The Preservation of Rule of Law in the Era of Artificial Intelligence (AI)

Chengzheng Chen

Fujian New-Stone Law Firm, Fuzhou, 350001, China

Keywords: Artificial Intelligence (AI), Machine Learning (ML), Rule of law, Judicial decision-making systems, Explainability

Abstract: The study of law and information technology comes with an inherent contradiction in that while technology develops rapidly and embraces notions such as internationalization and globalization, traditional law, for the most part, can be slow to react to technological developments and is also predominantly confined to national borders. However, the notion of the rule of law defies the phenomenon of law being bound to national borders and enjoys global recognition. However, a serious threat to the rule of law is looming in the form of an assault by technological developments within artificial intelligence (AI). As large strides are made in the academic discipline of AI, this technology is starting to make its way into digital decision-making systems and is in effect replacing human decision-makers. A prime example of this development is the use of AI to assist judges in making judicial decisions. However, in many circumstances this technology is a 'black box' due mainly to its complexity but also because it is protected by law. This lack of transparency and the diminished ability to understand the operation of these systems increasingly being used by the structures of governance is challenging traditional notions underpinning the rule of law. This is especially so in relation to concepts especially associated with the rule of law, such as transparency, fairness and explainability. This article examines the technology of AI in relation to the rule of law, highlighting the rule of law as a mechanism for human flourishing. It investigates the extent to which the rule of law is being diminished as AI is becoming entrenched within society and questions the extent to which it can survive in the technocratic society.

1. Introduction

The study of law and information technology comes with an inherent contradiction in that while technology embraces notions such as internationalization and globalization, the law, for the most part, is to a certain extent still confined to national borders. Transgressing this contradiction to a certain degree is the notion of the rule of law that carries within it the ideal that 'the "rule of law" is good for everyone', an attitude that seemingly enjoys international support.It is conceded that this overwhelming support for the rule of law is based on differing interpretations of what the rule of law is and in some cases it may even be hijacked by those who wish to use it as a smokescreen to hide practices that in fact contradict its ideals. Nevertheless, for the most part, the rule of law still carries the ideal of being, 'analogous to the notion of the "good", in the sense that everyone is for it, but have contrasting convictions about what it is'.If the rule of law, therefore, is a notion that is worth retaining as a measurement of a 'good' that is worth striving after, it should be disconcerting that for a second year running, it declined in more countries than it improved in, denoting an overall weakening of the rule of law worldwide[1].

However, a second more concealed threat is increasing as society becomes increasingly digitalizes. This is the threat from technology, more specifically technology containing elements of artificial intelligence (AI). As large strides are made in the academic discipline of AI, this technology is starting to make its way into digital decision-making systems, which in turn are replacing human decision-makers, institutions, both public and private, seeking increasing effectivity. Human decision-making is currently being assisted by digital decision-making systems and this function is increasingly being given to machines, the sphere of governance no exception. The threat to the rule of law lies in the fact that most of these decision-making systems are 'black boxes' because they incorporate extremely complex technology that is essentially beyond the

  DOI: 10.25236/ieesasm.2023.041

cognitive capacities of humans and the law too inhibits transparency to a certain degree. It is here that the demands of the rule of law, such as insight, transparency, fairness and explainability, are almost impossible to achieve, which in turn raises questions concerning the extent to which the rule of law is a viable concept in the technocratic society.

Section 2 of this article provides a brief description of the rule of law in order to provide a general overview of this complex concept and on which a subsequent analysis will be based. Section 3 focuses on the technological concept of AI, illuminating the complexity and opaqueness of these technologies. Section 4 provides examples of applications using this technology, the justice system just one such example. Section 5 provides an analysis wherein the extent to which AI is challenging the ideals of the rule of law is depicted. Finally, Sect. 6 concludes this article.

2. The rule of law as an ideal

There are many interpretations of what the rule of law actually is, which in turn complicates defining it precisely. A classical dictionary definition of the rule of law describes it as, '...the mechanism, process, institution, practice, or norm that supports the equality of all citizens before the law, secures a nonarbitrary form of government, and more generally prevents the arbitrary use of power'[2].That everyone is equal in the eyes of the law means that everyone is subject to the law and no-one is above the law, in other words that everyone, no matter who you are, is subject to the laws of a state. Besides having the function of curtailing state power, another perspective of the rule of law defines it not only in terms of the characteristics that a legal system should encompass but also in terms of justice in society in general, human rights being one such value[3].

A central tenet of the rule of law is that it embodies a notion of reciprocity between those that govern and those that are governed. On the one hand, those in positions of authority must exercise this authority according to established public norms and not arbitrarily (government must act within the confines of the law) and on the other hand, 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[4].

The rule of law discourse is often defined by the distinction made between the rule of law's formal requirements and the material aspects that it is purported to encompass. This is reflected in the numerous theories of the rule of law, where some view the rule of law as a concept comprised purely of formal structures of governance, these theories reflecting the concept of legal positivism, while others recognize it as including moral considerations.

Dworkin illuminates the distinction between the rule of law as comprising the existence of formal institutions of governance against the notion of it comprising considerations of morality, the first referred to as the 'rule-book' conception and the latter as the 'rights' conception. The former states that the power of the state should be exercised against individuals only where this is based on rules that have been made public. Both government and citizens must then abide by these rules until they are changed according to the rules for change that have also been made public. This conception does not say anything about the nature of the rules in the 'rule-book', this being related to substantive justice. The rights conception assumes that people have moral rights and duties with respect to one another and political rights against the state. These moral rights are required recognition in the positive law in order that people may enforce them through the courts or other institutions. It therefore disregards the distinction between formal requirements and the requirements of justice, requiring the rules in the rule-book to take heed of substantive and moral requirements[5]. The rule of law can also be described in terms of function, where it is argued that there is a limitation to studying the notion of the rule of law as an object, the question of its importance for the goals of development paramount as well as how these are to be achieved[6]. Simmonds uses the metaphor of the spoon: in explaining what a spoon is, the formal features are only intelligible in light of a description of what the spoon does, that is its purpose, and a spoon that has a bad purpose is a bad spoon.

