Artificial Intelligence Destroyed the Rule of Law?

Stanley Greenstein*

* Stanley Greenstein, LL.M., Dr., Faculty of Law, Stockholm University. Email: Stanley.greenstein@juridicum.su.se.

1 Introduction

We are living in the era of artificial intelligence, where many decision-making processes within society are increasingly being transferred to highly complex computerised systems. These systems comprise elements of artificial intelligence in the form of machine learning capabilities – they are mathematically complex, data-driven, self-learning and inhibit various degrees of autonomy required to achieve the goals that they have been assigned. The term 'artificial intelligence' is an umbrella terms incorporating many different types of technologies that share the goal of creating machines that are endowed with a level of intelligence comparable to that of humans. A major challenge is defining artificial intelligence as this technology is connected to time, it is dynamic and it is subjective – what artificial intelligence is for one person is not artificial intelligence for another, what was artificial intelligence say fifteen years ago is today mainstream technology and even trying to fathom out what 'intelligence' is can be fraught with endless discussion.

A central aspect of legal informatics is the examination of the relationship between technological development and the law, more specifically how technological development influences established legal concepts that have traditionally been forged over time in order to regulate society. A challenge, considering the slower pace at which the legal domain develops, is that legal concepts sometimes lose their regulatory power in relation to the fast-evolving technical environment. This is of utmost relevance in relation to the notion of the rule of law. The concept 'rule of law', bears similarities with that of artificial intelligence, to the extent that its composition is not entirely clear nor set in stone. It may mean different things to different people, which in turn is an attractive quality but which also results in feverous discussions of what it actually is. Nevertheless, the rule of law does seem to have a universal quality that has found traction in many parts of the world. The study of the notion rule of law reveals that many descriptions of its composition take the form of concrete, procedural, instrumental features or characteristics. While much of the focus is on these procedural features, it is argued that the rule of law has higher-level objectives or goals, namely, attaining a good society – a society that humans flourish in.

The problem addressed by this article has at its core the collision between two systems of rules. The first system is the law, a system comprising normative rules, developed via a process of human thought and infused with values that humans cherish, here represented by the rule of law. The second system is that of technology, which too has rules, however, these rules are based on mathematics, statistics, algorithms and the data science. It is argued that at present, these two systems of rules are not compatible to extent that it is challenging to represent human values based on morality and ethics, in the actual technology. It is also argued that in this battle between law and technology, technology is bound to be victorious. The main argument for implementing artificial intelligence in society is effectivity – there are simply economic gains to be made from the implementation of artificial intelligence and decisions can be made with greater speed. It is in this context that the notion of technological determinism is relevant, in other words technological development is unstoppable and has a life of its own. While artificial intelligence has multiple

benefits for society, it must be noted that it is eroding many aspects of the analogue world, one being how we represent and apply the law, more specifically the rule of law. An essential precondition for the application of the rule of law is its existence in the form of procedural features – legality, access to justice, equality before the law, absence of the misuse of power a few examples. However, technological development, represented by artificial intelligence, is eroding the existence of the procedural features of the rule of law, in turn essentially negating its existence and eliminating it as a legal mechanism for accomplishing the higher-level goals that it was intended to achieve – attaining a society that promotes human flourishing.[1]

In perusing the above goal, it is important to note the following. Technology can be used both for positive and negative purposes, often described as a double-edged sword. This article is limited to an examination of the effect of the increased use of artificial intelligence on the notion of the rule of law in its current form. The conclusions may be interpreted as negative to the extent that, as will be shown, technological development in general and artificial intelligence more specifically is challenging the rule of law. Considering this, some clarifications are necessary. First, artificial intelligence per se is not a phenomenon to be construed as a threat only and there are countless examples of the benefits of this technology to society at large, for example, to saving lives of prematurely born babies.[2] Second, artificial intelligence can in fact enhance the rule of law, for example, by increasing access to the law and legal tools, legal tech a facilitator in this respect.[3] However, while the rule of law maintains its current form, it is certain to be eroded by artificial intelligence.

2 The Rule of Law

The rule of law is a vague and elusive concept. This has resulted in the fact that its fluidity lends itself to multiple interpretations that are not always necessarily aligned. Yet the rule of law is a concept that enjoys a certain level of global acceptance and remains a symbol for the characteristics embodying a society to strive after. Its universal recognition makes it a concept that is entrenched in many legal systems, arguably because 'the "rule of law" is good for everyone', an attitude that seemingly enjoys international support.[4] The fluidity of the concept rule of law, while promoting its global application, does create

---

1 The theme of this article has its origin in a short piece published on Verfassungsblog, *Artificial Intelligence, Human Flourishing and the Rule of Law*, published 30 March 2022, available at https://verfassungsblog.de/roa-human-flourishing/. It is also based on Greenstein, Stanley, *Our Humanity Exposed: Predictive Modelling in a Legal Context*, dissertation, Stockholm University, 2017 available at http://su.diva-portal.org/smash/record.jsf?pid=diva2%3A1088890&dswid=-7446.

2 Siegel, Eric, *Predictive Analytics – The Power to Predict Who Will Click, Buy, Lie or Die*, John Wiley & Sons, 2013, pp. 265-289.

3 Thread, *The Lawyer's Practical Guide to: Legal Chatbots (and Whether They're Really Coming to Steal Your Job)*, available at https://www.threadsoftware.com/the-lawyers-practical-guide-legal-chatbots/.

4 Tamanaha, Brian Z., *On the Rule of Law*, Cambridge University Press, Cambridge, 2004, p. 1.

challenges when attempting to define its meaning and realm of application. Discussions on the function and composition of the rule of law often culminate in the distinction made between the rule of law's formal requirements (what conditions are required for a system to be considered compliant with the rule of law) and its material or moral objectives. Some view it as a concept comprised purely of formal structures of governance, while others recognize it as embodying moral considerations.[5] Ultimately this distinction is academic to the extent that both viewpoints are true. The rule of law is represented by procedural features that allow it to promote higher-level goals. Illuminated here, however, is that the procedural features of the rule of law are a necessary precondition to attain the higher-level goals of attaining a society within which individuals can excel. Without their existence there is no rule of law and thus their importance is primary. In other words, the features argued to comprise the rule of law – transparency, the making public of laws, the prohibition of retroactive laws, legality – are not necessarily ends in themselves, but rather their existence promotes higher level goals, such as attaining a better society, the notion of 'better' admittedly also a subjective notion in itself.

