THEMIS COMPETITION

SEMI-FINAL D: JUDICIAL ETHICS AND PROFESSIONAL CONDUCT

ARTIFICIAL INTELLIGENCE AND THE JUDICIAL RULING

TEAM PORTUGAL I

ELISA ALFAIA SAMPAIO, JOÃO J. SEIXAS, PAULO JORGE GOMES

TUTOR: PATRÍCIA HELENA COSTA

SOFIA, BULGARIA
2-4 JULY 2019

INTRODUCTION
FRANKENSTEIN'S MONSTER'S SOUL

We, in our wealthy western societies, live in a world made possible by Science. We live longer than our ancestors, in better health, eating better food and, despite sensationalist newspapers or populist politicians, in less violent, safer, societies than ever before in human History.

However, especially in Europe, Science, and its main by-product – technology – are perceived with growing mistrust. And, after nuclear energy, the cloning of mammals, and genetically modified organisms, the main threat seems to stem from developments in Artificial Intelligence. As more and more sectors of human life and comfort increasingly rely on robots and AI-operated machines, from medical diagnostics to self-driving cars, we hear growing calls for alarm: "Bank of England Economist Warns Thousands of Jobs at Risk from Robots"; "How Artificial Intelligence Could be Violating our Human Rights", are some recent newspaper headlines that illustrate the fact that Artificial Intelligence is the scare word of the day.

As it has been for some time, in one form or another, at least since the Luddite movement of the first half of the 1810s, the fear of AI is but another expression of technophobic thought. The revolutionary idea of Robert Owen to mechanize the weaving looms in his factories, announced by The Hull in 1817, found in Darwin's theory of Evolution the fuel that would ignite what became a perceived dispute between Man and Machine, as prefigured in an 1863 essay, "Darwin Among the Machines", by Samuel Butler. One could indeed see the rise of the machines in factories as a threat to one's livelihood, for an untiring machine could physically outperform any man, for hours on end; is it surprising, then, that a machine perceived to be as, or more, intelligent than any human being, should be feared as a possible substitute for Mankind itself?

In this brief essay, we intend to explore a specific context where AI can be expected both to reveal itself as an indispensable tool, and to raise more objections to its deployment: that of the Judiciary. With the possible exception of the various art forms, the Judiciary is undoubtedly that area of human activity where one expects human nature to manifest itself to its fullest: one expects the judge to apply both reason and emotion to his judgments, one complementing the other so that neither prevails. Balance is not only expected, it is to be desired. But are reason and emotion a part of intelligence? Can a machine be reasonable in its decisions? Or can it never be more than merely logical?

Not surprisingly, the ethical questions raised by the growing implementation, worldwide, of AI systems in several capacities in the Judiciary, was promptly addressed by the EU in the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment, issued by the European Commission for the Efficiency of Justice. Its tentative responses to an essentially unpredictable technological field will be part of the analysis to be found herein, as it seems to hesitate between embracing the

science and technology of AI, or succumbing to the alarmist potential popularized in science-fiction films. As such, it reflects some of the European's suspicion about science. As the then European Union's Science Adviser microbiologist Anne Glover stated in a 2012 interview with Science Insider (February 14, 2012), *"If you take people's opinions, for instance by looking at the Eurobarometer, people seem to be reluctant to accept innovative technologies. They are suspicious almost just because it's new. (…) There should be more communication about the rewards of the technologies"*.

Artificial Intelligence is one technology that shows the promise of the greatest rewards. As such, it demands great responsibility while being dealt with. Neither undue optimism, nor unjustified fear, must be allowed to prevent us from exploring this brave new world, or to benefit from its fruits.

As scientists and programmers try to perfect artificially intelligent systems that can operate in the judiciary, helping judges render better decisions, providing people with fairer and more equal justice, or even handing down decisions in a human case, to a human party, they are trying to get such systems to choose an action that best satisfies conflicting goals: and such ability, the ability to choose such an action "is not an add-on to intelligence that engineers might slap themselves in the head for forgetting to install; it is intelligence" [1].

CHAPTER I
A THEORETICAL FRAMEWORK

1.   The AI in the Fourth Revolution

We're living in a hyperhistory, described by LUCIANO FLORIDI as *"(…) the stage of human development when third-order technological relations become the necessary condition for development, innovation, and welfare"*, based on *"(…) technologies as users interacting with other technologies as prompters, through other in-between technologies (…)"*. This shapes the self-understanding of human identity as informational organisms or inforgs [2]. In this century, we are witnessing a new spring for AI on a daily basis: it is in our smartphones, in commercial logistics, search engines, electronic games, social networks, aviation, health and bank systems, or in legal and judicial contexts.

Roughly speaking, AI is known as intelligence demonstrated by a computer machine or software. This notion leads to other questions: what is intelligence? Where are the boundaries between thought and computing [3]? Does the human brain function as a computer or are the mental processes indivisible? Can machines learn as humans do? These questions pose deep philosophical problems (epistemological, ethical,

---

[1] STEVEN PINKER, *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*, Penguin Books, 2018, p.300.

[2] L. FLORIDI, *The Fourth Revolution: how the infosphere is reshaping human reality*, Oxford University Press, 2014, p.31.

[3] See WILLIAM J. RAPAPORT, «What is a Computer? A Survey», *Mind & Machines*, 2018.

metaphysical), aggregate different areas of science (cognitive, biology, logic, psychology, linguistics, mathematics, cybernetics, engineering) and energise several schools of thought (evolutionists, symbolists, computationalists, bayesians, analogisers). Naturally, this is not the place for, nor do the authors have the presumptuousness to, face all these enquiries. Nonetheless, the task proposed in this paper will demand some small detours into other less familiar issues for trainee judges, and try to clear up some miscomprehensions.

2. Big Data

It has been estimated that humanity had accumulated approximately 12 exabytes [4] of data in its entire history [5]. In the 21st century, between 2006 and 2011 alone, the data available had grown to over 1.600 exabytes. Nowadays, we live in the zettabyte Era [6], and it tends to grow exponentially since the use of data will generate more data [7]. Some estimations purport that by around 2020, for every person on earth, 1,7 megabytes of data will be created every second [8]; the IDC White Paper (2018), for instance, predicts for 2025 an increase of global data in the world of 175 zettabytes [9]. This ocean of data needs to be collected, stored, managed, and analysed computationally. In other words, data, to be big, needs models, through which algorithms extract inferences about patterns, trends and correlations [10].