One of the most well-known theories describing the rule of law is attributed to Lon Fuller in his work The Morality of Law.Fuller too perceives the rule of law as a combination of the formal

institutions of society together with what he terms ‘the inner morality of law’. Fuller’s conception of the rule of law is based on eight principles that are formalistic in nature: (1) there must be rules, (2) they must be prospective, not retrospective, (3) the rules must be published, (4) the rules must be intelligible, (5) the rules 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.Nevertheless, Simmonds, in his interpretation of Fuller’s outwardly formalistic depiction of the rule of law, argues that the ‘inner morality’ aspect of Fuller’s eight principles comes to the fore in relation to two further concepts, namely ‘the morality of duty’ and ‘the morality of aspiration’.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.The eight principles (representing the morality of duty in their rationale) provide a degree of regularity and order necessary in order 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.Furthermore, the attainment of the morality of aspiration requires that there be rules and orderliness, created by the morality of duty, and that eventually allow us 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’. The rule of law therefore is an instrument allowing us to ‘value the projective capacities of men and women’, an ideal that is achievable only where there are clear and notified rules.

Simmonds, in referring to the eight principles, 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.’

Consequently, there are many interpretations of the rule of law that find an expression in theories, which usually reflect the interwoven nature of both the functional and moral aspect of the rule of law[7].Wennerström, shedding light on the practical manifestation of the rule of law, states that it is usually used in national and international relations as a reference to a, ‘general order and predictability of events. It can refer to the state of affairs in a particular country or to the way in which a country conducts its international relations’. In addition to the formal and substantive divide, Wennerström refers to a third conception of the rule of law, namely the ‘functional’ conception, measuring the quality and also quantity of specific functions of a legal system, for example, the predictability of judicial decisions or the waiting period for access to the judiciary.It is with the emphasis on functionality the rule of law that is measured in regard to its manifestation within a state[8].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’[9].

Brownsword describes 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). In its essence, the contract entails that the actions of the governors always be 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[10].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, fairness in applying the law, separation of powers, participation in decision making, legal certainty, avoidance of arbitrariness and procedural and legal transparency[11].

The rule of law, therefore, is a political ideal, although its content and composition does remain a point of discussion and to a certain degree controversial[12]. In defining the rule of law in relation to its purpose, Krygier, in simple terms, stresses the fact that the rule is a solution to a problem, the problem being how to make the law rule[13].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.Associated with the notion of how power is exercised is 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 are general, equal and certain.

Associated with the notion of publicity is that of transparency. It has been argued that the rule of law is based upon two-pillar transparency principle, where the rule-making process should be open to people through political representation and that enforcement should allow procedural safeguards in the form of the ability to contest decisions.The transparency of the rule-making process is important in respect of this inherent function of the rule of law, namely a mechanism ensuring the ability to contest decisions.It is therefore in the eye of the beholder as to whether is defined more in terms of the formal structures necessary for making law or more as a concept requiring substantive morality. It can also be expressed in theoretical or practical terms, the latter coming to the fore in statements that it is a practical instrument that caters for society's need for predictability and that orders an otherwise chaotic society, thereby answering the question of what tomorrow brings.

As illuminated in this section, the rule of law is an allusive concept that comprises multiple interpretations, ranging from its function as a mechanism for curtailing arbitrary state power to a mechanism for describing the attributes necessary for attaining a just society that takes cognisance of various ideals and values, for example, human rights. Considering that not all these perspectives can be examined simultaneously, the following sections examine the rule of law from the perspective of its role as mechanism for determining rules, which if followed, create the conditions for allowing individuals to reach their potential in terms of the goals that they set for themselves and to achieve the ideals that they pursue. This is particularly relevant considering that the technology described below, discussed under the umbrella term called AI, can be described as especially inhibiting to the extent that individuals are made more susceptible to being manipulated and essentially categorized by the technology, albeit in a rather blunt manner. The notion of power thereon also elevates the function of the rule of law as a mechanism for minimizing the abuse of power.

3. The advent of artificial intelligence

What constitutes AI is subjective and best described as moving target. What AI is for one person may not necessarily be AI for another, what was considered AI say fifteen years ago is nowadays considered commonplace and even the question of 'what is intelligence?' is contested and debated. Popular culture has also played a role in the way AI is generally perceived.

Dartmouth College is the institution accredited with the birth of AI, where in 1951 John McCarthy, brought together a number of researchers at a workshop in order to study automata theory, neural nets and the study of intelligence[14]. This new academic discipline called 'artificial intelligence', in addressing problems, sought solutions inspired from a number of fields, such as neuroscience, mathematics and information theory control theory (cybernetics), all which coincided with the development of the digital computer.This new field transcended conventional subjects, such as mathematics, focused on topics such as duplicating human faculties (creativity, self-improvement and language use) and attempted to build machines that could function autonomously in complex, changing environments[15].

The above highlights that AI is not just about technology—rather, it incorporates multiple

disciplines in attempting to create machines that think like humans. Therefore, it is natural that these machines, in social contexts where they 'think' as well as or even better than human beings, are increasingly being used to assist and even replace humans in decision-making processes, or parts thereof. And this is in fact happening. Commercial actors and public authorities are increasingly starting to use machines to mediate their interaction with clients and citizens via models embedded in digital decision-making systems. This increases effectivity, cuts costs and optimizes processes as humans are gradually replaced by machines. However limited access to this technology also provides these actors with incredible power as the technology provides insight into human behaviour that only machines can gauge, this access restricted to those that own the technology. Finally, what the below section that examine the notion of AI will illuminate is the fact that in order to make decisions about people, they are essentially reduced to data points that are correlated with and mathematically weighted against each other. This in turn results in the models making the decisions, treating people based on the manner in which they are represented in the data as determined by the model incorporating the algorithm. This technology may prone to bias and mistakes and the digital representations of people may not reflect reality. However, probably the main harm with the technology is the fact that a model cannot be trained to foresee each and every personality it must decide about, resulting in the person having to be fitted to an existing set of factors. Not only is this problematic from a fundamental rights perspective, but it potentially prevents a person from being allowed to achieve his or her potential or desires in relation to identity creating, in turn inhibiting him or her from achieving the desired ideals, the manipulative effect models can have exacerbating this problem.

