brother' in Billis, Knust and Rui (n 17) 317ff.

[31] *cf Big Brother Watch and Others v the United Kingdom* App nos 58170/13, 62322/14 and 24960/15 (ECtHR, 13 September 2018) para 316: '(…) due to recent technological developments the interception of communications now has greater potential than ever before to paint an intimate and detailed portrait of a person's private life and behaviour.'

[32] A Barak, *Human Dignity: The Constitutional Value and the Constitutional Right* (Cambridge, Cambridge University Press, 2015) 156–67.

[33] C Dupré, 'Article 1 – Human Dignity' in S Peers et al (eds), *The EU Charter of Fundamental Rights. A Commentary*, 2nd edn (Oxford, Hart Publishing, 2021) 3, 6.

[34] See EU Charter, Art 52 and ECHR, Art 8.

[35] For example, (bulk) interception, storage, processing, examination and use of communications constitute interferences with Art 8(1) ECHR rights, the permissibility of which depends on the fulfilment of the terms defined in Art 8(2), see *Big Brother Watch and Others v the United Kingdom* App

Interference with well-established privacy rights through the unrestrained and unsupervised use of AI may undermine existing protective limitations, democratic safeguards and rule of law guarantees, even if the intrusive measures and applications are programmatically aimed at keeping populations and societies safe and secure. At a practical level, novel information and interception systems consisting of advanced monitoring mechanisms and surveillance algorithms may well be significantly contributing to the preventive and suppressive efforts of intelligence and law enforcement agencies against new forms of organised and transnational crime and other types of global threat. Continuous progress in technology can be instrumental in securing wider protection for fundamental freedoms while also enhancing the effectiveness in administering criminal justice. However, a series of problematic issues involving the danger of questionable policies and strategies towards achieving greater operational benefits may arise in parallel with this process. One example pertains to the possibility of a constant, rapid, multi-layered, multi-purposed, cross-sectoral/-departmental and, hence, practically uncontrolled flow and exchange of vast amounts of massively and automatically collected and analysed personal data between cooperating agencies of different legal orders. Any opportunity for public authorities to acquire and invest in spy software developed and/or operated by private companies or third-party organisations in non-transparent terms can, in many respects, also be open to questions. The same goes for the discretion agencies enjoy in avoiding external regulatory and 'bureaucratic' obstacles by developing their own, tailor-made big data algorithms, exclusively designed and trained by internally appointed computer experts absent any prior public consultation with human rights scholars and institutions.

In any case, the stakes are already high for the rule of law and the effective protection of self-determination and privacy rights. Personal data constitute the foundational element of AI-based applications employed for the purposes of security and crime control. As we have seen, such applications are not free from the risks of manipulation, discrimination and non-transparency or from the problems of insufficient oversight and accountability.[36] Automatic processing and analysis of collected and stored data may produce new or reveal further

nos 58170/13, 62322/14 and 24960/15 (ECtHR GC, 25 May 2021) paras 324ff, 330; see also S Brinkhoff, 'Big Data Data Mining by the Dutch Police: Criteria for a Future Method of Investigation' (2017) 2 *European Journal for Security Research* 57. Private life under Art 8 regards 'personal data', which are defined as 'any information relating to an identified or identifiable individual', see *Benedik v Slovenia* App no 62357/14 (ECtHR, 24 April 2018) para 102. Personal data include 'anything from a name, photo, email address, bank details, GPS tracking data, posts on social networking websites, medical information or a computer's IP address', see Handbook on European Data Protection Law (n 3) 350.

[36] *cf*, eg, European Parliament's resolution on AI (n 14) at 1, 4, 6, 8, 9, 11, 13, 16, 17, 22, 23, 25, 28; Artificial Intelligence Act (n 6) Arts 1, 10, 13, 14, 15. See also the references in n 5 above. In the context of the tracing measures adopted during the Covid pandemic, *cf* European Institute of Innovation & Technology, 'The European Struggle with COVID-19 Contact Tracing Apps' (29 April 2020), available at eit.europa.eu/news-events/news/european-struggle-covid-19-contact-tracing-apps.

sensitive information on individuals. A main concern with respect to the use of algorithms for producing or acquiring such 'new knowledge' refers to the 'black box' problem: the public and interested parties may have knowledge of the information fed into the system and the system's concrete output, but they rarely know how the decision-making process actually works in between. Thus, unpredictable correlations and unexplainable inferences cannot be excluded. Further, it is rather unlikely for authorities or enterprises developing AI programs for purposes of security and crime control to divulge the exact functionality of their algorithms. The problem is exacerbated by the fact that today's algorithms are trained and operate mainly on the basis of statistical observations and correlations. 'No matter how sophisticated, predictive algorithms and their users can fall into the trap of equating correlation with causation.'[37] As a result, there is always the danger that the information and knowledge obtained by current algorithmic systems can be distorted, discriminatory and unreliable.[38]