2.1 The Procedural Features of the Rule of Law

One of the more well-known descriptions of the rule of law is provided by Lon Fuller in his work entitled *The Morality of Law*, which perceives the rule of law as a combination of the formal institutions of society together with what he terms 'the inner morality of law', describing it by means of eight formalistic principles of a procedural nature: 1) there must be rules, 2) they must be prospective, not retrospective, 3) they must be published, 4) they must be intelligible, 5) they must not be contradictory, 6) compliance with the rules must be possible, 7) the rules must not be constantly changing and 8) there must be congruence between the rules as declared and as applied by officials.[6]

Emphasising the procedural nature of the rule of law, The World Justice Project states:

> Effective rule of law reduces corruption, combats poverty and disease, and protects people from injustices large and small. It is the foundation for communities of justice, opportunity, and peace – underpinning development, accountable government, and respect for fundamental rights'.[7]

The procedural features comprising the rule of law are represented by four principles each containing sub-features. The World Justice Project separates the rule of law into four principles, namely, 'accountability' ('the government as well as private actors are accountable under the law'), 'just law' ('the law is clear, publicized, and stable and is applied evenly and ensures human rights as well as property, contract, and procedural rights'), 'open government' ('the

---

5 Simmonds, Nigel E., *Central Issues in Jurisprudence- Justice, Law and Rights*, Sweet and Maxwell, 2008, p. 115.

6 Fuller, Lon, *The Morality of Law*, Revised Edition, Yale University Press, New Haven and London, 1965.

7 World Justice Project, *World Justice Project Rule of Law Index 2019 Insights*, 2019, p. 7.

processes by which the law is adopted, administered, adjudicated, and enforced are accessible, fair, and efficient') and 'accessible and impartial justice' (justice is delivered timely by competent, ethical, and independent representatives and neutrals who are accessible, have adequate resources, and reflect the makeup of the communities they serve').[8]

The procedural nature of the rule of law is also reflected by The European Commission for Democracy Through Law (Venice Commission).[9] The Venice Commission adopted a report on the rule of Law in 2011.[10] It subsequently published a checklist in order to help determine to what extent countries live up to the rule of law. This checklist is in turn based on five core elements required for the rule of law, namely 'legality', 'legal certainty', 'prevention of abuse/misuse of powers', 'equality before the law and non-discrimination' and 'access to justice'.[11]

The existence of the rule of law and the extent to which it is upheld by states is therefore determined by procedural factors. This is evident by the fact that the rule of law can be compressed into a checklist. These features are not necessarily the higher-level goals of the rule of law. Rather, it can be argued that the rule of law has higher-level goals, which are automatically attained by ensuring the existence of the procedural features. These high-level goals are discussed in the next section.

2.2 The Higher-Level Goals of the Rule of Law

The rule of law is said to embody a notion of reciprocity between those that govern and those that are governed, where those in positions of authority must exercise this authority according to established public norms and not arbitrarily and where citizens are expected to comply with legal norms, the law should be the same for everyone, no one is above the law and finally, everyone should be protected by the law.[12] It is also described as a solution to the problem of how to make the law rule, addressing how power is exercised, more specifically its abuse by exercising it in an arbitrary manner.[13] There is also the idea that the source of authority to rule originates from a moral right to rule, where this moral dimension dictates that rules be publicly declared in a perspective manner and

---

[8] World Justice Project, *What is the Rule of Law?*, available at https://worldjusticeproject.org/about-us/overview/what-rule-law.

[9] European Commission for Democracy Through Law, available at https://www.venice.coe.int/WebForms/pages/?p=01_Presentation.

[10] Venice Commission, available at https://www.venice.coe.int/webforms/documents/?pdf=CDL-AD(2011)003rev-e.

[11] Ibid.

[12] Stanford Encyclopedia of Philosophy, *The Rule of Law*, available at https://plato.stanford.edu/entries/rule-of-law/, p. 1.

[13] Krygier M, *What's the point of the rule of law?*, Buffalo Law Rev 67(3):743–791, 2019, p. 758-760.

are general, equal and certain.[14] The rule of law, therefore, is a political ideal, although its content and composition remains a point of discussion and to a certain degree controversial.[15]

Fullers eight principles above are outwardly procedural yet of interest is Simmonds's interpretation of Fuller's outwardly formalistic depiction of the rule of law, where he argues that the 'inner morality' aspect of Fuller's eight principles promotes 'the morality of duty' and 'the morality of aspiration'.[16] The former involves a duty to abide by laws that are obligatory and either one does this or not, whereas the latter concept is not an 'either/or' notion but rather a question of degree, where one strives towards this ideal to the best of one's ability.[17] The eight principles provide a degree of regularity and order necessary to attain the morality of aspiration, and they represent the morality of aspiration in that they represent an ideal to which a legal system should strive towards.[18] Furthermore, the attainment of the morality of aspiration requires that there be rules and orderliness, created by the morality of duty, and that eventually allows humans to attempt to attain that situation as depicted by the concept rule of law. Accordingly, Simmonds argues that the morality of duty and the morality of aspiration differ in their goal, where the latter concerns the attainment of the 'good life' in a context where 'people can meaningfully formulate and pursue personal projects and ideals'. It is here that the rule of law can be argued to fulfil its function as an instrument promoting the greater goal of human flourishment and 'value the projective capacities of men and women', an ideal that is achievable only where there are clear and notified rules.[19] Simmonds states:

> These values are internal to the law in the sense that they form a part of the concept of law itself. We understand what the law is only by reference to its purpose; and its purpose is an ideal state of affairs (the rule of law) represented by the eight principles. […][The law] carries a commitment to the idea of man as a rational purposive agent, capable of regulating his conduct by rules rather than as a pliable instrument to be manipulated; and it carries a commitment to the values of the rule of law as expressed in the eight principles.'[20]

---

[14] Bayamlioglu E, and Leenes R, *Data-driven decision-making and the 'Rule of Law,* TILT Law Technology, Working Paper, Tilburg University, 2018, available at https://www.researchgate.net/publication/329873662_Data-Driven_Decision-Making_and_The_'Rule_of_Law'_Data-Driven_Decision-Making_and_The_'Rule_of_Law', p. 29.

[15] Matsou, *Let the Rule of Law be Flexible to Attain Good Governance*, in: Berling P, Ederlöf J, Taylor V (eds.), Rule of Law Promotion: Global Perspectives, Local Applications, Skrifter från juridiska institutionen vid Umeå universitet Nr 21, Iustus Förlag, Uppsala, pp 41–56, 2009, p. 41.

[16] Simmonds, Nigel E., *Central Issues in Jurisprudence- Justice, Law and Rights*, Sweet and Maxwell, 2008, at p. 118.

[17] Ibid.

[18] Ibid, p. 121.

[19] Ibid, p. 120.

[20] Ibid, p. 122.

This interpretation promotes the rule of law as a mechanism creating stability, allowing individuals to achieve their potential and flourish as human beings. This role of the law is not novel and it has traditionally been viewed as a mechanism for achieving human flourishing.[21] The law as a tool for shaping a society is also evident in the statement that it is a mechanism whereby, 'ordinary, everyday ways in which situated, embodied subjects experience their culture and their own evolving subjectivity, and when they consider the ways in which networked information technologies reshape everyday experience'.[22] It is argued that the factors that best achieve a position of human flourishing in the networked society are access to knowledge, operational transparency and room for play of everyday practice.[23]

The law's importance, therefore, is as a mechanism creating the environment within which the individual enjoys the freedom to make decisions ('right of subjective freedom') and pursue his or her notion of a good life in a sphere of liberty, free from interference. In this context, it is argued that notions such as 'rights', 'justice' and 'equality' gain importance.[24] Considering the difficulty associated with defining the rule of law, its existence is sometimes measured in terms of the existence of 'outcomes', whereby, for example, the rule of law is measured in terms of actual access to courts and not necessarily how this is represented in law.[25] This can be viewed as a means of measuring the rule of law by focusing on the higher-level goals it achieves.