3.1 Achieving artificial intelligence

AI is an academic discipline within the realm of computer science. It has been described as 'the field devoted to building artefacts capable of displaying, in controlled, well-understood environment, and over sustained periods of time, behaviours that we consider to be intelligent, or more generally, behaviours that we take to be at the heart of what it is to have a mind [...] any insight into human thought might help us to build machines that work similarly'.AI is an academic discipline that covers many subjects: philosophy, mathematics, economics, neuroscience, psychology, computer engineering, control theory and cybernetics and linguistics[16].More specifically, it encompasses topics such as knowledge representation, heuristic search, planning, expert systems, machine vision, machine learning, natural language processing, software agents, intelligent tutoring systems and robotics.A more formal definition describes AI as: [a] interdisciplinary approach, understanding, modeling, and replicating intelligent and cognitive processes by invoking various computational, mathematical, logical, mechanical, and even biological principles and devices.It forms a key branch of cognitive science as it typically focuses on developing models that explain various dimensions of human and animal cognition.

A test that became the yardstick for determining the presence of AI is the Turing test, which simply put states that a human being, addressing written questions to a hidden entity, cannot determine whether the written responses originate from a human or from a computer.AI technologies are characterised by two main attributes, namely, 'autonomy' i.e., '[t]he ability to perform tasks in complex environments without guidance by a user' and 'adaptivity' i.e., '[t]he ability to improve performance by learning from experience'.Presently there is no legal definition of AI. However, a recent draft of a new regulation on AI was presented by the European Commission, where AI is defined as, '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' (Article 3) and where Annex 1 refers to, '(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[17].'

While AI has experienced periods of greater and lesser interest over time, the recent increase in its popularity can be attributed to a greater availability of data, increased processing power and more advanced mathematical algorithms which can be used to gain greater insight into data as well as allow AI to operate autonomously. This has invigorated the research area of machine learning. Machine Learning is a sub-category of AI and which broadly speaking concerns the use of algorithms to learn from training data in both a 'supervised' and 'unsupervised' (self-organized manner)[18].It is also described as: an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Machine learning is described as, 'a subfield of artificial intelligence concerned with the computerized automatic learning from data of patterns. The aim of machine learning is to use training data to detect patterns, and then to use these learned patterns automatically to answer questions and autonomously make and execute decisions'.At the heart of machine learning is the mathematical algorithm, which can be described as, '[a] process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer'.Whereas a regular computer system's logic is created by a human programmer, the logic of a system using machine learning is created by an algorithm. Machine learning is essentially the application of mathematical algorithms on data to produce a model that can be incorporated into decision-making systems, the model autonomous to the extent that it can update itself based on new data. A model can be conceptualized in two ways. First, on an abstract level, models explain various dimensions of human and animal cognition and where the focus can be on the engineering of smart machines and applications.Second, models are the core technical component of a system used to make decisions, having been equipped with the insights into data gained by an algorithm.Models can also have a predictive aim in that having gained insights from data they can make predictions concerning human behaviour: '[a] predictive model captures the relationships between predictor data and behaviour [...][o]nce a model has been created, it can be used to make new predictions about people (or other entities) whose behaviour is unknown'.The algorithm and accompanying knowledge learnt is then incorporated in a computer model and rolled out as part of a decision- making system.In attempting to optimize the giving of credit, by minimizing the risks, an algorithm will analyse the historical data in order to produce a predictive model that will be incorporated in any decision-making system, which is then set to work, predicting the likelihood or probability of new credit applicants successfully being able to repay their loans in the future. This learning capability of algorithms is central as they start to operate more autonomously and in increasingly complex and dynamic areas of application.

The modelling process can be described by means of an example: a credit institution wants to increase profitability by identifying risks in the form of potential clients who will default on the credit repayments; the credit institution holds huge amounts of historical data about clients' repayment behaviour and their associated circumstantial characteristics; an algorithm is used to make correlations between these data points in order to discover rules (who defaulted and why); a model is created incorporating these rules; a prospective client applied to the credit institution for a loan and the model incorporated into a system operated by the credit institution makes a determination of the probability that the prospective client will default in his or her credit repayments and finally a determination is made by the model, which determination may be followed by the credit institution. Quite simply, this data includes information on both people who have successfully repaid their loans and those that have defaulted. The creation of a system for making decisions can be described as follows:

Two concepts of relevance within machine learning are ‘supervised’ and ‘unsupervised’ learning. Supervised learning is the process whereby a programme is provided with labelled input data as well as the expected results and the algorithm, learning the underlying patterns, is then able to identify these learned patterns when confronted with new data. With unsupervised training, the algorithm is provided with the input data only, and it is then able to freely analyse the data in order to find interesting groupings of data of its own accord and without the data being labelled. Supervised algorithms are essentially taught using historical data training sets, and once they have achieved a certain level of capability, are then applied to novel situations, where predictive decisions can be made.

3.2 A technology inspired by nature

Biological inspiration has always been at the core of AI. If the end goal has been to achieve a mechanical intelligence on a par with human intelligence, then nature has also been an inspiration for attaining that goal. In other words, the art of learning has also occupied a central role in AI research. In attempting to develop more effective technology, nature has been a source of inspiration, with two sources of inspiration dominating, namely, the human brain (neurocomputing) and evolution (evolutionary computing).

Turning to neural computing, a term that has gained attention within the AI discourse is that of ‘artificial neural networks’. Natural Computing is an interdisciplinary field of study in computer science that is concerned with computation and biology, a sub-category of which is Biologically Inspired Computing (the study of biologically motivated computing for problem solving originating in the natural world). It is here that artificial neural networks come to the fore as an architecture that is modelled on the neurons of the human brain, has adaptive learning processes, are used in pattern recognition and where the feedback from the environment is divided into either supervised or unsupervised learning strategies.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 (supervised learning).The architecture of deep neural networks consist of many layers of nodes, to which data are sent. Typically, there is the ‘input layer’ (accepts data to the network), ‘output layer’ (delivers the output) and between these two layers there may be many ‘hidden layers’ (where the mathematical calculations are performed on the input data).Neural networks learn just as children learn: wishing to teach a neural network to identify a picture of a cat, it is fed thousands of pictures of cats, with the pictures of cats ‘labelled’ as representing a cat (this requires a human to teach the algorithm what a cat looks like); then testing the ability of the neural network to recognize pictures of cats, it is fed pictures of all types of animals with the task of identifying the pictures with cats in them; where the neural network is informed that it got it wrong, it automatically adjusts the complex mathematical weighting structure at its inner layers in order to ‘learn’ and increase its accuracy in the future; when the accuracy is deemed good enough, it is put to work in the digital environment in order to identify cats with a certain degree of probability.