As Professor JACK M. BALKIN highlighted, *"Big Data is the fuel that runs the Algorithmic Society; it is also the product of its operations"* [11]. By contrast to humans, AI systems are comfortable with a large number of data sets. It allows the AI system to find new patterns and to label new examples, expanding the collection of all perceived history. For instance, to surpass the ambiguity of natural language, a translation algorithm will have a better performance if it has billions of words stored in its training set, instead of just a couple million. The increase in data will improve the machine's ability to achieve its goals.

3. Human *versus* artificial intelligence

As JAMES H. MOORE pointed out in 1985, the advances of new technologies involve not only *policy vacuums* but also *conceptual vacuums* which needed to be filled [12]. Generally speaking, intelligence is

---

[4] Putting in perspective, 1 exabyte corresponds to a 50,000 year-long video of DVD quality.

[5] L. FLORIDI, *The Fourth Revolution*…p. 13 (note 2).

[6] One zettabyte corresponds to 1000 exabytes.

[7] Again, putting the power of exponentially in perspective, if one takes 30 normal steps forward it will be moved around 30 meters. If one takes 30 exponential paces, doubling the length each time (first step one meter, second step two meters, third step four meters…) at the 29th step one would had reach the moon. The 30th step would bring the traveller back to earth, C. CHACE, *The Artificial Intelligence and the Two Singularities*, CRC Press. 2018, p. 45.

[8] See https://www.domo.com/learn/data-never-sleeps-6 (last access on 29 May 2019).

[9] IDC White Paper, *The Digitization of the World From Edge to Core*, 2018, available at https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf, last access 10.06.2019.

[10] Consultative Committee of the Convention for the Protection of individuals with Regard to Automatic Processing of Personal Data, Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data, Council for Europe, 2017, p. 2.

[11] J. M. BALKIN, *The Three Laws of Robotics in the Age of Big, Data*, 2017, p. 6, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2890965 (last access 13 June 2019).

[12] JAMES H. MOORE, «What is Computer Ethics?» *Metaphilosophy*, 16(4), 1985, p. 266.

associated with reasoning, memory, understanding, learning and planning: a «general intelligence» with the ability to perform intellectual tasks. There are many disputed definitions of intelligence in psychology, including the «multiple intelligences» theory, proposed by Howard Gardner [13]. SHANE LEGG & MARCUS HUTTER proposed a definition of intelligence as a measure of *'an agent's general ability to achieve goals in a wide range of environments'*. However, there is no agreed definition[14].

Ontologically, AI is based on a previous design to develop a task and achieve specific goals. The AI is limited to a set goal, even if it has astonishing learning capacities, which nevertheless are still aimed at obtaining a specific result, usually based on an inductive approach. Due to the realm of perception and its meaning, human intelligence is epistemologically broader [15]. AI systems have shown difficulties in dealing with semantic content, mainly with the open texture of natural and legal language. In the latter case, *ambiguity* (when the same legal concept have different meanings in different contexts), *vagueness* (neutral concepts that can possess intrinsic properties which are by themselves sufficient condition both to assign and not to assign the specific term), *variable standards* or *evaluative-open concepts* ("due care", "public interest"), or *defeasibility* (the forbidding rule about motor vehicles in the park does not apply to ambulances) [16], are some attributes of legal norms and language that require special attention to the particular case. As MIREILE HILDEBRANT pointed out, *"meaning depends on the entanglement of self-reflection, rational discourse and emotional awareness that hinges on the opacity of our dynamic and largely inaccessible unconscious"* [17].

However, if one accepts that meaning in law is normative and objective, in the sense of being reference-related and inter-subjectively valid [18], AI systems would be an undeniably helpful tool in this quest. Some small-scale algorithms have already been successful in resolving the open texture problem, such as the 1980's Case-Based Reasoning, although it hasn't advanced substantially since. Considering the growth of Big Data and the integration of suitable models and data sets with the deep learning capacities of AI, it is conceivable that significant advances are still to come in this area. The main problem, in our view, is not in dealing with semantic content, where AI could be a useful instrument. In the judicial point of view, the core

---

[13] Namely, linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic, intrapersonal, and interpersonal intelligences (see HOWARD GARDNER, *The Theory of Multiple Intelligences*, Basic Books, 2011).

[14] SHANE LEGG, & HUTTER MARCUS, Universal Intelligence: A Definition of Machine Intelligence, 2007, p. 12. Available at: https://arxiv.org/pdf/0712.3329v1.pdf (last visit 10 April 2019).

[15] WENCESLAO J. GONZALEZ, «From Intelligence to Rationality of Minds and Machines in Contemporary Society: The Sciences of Design and the Role of Information», *Minds & Machines*, 26 June 2017, pp.8-12

[16] Generally speaking, defeasibility is what happens when even though the scope of the rule is correctly determined and its applied to a given case to produce the conclusion C, it is possible to formulate the reason R and reject the conclusion C, cfr. F. BÉLTRAN & G. B. RATTI , «Validity and Defeasibility in the Legal Domain», *Law and Philosophy*, 29, 2010, pp. 601-626.

[17] MIREILLE HILDEBRANT, *Law as Computation in the Era of Artificial Intelligence. Speaking Law to the Power of Statistics*, 2017, p. 10, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2983045 (last visit 23 May 2019).

[18] MATTHIAS KLATT, *Making the Law Explicit - The Normativity of Legal Argumentation*, Hart Publishing, 2008, pp. 211 ff.

issue is to be found in the fundamental externally justificatory demands of legal discourse, where the opacity of the AI reasoning systems poses justified fears. In short, one should not overestimate human intelligence or underestimate the AI potential. Being two different kinds of intelligence they are not commensurable. As LUCIANO FLORIDI pointed out, AI pursues neither a *descriptive* nor a *prescriptive* approach to the world. It *inscribes* new artefacts that interact with nature, becoming part of it [19].