It is therefore logical that the public debate focuses on both the predictability and explicability problems in AI uses and outputs as well as on questionable practices and dubious cooperation methods followed by enforcement and security agencies, national and supranational.[39] In view of the current technological developments and the legal guarantees put in place thus far, it is necessary to build more coherent and consistent models enabling transparent, predictable, accurate, proportional and non-discriminatory uses of AI technology in the personal and big data fields as well as systems designed to overcome the legal and pragmatic issues causing obstacles to independent oversight and judicial control. Again, maintaining direct communication channels and developing improved methods of collaboration between competent national and supranational authorities, computer scientists, AI developers and legal experts in the areas of privacy and data protection are essential steps in strengthening both the effectiveness of law enforcement and the rule of law.

V. EQUALITY AND PROCEDURAL JUSTICE

The problem of potential discriminatory outputs is a common topic of concern in the context of automated decision-making and AI predictions, as already noted. This issue depends mainly on such factors as the data quality, any pre-existing biases fed into the algorithmic system by its designers and trainers

[37] SK Sgaier, V Huang and G Charles, 'The Case for Causal AI' (2020) 18 *Stanford Social Innovation Review* 50.

[38] See Billis, Knust and Rui (n 2) 706–707 with detailed references and examples. See also European Parliament's resolution on AI (n 14).

[39] See, eg, for a summary of recent worrying developments in Europe, the report by A Fotiadis, L Stavinoha, G Zandonini and D Howden, 'A Data "Black Hole": Europol Ordered to Delete Vast Store of Personal Data' *The Guardian* (10 January 2022), available at www.theguardian.com/world/2022/jan/10/a-data-black-hole-europol-ordered-to-delete-vast-store-of-personal-data.

and/or the system's incapacity for true causal reasoning.[40] Thus, algorithmic decisions 'may end up reproducing historical patterns of discrimination'.[41]

The overall prohibition of discrimination derives from the principle of equality, and both equality and non-discrimination 'are part of the foundations of the rule of law'.[42] Equality rights such as those enshrined in Articles 20–26 EU Charter (equality before the law; non-discrimination; cultural, religious and linguistic diversity; equality between men and women; the rights of the child and of the elderly; integration of persons with disabilities) and the provision of Article 14 ECHR (prohibition of discrimination) also tie in with the primary demand for respecting and protecting human dignity.

In the various constellations of its use, AI may effectively promote equal treatment and fairness.[43] But it may also impair them. Relevant considerations have focused on administrative law, labour/civil law and healthcare matters. In terms of law enforcement and criminal justice, the risk of algorithmic systems generating discriminatory outputs is usually discussed in the context of identification, predictive policing and recidivism risk assessment applications.[44]

Furthermore, equality is related to procedural justice. Essential elements of procedural justice are impartiality, consistency, accuracy and truthfulness, representativeness, ethical appropriateness, transparency, the possibility to contest and review a decision and the respectful treatment of those affected by the conflict and the procedure.[45] Sometimes, these elements conflict with each other – even more so in the AI and criminal justice field.[46] In any case, in liberal rule of law orders, with respect to both conventional mechanisms and alternative means of prevention, enforcement and conflict resolution, justice and social legitimacy require the application of formal standards, adequate due-process guarantees and minimum fair-trial rights providing substantial and equal opportunities of

[40] See the European Parliament's resolution on AI (n 14) at 8, 22, 23, 24.

[41] J Ryberg and J Roberts, 'Sentencing and Artificial Intelligence – Setting the Stage' in J Ryberg and J Roberts (eds), *Sentencing and Artificial Intelligence* (Oxford, Oxford University Press, 2022) 1, 8.

[42] United Nations and the Rule of Law, 'Equality and Non-Discrimination', available at www.un.org/ruleoflaw/thematic-areas/human-rights/equality-and-non-discrimination/.

[43] See, eg, M Bagaric and D Hunter, 'Enhancing the Integrity of the Sentencing Process through the Use of Artificial Intelligence' in Ryberg and Roberts (n 41) 122ff, 131ff with further references.