Natural evolution is also used for inspiration to determine which AI solutions achieve the best results, given a particular problem. The natural evolution approach is prevalent when building predictive models and determines which solutions are best suited in order to solve a particular problem. It is based on the Darwinian theory of evolution and survival of the fittest.Therefore, where the aim is to learn the best solution to a problem, the rationale is that competition among the potential solutions will ultimately produce the winning (most optimal) solution.The rationale is simply that by using the process of trial-end-error, the solutions that best solve the required problem will be retained and in turn used to construct new solutions.In other words, only a limited number of solutions can exist in an environment and that those that compete most effectively for resources are the best suited for that environment:

Phenotypic traits are those behavioural and physical features of an individual that directly affects its response to the environment (including other individuals), thus determining its fitness. Each individual represents a unique combination of phenotypic traits that is evaluated by the environment. If it evaluates favourably, then it is propagated via the individual’s offspring, otherwise it is

discarded by dying without offspring small, random variations – mutations – in phenotypic traits occur during reproduction from generation to generation new combinations of traits occur and get evaluated. The best ones survive and reproduce and so evaluation progresses.

Bedau provides an example that explains how genetic algorithms works in practice. A problem may be to find the shortest route between two cities and in this regard, an itinerary may be suggested. A 'fitness function' will be used to calculate the 'fitness' of a proposed solution. In this example, such a fitness function will be the sum of all the segments of the itinerary, the fitness function in effect being the environment to which the solution must adapt. The more effective solutions in turn are used to model new solutions by means of randomly creating 'mutations' comprising elements of the more successful solutions, the ensuing 'generations' of solutions becoming more and more effective.

Without delving too deeply into the intricate workings of solutions (algorithms), the point wishing to be stressed here is that decision-making algorithms, be they deep neural networks or biologically inspired, are built using mathematical complexities and statistical rules that far exceed the cognitive capabilities of most humans.

3.3 A question of design

An initial point when discussing design aspects of AI is that a distinction must be made between rule-based systems and ML systems. The former can be described as static systems, with stand-alone characteristics, where the rules are determined by humans with ML systems, which are dynamic and heavily integrated with other systems.

Concerning philosophies of AI design, two types prevail. The first philosophy is known as 'the traditional approach', ('good old fashioned AI' or the 'neat approach'). This philosophy employs a symbolic basis for studying these mechanisms, symbolic knowledge representation and logic processes that can explain why the systems work. Neat AI approaches are prescriptive in nature, which means that they provide an explanation as to why they work. The main drawback of this form of AI is in relation to scalability, that is, as the size and complexity of problems requiring solving increase, they require increased resources (if this AI is to continue providing a guarantee in relation to the most 'optimal', 'precise' or 'true' solution). For example, algorithmic decision trees are easy to understand and to explain.In other words, as an observation falls through the decision tree branches, the logic used to determine which branch of the tree to send it on to or what it was that contributed to a certain result, is identifiable and explainable.

The second and newer philosophy is called 'scruffy AI' and has been described as 'less crisp technique[s] that are able to locate approximate, imprecise or partially/ true solutions to problems with a reasonable cost of resources'. Instead of a symbolic base, it uses inference strategies for adaption and learning and bases them on biological or natural processes. Scruffy AI is descriptive (as opposed to neat AI) which means that it reveals *how* a solution was arrived at (the process for achieving a solution) but not why. The biggest difference between these two methods, therefore, is that the former can explain why a solution was suggested while the latter can explain how it was reached (but not why). Another distinction between the above two philosophies is that scruffy AI involves,'[...] the incorporation of randomness in their processes resulting in robust, probabilistic and stochastic decision-making contrasted to the sometimes more fragile determinism of the crisp approaches'. Finally, neat AI adopts a deductive approach to problem solving whereas scruffy AI incorporates an inductive approach.

The above illustrates that a driving force propelling technology are market forces, where the demand for intelligent problem-solving automation has surpassed the supply of problem-solving products, spurring technologies that are faster but less explainable. Consequently, as time and resources become a scarce commodity, more precise tailor-made algorithms are discarded in favour of robust solutions, that work across a wide array of problems in a satisfactory manner.In other words, the technology can be made more explainable but this comes with a financial cost, a cost that many commercial entities may be reluctant to take. In addition, the problems to be addressed using AI solutions are becoming more complex, self-driving cars being one example. The presumed

benefits of AI take centre stage, so does the faith in AI to solve these complex problems.

4. AI in the criminal justice system

Artificial intelligence is increasingly being incorporated into systems that are intended to assist actors within the criminal justice system in their decision-making responsibilities. This section examines one such initiative.One such example is the use of systems incorporating models to assist judges in making determinations about people in various circumstances, for example whether a person should be released pending trial but also the severity of a sentence.

The criminal justice system in the United States is an example where AI is being used in order to mediate between state and accused. It is becoming commonplace that 'pretrial risk assessment algorithms' are being consulted when setting bail, determining the duration of prison sentences and contributing to decisions concerning guilt and innocence.The basis for decisions made by these algorithms are factors such as age, sex, geography, socioeconomic status, family background, neighbourhood crime and family status. The intelligent aspect of any such system, and which is concealed in technical complexity, is the manner in which the selected factors are mathematically weighted in relation to each other in order to form a behavioural profile of the accused.

In the matter of *State v. Loomis*, where a Wisconsin trial court sentenced a defendant to six years in prison for a criminal act, the corresponding sentence was in part determined by 'algorithmic risk assessment'.In the United States criminal context, 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 the above matter, the PSI incorporated an algorithmic assessment report. The software, called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), was developed by Northpointe Inc., a private company, where the output comprises a number of bar charts depicting the risk that the accused will commit crimes in the future.Accompanying the PSI was also a procedural safeguard in the form of a written statement to the judges concerning the risks associated with pretrial risk assessments. The Wisconsin Supreme Court subsequently upheld the lower court's decision, stating that the use of the algorithmic risk assessment software did not violate the defendant's right to due process even though it was not made available either to the court or Loomis. COMPAS assesses variables under five main areas: criminal involvement, relationships/lifestyles, personality/attitudes, family, and social exclusion.Subsequently, and upon request, Northpointe Inc. refused to make the software available citing that it was proprietary and a core business secret.