The difference between human intelligence and AI is about the same as the difference between «general, full or strong AI» and the «narrow or weak AI». The *strong AI* is related to AI systems which could carry out the same cognitive functions as humans (only probably better), applied to all problem solving or human activities. In other words, a strong AI would be an emulation, not just a simulation, of human intelligence, with volition and maybe even consciousness [20]. Nowadays, it is still science-fiction, despite the growing optimism that AGI will be achieved in this century. The actual AI systems are weak, or narrow, focusing on single subsets or in a pre-programmed way of working. Although the prevalent AI is narrow, it is nonetheless getting stronger and increasingly raising ethical concerns that could shake some basic foundations of human knowledge[21].

4. Trustworthy AI

In obtaining a reliable model of an AI agent, the quantity of the data fed to it is a crucial factor. But the quality of said data is even more so, for a lack of attention to data quality could easily lead to take correlation for causation, thus wrongly predicting a link between two unrelated phenomena and creating false positives/negatives. Indeed, knowledge is more than information; it requires explanation and understanding, not just truth or correlation [22]. The European Union took the lead with the Draft Ethics Guidelines for Trustworthy AI by the High-Level Expert Group on Artificial Intelligence (henceforth AI-HLEG), proposing the cornerstone concept of «trustworthy AI», admittedly influenced by the paper of LUCIANO FLORIDI *et alli* [23]. The «trustworthy AI» should be based on an ethical purpose, based on the respect for fundamental rights. The AI HLEG lists five principles and values for a human-centric AI: (1) *beneficence* ("do

---

[19] L. FLORIDI, *The Fourth Revolution...* (note 2), p.142.

[20] Christof Koch, chief scientific officer of the Allen Institute for Brain Science in Seattle, considers that consciousness is a property of matter well organized, just like mass or energy. If one could emulate a human brain, there would be consciousness, cfr. https://www.technologyreview.com/s/531146/what-it-will-take-for-computers-to-be-conscious/ (last visit 06.06.2019).

[21] Such as the Human Brain Project, launched in October, 2013, as an interdisciplinary European project involving several researchers of more than 100 institutions of 24 countries. This European project seeks to leverage cutting edge information and communication technologies, creating a multi-level brain simulation platform (see: https://www.humanbrainproject.eu/en/brain-simulation/, last visited 05 May 2019). This project raises medical hopes for the diagnosis and treatment of brain diseases, but also some ethical apprehensions. As Daniel Lim puts it, if we could emulate a human brain in a computer, there would be a new personhood, DANIEL LIM,. «Brain simulation and personhood: a concern with the Human Brain Project», *Ethics and Information Technology*, 20 October, 2013.

[22] L. FLORIDI, *The Fourth Revolution...* (note 2), p. 130.

[23] LUCIANO FLORIDI/JOSH COWLS/MONICA BELTRAMETTI/RAJA CHATILA/PATRICE CHAZERAND/VIRGINIA DIGNUM/CHRISTOPH LUETGE/ROBERT MADELIN/UGO PAGALLO/FRANCESCA ROSSI/BURKHARD SCHAFER/PEGGY VALCKE/EFFY VAYENA, «AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations», *Minds and Machines*, 2018.

good"); (2) *non-maleficence* ("do no harm"); (3) *autonomy* ("preserve human agency"); (4) *justice* ("be fair"); (5) and *explicability* ("operate transparently"). The AI algorithms should also be technically robust and reliable to deal with errors and inconsistencies, permitting corrections and calibrations [24].

In an auxiliary paper, the AI-HLEG offered a comprehensive definition for AI [25]:

"Artificial 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."

This definition assumes a new kind of intelligence through machine processing or computation, aimed at achieving set goals. Just as aeroplanes do not fly like birds, or submarines do not swim, so AI is not human intelligence *redux*. As the AI-HLEG points out, rationality does not exhaust the notion of intelligence, even though it is a significant part of it. This has a significant symbolic effect when it comes to the process of judicial decision: even if the judge is bound by the law, where his authority is delegated by the State, he is not exercising its power as an automaton, but as a human being before another human being.

5. Machine learning and deep neural networks

STUART RUSSEL and PETER NORVIG stated that the computational learning theory relies on this fundamental principle: *"any hypothesis that is seriously wrong will almost certainly be "found out" with high probability after a small number of examples because it will make an incorrect prediction. Thus, any hypothesis that is consistent with a sufficiently large set of training examples is unlikely to be seriously wrong, that is, it must be probably approximately correct"* [26]. Machine learning involves a set of techniques mostly dealing with a mix of statistics and computer engineering, from which the required computational algorithms are developed. It uses mathematical models with data sets, mainly obtained from the Big Data, where the parameters are configured during the learning phase, through different learning methods [27].

---

[24] High Level Expert Group on Artificial Intelligence, Draft Ethics Guidelines for Trustworthy AI, 2018. Available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelin-trustworthy-ai (last access 09 June 2019)

[25] High-Level Expert Group on Artificial Intelligence, A Definition of AI: Main Capabilities and Disciplines. 2019. Available at: https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines (last visit 09 June 2019).

[26] STUART J. RUSSELL & PETER NORVIG, *Artificial Intelligence: A Modern Approach*, 3 ed., Pearson Education Limited, 2016, p. 714.

[27] CONSTANCE DE SAINT-LAURENT, «In Defence of Machine Learning: Debunking the Myths of Artificial Intelligence», *Europe's Journal of Psychology*, 14(4), 2018, p. 737; XÁVIER ROSIN & VASILEIOS LAMPOS, «In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data», in *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment. Strasbourg*, CEPEJ - Commission Européene pour l'Efficacité de la Justice, 2018, p. 51, available at https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (last visit 16 June 2019).

Algorithms are not able to create neutral or non-discriminatory and independent predictions about future events since they are contingent from its previous design.

There are three main types of learning: *supervised*, *unsupervised* and *reinforcement learning*. In *supervised* machine learning the AI system is given pre-labelled data and required to work out the rules that connect them. Thus the agent observes a data set, interprets it as a set of possible input-desired output examples and creates a model of the underlying function so that the difference between the desired and predicted outputs is as small as possible for previously unseen patterns. The supervisor then compares the desired and the predicted outputs, adjusting the model.

In the *unsupervised learning* or self-organising systems, the machine is given no pointers and has no desired outputs. It has to identify the inputs and the outputs as well as the rules that connect them, even though there is no specific feedback. This type of learning is used in detecting potential useful clusters (grouping) of input examples: a self-driving taxi can develop the concept of «good traffic days» and «bad traffic days» without any previous labelled examples.