Brownsword defines the rule of law as a combination of the condemnation of arbitrary governance on the one hand and the irresponsible citizenship on the other. According to this view, the rule of law represents a contract between, on the one hand, lawmakers, law-enforcers, law-interpreters and law appliers and on the other hand citizens (including lawmakers, law-enforcers, law-interpreters and law appliers), whereby the actions of the governors are in accordance with the law and that the citizens abide by decisions made in accordance with the legal rules, the result being that no one is above the law.[26]

The Council of Europe has also weighed in on defining the rule of law:

> The rule of law is a principle of governance by which all persons, institutions and entities, public and private, including the state itself, are accountable to laws that are publicly promulgated, equally enforced, independently adjudicated and consistent with international human rights norms and standards. It entails adherence to the principles of supremacy of law, equality before the law, accountability to the law,

---

21 Nussbaum, Martha C., *Creating Capabilities*, The Bellknap Press, 2013, pp.ix-xi and 186.

22 Cohen, Julie E., *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice*, Yale University Press, 2012, p. 6.

23 Ibid.

24 Simmonds, Nigel E., (n 14), p. 7.

25 World Justice Project, *Rule of Law Index, 2019*, 2019, p. 7.

26 Brownsword, Roger, *Technological Management and the Rule of Law*, Law Innovation and Technology, 8:1, 100-140, Routledge, 2016, p. 106.

> fairness in applying the law, separation of powers, participation in decision making, legal certainty, avoidance of arbitrariness and procedural and legal transparency.[27]

Krygier stresses the fact that the rule is a solution to a problem, the problem being how to make the law rule.[28] The reason for striving to make the law rule are concerns regarding the way power is exercised, more specifically the abuse of power by exercising this power in an arbitrary manner.[29] Finally, a practical function of the rule of law is that it is a functional instrument that caters for society's need for predictability and that orders an otherwise chaotic society, thereby answering the question of what tomorrow brings.[30]

The above discussion of the rule of law reveals a rather dynamic concept with multiple definitions, some choosing to focus on the procedural features defining the rule of law while others choose to focus on its higher-level goals, which are not necessarily as explicit. Despite the sometimes-distinct separation between these two sides to the rule of law, many seem to acknowledge their correlation. For example, the increased existence of the procedural features of rule of law in some societies and their increased economic growth, increased peace, less inequality, increased health and more education is used to illustrate this connection as well as the degree to which increased formal manifestations of the rule of law lead to a better society.[31]

Having considered the composition of the rule of law, in terms of both instrumental features and higher-level goals, the next section examines the technology of artificial intelligence.

3 Artificial Intelligence

Artificial intelligence is a concept that is extremely difficult to get a grip of. This is potentially due to the fact that artificial intelligence is a term used to describe so many sub-categories of technologies, all of which are technologically different, yet which share a common goal – namely striving after the ultimate goal of re-creating human intelligence in machines. This section puts forward a number of definitions of artificial intelligence, it then seeks to describe why this

---

[27] Council of Europe Commissioner for Human Rights, Issue Paper, *The Rule of Law on the Internet and in the Wider Digital World*, at p. 8. The Council of Europe also refers to the rule of law test established by the European Court of Human Rights, which has the following formulation: 'To pass these tests, all restrictions on fundamental rights must be based on clear, precise, accessible and foreseeable legal rules, and must serve clearly legitimate aims; they must be "necessary" and "proportionate" to the relevant legitimate aim (within a certain "margin of appreciation"); and there must be an "effective [preferably judicial] remedy" against alleged violations of these requirements'.

[28] Krygier, Martin, *What's the Point of the Rule of Law?*, Buffalo Law Review, Vol. 67, Number 3, May 2019, p. 758.

[29] Ibid, p. 760.

[30] Sannerholm, Richard, *Rättsstaten: skandaler, kriser, politik*, Timbro, 2020, p. 12 (informally translated by the author).

[31] World Justice Project, *What is the Rule of Law?*, available at https://worldjusticeproject.org/about-us/overview/what-rule-law.

technology is often referred to as a 'black box' by focussing on the role of data in decision-making systems as well as by illustrating how a specific form of artificial intelligence, the neural network, learns from data. In this manner the incompatibility between artificial intelligence and the rule of law will be highlighted. This is necessary in order to identify some of the risks associated with this technology and address these in relation to the rule of law.

3.1 In Search of a Definition of Artificial Intelligence

From the outset, the complexities of describing artificial intelligence become apparent when consulting the description provided by Russell and Norvig.[32] They view artificial intelligence from the point of view of four different approaches. In graphically enhancing the different approaches, the various approaches are separated by a vertical and horizontal axis. The definitions above the horizontal axis relate to thought processes and reasoning ('thinking humanly' and 'thinking rationally') and below the horizontal axis to behaviour ('acting humanly' and 'acting rationally'). The definitions to the left of the vertical axis relate to success in terms of fidelity to human performance in other words exactness compared to human performance ('thinking humanly' and 'acting humanly'), and rationality ('thinking rationally' and 'acting rationally'), that is, to what extent artificial intelligence does the 'right thing' given what it knows).[33] Which of these different approaches have traditionally been followed has depended on the academic affiliation - a human-centered approach will require observations and empirical investigations of what human behaviour is while a rationalist approach will typically involve a greater emphasis on mathematics and engineering.[34] Russell and Norvig advocate the approach that defines artificial intelligence in terms of rational action, that is where intelligent agents take the best possible action in a situation.[35] In other words, their definition breaks artificial intelligence down in to two main categories, namely, rational thinking versus acting. This definition of artificial intelligence highlights the fact that artificial intelligence is a dynamic concept to be studied from different scientific approaches and it is defined will depend on the approach it is studied from.

---

[32] Russell, Stuart and Norvig, Peter, *Artificial Intelligence: A Modern Approach*, third edition, Pearson, Education Limited, 2016, pp. 1-2.

[33] Ibid.

[34] Ibid.

[35] Ibid, p. 30. Here it is also worth mentioning that the term 'intelligent agent' too has a certain connotation. According to Russell and Norvig, an intelligent agent, 'is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators … a human agent has eyes, ears, and other organs for sensors and hands, legs, vocal tract, and so on for actuators. A robotic agent might have cameras and infrared range finders for sensors and various motors for actuators. A software agent receives keystrokes, file contents, and network packets as sensory inputs and acts on the environment by displaying on the screen, writing files, and sending network packets', Russell, Stuart and Norvig, Peter, *Artificial Intelligence: A Modern Approach*, third edition, Pearson, Education Limited, 2016, p. 34.

Another description reflecting the academic nature of artificial intelligence is that it is, '[t]the attempt to make computers do the sorts of things human and animal minds can do – either for technological purposes and/or to improve our theoretical understanding of psychological phenomena'.[36] Artificial intelligence is also defined in the following manner:

> AI is the field devoted to building artefacts capable of displaying, in controlled, well-understood environments, and over sustained periods of time, behaviors that we consider to be intelligent, or more generally, behaviors that we can take to be at the heart of what it is to have a mind'.[37]

Over time, and with the incorporation of technologies that comprise the sub-fields of artificial intelligence, such as machine-learning, natural language processing and robotics a few examples, complex technologies encompassing capabilities that resemble elements of artificial intelligence have gained traction in society. These are increasingly being used in autonomous decision-making systems, both of a commercial nature but also mediating the relationship between citizens and public authorities. An example of the former is the autonomous vehicle while the automated payment of social benefits is an example of the latter. This increasing use of artificial intelligence in mainstream society has in turn led to definitions of artificial intelligence that reflect this development, where artificial intelligence is described in terms of systems which are self-learning and autonomous in relation to pre-determined goals. Consequently, this more functional approach to artificial intelligence and its increased use in society has resulted in a more functional definition, especially in sources that have a soft law nature. A definition of artificial intelligence is provided by a working group at the EU level that was also given the task of developing an ethical code for the development and use of artificial intelligence. The definition of artificial intelligence put forward by this group is as follows:

> [a]rtificial intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behaviour by analysing how the environment is affected by their previous actions.[38]

---

[36] Frankish, Keith and Ramsey, William, M., (eds.), *The Cambridge Handbook of Artificial Intelligence*, Glossary, Cambridge University Press, 2014, p. 335.