Another system that became the object of a court case is that of System Risk Indication (SyRI), which is a digital system used by the authorities in the Netherlands to identify benefit fraud in poorer socioeconomic communities. The system received attention due to a decision from a District Court in the Hague, Netherlands, that ruled that the system nestopped due to human rights violations. The algorithm in SyRI was used to spy on entire neighbourhoods classed as low-income, even though there was no prior suspicion of fraud being committed.The system was enabled by legislation called the SyRI Act of 2013 that enabled the collection of data from various public databases, with data concerning income, house ownership, benefits, address, family relations, debts, and data on the use of water and energy being correlated and weighed by the algorithm, the final output a score, where the highest level received the label 'worth investigation'.The harms of this system were multiple, privacy, discrimination and stigmatization being some clear examples. However, another harm was the rule of law to the extent that the system allowed for the indiscriminate wielding of power by the executive:

The laws introducing SyRI do not meaningfully limit the power of the executive bodies involved. The law is filled with vague language that refers to the 'necessity' of the system, the 'safeguards' that are in place, and a set of extremely broad purposes. The administration has a very broad margin to decide which data they will collect and have the freedom to use secret risk models to analyse this data, co-opting the rationale behind operating practices of intelligence agencies. To my knowledge, this is an unparalleled expansion of power of the executive branch, which exposes every adult

citizen in the Netherlands to the danger of arbitrary and opaque interferences. All of this also has a tremendous damaging effect on the relationship of trust that you should have between citizens and the state.

An aspect associated with the SyRI matter is that the system was based on a piece of enabling legislation that made its existence and operation public. This allowed for the investigation into the system. It is argued that there are many systems out there that are not based on legislation in this manner thus their existence is not common knowledge.With the increased use of systems like COMPAS and SyRI, there is a risk that we may be stumbling head-on into the 'digital welfare dystopia'.

The technology used in algorithmic decision-making systems is complex and the next section delves a little deeper into how these systems work.

5. The erosion of the rule of law in the age of AI

Technology is often described as a 'double-edged sword' as its effects on society can be both beneficial but also risky. For example, technology may curtail freedom of expression but at the same time facilitate it. The inherent nature of AI is without doubt a threat to the rule of law and these must therefore be addressed. It is therefore necessary first to highlight some of the risks to the rule pf law.

The Loomis case raises some important issues. The Supreme Court, while dismissing the matter on appeal, provided five reasons for caution: first, COMPAS was proprietary software and void of transparency; second, the COMPAS system calculated the recidivism risk for groups, not individuals; third, COMPAS relies on national data and not on data from Wisconsin; fourth, studies question the extent to which sentencing algorithms disproportionately classify minority offenders as having a higher risk and fifth, COMPAS was developed to assist the Department of Corrections and not necessarily to be applied in the criminal court system.In order to limit the scope of discussion, these points act as a point of departure.

5.1 The notion of accessibility to the law

A core element of the rule of law is that laws should be accessible in order that people can abide by them and know what is expected of them, predictability being paramount.It is for this reason that the notions of publication and intelligibility as promoted by scholars such as Fuller are depicted as central to the rule of law. Undermining these prescribed attributes of the rule of law is the lack of accessibility that AI presents. The technological complexity associated with AI does not make it suitable to human comprehension, insight or transparency. For example, the mathematical calculations taking place at the hidden layers of neural networks or the mutating capabilities of genetic algorithms are beyond human cognitive comprehension and for the most part human explanation. Using the above example of teaching a neural network what a cat is, it is highly unlikely that the mathematical complexities of this operation can be fully explained in natural language. For example, it can be stated that a neural network was employed to solve a problem and that it can identify cats with a certain degree of accuracy. However, the 'why' in relation to the output cannot be explained.It is here that the notion of natural language versus the language of AI gains importance. AI has rules, however these rules are the rules of mathematics and statistics. To exacerbate matters, these rules are hidden either in the proprietary 'black box' or hidden also to the extent that they cannot be understood—they cannot be read, they cannot be discussed, they cannot be analysed and they cannot be reasoned. The rule of law up until now has been dependent on its form being in the format of natural language—it entails a governance by natural language as compared to the governance of the algorithm.The rule of law is dependent on natural language in order to be comprehended. This is not necessarily the case for all areas of law, where some legal processes are easier to automate. For example, the levying of a congestion tax in the city of Stockholm, has been successfully fully automated.Therefore, as governance increasingly finds its expression in computer code, its comprehension by a country's citizens is bound to decrease. This in turn relates to the notion of what is 'intelligible', with neither regular people nor judges applying

systems like COMPAS having the cognitive ability to actually understand them. Even their creators do not really understand them, these self-learning and self-evolving algorithms taking on a life of their own. This is especially so considering that algorithms potentially mutate themselves and updated their processes multiple times a second, the mathematical balancing of the attributes that the algorithm considers constantly being altered.It is here that the technical concept of interpretability comes to the fore.It is also argued that with complexity comes the potential for error. It has been suggested that a rule of thumb for working with the technology of AI should be that the technology be deemed incorrect until proven correct.This highlights the importance of the right to complain, where important decisions are taken automatically by autonomous machines.

It seems that there is a psychological line, that when crossed, places the rule of law on a collision course with AI. This is the moment at which we allow machines to make decisions over human beings without humans really understanding how these machines work. Public services may be automated and algorithms may be used to streamline government. However, the moment one uses opaque technology that is incomprehensible, our trust in the technology is nothing more than the inability to understand it. AI is an existential threat to the rule of law and a question that has been put is whether the future will bring with it a rule of law or a rule of algorithm?

It is in this regard that the notions of legality and accessibility of the rule of law as expressed in the Venice Commission are triggered. It requires that laws are accessible, that court decisions are accessible and that the effects of laws are foreseeable.However, the use of AI by the judiciary can be seen as undermining these principles as illustrated above.

5.2 A black box created by law

The blame for the erosion of the rule of law cannot be put squarely at the foot of technology. Sometimes the law itself, as a mechanism for balancing conflicting interests, reaches a balance between these interests in the form of balancing the various rights and obligations of different stakeholders in traditional legal documents. However, technology is a disruptor of society in many ways and one manifestation of this disrupting characteristic is its putting out of synch the balancing of interests that may have occurred by means of traditional law.