[44] See, eg, the European Parliament's resolution on AI (n 14) at 9, 24, 27. On equality, discrimination and algorithmic risk assessment instruments (such as the infamous COMPAS system), see J Angwin, J Larson, S Mattu and L Kirchner, 'Machine Bias' (*ProPublica*, 23 May 2016), available at www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; B Davies and T Douglas, 'Learning to Discriminate – The Perfect Proxy Problem in Artificially Intelligent Sentencing' in Ryberg and Roberts (n 41) 97; K Lippert-Rasmussen, 'Algorithm-Based Sentencing and Discrimination' in Ryberg and Roberts (n 41) 74.

[45] M Rehbinder, *Rechtsoziologie*, 6th edn (Munich, Beck, 2007) 118.

[46] See, eg, on the possibility of a negative relationship between the transparency of an algorithm and its accuracy in criminal sentencing (complicated and opaque algorithms producing more accurate predictions of future criminality *versus* algorithms which are more transparent, but produce less accurate predictions), J Ryberg and T Petersen, 'Sentencing and the Conflict between Algorithmic Accuracy and Transparency' in Ryberg and Roberts (n 41) 57.

procedural inclusion, participation and remedy.[47] Particularly significant in this regard are the provisions of Articles 5 and 6 ECHR (right to liberty and security; right to a fair trial) and Articles 47–48 EU Charter (right to an effective remedy and to a fair trial; presumption of innocence and right of defence).

In line with this, the European Parliament stressed in its resolution of 6 October 2021 that AI may only be used in law enforcement and criminal justice if it fully complies with the principles of equality before the law and equality of arms as well as the rights to an effective remedy and a fair trial. The Parliament highlighted the power asymmetry between those who employ AI technologies and those who are subject to them and noted that the use of AI must not become a factor of inequality, social fracture or exclusion.[48] It further underlined 'the impact of the use of AI tools on the defence rights of suspects, the difficulty in obtaining meaningful information on their functioning and the consequent difficulty in challenging their results in court, in particular by individuals under investigation'.[49]

Finally, the resolutions on the need to limit AI usage with rule of law safeguards and to be able to effectively contest automated decision-making[50] on the grounds of equality, non-discrimination and procedural justice are fundamentally important for a variety of criminal justice and procedure matters such as sentencing and evidence.[51] The European Parliament, having considered that 'decisions in the field of law enforcement are almost always decisions that have a legal effect on the person concerned' and that 'the use of AI may influence human decisions and have an impact on all phases of criminal procedures', *inter alia* noted that:

- in judicial and law enforcement contexts, the decision giving legal or similar effect always needs to be taken by a human, who can be held accountable for the decisions made;
- those subject to AI-powered systems must have recourse to remedy;

[47] See E Billis and N Knust, 'Alternative Types of Procedure and the Formal Limits of National Criminal Justice: Aspects of Social Legitimacy' in U Sieber, V Mitsilegas, C Mylonopoulos, E Billis and N Knust (eds), *Alternative Systems of Crime Control. National, Transnational, and International Dimensions* (Berlin, Duncker & Humblot, 2018) 39, 57, with references to J Rawls, *A Theory of Justice* (Cambridge, Massachusetts, Belknap Press of Harvard University Press, 1971) and JW Thibaut and L Walker, *Procedural Justice: A Psychological Analysis* (Hillsdale, NJ, Erlbaum, 1975). See also TR Tyler, *Why People Obey the Law* (Princeton, Princeton University Press, 2006) 163–65; TR Tyler, 'Does the American Public Accept the Rule of Law?: The Findings of Psychological Research on Deference to Authority' (2007) 56 *DePaul Law Review* 664.

[48] European Parliament's resolution on AI (n 14) at 2 and 10.

[49] ibid at 10.

[50] On the right to contest adverse decisions, which directly derives from the rule of law principle, and the challenges raised by automated decision-making systems in this context, which are associated with the 'right to a reasoned decision and the procedural equality between parties during the review', see FP Ettorre, 'The Right to Contest Automated Decisions' *The Digital Constitutionalist* (14 February 2022), available at https://digi-con.org/the-right-to-contest-automated-decisions/.

[51] AI may serve as a novel and alternative means for automatically processing/producing evidence with the purpose of enhancing effectiveness, objectivity, fairness and procedural economy in traditional criminal trials. However, mostly due to the noted asymmetry of power, this could also result in compromising the effective participation of the defendant in the process. On the merits and challenges of admitting, reviewing and assessing AI-based evidence in criminal proceedings and on the relevant risks for the defendant's rights connected with such applications, see chapter 15 by Eftychia Bampasika in this volume.