[37] Arkoudas, Konstantine and Bringsjord, Selmer, *Philosophical Foundations*, in The Cambridge Handbook of Artificial Intelligence, Frankish, Keith and Ramsey, William, M., (eds.), Cambridge University Press, 2014, p. 34.

[38] High-Level Expert Group on Artificial Intelligence, *A definition of AI: Main capabilities and scientific disciplines*, European Commission Directorate-General for Communication, 18 December, 2018, available at https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf.

The OECD defines 'artificial intelligence system' as, 'a machine-based system that can, for a given set of human defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments'.[39] In addition, artificial intelligence is defined in terms of seven different use cases, also known as patterns, that can coexist in parallel within the same AI system. These are hyper-personalisation, conversation and human interaction, recognition, predictive analytics and decisions, goal-driven systems, autonomous systems and patterns and anomalies.[40] Two points are noteworthy here. Firstly, the OECD is seen as that organisation whose definition of artificial intelligence enjoys widespread acceptance. Secondly, the definition of artificial intelligence includes the term 'system'. In other words, the definition progresses from artificial intelligence as an academic subject to systems that are implemented in society.

It is then noteworthy that the European Union has followed the OECD in defining artificial intelligence. However, it has taken the bold step of attempting to define the concept of artificial intelligence in a hard law legal document. Here, reference can be made to the European Commission draft proposal for a regulation on AI that does include a definition of artificial intelligence. According to the regulation, 'artificial intelligence system' means:

> software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.[41]

A proper understanding of this definition in turn requires an examination of Annex I, wherein which a rather technical and broad definition of artificial intelligence is provided:

> (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
>
> (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
>
> (c) Statistical approaches, Bayesian estimation, search and optimization methods.[42]

---

[39] OECD, *Recommendation of the Council on Artificial Intelligence*, adopted on 22 May, 2019, available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[40] OECD, *Artificial Intelligence and Responsible Business Conduct*, available at https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf.

[41] Article 3, European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain European Union Acts, Brussels, 21.4.2021 COM(2021) 206 final, available at 1ProposalforaRegulationlayingdownharmonisedrulesonartificialintelligencepdf.pdf .

[42] Annexe I, European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain European Union Acts, Brussels, 21.4.2021 COM(2021) 206 final, available at 1

The above definition is by no means final, as this legal document is still in its drafting stage. Yet is does provide a hint of which technologies are to be included in a definition of the concept of artificial intelligence system. From a definition of artificial intelligence perspective, the technique of using an annex to describe artificial intelligence is relevant to the extent that an annex is easier to change and this is necessary in order to keep abreast with the changing nature of the technology.[43] It signals that the realm of artificial intelligence is in constant flux and that there may very well be new and exciting emerging technologies incorporating elements of artificial intelligence, around the corner and that require regulating.

Reflecting on the above discussion, it becomes apparent that artificial intelligence is subjective and dynamic. The discussion is rather academic and to a certain degree rather abstract, yet it raises an important issue, namely, that any discussion of artificial intelligence requires the initial question, 'what do you mean by artificial intelligence?'. Another consideration to reflect upon when considering technological development and emerging technologies incorporating elements of artificial intelligence, is what exactly is trying to be achieved. For example, attempting to create technology with rational thought may come into conflict with human thought, to the extent that humans do not necessarily think rationally, or with the same rationality as artificial intelligence. It may also account for the inability for humans to accept certain decisions by artificial intelligence and it can also account for the fact that the decisions made by artificial intelligence are not necessarily aligned with the manner in which legal norms and principles have come about.

3.2 Artificial Intelligence a Black Box

The above section provided some definitions of artificial intelligence, yet these do not really describe how this technology works. In other words, what is the essence of this technology? In this regard, two issues require attention. Firstly, artificial intelligence is essentially all about data and algorithms. While this description of artificial intelligence seems rather simplified, it is these two components that lie at the core of machine learning.[44] This simplified explanation is misleading to the extent that it fails to emphasise the quantity of data that lies behind machine learning techniques as well as the mathematical complexity associated with modern algorithms. The next section delves deeper

file://storage/home/lylu/Downloads/regulation_annex_ai_875FDD6D-CC6A-E50A-8E48824677EFED42_75789%20(1).pdf.

[43] Center for Strategic and International Studies, *The European Approach to Regulating AI*, interview with Dragos Tudorache, Member of the European Parliament and Co-Rapporteur of the EU AI Act, available at https://www.youtube.com/watch?v=BBmq4T_550U.

[44] This is confirmed in a White paper put forward by the European Commission, which states that this technology entails the extraction of knowledge from data using mathematical algorithms. European Commission, White Paper on Artificial Intelligence – A European Approach to Excellence and Trust, Brussels, 19.2.2020 COM(2020) 65 final, available at https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf, p. 16.

into the reasons for associating artificial technology with a black box, both legal factors and technical reasons relevant in this regard.

3.2.1 Opacity Created by Law

An initial reflection is the extent to which the law is a factor in creating the black box of technological development, including artificial intelligence. In general, the laws promotion of intellectual property rights plays a role in this. Generally, a large percentage of technological development takes place within private companies and this proprietary technology for the most part enjoys legal protections via intellectual property. An example of this trend is illustrated by means of referring to the General Data Protection Regulation (GDPR). Article 22 of the GDPR is that article that regulates artificial intelligence. Without delving deeper into this article, what is relevant is the manner in which Recital 63 (expanding on Article 22), affords the data subject a right to know and receive communication regarding the logic behind any data processing in relation to automated decision-making. This right to an explanation potentially grants the data subject a transparency right in order to access the technology on which an intelligent agent is based. However, this right is watered down in the very same recital where trade secrets and intellectual property rights take precedence over transparency.[45]

While this can be problematic in general, the challenges increase when this technology is used in the public administration as well as justice system. In this context the potential for the violation of multiple basic rights afforded citizens in relation to state functions increases.

3.2.2 A Data-Driven Science

Data is an important element driving artificial intelligence and its sub-domain of machine learning. Typically, artificial intelligence has seemingly predictive capabilities. However, it is not a crystal ball that can see into the future. Rather, artificial intelligence has the ability to learn from past examples and apply this knowledge to the future. In this regard, the examples of past action are in the form of data and it is data that is at the heart of the machine-learning capability of artificial intelligence.

This in turn has led to a novel approach for extracting knowledge from data, namely the 'data driven' approach. A traditional approach to gaining insights from data has been a 'theory-based' approach, where a hypothesis is devised (typically by a human) and then this hypothesis is tested against a data source, this approach useful in finding a reason for a particular event.[46] For example, a

---

[45] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4.5.2016.