The General Data Protection Regulation (GDPR) is an example of this balancing act performed by traditional law in the context of data protection. This legal framework attempts to balance intellectual property rights against rights associated with privacy in relation to AI. Recital 63, expanding on Article 22 which deals with AI, affords the data subject a right to know and receive communications regarding the logic behind any data processing in relation to automated decision-making. This potentially grants the data subject the right to an explanation of the technology. However, this right is watered down in the very same recital where trade secrets and intellectual property rights take precedence over transparency.This can be seen in the light of a right to information concerning the processing of personal data can be found in Article 15(1)(h) GDPR that provides the data subject with a right to information about the logic involved in automated processing and the consequences thereof, although it has been argued that a ‘right to information’ is the same as a ‘right to explanation’.Here it can be argued that the balancing of privacy against intellectual property is disrupted by the nature of the technology itself—AI cannot be compared to any technology preceding it and it can be argued that transparency into its inner working is absolutely necessary in order to provide adequate adequate protection from its harms. Hence the argument that intellectual property rights are potentially assisting in the creation of the black box of technology.

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 the SyRI case too, the proprietary nature of the technology, creating a black box in terms of insight into the technology, is argued to have weighed against the state in that the court was not able to verify the state's claims as to how the technology works.

Considering the wide ambit of the applicability of the GDPR, it is not inconceivable that it will be relevant in many circumstances where AI is used to make administrative decisions or even in the justice system. Here too the SyRI case illuminated the GDPR by referencing the data protection principles of transparency, purpose limitation, and data minimization, the latter two making up the proportionality principle.

It is acknowledged that there are good reasons as to why intellectual property rights and trade secrets are protected in law. For example, intellectual property rights have the goal of encouraging creativity and providing an economic incentive for creativity. However, considering the potential for errors or inaccuracies in the decisions made by AI in the public domain, it is not inconceivable that a clash between the rule of law and the values it carries (openness, transparency, right to explanation and check on the abuse of power) and other areas of law may be brought to the fore by the increased reliance on AI.

Finally, it should be noted that the imbalance created by AI in relation to differing interests can be rectified by various mechanisms. In this regard, trusted third parties may have a constructive role in ensuring that algorithms are developed and applied in accordance with the values of the rule of law. For example, in the United Kingdom, The Law Commission on the Use of Algorithms in the Justice System recently published a report where one of the recommendations was the creation of a National Register of Algorithmic Systems, where various aspects in relation to the algorithms being used in the criminal justice system could be checked and verified.This idea is also reflected in the draft regulation on AI made public recently by the European Commission. Here, in relation to high-risk AI, the draft regulation creates the mechanism whereby these AI systems must be registered in an EU database (Article 51) established by mean of collaboration between Member States (Article 60).It is argued that the use of trusted third parties potentially increases insight into the complexity of AI while at the same time preventing general public insight, thereby maintaining the interests protected by intellectual property rights.

5.3 Detecting bias and discrimination in data

The notion of bias is an inherent aspect of data science and therefore technologies AI. In other words, the second you handle data it automatically brings with it bias. The act of choosing one dataset over another will potentially reflect a certain bias. Bias can be both intentional and unintentional and a rule of thumb should always be that a data set incorporates some degree of bias. Bias is present in almost all data sets and biased data will invariably lead to a biased output by the models that are trained on this biased data.A definition of bias is that, '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'.The problem with bias and discrimination in a data context is that 'masking' can occur: this occurs 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.Typical example are using area code (zip code) to denote health status, where socioeconomic factors may play a role or using area code instead of race. Bias should also be distinguished from discrimination, which is a legal concept that can be described as, 'the prejudiced treatment of an individual based on their membership in a certain group or category', where the attributes encompassing discrimination include race, ethnicity, religion, nationality, gender, sexuality, disability, marital status, genetic features, language and age.Consequently, a model is said to be 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.Bias and discrimination are therefore related to the extent that bias in data can lead to discriminatory effects, but may not necessarily do so in all cases.

Subsequent to the Loomis case, the use of AI 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.First, examining 7000 decisions, the results showed that the algorithm is only 20 percent successful in accurately predicting recidivism. Second, the algorithm incorrectly flagged African Americans at twice the rate of White people.The Venice Commission treats equality before the law and non-discrimination as an essential element of the rule of law. 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.

One example of how bias can creep into data is via context. For example, a data set from one context used to train an algorithm may not work as expected in another context. Data is contextual and using data from one context in another context can lead to incorrect decisions. Data is described as 'spatiotemporal' with data having a certain meaning in a defined situation but this meaning can vary in another situation, as well as over time.Once again the Loomis case is an example of this. COMPAS was developed to be used by the Department of Correction but is now being deployed for use in sentencing. These are different contexts and while an algorithm learnt from one context the conclusions it draws may not be as relevant in another context. Here attention must be drawn to the issue of contextual data or systems being developed for one context being used in another. In this regard mention can be made of other research on the degree with which predictive models used in the criminal justice system can lead to unfair outcomes. In one paper, Machine Learning algorithms were compared against conventional systems in relation to the prediction of juvenile recidivism. The main conclusions were that the Machine Learning models scored slightly higher with regards to accuracy whereas with regards to fairness, the Machine Learning models tended to discriminate against males, foreigners and specific national groups.

It is within the realm of the Venice Commission that discrimination is viewed in opposition to the rule of law, where not only is non-discrimination demanded but also 'equality in law' and 'equality before the law'. This is a relevant distinction within the context of decision-making systems incorporating elements of AI, where it can be difficult to identify an inequality or specific instance of discrimination. Put another way, the existence of the prerequisites demanded by a traditional law may be difficult to identify within the complex mathematical processes of AI. Compounding the situation is the notion that the mathematical rules of the decision-making models have not necessarily been exposed to traditional law-making procedures, but rather are 'promulgated' by private corporations.

5.4 The potential for the abuse of power

The argument of Krygier, mentioned above, is that the rule of law essentially concerns power, where its main goal is to make law rule in order to curb the potential for abuse of power by those who use this power in an arbitrary manner. He states that there are many ways to exercise power and that the arbitrary ways should be shunned.It is in this context that the Venice Commission benchmark of 'prevention of abuse (misuse) of power' is relevant. It is submitted that there is a correlation between on the one hand defining the rule of law through the lens of power and on the other hand the notion of reciprocity. For reciprocity to flourish, a certain equilibrium in the power relationship between those that govern and the governed is required. However, it is argued that the transfer of governance to technology, such as witnessed in the Loomis case, brings with it a monopoly in terms of access to the technology. It is essentially only those who govern that have the resources to produce or purchase the technology that is used to make decisions about citizens. This continually increasing imbalance is disempowering the governed in favour of those who govern. For example, with the monopolization of the power over technology in the hands of those that govern, the risk that executive discretion becomes unfettered increases, this being contrary to the rule of law as expressed by the Venice Commission.In addition, an aspect of the abuse of power as identified by the Venice Commission is irrational decisions.However, to what extent can the decisions taken by AI ever be challenged as 'irrational' when they firstly cannot be comprehended

but also when human rationality is not necessarily a prerequisite for an algorithmic solution? A final complexity in the power equilibrium is the fact that the producers of AI technology are private actors, the power equilibrium essentially having to be achieved between three entities, namely those that govern, those that are governed and the private corporations developing the technology for mediation. The Venice Commission does recognize that there may be situations where private actors exercise powers that traditionally have been exercised by states.However, the examples provided include the management of prison services, and it is argued that situations where private actors take over the judicial discretion of judges was never envisaged.