- under EU law, a person has the right not to be subjected to a decision which produces legal effects concerning them or significantly affects them and is based solely on automated data processing;
- automated individual decision-making must not be based on special categories of personal data, unless suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place;
- EU law prohibits profiling that results in discrimination against natural persons on the basis of special categories of personal data;
- authorities making use of AI systems need to uphold extremely high legal standards and ensure human intervention, especially when analysing data deriving from such systems, and;
- the sovereign discretion of judges and decision-making on a case-by-case basis need to be upheld.[52]

VI. CONCLUDING THOUGHTS: THE CASE OF AI IN SENTENCING

In the context of security, crime control and criminal justice, the relationship between law and new technologies is a tense and multi-layered one: from the exploitation of advanced electronic tools to facilitate illegal behaviour to the employment of machine learning as an alternative means to enhance law enforcement and the administration of criminal justice. And from legal scholars and policy planners studying and adapting to the evolution of AI by re-evaluating old and designing new regulatory frameworks according to current ethics and societal values to computer scientists collaborating closely with lawyers in the interpretation and realisation of the most complex concepts and ideas. An integral part of this phenomenon is the increasing reliance of state authorities and supranational agencies on machines and automated algorithmic systems for the prevention, detection and suppression of security threats and (possible) crimes and for an improved criminal justice delivery. Such mechanisms can make the exercise of public authority more effective and efficient and legal decisions more objective and less arbitrary and noisy. At the same time, their use has the potential to radically expand the states' powers of surveillance, intrusion and coercion and entails, as we have seen, various rule of law and human rights concerns.

A characteristic example applicable to many of the considerations examined in this chapter is sentencing in criminal trials. Theory and recent practice in various legal orders have shown that the new types of AI and machine learning – praised for their presumed mathematical ability to reduce factual complexity and improve quality (especially: consistency, accuracy, objectivity) and speed in decision-making but also criticised for potentially posing risks to traditional protective principles and rule of law constraints – are already transforming the conventional criminal trial.[53] Sentencing issues have been diachronically

[52] European Parliament's resolution on AI (n 14) at 16.

[53] See, eg, with respect to the procedural stage of sentencing, the different applications and topics of concern analysed in Ryberg and Roberts (n 41).

at the core of theoretical and legal policy debates in the field of criminal law and procedure. Sentencing, which is determined by a series of concurrent (and, sometimes, conflicting) aims, consists of many layers and components (substantive and procedural) and varies from one legal tradition or system to the next in terms of structure and (division and scope of) institutional powers, has also been one of the first criminal justice sectors to be subjected to the AI experiment.[54] As Ryberg and Roberts emphasised:

> A sentencing decision involves balancing multiple objectives and many factors. (…) The use of AI at sentencing is therefore open to different interpretations and applications. At one extreme, the use of AI at sentencing may refer to the application of a simple algorithm implemented to inform a judge in the determination of a single factor [eg, offender's risk profile] that should be included in the sentencing decision. At the other extreme, it may refer to a fully automated 'Robo judge' that specifies sentences without any human involvement.[55]

Regardless, the determination of punishment is usually characterised by the application of objective indicators and a set of *a priori* defined 'measurable' criteria, even in legal orders where sentencing judges have a greater margin of appreciation in evaluating the specific circumstances of cases and offenders. A more extended use of algorithms aligned with this rather straightforward method is therefore to be expected in sentencing issues compared to other, more complex and multi-factored decision-making processes involving, for example, the systematic assessment of different types of evidence during trial and the rendering of reasoned judgments about guilt.

Notwithstanding the presumed potential of AI in minimising human error and reducing the arbitrary exercise of power, it does not mean that we have already reached the point of replacing human sentencing judges with machines, nor that we should be considering such a possibility in the first place. Even if machine-based calculations of punishment are not directly violating human dignity, the normative imperatives of the human-centric approach do not allow for the use of AI as a substitute for human decision-making but only as a supplement. In fact, we can contemplate a variety of reasons against the 'dehumanisation'[56] of sentencing, most importantly: the serious consequences of punishments for people's lives and the social necessity of maintaining clear accountability and review systems for sentencing decisions; the significance of not only objective but also subjective, moral and emotional (and, thus, not easily computable) factors, such as mercy and leniency, for effectively fulfilling

[54] ibid 1, 3ff.

[55] ibid at 5–6.