[46] Custers, Bart, *Data Dilemmas in the Information Society: Introduction and Overview*, in Custers, Bart et al. (eds.), Discrimination and Privacy in the Information Society – Data Mining and Profiling in Large Databases, Springer, 2013, at p. 9.

hypothesis may be that milk will sell better if placed as far back in a store as possible and after moving the milk, sales data can prove or disprove this hypothesis. However, with the advent of big data and increasingly complex machine learning algorithms used to analyse this data, the 'data-driven' approach to knowledge extraction has come to the fore. Instead of inventing a hypothesis to be tested against the data, the data is randomly analysed for connections and correlations between data points. The correlations may be novel in that humans may never have thought of making such connections and it was only by utilizing tools from within the spheres of machine learning and artificial intelligence, that one was able to extract this knowledge.[47] A hypothetical example is where machine learning is used to attempt to reduce crime. Law enforcement may require a starting point to solve a murder. Analysing a data set, machine learning techniques may learn that 'men who have a size 42 shoe and blue eyes are more likely to commit the crime of murder'. This statement may be illogical from a human rationality perspective and even more so from a legal point of view, a clear causal connection usually demanded by most established legal systems missing. However, despite the inability to back this statement using human thought processes, it may be logical from the techno-rational point of view, to the extent that this is a fact or truth as identified by an algorithm and learned from the data. Quite simply, the data used to train a model for solving crime may have contained examples where the murderer was a 'male who had a size 42 shoe and blue eyes'. This has also changed the working methods of law enforcement to the extent that an initial step to solving a crime is to use the insights from data as a starting point, in this case starting with all men with a shoe size 42 and blue eyes. However, herein lies a major problem: data is full of bias and the humans who work with the data may also be biased. If data is collected from a source that is biased, that bias will be reflected in the data and consequently in any product developed using that data set. For example, data collected about criminals in a correctional facility existing in an unjust society will reflect any societal bias that led to those prisoners being incarcerated in the first place, any eventual technology developed using that data, perpetuating this bias and injustice reflected in the justice system. Also, there is an element of bias already at the stage where the learning data set is chosen – it may just be that that particular data set includes more examples where murderers happened to be 'men who had a size 42 shoe size and blue eyes'. And just to complicate matters and make the technology even more inaccessible, the correlations between data points is not based on fact but rather on probability and propensity. In other words, referring to the example above, artificial intelligence can only state that men with a 42-shoe size and blue eyes are more likely to commit a crime and express that as a percentage – but it cannot predict with one hundred percent certainty that a person will commit a crime. The problem is that this knowledge is just a probability but is treated as a complete truth and not just an estimation of propensity. Ellul argues that technology has acquired an autonomy from its

[47] Ibid.

association with the legitimacy of scientific progress in general and technology has a legitimacy due to the perception that it is scientific and objective.[48]

The challenge with bias is its close proximity to discrimination. Bias is said to be present when:

> the available data is not representative of the population or phenomenon of study […][that] [d]ata does not include variables that properly capture the phenomenon we want to predict [and that] [d]ata includes content produced by humans which may contain bias against groups of people'.[49]

The problem with bias and discrimination in a data context is 'masking', that is where two characteristics are correlated, the one trivial and the other sensitive, and where the former is used to indicate the presence of the latter.[50] A typical example is using area code (zip code) to denote health status, socio-economic factors determining factors that are common for a particular group. Bias should also be distinguished from discrimination, which is a legal concept described as, 'the prejudiced treatment of an individual based on their membership in a certain group or category', where the features encompassing discrimination include race, ethnicity, religion, nationality, gender, sexuality, disability, marital status, genetic features, language and age.[51] Consequently, decision-making systems are discriminatory in situations where two individuals have the same characteristic relevant to a decision making process, yet they differ with respect to a sensitive attribute, which results in a different decision produced by the model.[52] Bias and discrimination are therefore related to the extent that bias in data can lead to discriminatory effects.

It has been argued that the data-driven characteristics of artificial intelligence cause friction with legal notions such as the rule of law. In this regard, Bayamlıoğlu and Leenes argue that three manifestations of this friction relate to law as a normative enterprise, law as a causative enterprise and law as a moral enterprise, the end result being a shift from the rule of law to the rule of technology.[53] Concerning the first point, they write that, '[r]ules, principles, standards and in general 'norms' provide uniformity, predictability, and social coordination for they inform individuals about their way of conduct, and explain

---

48 Ellul, Jacques, *The 'autonomy' of the technological phenomenon*, in Scharff, RC and Dusek V., (eds), Philosophy of Technology: The Technological Condition, Malden, Blackwell Publishing Ltd, 2003, pp. 386-395.

49 For an in-depth description of the different types of bias that can occur, see Prabhakar Krishnamurthy, *Understanding Data Bias*, Towards Data Science, https://towardsdatascience.com/survey-d4f168791e57, 2019.

50 Custers, Bart, (n 44), pp. 9-10.

51 Calders, Toon and Zliobaite, Indre, *Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures*, in Custers, Bart, et al. (eds.), Discrimination and Privacy in the Information Society – Data Mining and Profiling in Large Databases, Springer, 2013, p. 44.

52 Ibid.

53 Bayamlıoğlu, Emre and Leenes, Ronald, *The 'rule of law' implications of data-driven decision-making: a techno-regulatory perspective*, Law, Innovation and Technology, 2018, VOL. 10, NO. 2, 295–313, available at https://doi.org/10.1080/17579961.2018.1527475.

the legal course of events in situations addressed by the Law'.[54] However, artificially intelligent systems, basing their decisions on data and correlation, lack these norms that have been painstakingly developed by the various institutions making up the legal system. Concerning the second point, the notion of legal causation is replaced by correlation. In other words, the connection between action (or lack thereof) and effect that can be explained, shown or proven are replaced by correlations that are hidden deep within the mathematical complexity. The authors write that, '[l]egal effects are not a matter of correlation between certain facts and effects, but of (legal) causation, or rather the law creates (constitutes) legal effects.[55] Finally, the third point referred to by the authors refers to the fact that not only does the normative character of law suffer but also the moral grounds that the norms are based on, the erosion of human autonomy one such example.[56]

The challenges associated with the data-driven approach are multiple. Data in general contains bias, which is difficult to eradicate. Any potential bias in those who work with the data or even in the context from which the data was collected, will result in a transfer of this bias to an end product. Also, considering the amount of data used to train algorithms, insight into the reasons for a decision is dramatically reduced. Finally, the ability to contest decisions is impeded, the correlations just that, and not an undisputed truth.

3.2.3 A Technology Fraught with Complexity

One of the reasons for using the metaphor of the black box is because of the mathematical complexity that artificial intelligence is composed of in relation to human cognitive abilities. Humans simply do not have the ability to follow the mathematical intricacies or understand precisely what is occurring deep within the black box. It is the complexity associated with artificial intelligence that forms the basis of its incompatibility with the procedural features of the rule of law.