By *reinforcement learning,* the system gets feedback from the environment through artificial punishments or rewards. The decision made before the reward is solely the agent responsibility; there is no supervisor or human intervention. [28]. The ability to learn provides the AI system with the adaptability for solving problems in a complex and rapidly changing environment, achieving significant breakthroughs and challenging the dividing line between creativity and reason made by machines [29].

Deep learning through artificial neural networks is the most challenging and unique among machine learning algorithms as it exhibits many similarities with the biological neural networks. The deep learning and neural networks require a large amount of data and are extremely efficient in finding complex patterns. They use several layers of processing, each taking data from previous layers through fundamental units

---

[28] EZEQUIEL LÓPEZ-RUBIO, «Computational Functionalism for the Deep Learning Era», *Minds & Machines*, 5 October, 2018. p. 4; STUART J. RUSSELL & PETER NORVIG, *Artificial Intelligence...*, pp. 694-695 (note 26).

[29]   In 1996, William McCune solved the Robbins axiom in Boolean algebra, with the help of the Equational Prover program, succeeding where the best mathematicians had failed for 60 years (see, THE NEW YORK TIMES, «Computer Math Proof Shows Reasoning Power», December, 10, 1996, stored in https://archive.nytimes.com/www.nytimes.com/library/cyber/week/1210math.html (last visit 05 May 2019). In 1997, the IBM's Deep Blue chess program succeeded in defeating world champion Gary Kasparov in a six-game match. In 2011, it was used the supercomputer Watson in the famous American TV quiz show Jeopardy, outperformed its two human opponents. The Watson program is currently used in healthcare, as a diagnosis and treatment assistant, and in several educational projects. In 2015, the AlphaGo, developed by Google, became the first computer program to win a 9-dan professional in the board game Go, by a score of 4 to 1. It learns by examining hundreds of thousands of online Go games played between humans, using it as data for a machine-learning algorithm. AlphaGo played against different versions of itself, fine-tuning its strategies by deep reinforcement learning. It was considered by the *Science* magazine one of the breakthroughs of 2016 (see https://www.sciencemag.org/news/2016/12/ai-protein-folding-our-breakthrough-runners, last visit, 03.06.2017). The AlphaGo Zero computer program, with only a little period of training, beat its predecessor AlphaGo with a 100-0 victory. In 2018, ALVIN RAJKOMAR and EYAL OREN signed the study «Scalable and accurate deep learning with electronic health records», Npj (Nature Partner Journals), Digital Medicine, 1, n.º 18, 201, May 2018, available at https://www.nature.com/articles/s41746-018-0029-1 (last visit 04 2018), where it shows the performance of a predictive deep learning algorithm, that analysed clinical records of 216,221. It predicted with a level accuracy between 75 and 94% the risk of in-hospital mortality (93%-94%); 30-day unplanned readmission (75-76%), prolonged hospital stay (85-86%) and discharge diagnosis (90%).

called artificial neurons and passing an output up to the next layer. The nature of the output may vary according to the nature of the input, which can be weighted and not just turn on or off [30].

CHAPTER II
THE IMPLEMENTED SYSTEMS

1. The experiences within the European Union

There are several possible classifications on AI reasoning methods and techniques. To pinpoint in what way those technical categories can be seen as judicial AI tools, one could tackle some examples such as advanced case-law search engines, online dispute resolution, tools of assistance in drafting deeds, analysis tools (predictive or scales), categorisation of documents (such as contracts), or chatbots to offer legal information or legal support. Not all of the pinpointed examples can be transposed to a judicial decision point of view. Although the practical examples of the use of AI in judicial decisions are rare, most of the examples (as some of those stated above) are of tools used in a judicial context but mostly by private companies or other judicial actors, namely by lawyers and law firms. Besides the tools put in action by the private sector and judicial actors, there have been some academic projects using reasoning methods to predict judicial decisions that are worthy of mention, as we will see further on.

Some EU Member States already have some sort of AI judicial tools implemented and/or have a public political strategy to develop AI technologies, including AI in the administration of justice. In 2016, the UK made public a report in *"Robotics and artificial intelligence"* [31]; in 2017, Finland launched a strategic plan to turn the country into a leader in the application of AI (*"Finland's age of artificial intelligence"* [32]); in 2018, France also made public a report *"For a meaningful artificial intelligence towards a French and European strategy"* [33].

In the UK one can find *Luminance*, a tool of text analysis based on machine learning technology (pattern-recognition, as pointed out by the company [34]) that reviews documents and learns from the interaction between lawyers and documents; or *HART* (Harm Assessment Risk Tool), the algorithm that predicts the level of risk of suspects committing further crimes in a certain period of time [35], through an algorithm of "random forest", combining certain values, the majority of which focus on the suspect's offending history,

---

[30] C. CHACE, *The Artificial Intelligence and the Two Singularities*, p. 14 (note 7).

[31] Available on-line at https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf (visited on the 29th of May of 2019).

[32] Reports available on-line at: http://julkaisut.valtioneuvosto.fi/bitstream/hadle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y and http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf (visited on the 31st of May of 2019).

[33] https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf (visited on the 29th of May of 2019).

[34] The website: https://www.luminance.com/.

[35] As described by Cambridge University (https://www.cam.ac.uk/research/features/helping-police-make-custody-decisions-using-artificial-intelligence, last visited on the 26th of March of 2019), this algorithm helps the police to decide, after taking someone into custody, whether let the person free on police bail or keep him/her locked until going to court.

as well as age, gender and geographical area. In France, there are some tools such as *Doctrine*, *LexisNexis* and *Dalloz*, simple search engines for court decisions and other legal texts. More interesting are the software tools *Prédictice* and *Case Law Analytics*, both analysis tools with the aim of predicting the outcome of a specific case [36] ("trend" analysis tools, in fact).

France also conducted an experiment to test predictive justice software (the *Prédictice* software tool) on various litigation appeals in 2017, in the two courts of appeal in Rennes and Douai. The results were not optimal. In fact, the aim of the experiment was to try to reduce excessive variability in court decisions in the name of the citizen's equality before the law. The result was that the experiment did not add any valuable insight as to the role of AI in decision-making. It seems that the software got confused between lexical occurrences and the causalities that had been decisive to the judges in the decisions used as "data fuel", leading to absurd results [37].