[56] See in this respect the interesting analysis by N Dagan and S Baron, 'The Compassionate Computer – Algorithms, Sentencing, and Mercy' in Ryberg and Roberts (n 41) 145, 146-47 with further references. See also M Schwarze and J Roberts, 'Reconciling Artificial Intelligence and Human Intelligence – Supplementing Not Supplanting the Sentencing Judge' in Ryberg and Roberts (n 41) 206, 208ff.

the purposes of humane punishment in contemporary rule of law states; and the legitimising effect of the human ability to comprehend and evaluate the contextual meaning of societal circumstances at the imposition of criminal penalties.

Important on similar grounds is the issue of achieving proportional results in sentencing, as prescribed, for example, in Article 49(3) EU Charter (principles of legality and proportionality of criminal offences and penalties), ie, in terms of assessing the gravity of criminal misconduct and properly determining the corresponding penalties, the severity of which must not be disproportionate to the offence.[57] Proportionality is already relevant at the legislative, regulatory and technical levels, specifically in the sense of defining *a priori* and *in abstracto*, on factors such as gravity and complexity, the types of crimes and cases to which AI-assisted or AI-based sentencing could/should be applied. Regarding proportionality assessments and proportional outputs in constellations where algorithms are used to assist/complement the responsible sentencing body in reaching the 'proper' decision, the judge and the parties need to be aware of any technical and inherent limitations of a non-human system in terms of comprehending and implementing proportionality criteria. The natural shortcomings and limitations of human sentencers cannot be ignored. Still, we must, before we entrust the electronic assistant – and, in a dystopic future, the fully automated and autonomously deciding machines replacing the human judge – with sentencing tasks, carefully consider to what extent disproportionate penalties specifically generated by an AI may further jeopardise social legitimacy of the judicial system and risk fracturing the rule of law. We also need to keep in mind that, as accepted by the European Court of Human Rights (ECtHR), a grossly disproportionate sentence could, on some occasions, even amount to *inhuman* or degrading punishment.[58]

Finally, a widely recognised goal of employing such applications in sentencing, as for other AI uses, is to secure not only swift but also more objective and impartial decisions, free from the 'restrains' of human emotions, prejudices and personal interests. However, any capabilities of new technologies in strengthening the administration of criminal justice and the rule of law must not overshadow the possible risks to equality, objectivity and impartiality caused by AI itself. Even in the case of computers used to *support* the human sentencer in his/her tasks, the question arises: to what extent can judges, consciously or unconsciously, remain independent and impartial if they, overly trusting in the technical superiority of machines or, for other reasons, overly relying on them,

[57] The EU principle of proportionality 'entails that the severity of penalties must be commensurate with the seriousness of the infringements', see Opinion of the Advocate General Bobek, Case C-384/17 *Dooel Uvoz-Izvoz Skopje Link Logistik N&N v Budapest Rendőrfőkapitánya*, EU:C:2018:494, 26 June 2018, para 32.

[58] Violating, thus, ECHR, Art 3, see *Vinter and Others v the United Kingdom* App nos 66069/09, 130/10 and 3896/10 (ECtHR GC, 9 July 2013) paras 83 and 102.

base their decisions mainly on information processed and knowledge generated by algorithms.[59]

We examined above the reasons why outputs of current algorithmic systems can be distorted, discriminatory and unreliable (systematic bias, technical difficulties in causal reasoning, etc). Designing and rigidly applying appropriate procedural safeguards to ensure, for all interested parties, equality of participation possibilities, information and resources symmetry, openness and transparency in the joint automated and human decision-making processes and, accordingly, adequate contest and review opportunities, may have a counterbalancing effect on the inherent shortcomings of AI applications in criminal justice and, at the same time, a reinforcing effect with respect to the social acceptance of the final outcome. Furthermore, it may prove fruitful to leave aside extreme and generalising positions such as the unproductive narrative about the overall inevitability of AI on the one side and the unrealistic propositions about outright AI bans on the other. Instead, focusing on how to improve the mutual understanding and cross-disciplinary collaboration between computer scientists, legal scholars and practitioners from all relevant fields and institutions could substantially contribute to the efforts to minimise the analysed systemic risks and weaknesses and to create new models of human-machine cooperation in criminal justice: models that do not challenge but strengthen the rule of law.

[59] *cf* European Parliament's resolution on AI (n 14) at 15: '(…) [If] humans only rely on the data, profiles and recommendations generated by machines, they will not be able to conduct an independent assessment; highlights the potentially grave adverse consequences, specifically in the area of law enforcement and justice, when individuals overly trust in the seemingly objective and scientific nature of AI tools and fail to consider the possibility of their results being incorrect, incomplete, irrelevant or discriminatory; emphasises that over-reliance on the results provided by AI systems should be avoided, and stresses the need for authorities to build confidence and knowledge to question or override an algorithmic recommendation'.