While artificial intelligence can potentially comprise a wide variety of technologies or combinations thereof, one technology that illuminates the complexity of artificial intelligence is that of deep learning, the neural network a form of deep learning that has catapulted artificial intelligence into mainstream use. Above, the importance of data and examples within that data was highlighted. For example, a credit institution will have data of historical examples of who was not able to repay (and defaulted on) a loan under various circumstances. By learning from these examples, a model can be built to predict who will default on their loan in the future. Deep neural networks are mathematical techniques for learning from examples that is modeled on the human brain. Just as the human brain has neurons connected to each other by means of synapses, the 'neurons' of the artificial intelligence neural network, which can also be described as general processing units, are also connected to

---

54 Ibid.

55 Ibid, p. 307.

56 Ibid, p. 309.

each other in the form of thousands of layers. Just as the human brain functions by increasing and decreasing the activity between the brain's neurons, so too is it with the mathematical connections between the deep neural networks 'neurons' that are at the core of the machine learning process – as these mathematical connections are formed, removed, strengthened and weakened the deep neural network learns, just as humans do.[57] The neurons of the deep neural network are also structured in the form of 'layers'. There is an input layer, an output layer and potentially thousands of hidden layers in-between. The input given to a deep neural network's input layer passes through each subsequent layer until finally the output is provided by the output layer.[58] Neural networks learn just as children learn: wishing to teach a neural network to identify a picture of a circle, it is fed thousands of pictures of circles, with the pictures of circles 'labelled' as representing a circle (this requires a human to teach the algorithm what a circle looks like); then testing the ability of the neural network to recognize pictures of circles, it is fed pictures of all types of shapes with the task of identifying the pictures with circles in them; where the neural network is fed a square and it answers with a circle, it is informed that it got it wrong and it automatically adjusts the complex mathematical weighting structures at its inner layers in order to 'learn' and increase its accuracy the next time; when the accuracy is deemed good enough, it is tested against a test data set to measure accuracy and eventually it is put to work in the digital environment in order to identify circles, which it will do with a certain degree of probability. Considering that a neural network may have thousands of hidden layers with many nodes, all connected to each other with differing mathematical weights that are altered as it learns from new iterations, it is impossible to say with a one hundred percent certainty how a neural network reached a decision, that is, how the output from the outer layer was reached. It is in this context that the terms 'deep learning' or 'deep neural networks' arise, networks that essentially learn by being fed data and information about this data.[59]

Considering the complexities of the technology, and the fact that the technology is based on probability and not certainty, one risk is that mistakes will occur. Palmiotto, analysing the use of artificial intelligence in courts of law, refers to mistakes as 'miscodes' and argues the point that these miscodes, in light of the inaccessible technology, impact the ability to guarantee a fair trial.[60] In classifying the mistakes that can occur, the author refers to the 'technical miscode' (errors affecting accuracy), 'scientific miscode' (algorithms containing dubious scientific methods or value judgements affecting validity) and 'legal miscode' (unlawful design affecting legality).[61] The lack of insight into the black

---

[57] David, Eli, *Deep Learning for Dummies*, John Wiley & Sons, Inc, New Jersey, 2018, pp. 22-23.

[58] Ibid, p. 26.

[59] Modin, Anton and Andrén, Jacob (eds.), T*he Essential AI Handbook for Leaders*, Peltarion AB, 2019, p. 63.

[60] Palmiotto, Francesca, *The Black Box on Trial: The Impact of Algorithmic Opacity on Fair Trial Rights in Criminal Proceedings*, in M. Ebers, M. Cantero Gamito (eds.), Algorithmic Governance and Governance of Algorithms, Data Science, Machine Intelligence, and Law 1, https://doi.org/10.1007/978-3-030-50559-2_3.

[61] Ibid.

box prevents any investigation of the technology and in the long run can create an environment of mistrust.

It becomes apparent that artificial intelligence it is an umbrella term for many sub-technologies. Artificial intelligence is a black box for many reasons, one of which is the complexity associated with it. Considering this complexity, and the opaqueness of the black box, it is impossible to know exactly how decisions are arrived at by artificial intelligence. Despite this opacity, artificial intelligence is being used in many realms of society, the judicial setting one example. The simple solution, it would seem, is to open up the black box. However, this approach fails to acknowledge the problem in relation to complexity. This is eloquently addressed by Karnow, who states the following:

> Negligence and strict liability were born and raised in a Newtonian universe, the universe of billiard balls hitting billiard balls, car hitting cars; force, mass and reaction; and machinery executing one step at the time. The risk discernible in this world are the consequences of Newtonian mechanics, which is linear: A causing B causing C … With autonomous robots that are complex machines, ever more complex as they interact seamless, porously, with the larger environment, linear causation gives way to complex, nonlinear interactions … the problem is not ignorance; the problem is the limits of knowledge.[62]

Even with transparency providing access to the black box of technological development, this is no guarantee that the technology would even be understood. From the perspective of the law, where the linear causal connection is an important notion, it is problematic that we rely on technology where the logic is hidden deep inside its mathematical core. This said, this technology is starting to infiltrate judicial settings. The characteristics of artificial intelligence – data-driven correlations, fraught with bias, feeding neural networks that are self-learning and steeped in complexity – lack compatibility with the procedural features of the rule of law. Nevertheless, it is being used in many contexts within society, courts of law just one such example.

4 Black Boxes in The Judiciary

Thus far the notions of the rule of law and artificial intelligence have been discussed independently. The next task is to connect these two notions in relation to the main goal of this paper, namely to illustrate how artificial intelligence is eroding the procedural features considered necessary for the existence of the rule of law. Considering the wide ambit of both these concepts, it is argued that the most effective means of bridging them and illustrating how the rule of law is diminished by artificial intelligence is by reference to the case of COMPAS, a criminal law case in the United States.

---

[62] Karnow, Curtis E.A., *The Application of Traditional Tort Theory to Embodied Machine Intelligence*, from the selected works of Curtis E. A. Karnow, available at https://works.bepress.com/curtis_karnow/9/, p. 15.

4.1 The Case of COMPAS

In the matter of *State v. Loomis*, a Wisconsin trial court sentenced the defendant to six years in prison for a criminal act, the corresponding sentence in part determined by 'algorithmic risk assessment'.[63] In criminal matters in the United States judicial system, it is common procedural practice that judges are provided with a presentencing investigation report (PSI) that provides background information about the defendant and includes an assessment of the risk of recidivism based on this report. In *State v. Loomis*, the PSI incorporated an algorithmic assessment report. This took the form of software which was referred to as COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) and was developed by Northpointe Inc., a private company, where the output comprises a number of bar charts depicting the risk for recidivism on the part of the accused, in other words the risk that the accused will commit crimes in the future.[64] Accompanying the PSI was also procedural safeguard in the form of a written statement to the judges concerning the risks associated with pretrial risk assessments of this kind. After being found guilty, Loomis appealed the sentence as well as the use of COMPAS. The Wisconsin Supreme Court subsequently upheld the lower court's decision, stating that the use of the algorithmic risk assessment software did not violate Loomis's right to due process even though it was not made available either to the court or Loomis.[65] It has also become apparent that COMPAS assesses variables under five main areas: criminal involvement, relationships/lifestyles, personality/attitudes, family, and social exclusion.[66] It is these variable that determine the recommendation suggested by COMPAS. Subsequently, and upon

---

63 State v Loomis, No 2015AP157-CR (Wis Ct App Sept 17, 2015), 2015 WL 1724741 (Supreme Court of Wisconsin). See also Harvard Law Review, Criminal Law (2017), State v. Loomis, 130 Harv. L. Rev. 1530, available at https://harvardlawreview.org/2017/03/state-v-loomis/.