In Austria AI has been used as a tool to structure information for the quick and efficient analysis and handling of documents [38]: the AI tool analyses incoming mail without any manual contact by the court's staff, extracting metadata, identifying and recognizing procedures to file documents and its categorization; it functions as a tool for digital file management (particularly important in the management of unstructured documents); as a tool for analysis in investigating data, namely analysing and classifying metadata from any form of data and recognition of communication flows and relationships; and, at last, as a tool for automatic anonymization of court decisions (personal data of the parties).

University College London (UCL) also conducted an investigation [39] to predict judicial decisions of the European Court of Human Rights using only the textual information extracted from relevant sections of ECtHR judgments. The investigators framed the task as a binary classification problem where the training data consisted of textual features extracted from given cases and the output was the actual decision made by the judges. The study predicted the outcome with 79% accuracy. The authors concluded that *"the information regarding the factual background of the case as this is formulated by the Court in the relevant subsection of its judgments is the most important part obtaining on average the strongest predictive performance of the Court's decision outcome"*, and that *"the rather robust correlation between the outcomes of cases and the text corresponding to fact patterns contained in the relevant subsections coheres well with other empirical work on judicial decision-making in hard cases and backs basic legal realist intuitions"*.

---

[36] https://predictice.com/ and https://www.caselawanalytics.com/

[37] XÁVIER ROSIN & VASILEIOS LAMPOS, «In-depth study on the use of AI in judicial system…p. 42 (note 27).

[38] *How is Austria approaching AI integration into judicial policies?*, a presentation from STAWA, Georg, President of the CEPEJ and Head of Department for Strategy, Organizational Consulting and Information Management, Federal Ministry for Constitution, Reforms, Deregulation and Justice, Austria, 2018, https://rm.coe.int/how-is-austria-approaching-ai-integration-into-judicial-poli-cies-/16808e4d81 (visited on the 12th of June 2019).

[39] NIKOLAOS ALETRAS, DIMITRIOS TSARAPATSANIS, DANIEL PREOŢIUC-PIETRO, VASILEIOS LAMPOS, *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective*, 2016, https://peerj.com/articles/cs-93/.

Another fruitful field of application of AI solutions in the judicial world is in small claims civil litigation. Many countries within the EU have already put in place – or are on the verge of doing so – some sort of Online Dispute Resolution (ODR) service. The Netherlands, United Kingdom, Latvia and Estonia are some of them. Estonia intends [40] to create a totally human-independent system that renders decisions in small claims up to €7.000,00. In theory, the two parties would upload documents and other relevant information and the AI technology (ODR) would issue a decision; that decision can be appealed to a human judge.

The UK ODR platform for small claims resolution is not a truly AI solution, since it is a human judge that decides the dispute. The main difference between this method and the traditional decision-making method is that all the contact between the user and the court is through the online platform. The other difference from a traditional approach is that there are *online facilitators,* that is, in Professor Richard Susskind's own words, *"individuals who will look at claims and bring the parties together negotiating and perhaps acting as mediators after some kind of guidance"* [41].

Latvia also has an ODR solution similar to the UK ODR in claims up to €2.100,00: it is a totally (or mostly) written procedure, submitted on-line by the claimant, and it only applies to small claims for recovery of money or for recovery of maintenance, and the application need to comply with specific rules on these proceedings (a certain form model or, for instance, the claimant has to indicate if he or she requests a court hearing to consider the matter). As the British ODR, the decision is rendered by a judge and not by any sort of AI tool[42].

The European Commission provides an ODR platform, as well, to help resolve consumer disputes on on-line purchases without going to court. It can be used for any contractual dispute arising from online purchases of goods or services where the trader and consumer are both based in the EU or Norway, Iceland, and Liechtenstein. This ODR is regulated by the Regulation (EU) no. 524/2013 of the European Parliament and of the Council of 21 May 2013. It is an ADR (alternative dispute resolution) and the platform merely works to facilitate communication between the parties and a dispute resolution body, without going to court. One of the biggest advantages of this ODR is that it provides automated translations between all EU languages, as well as information and support throughout [43].

The Netherlands ODR is the oldest one in Europe that we are aware of. The *e-Court* is a private initiative ADR launched in 2010 and, as the model intended by Estonia, it is a fully automatic AI decision render.

---

[40] For more details, the article on https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/ (visited on the 26th of March of 2019).

[41] For a brief and clear explanation of UK's ODR, https://www.judiciary.uk/reviews/online-dispute-resolution/what-is-odr/ (visited on the 5th of June 2019).

[42] For a more detailed analysis, https://e-justice.europa.eu/content_small_claims-42-lv-en.do?member=1 (visited on the 12th of June of 2019).

[43] For more details, the on-line address of the ODR platform https://ec.europa.eu/consumers/odr/main/?event=main.home2.show (visited on the 12th of June of 2019).

The creditor submits the required information (documents) and the decision is rendered without any human intervention. Nevertheless, to initiate enforcement proceedings, the users of e-Court still have to obtain an enforceable title, and this title is issued by humans. In fact, the automated on-line-made decisions are sent to a public court, where the clerks manually recalculate the awarded amounts [44].

Also worth mentioning is *Rechtwijzer*, another Dutch-made ODR solution: its mission was to reduce the burden of the legal process of divorce by reducing the adversarial nature of the process. The process started with a diagnosis phase, then the intake phase for the initiating party and, at last, the other party was invited to join and undertake the intake process. This platform was a channel of communication between the parties to work on agreements (and hopefully achieve them) on the topics needing resolution. Even though it was a solution based on the negotiation of the parties, they were also informed about the legal rules concerning the agreements negotiated (dividing property, child support, etc.) and, at the end of the on-line process, these agreements would be reviewed by a neutral third party (a lawyer) [45]. The *Rechtwijzer* project ended in 2017 and there seems to be no official explanation for its demise.