5.5 Challenging traditional legal protections

The increased use of AI to predict human behaviour, more specifically criminal behaviour, is also challenging some traditional legal notions. One legal notion being challenged is that of an accused being regarded 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. This is recognized in the principles of 'nullum crimen sine lege' and 'nulla poena sine lege' which recognize that there is no crime or punishment without a law, these principles also incorporated in the Venice Commission.The presumption of innocence and right to a fair trial are encompassed in the benchmarks regarding the 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 AI 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'.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. Ellul argues that technology has acquired an autonomy from its association with the legitimacy of scientific progress in general. In other words, technology has a legitimacy due to the perception that it is scientific and objective.

5.6 The right to contest decisions

One of the core characteristics of the rule of law, as discussed above, is the notion of a right to contest decisions. Considering the black box nature of AI—due its complexity as well as due to legal constructions associated with intellectual property law—it becomes apparent that the right to contest decisions weakens considerably. It is argued that, '... in techno-regulatory settings, the three phases of legal process as: direction (rule making), detection and correction collapse on top of each other and become an opaque inner process imbedded in the systems'.One potential solution is that of making contestability part of the design process.However, one problem revolves around the fact that in order to contest a decision, for example an automated decision, one first needs to know that a decision has been taken about oneself. This my not be that challenging, for example, in the COMPAS situation where it is rather clear that a person has been subject to a decision by a black box. However, there are decisions taken about people every day that do not reach any formal forum, such as a court of law, a consequence being that we are never enlightened about the fact that a decision was actually taken.

Relevant in this regard is how one could be notified that one has been the object if a decision taken by AI, for example in the form of a predictive model. One suggestion is the creation of a right of access to the knowledge that a decision was taken, this referred to as a 'right to know'. Branscomb advances such a right, calling it 'the right to know', stating that this right can be complex and also can take on various forms. In other words, the enforcer of the right can be a different protagonist depending on the circumstances. For example, it could be a right for an individual to know his or her origins or it may be the right of the public to know the basis for decisions of a public nature.The idea is that some form of notification mechanism would alert a

person to the fact that an entity has taken a decision about him or her by means of AI.

It is also important to consider the manner in which technology can assist with the right to contest decisions, thereby fortifying the rule of law. Here reference is made to the 'chatbot lawyer' called DoNotPay. This application uses artificial intelligence and provides a free service that assists individuals, who have received a parking fine, to appeal that parking fine via a user-friendly interface. Having received a parking fine, the individual accesses the chatbot lawyer and is prompted in an interactive manner to provide certain details surrounding the circumstances under which the fine was received. Thereafter the chatbot lawyer seeks a legal bases upon which to file an appeal against the fine. For example, it could be that there were no signs reflecting that it was illegal to park in a particular manner. Having provided the necessary information, the user merely presses a button and the appeal is automatically sent off to the authorities.

6. Analysis: constraining human flourishing

Having examined the notion of the rule of law, it is submitted that one of its main functions is to allow human beings to flourish. In other words, it allows individuals to attain their desired goals and be creative in deciding what a good life is, this state also referred to as having agency. One of the main harms of AI is that this technology curtails human agency thereby diminishing human flourishing, the promotion of which is argued to be a goal of the rule of law.

A question that can be asked is if one of the aims of law in general is to condition a desired type of behaviour in society, what then is the difference between a system of governance under the rule of law and say another system of social control that uses technology to designate the 'model citizen', thereby encouraging citizens to live up to this measure?The answer potentially is provided by Simmonds in his interpretations of Fullers eight principles, addressed above. Simmonds refers to the fact that the law is not the only form of governance, other forms being coercion, social conditioning or mediation and compromise. However, the law stands out in relation to these other forms of social control to the extent that it bears a commitment to the idea that people are rational purposive agents capable of regulating their conduct in relation to rules as well as a commitment to the rule of law as expressed by Fuller in his eight principles.In other words, the rule of law is an instrument that allows people to adjust their behaviour in relation to the law, it allows them to be free agents in effect enhancing personal autonomy and it is to a certain extent empowering. It is argued that the notion of reciprocity, as encompassed by the rule of law, is that notion that allows individuals to attain a certain level of agency. According to Murphy, the rule of law specifies certain requirements that lawmakers must abide by in order to govern legally, in other words, restricting the extra-legal use of power, continuing that the rule of law ensures that the political relationships structured by the legal system express the moral values of reciprocity and respect for autonomy.She continues that citizens experience resentment when the law is not clear, if the law is contradictory or if it is not properly enforced and in general when citizens follow the law without this being reciprocated by government.Fuller states:

Certainly there can be no rational ground for asserting that a man can have a moral obligation to obey a legal rule that does not exist, or is kept secret from him, or that came into existence only after he had acted, or was unintelligible, or was contradicted by another rule of the same system, or commanded the impossible, or changed every minute.

It is therefore argued that one of the greatest concerns with AI is in relation to algorithmic governance, where the notion of reciprocity is diminished. Brownsword refers to the regulatory environment that is technologically managed as opposed to rule-based and makes the distinction between traditional normative rule-based regulatory instruments and non-normative technological management.The above notions are succinctly distinguished by referring to the former as speaking in terms of'oughts' and the latter in terms of 'can' and 'cannot'.Consequently, the traditional rule-based regulatory environment provides the citizen with a choice as to follow this rule or not whereas the latter provides no such option—it is a take it or leave it situation, where the programming code determines behaviour and there is no leeway for considering the degree to which one wants to live up to a rule. It is precisely this that places the notion of reciprocity in danger.