64 There are currently three systems being used for predicting criminal behaviour in the United States context: Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), Public Safety Assessment (PSA) and Level of Service Inventory Revised (LSI-R). LSI-R, developed by Canadian company Multi-Health Systems, takes factors such as criminal history and personality patterns into account, Electronic Privacy Information Centre, Algorithms in the Criminal Justice System: Pre-Trial Assessment Tools, https://epic.org/algorithmic-transparency/crim-justice/. Another tool used in the judicial system is that of the Harm Assessment Risk Tool (HART), an algorithmic decision-making support system used in the United Kingdom for assisting in custody matter, one accusation levied at the system being that it is discriminatory against the poor, see Wired, *UK police are using AI to inform custodial decisions – but it could be discriminating against the poor*, available at https://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit.

65 Harvard Law Review, Criminal Law (2017), *State v. Loomis*, 130 Harv. L. Rev. 1530, available at https://harvardlawreview.org/2017/03/state-v-loomis/.

66 Electronic Privacy Information Centre, *Algorithms in the Criminal Justice System: Pre-Trial Assessment Tools*, https://epic.org/algorithmic-transparency/crim-justice/.

request, Northpointe Inc. refused to make the software available citing that it was proprietary and a core business secret.[67]

4.2 COMPAS and the Rule of Law

The case of COMPAS is useful to the extent that it highlights many different ways in which the procedural features of the rule of law are to varying degrees undermined by technology. It also casts light on the risks associated with such technologies by virtue of the composition of artificial intelligent technologies. Without precise knowledge of the composition of the technology underlying COMPAS (it is after all a black box), in this context it is used to illustrate the challenges that technologies incorporating elements of artificial intelligence pose to traditional legal notions, such as the rule of law.

Just by considering the nature of the technology of artificial intelligence, the use of COMPAS becomes problematic. It is uncertain how COMPAS is technologically designed and just exactly what machine learning techniques it is comprised of. It is uncertain just how complex the technology is as the law itself renders it a black box. The obstacle of proprietary software arose in the Loomis case where the applicant asserted that he had the right to information that the trial court had used at sentencing, but that the proprietary nature of COMPAS prevented this. The reply of the Supreme Court was that,'[…] Northpointe's 2015 Practitioner's Guide to COMPAS explains that the risk scores are based largely on static information (criminal history), with limited use of some dynamic variables (i.e. criminal associates, substance abuse).' In addition, the court argued that the COMPAS score was based on questions that the appellant himself had answered, which gave him access to information upon which the risk assessment was made.

In addition, data is contextual and reflects societal values, whether these are positive or negative. For example, the data used to develop COMPAS came from the correctional services context and applied in another context, namely the judicial context. Also, should there be any kind of bias in the penal system as a whole, these will be mirrored in a system based on that data from that penal system. For example, should a certain ethnic group be treated differently by the system, this anomaly will be replicated in the technology developed upon that system.

Subsequent to the *Loomis* case, the use of artificial intelligence in the justice system in the United States has received increased media attention. This especially since ProPublica, having examined the outcomes of cases where algorithmic risk assessments have been used, has claimed that the statistics are starting to identify a racial bias in decisions, where White people were treated more favourably than African Americans.[68] First, examining 7000 decisions, the results showed that the algorithm is only 20 percent successful in accurately

---

[67] Liptak, Adam, *Sent to Prison by a Software Programe's Secret Algorithm*, The New York Times, 2017, available at https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html

[68] Angwin, Julia, Larson, Jeff, Mattu, Surya and Kirchner, Lauren, *Machine Bias*, ProPublica, available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

predicting recidivism. Second, the algorithm incorrectly flagged African Americans at twice the rate of White people.[69] Here one can argue that technologies such as COMPAS breach the Venice Commissions principle of 'equality before the law and non-discrimination'. The Venice Commission treats equality before the law and non-discrimination as an essential element of the rule of law.[70] Here grounds for discrimination include race, colour, sex, language, religion, political or other opinion, national or social origin, association with national minority, property, birth or other status.[71]

In addition, one can analyse the use of COMPAS in accordance with the Venice Commission principles addressed above. One principle is that of 'legality', which establishes supremacy of the law.[72] One can contemplate to what extent the law is supreme when private companies are given the task of transposing it into black boxes to be used in the judiciary. This not only places these companies in a position of supremacy, but is also challenges the principle of 'legal certainty', where the law is not certain (the law is unknown as it resides in the black box) , it may not be foreseeable (artificial intelligence is self-learning and the black box is designed to update itself) and even if the public has access to the black box, even data scientists do not understand the inner workings of deep neural networks, so how could the law be easy to understand? According to the Venice Commission, 'equality before the law and non-discrimination' are central to the rule of law. However, research focussing on decisions made by COMPAS have revealed that there seems to be built-in discrimination, thereby challenging the system to the extent that not everyone is equal before the law. In addition, 'access to justice' implies an independent and impartial judiciary and the right to have a fair trial. To what extent can the judiciary be deemed independent when the tools it uses to help assess recidivism are proprietary software that is not open to public scrutiny? Another aspect of 'access to justice' is that justice be seen to be done, but how is this possible when the algorithms and data of tools like COMPAS are black boxes?

In addition, access to justice recognizes the principles of 'nullum crimen sine lege' and 'nulla poena sine lege' which recognize that there is no crime or

---

69 Ibid. The statistics showed that of those labelled 'high risk but did not re-offend', 23.5% were White and 44.9% African American and of those labelled 'low risk yet did re-offend, 47.7% were White and 28% African American.

70 European Commission for Democracy Through Law (Venice Commission) 2016, p. 18.

71 Ibid. Here mention can also be made of a study that compared the use of machine learning algorithms and SAVRY ((Structured Assessment of Violence Risk in Youth), a regular tool used for recidivism prediction. The two indicators compared were that of predictive performance and fairness. Here research showed that while the machine learning techniques were slightly more effective, SAVRY was in general fairer, the machine learning models tending to discriminate against male defendants, foreigners, or people of specific national groups. For example, foreigners who did not recidivate were almost twice as likely to be wrongly classified as high risk by the machine learning models than Spanish nationals. See Tolan, Songül, et. al., *Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia*, Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law - ICAIL '19, ACM Press, 2019, available at https://chato.cl/papers/miron_tolan_gomez_castillo_2019_machine_learning_risk_assessment_savry.pdf.