2. What is being done in the EU

Aside from what is already put into practice, in April 2018 the UE Member States signed a declaration of *"Cooperation on Artificial Intelligence"* [46], where the countries agreed to build a EU towards achievements and investments in AI, as well as progress towards the creation of a Digital Single Market. The same month the European Commission issued a communication on *"Artificial intelligence for Europe"* [47], addressed to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. In the communication, the Commission argues that UE *"should have a coordinated approach to make the most of the opportunities offered by AI and to address the new challenges that it brings"* [48], granting explicit support in AI research on *inter alia "public administrations (including justice)"* [49]. Later that year, the European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe launched an *"European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment"* [50]. Despite the path taken by the EU over these past

---

[44] For more details, the analysis of H.W.R. (HENRIËTTE) NAKAD-WESTSTRATE, H.J. (JAAP) VAN DEN HERIK, A.W. (TON) JONGBLOED AND ABDEL-BADEEH M. SALEM, *The Rise of the Robotic Judge in Modern Court Proceedings*, conference paper on the 7th International Conference on Information Technology, 2015, pp. 59-67.

[45] For more details, https://law-tech-a2j.org/odr/rechtwijzer-why-online-supported-dispute-resolution-is-hard-to-implement/. (visited on the 6th of June 2019).

[46] Available on-line at https://ec.europa.eu/jrc/communities/en/node/1286/document/eu-declaration-cooperation-artificial-intelligence (visited on the 6th of June 2019).

[47] Available on-line at https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe, (visited on the 6th of June 2019).

[48] Ibid, p. 3.

[49] Ibid, p. 8.

[50] Available at https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (visited on the 6th of June 2019).

years, there is still a long way to go concerning the use of AI technology in judicial decisions within the EU.

CHAPTER III
SOME ETHICAL AND LEGAL CHALLENGES

1. The European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment

Bearing in mind that the implementation of AI is not something for the far future but something for our time, the European Commission for the Efficiency of Justice (CEPEJ) formally adopted the five fundamental principles on the use of AI in judicial systems and their environment previously mentioned in Chapter 1 [51]. These principles aim to guarantee the respect of the European Convention on Human Rights (ECHR) and the Convention on the Protection of Personal Data (CPPD) by framing public policies on this field, and assuring that the processing of AI respects principles such as the transparency, impartiality and equality, certified by an external and independent expert assessment.

These principles, however, are not to be written in stone. The CEPEJ intends to subject them to monitoring and supervision with the aim of a continuous improvement of practices. For now, the five principles are: 1. *Respect for fundamental rights*: ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights [52] ; 2. *Non-discrimination*: specifically prevent the development or intensification of any discrimination between individuals or groups of individuals [53]; 3. *Principle of quality and security*: with regard to the processing of judicial decisions and data, use certified sources and intangible data with models elaborated in a multi-disciplinary manner, and in a secure technological environment [54]; 4. *Principle of transparency, impartiality and fairness:* makes data processing methods accessible and understandable, and authorise external audits [55]; and 5. *Principle "under user control":* precludes a prescriptive approach and ensures that users are informed actors and in control of the

---

[51] European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, pp. 7-12.

[52] The processing of the data must serve clear purposes, in compliance with the ECHR and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data; the use of AI to assist in judicial decision-making must not undermine the guarantees of the right of access to the judge and the right to a fair trial, i.e., equality of arms and respect for adversarial process; ethical-by-design approach, meaning that the ethical choices are made in the design phase and never left to the user

[53] The AI users must ensure that the methods do not reproduce or aggravate such discrimination; there must be taken measures in the development and deployment phases when processing sensitive data, ensuring that when discrimination has been identified, must be taken measures to limit or neutralise these risks, as well as awareness-rising among stakeholders; AI use to combat discriminations is encouraged

[54] Through a multidisciplinary approach – designers of machine learning, justice system professionals and researchers in the fields of law and social sciences; data used on the machine learning process should come from certified sources and should not be modified until they have been used, and the whole process must be traceable; secure environments to ensure system integrity and intangibility

[55] A balance between the intellectual property, the need for transparency, impartiality, fairness and intellectual integrity, applying to the whole process; it should be able to be certified and audited by independent authorities; public authorities should grant certification, regularly reviewed

choices made [56]. These five principles tackle some of the main ethical issues posed by the use of AI tools in a judicial system and their environment, as well as the principles and legal barriers that surround this field within the EU.

2. The use and automatic treatment of personal data

Article 9(1)(a) of the CE's Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data [57] provides the principle that *"Everyone has the right not to be subject to a decision affecting him significantly, which shall be taken solely on the basis of automatic processing of data, without his point of view being taken into account. Notwithstanding this principle of prohibition, Article 9(2) states that "paragraph 1(a) shall not apply if the decision is authorised by a law to which the controller is subject and which also provides for appropriate measures to safeguard the rights, freedoms and legitimate interests of the data subject"* (in a similar sense, see article 22 of the General Data Protection Regulation).

In *Z. v. Finland* [58], concerning Article 8 [59], the court stated that the protection of personal data is of fundamental importance to a person's enjoyment of his or her right to privacy and family life, just as *"[r]especting the confidentiality of health data is a vital principle in the legal systems of all the Contracting Parties to the Convention. It is crucial not only to respect the **sense of privacy** of a patient but also to preserve his or her **confidence** in the medical profession and in the health services in general."*[60]

Recently, the court stressed [61] the fact that it has consistently held that systematic storage and other use of information relating to an individual's private life by public authorities entails important implications for the interests protected by Article 8 of ECHR. Thus any interference will be in breach of the ECHR unless it is in **accordance with the law** and shows itself to be necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others

---

[56] User autonomy should be increased; the possibility of review judicial decisions and the data used to produce the result; informed consent, meaning that the user must be informed in a clear way if the AI tools are binding, the alternative options available, the right to legal advice and the right to access a court within the meaning of Article 6 of the ECHR; literacy programmes on the use of the AI tools

[57] As amended by the Protocol adopted in May 2018.

[58] *Z. v. Finland*, n. º 22009/93, §§95, 25 February 1997

[59] *1. Everyone has the right to respect for his private and family life, his home and his correspondence. 2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.*

[60] As to regarding public access to personal data, the court recognises that "a margin of appreciation should be left to the competent national authorities in striking a fair balance between the interest of publicity of court proceedings, on the one hand, and the interests of a party or a third person in maintaining the confidentiality of such data, on the other hand. The scope of this margin will depend on such factors as the nature and seriousness of the interests at stake and the gravity of the interference (see, for instance, the Leander v. Sweden judgment of 26 March 1987, Series A no. 116, p. 25, para. 58; and, mutatis mutandis, the Manoussakis and Others v. Greece judgment of 26 September 1996, Reports 1996-IV, p. 1364, para. 44)."