Brownsword argues that it is still important to recognize the link between the regulators' normative intentions and the technology that embeds these intentions in that it enables a testing of the technology against the rule of law.In other words, if the rule being enforced by the technology lives up to the rule of law, then the technology too lives up to the rule of law. However, it is argued that it is exactly here that it is crucial to make a distinction between the different types of technology, or more precisely, between the technology where normative rules have been transformed into code by human programmers (regular programming) and the technology of AI. For it is within the context of the latter that the rules that regulate may have been identified by the technology itself, may fluctuate from one second to the next and may operate differently if the input data changes with just one single data point. In addition, in the age of technological management, the regulators are private companies who make the rules that are locked away in black boxes. Probably affecting the notion of reciprocity, the most is the fact that in the age of technological management, to what extent do people even know that they have been the subject of a decision, the technologies of regulation hidden where we may not expect them to be. Considering the notion of a contract between the regulators and citizens, Brownsword himself alludes to the fact that any defection on either side can lead to a downward spiral of diminished trust.

This raises a number of questions in the age of AI: to what extent is it possible for citizens to enter into this contract of reciprocity with machines, to what extent are citizens increasingly being expected to adhere to the notion of reciprocity where the opposing partner is not Government but rather private corporations that produce the technology, to what extent can human agency and autonomy exist as values in societies characterised by technocratic governance, to what extent can the governance of the algorithm be considered fair when it is hidden in the 'black box' and finally, how can we be expected to abide by rules that potentially change in a micro-second?

It is argued that human agency is a concept that runs contrary to AI processes. A central questions is that of equal treatment and fairness, where being incorporated into a group does not necessarily mean that one shares all of that groups characteristics.It has been suggested that, 'persons should always be treated as persons with interests and a voice that needs to be heard' and one can question the extent to which AI diminishes people to data points, hidden in the black box of complexity, effectively taking away peoples' voices.

The notion of human flourishing is addressed by Floridietal.Here the notion of promoting human flourishing in the light of AI developments is reflected upon. Here human flourishing is described in terms of, 'who we can become (autonomous selfrealisation); what we can do (human agency); what we can achieve (individual and social capabilities; and how we can interact with each other and the world (societal cohesion).'The authors argue that AI's predictive power should be used for fostering self-determination and social cohesion instead of undermining human flourishing. According to the Report of House of Lords Select Committee on Artificial Intelligence, people should be able to 'flourish mentally, emotionally and economically with artificial intelligence'.

7. Conclusions

The rule of law as a legal notion is elusive to the extent that the more one attempts to define it, the more diffuse it appears to become. The spectrum describing the rule of law is long—it is viewed as a political ideal, a mechanism for curtailing the abuse of power as well as a mechanism for ensuring that society uphold certain values, for example, human rights. A common denominator of the rule of law is that it is viewed as a notion that is worth protecting despite its susceptibility to political abuse.

Modern technologies are increasingly being used within society, AI a prime example. As Machine Learning techniques are improved, so to are AI systems being used to assist human decision-makers in almost all fields. It should be anticipated that as these technologies become better at assisting with decisions, more control and responsibility will be transferred to them. It is therefore important that heed should be taken to the fact that these technologies are challenging the ideals associated with the rule of law as a concept of traditional law. In addressing the harms associated with AI in relation to the rule of law, a common denominator that stands out is the

manner in which it potentially inhibits the flourishing of humans. While this may traditionally not be the first association in relation to the rule of law as a concept, it is nevertheless important to address as human agency can be argued to be a cornerstone of society.

A challenge for the future will be how to reap the benefits of AI for society while at the same time protecting society from its harms, essentially promoting innovation while at the same time balancing it against the interests of society. A challenge will be to determine which values to balance technology against. In this regard, it is argued that the values enshrined in the rule of law operate as a good starting point in determining the fabric of any society. Herein lies the value of protecting the rule of law from technologies incorporating AI.

References

[1] World Justice Project *Insights*, p.7, 2019.

[2] Choi (2021), [online] Available: https://www.britannica.com/topic/rule-of-law (Accessed 11 June 2021).

[3] Organization for Security and Co-Operation in Europe (OSCE), Rule of law, Rule of law | OSCE (Accessed 11 June 2021).

[4] Stanford Encyclopedia of Philosophy, The Rule of Law,[online] Available: https://plato.stanford.edu/entries/ rule-of-law/ (Accessed 20 December 2019), p. 1.

[5] Dworkin, Ronald. A Matter of Principle, Cambridge Mass: Harvard University Press, 1985, pp. 11–14.

[6] Matsuo. Let the rule of law be flexible to attain good governance. Rule of law promotion: global perspectives, local applications. Iustus, Uppsala, 2009, p. 42–53.

[7] Ziegert K. Is the rule of law portable: A socio-legal journey from the Nordic Mediterranean Sea via the silk road to China, 2009, p. 29.

[8] Wennerström E. Measuring the rule of law, 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, 2009, pp. 58-62.

[9] World Justice Project, Rule of Law Index, 2019, (2019a), p. 7.

[10] World Justice Project, World Justice Project Rule of Law Index 2019 Insights, (2019b), p. 7.

[11] Brownsword R. Technological Management and the Rule of Law. Law Innovation Technol, Routledge 8(1):100–140, 2016.

[12] Report of the Secretary General of the United Nations, The Rule of Law and Transitional Justice in Confict and Post Confict Countries, United Nations Security Council, p. 4, 2004.

[13] Matsuo. Let the rule of law be flexible to attain good governance. Rule of law promotion: global perspectives, local applications. Iustus, Uppsala, 2009, p. 41.

[14] Sannerholm R. Rättsstaten Sverige: skandaler, kriser, politik, Timbro, 2020, p. 12.

[15] Brownlee J, Clever Algorithms: Nature-Inspired Programming Recipes, [online] Available: http://github.com/cleveralgorrithms/CleverAlgorithms, 2011, p. 3.

[16] Arkoudas K, Bringsjord S. Philosophical Foundations. In: Frankish K, Ramsey WM (eds) The Cambridge Handbook of Artifcial Intelligence, Cambridge: Cambridge University Press, 2014,p.34.

[17] Elements of AI, Chapter 1, [online] Available: https://www.elementsofai.com/.

[18] European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artifcial Intelligence (Artifcial Intelligence Act) and Amending Certain European Union Acts, Brussels, 21.4.2021 COM(2021) 206 final, [online] Available: europa.eu (last accessed 2021-04-29).