72 Venice Commission, available at https://www.venice.coe.int/WebForms/pages/?p=02_Rule_of_law&lang=EN.

punishment without a law. Taking the greater justice system into account, artificial intelligence is being used to determine who has the propensity to become a criminal even before they have actually performed a criminal act. For example, in the US, artificial intelligence is being used in a joint project between the Memphis Police Department and the University of Memphis. Called 'Blue Crush' (Crime Reduction Using Statistical History), a new methodology is being used to detect crime in advance, whereby police investigators, sociologists and mathematicians are combining their expertise and utilizing technology in order to predict what crime will take place, when and where.[73]

While the above initiative is on a general level, incentives are already in place to judge people on an individual level. For example, artificial intelligence is being used to assign individuals a 'threat score', that is the extent to which a specific person poses a threat to their surroundings and to law enforcement that may be called to a disturbance. One such initiative from Fresno California, USA, uses the software called 'Beware' in order to give specific individuals a score, the data fed to the predictive algorithm originating from official public data sources and combined with searches of social media and other internet websites. Should an individual contact law enforcement in connection with a disturbance, the software automatically creates a profile of that person in connection with that specific address (the same applicable to vehicles). In addition, Beware can make a profile of all the inhabitants in surrounding areas from which a call was made, which means that if a person calls the police in connection with a disturbance, that person's neighbours too will inevitably be associated with a score, even if they had absolutely no involvement with the disturbance.[74]

Another example is from Chicago, where police have embarked on a violence reduction strategy and consequently use predictive software including algorithms in order to create a 'Strategic Subject List' (SSL), which is also referred to as the 'heat list' and which contains the names of 400 individuals considered most dangerous in the city. The list is created by algorithms that use historical crime data, disturbance calls, suspicious person reports but also social media activity in order to compile the list. The people on the list are not necessarily violent criminals. Rather, they have made it onto the list because they have been arrested with a person that has previously been arrested for a violent act. For example, a person may be arrested for a minor offence but is in the company of a person previously arrested for a violent crime. This connection will automatically ensure that this person lands on the list of potential violent people even though he or she has never committed a violent act in his or her life. What is interesting here is the technology behind the 'heat list', in that it is not necessarily only the actual names of a person's acquaintances in the social media that result in a person landing on the 'heat list', but the underlying structure of the social media networks is also relevant in this regard.[75] These examples highlight the extent to which access to justice and fairness are challenged and raise fundamental questions concerning the state of established principles

---

[73] Vlahos, James, *The Department of Pre-Crime*, Scientific American, January 2012, at p. 50.

[74] Stanley, Jay, ACLU, *Eight Problems with Police 'Threat Scores'*, available at https://www.aclu.org/blog/privacy-technology/surveillance-technologies/eight-problems-police-threat-scores.

[75] Ibid.

associated with the rule of law, such as innocent until proven guilty. For example, the use of algorithmic risk assessments in criminal trials in order to determine recidivism raises the question of whether the accused is deemed guilty of a potential crime, that is the propensity to commit a crime before it has actually occurred.[76] The presumption of innocence and right to a fair trial are encompassed in the checklist regarding access to justice of the Venice Commission.

Another challenge to the traditional view of the rule of law is the extent to which the judiciary, relying on artificial intelligence developed by private corporations, can be deemed independent. The Venice Commission demands that there should be legal guarantees in order to secure the independence of the judiciary. Independence, according to the Venice Commission is taken to mean a judiciary 'free from external pressure'.[77] While the corporations that produce algorithmic risk assessments may not directly exert pressure on judges, a question that requires raising is to what extent people (judges, jurors, and parole officers) will dare go against a risk assessment made by technology. This in turn brings to the fore issues of a philosophical nature where technology is granted a degree of autonomy.

The above section illustrates the extent to which technological developments are negating the existence of the procedural features of the rule of law. This can be illustrated in relation to some of the procedural features put forward by Fuller and referenced above: feature three states that rules require publishing, however, the rules deep inside the black box remain copyright protected and a trade secret, their proprietary nature preventing any form of transparency or oversight, the situation compounded where artificial intelligence is used within the justice system; feature four requires rules to be intelligible, however, the rules of artificial intelligence in their mathematical form are hardly intelligible to data scientists and mathematicians, the inner layers of the deep neural network highly inaccessible in relation to human cognitive ability; feature seven requires rules that do not constantly change, however, the rules of artificial intelligence are designed to be just that – in constant change and in less than a fraction of a second – and as the self-learning algorithms constantly update themselves and as they grind through an ever-increasing volume of data where one single new data point can radically alter a decision, the mathematical rules of artificial intelligence change in order to allow them to achieve their pre-defined goals with ever-increasing accuracy, and this all in the name of creating a more effective society; and finally feature eight requires congruence, however, the notion of congruence itself seems outdated where either the decision-making processes are transferred from public officials to machines or where the suggestions offered by the black box result in a blind signing-off on decisions essentially taken by the technology, access to the black box inhibited by complexity and as well as legal constraints. In putting forward his eight features of the rule of law, Fuller states that a failure to uphold any one of these features does not result in a bad system of law – rather, it cannot be called a legal system at all.[78]

---

[76] European Commission for Democracy Through Law (Venice Commission), 2016, p. 16.

[77] Ibid, p. 20.

[78] Fuller, Lon, *The Morality of Law*, Revised Edition, Yale University Press, New Haven and London, p. 39.

It is with this development in mind that that one is required to contemplate the extent to which one is able to uphold the procedural features of the rule of law as the rules of the black box of artificial intelligence are allowed to take precedence.

5 Concluding Remarks

This article essentially described the relationship between technology, represented by artificial intelligence, and law, represented by the notion of the rule of law. In describing the rule of law, focus was placed on the procedural features of the rule of law, the existence of which ensure the attainment of the higher-level goals of the rule of law in relation to a desirable society. The discussions of what these higher-level goals of the rule of law are, are aplenty, however it can be argued that ultimately the rule of law takes the form of tangible procedural features. The technology of artificial intelligence is difficult to define considering the many sub-categories of technologies it represents and there are also conflicting ideas in relation to what artificial intelligence actually should be trying to achieve – are we striving after intelligent agents that are rational or intelligent agents that think like humans? This discussion, while academic in nature is relevant, and just as the rule of law as a legal notion revolves around the idea of striving after a society that allows humans to flourish, so too is a discussion required surrounding what weight should be given to values that humans embrace. Do we want to be surrounded by intelligent agents the logic of which we do not understand or do we rather want to be surrounded by intelligent agents that think like humans, in effect taking human moral and ethical values into account? It is this perspective that seems to get lost in the hype surrounding the implementation of artificial intelligence for the sake of making society more cost effective and building technology with a rule of law-perspective would no doubt result in more human-centric technology.

However, at present the black box of artificial intelligence is not calibrated with the procedural features that characterise the rule of law. The main argument of this paper can be summarised in the following manner. A precondition for the existence of the rule of law is the existence of a number of procedural features that have been promoted by scholars such as Lon Fuller. The problem is that the nature of the black box of artificial intelligence accompanied by the legal protection it is afforded, makes it almost impossible for these procedural features to exist. The rules of law in effect become the rules of technology and the rules of technology are not public, they are in constant change, they do not promote fairness and in effect dictate the conditions of application, the notion of congruence becoming obsolete. If the rule of law, despite its outwardly procedural form, is arguably a mechanism to attain a certain type of society, where individuals can flourish and be themselves, a conclusion must surely be that artificial intelligence is an existential threat – a threat to the rule of law in its current analogue and procedural form, and in the long run to society at large.

It seems that we are at a crossroads – a junction where some decisions are required about what type of society we want to live in. Do we want to live in the black box society, where the hidden rules of technology govern or do we want to live in a society where humanity is at the centrum and allowed to flourish? If it is now that the procedural features of the rule of law are slowly being eroded

by artificial intelligence, yet the higher-level values that the rule of law promotes are still desired, what then is the response? It seems that there are two main routes forward: either the procedural features of the rule of law are embedded into the technology, indirectly promoting the higher-level values of the rule of law or the technology is developed in a manner that secures the higher-level values promoted by the rule of law, which may not necessarily require the adherence to the procedural features to the extent that the higher-level values can be secured by novel mechanisms. However, unless the rule of law is in some manner designed into the technology, there is a risk that the rule of law as a legal notion will fade away and with it a society in which humans are allowed to flourish.