[61] *Surikov v. Ukraine*, n. º 42788/06, §§70-74, 26 January 2017.

(Article 8 § 2 ECHR). The court also stressed that it is well established case law that *accordance with law* requires it to be **accessible**, **foreseeable** and accompanied by necessary **procedural safeguards** affording adequate legal protection against arbitrary application of the relevant legal provisions [62].

3. Some recent decisions of the ECtHR and the ECJ

One of the biggest challenges put forward by the use of AI in judicial systems and their environment is the compliance with the rights and principles enshrined within the ECHR. As stated in the *European Ethical Charter*, these solutions have to comply with such individual rights as *"the right to a fair trial (particularly the right to a natural judge established by law, the right to an independent and impartial tribunal and equality of arms in judicial proceedings) and, where insufficient care has been taken to protect data communicated in open data, the right to respect for private and family life"* [63].

In the judgement *Vernes v. France* [64], the court found several violations of Article 6 §1, but as far as what concerns us most right now, it found a violation due to the impossibility for the applicant to request a public hearing [65] and a violation due to the lack of impartiality of the administrative body resulting from the absence of an indication of its composition [66]. In its judgement, the Court recalled that **public hearing is a fundamental principle** enshrined in Article 6 § 1 of the Convention. This principle may suffer from adjustments justified in particular by the interests of the private life of the parties or the safeguarding of justice (*Diennet v. France*, n. º 18160/91, 26 September 1995) or by the nature of the matters submitted to the judge in the context of the proceedings in question (*Miller v. Sweden*, n. º 55853/00, 8 February 2005, *Göç v. Turkey*, n. º 36590/97, § 47,). The Court concluded that, in the absence of a public hearing, the applicant's right to a fair trial was not ensured. The Court also recalled that for the purposes of Article 6 § 1, **impartiality** must be assessed on the basis of a subjective approach, allowing to determine the personal conviction of a judge on such an occasion, and also according to an objective approach such as to ensure that it offered sufficient safeguards to exclude any legitimate doubt in this respect, and, thus being so, the Court agreed with the applicant that the failure to state the identity of all the members of the administrative body who deliberated was such as to cast doubt on its impartiality, which implies that impartiality is also assured by the **identification of the judges** who rendered the decision. This is a factor of undeniable relevance if and when a case should be judged by an AI.

---

[62] *Ibidem*, §71.
[63] *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, CEJEP, §8, p. 15.
[64] *Vernes v. France*, n. º 30183/06, ECHR, 20 January 2011.
[65] *Vernes v. France*, n. º 30183/06, §§30-31, 20 January 2011.
[66] *Vernes v. France*, n. º 30183/06, §§41-44, 20 January 2011.

In *Golder v. United Kingdom*, the court recognized a **right of access to a court** but also stated that it is not absolute, admitting some **implied limitations** [67] [68]. The case law established in *Deweer v. Belgium* [69] made clear that the **right to a court** is perceived as an element of the right to a fair trial, enshrined in Article 6 §1, and it is **no more absolute in criminal than in civil matters**.

In *Kontalexis v. Greece* [70], the ECHR recalls that under Article 6 § 1, a court must always be **established by law**, which reflects the principle of the rule of law, inherent in the entire system of the Convention and its protocols (§38). The court states that a body that has not been established in accordance with the will of the legislator, would necessarily lack the legitimacy required in a democratic society. In *DMD GROUP, a.s, v. Slovakia*, the court reiterated the notion of a court established by law, and that the paramount importance of **judicial independence** and legal certainty for the rule of law calls for particular clarity of the rules applied in any case and for clear safeguards to ensure **objectivity and transparency**, as to avoid any appearance of arbitrariness in the assignment of particular cases to judges [71]. And, more recently [72], the Court considered that where the assignment of a case is discretionary in the sense that the modalities are not prescribed by law, it puts at risk the appearance of **impartiality**, by allowing speculation about the influence of political or other forces on the assignee court and the judge in charge, even where the assignment of the case to the specific judge in itself follows transparent criteria.

Regarding the independence and impartiality of the court, as required by Article 6 §1, it is settled case law that impartiality must be determined *"according to a **subjective test**, where regard must be had to the personal conviction and behaviour of a particular judge, that is, whether the judge held any personal prejudice or bias in a given case* [which must be presumed until proved otherwise]*; and also according to an **objective test**, that is to say by ascertaining whether the tribunal itself and, among other aspects, its composition, offered sufficient guarantees to exclude any legitimate doubt in respect of its impartiality (see, for example, Morice v. France [GC], no. 29369/10, § 73, 23 April 2015 and the cases cited therein)"* [73].

Concerning the principle of equality of arms, the court held recently [74] that *"the **adversarial principle** and the **principle of equality of arms**, which are closely linked, are fundamental components of the concept of a "fair hearing" within the meaning of Article 6 § 1 of the Convention. They require a "fair balance"*

---

[67] *Golder v. The United Kingdom*, n. º 4451/70, §38, 21 February 1975: *"As this is a right which the Convention sets forth (see Articles 13, 14, 17 and 25) (art. 13, art. 14, art. 17, art. 25) without, in the narrower sense of the term, defining, there is room, apart from the bounds delimiting the very content of any right, for limitations permitted by implication."*

[68] But even where there are implied limitations, some other aspects of the right enshrined on Article 6 § 1 must be observed, such as the right to be heard before a court within a reasonable time (cfr. *Kart v. Turkey*, n. º 8917/05, §67-70, 3 December 2009).

[69] *Deweer v. Belgium*, n. º 6903/75, §49, 27 February 1980.

[70] *Kontalexis v. Greece*, n. º 59000/08, 31 May 2011.

[71] *DMD GROUP, a.s, v. Slovakia*, n. º 19334/03, § 66, 05 October 2010.

[72] *Miracle Europe KFT v. Hungary*, n. º 57774/13, §58, 12 January 2016.

[73] *Ivanovski v. "the Former Yugoslav Republic of Macedonia"*, n. º 29908/11, §§136-141, 21 January 2016.

[74] *Prebil v. Slovenia*, n. º 29278/16, §§42-45, 19 March 2019.

between the parties: each party must be afforded a reasonable opportunity to present his case under conditions that do not place him at a substantial disadvantage vis-à-vis his opponent or opponents (see Regner, cited above, § 146)."

One of the most important rights enshrined in Article 6 §2 is the **presumption of innocence**. The ECHR perceives it as the right to be presumed innocent until proven guilty according to the law. It is *"viewed as a procedural guarantee in the context of a criminal trial itself"*, but the presumption of innocence also *"imposes requirements in respect of, inter alia, the burden of proof; legal presumptions of fact and law; the privilege against self-incrimination; pre-trial publicity; and premature expressions, by the trial court or by other public officials, of a defendant's guilt (see Allen v. the United Kingdom [GC], no. 25424/09, § 93, ECHR 2013, and the references cited therein)"* [75].

Combining almost all provisions of the rights enshrined in Article 6, very recently, the court was challenged in *Sigurdur Einarsson a. o. v. Iceland* [76] with potential violations of said article in a criminal proceeding where the defendant alleged, *inter alia*, that he had been denied full access to the file held by the prosecution. The criminal proceedings concerned a potential criminal conduct in connection with the collapse of one of the country's largest banks during the financial crisis that hit Iceland in 2008. The investigation lasted almost three years and led to an extensive collection of data (inclusive data seized due to a court search warrant). To conduct a search of the electronic data, the prosecution used a AI tool called "Clearwell", an e-discovery system, whose results were exported and tagged as "investigation documents". The applicants complained that they never had the opportunity to review the documents submitted to the court and that they had been denied the possibility of searching the same data using the electronic system applied. This substantiates, in their view, a violation of the principle of equality of arms (relying on Article 6 §1 and §3(b)) because they should have had the same opportunities as the prosecution to access and select evidence from the collection of documents gathered by the police during the investigation.

The court didn't find any violation of Article 6 on mass data that was not tagged, stating that to that extent the prosecution did not hold any advantage over the defence (it was not a situation of *non-disclosure*). Regarding the tagged data, this was reviewed by the investigators (manually and through "Clearwell") in order to pick which material should be in the investigation file. The court recognized that this selection was made by the prosecution alone without the involvement of the defence or any judicial supervision, as well as that further searches by the defence in the data would have been technically possible and appropriate for a search for potential disculpatory evidence. The court thus concluded that *"any refusal to allow the defence*

---

[75] *Kangers v. Latvia*, n. ° 35726/10, §50, 14 March 2019. Also, *Lolov v. Bulgary*, n. ° 6123/11, 21 February 2019; *Allenet de Ribemont v France*, 10 February 1995; *Viorel Burzo v. Romany*, n. s 75109/01 and 12639/02, 30 June 2009; *Lizaso Azconobieta v. Spain*, n. ° 28834/08, 28 June 2011.

[76] *Sigurdur Einarsson a. o. v. Iceland*, n. ° 39757/15, 4 June 2019.

to have further searches of the "tagged" documents carried out would in principle raise an issue under Article 6 § 3(b) with regard to the provision of adequate facilities for the preparation of the defence" [77]. The case law of *Sigurdur Einarsson a. o. v. Iceland* is paramount in the combination of AI tools and the rights enshrined in Article 6. Even though it was not held in violation of the article due a procedural formality, it established a clear principle that where AI tools are used to deal with massive data and information is extracted through that mechanism, the **principle of equality of arms** (Article 6 §1) and the **right to have adequate time and facilities** for the preparation of defence demands that the defendant (in a criminal case or in a civil claim, as stated in *Deweer v. Belgium*) has the right to participate in the cherry picking of information and has the right to conduct his/her own search through the data using the same tool as the prosecution.

Regarding the European Union, in the joined cases C-293/17 and C-294/17, the Council of State of the Netherlands requested a preliminary ruling from the ECJ, where it was asked whether Article 6(2) of the Habitats Directive [78] could be interpreted as meaning that measures such as procedures for the surveillance and monitoring of farms whose activities cause nitrogen deposition and the possibility of imposing penalties, up to and including the closure of those farms, are sufficient for the purposes of complying with that provision. The answer was positive [79].

4. Ethical concerns: opacity and anchoring

The black box problem, arising mainly in deep neural networks, is what happens when the AI agent gives a result in a way that humans or even its creators cannot understand or explain how it was achieved, even though the accuracy outperforms human decisions or predictions. Its use in judicial ruling could be a threat to some nuclear concepts in judicial decision, such as causation and intention [80]. Moreover, it could lead to suspicions about the parameters or variables used in the AI agent, casting doubts on judicial independence. If we want to preserve the essential core of judicial ruling, we must not accept a simple "computer says no" answer. However, one cannot ask the human judge to open up his brain and describe how

---

[77] *Ibidem*, §85-91. Despite the mentioned conclusion, the court dismissed a violation of Article 6 §3(b) because it found that the applicant did not formally seek a court order to have access to those documents, and thus it was not denied a fair trial overall. The judgement had a dissenting opinion on this part from Judge Pavli. Particularly interesting, §10: *"It is worth recalling at this point that what is at stake in this case is a fundamental tenet of criminal due process, namely equality of arms. In the light of this cardinal principle, the majority's overall approach seems insufficiently attuned to the complexities of electronic disclosure in criminal (or for that matter, civil) proceedings involving high-volume data; to the use of modern technological tools in this context; and to their combined implications for equality of arms. The assumption that standard rules of disclosure ought to apply unchanged in this context is one that, at the very least, needs to be tested."*

[78] Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora.

[79] *"In the light of the foregoing, the answer to the eighth question in Case C-293/17 is that Article 6(2) of the Habitats Directive must be interpreted as meaning that measures introduced by national legislation, such as that at issue in the main proceedings, including procedures for the surveillance and monitoring of farms whose activities cause nitrogen deposition and the possibility of imposing penalties, up to and including the closure of those farms, are sufficient for the purposes of complying with that provision." (§137)*

[80] YAVAR BATHAEE, «The Artificial Intelligence Black Box and the Failure of Intent and Causation», *Harvard Journal of Law & Technology*, vol. 31, n.º 2, 2018, p. 938.

he got to his ruling, just as we cannot expect full transparency from AI algorithms that can only be achieved at the expense of its performance. The full transparency of the AI agent would probably not be understandable for the majority of people; to which we should add the issues concerning the intellectual property rights over the algorithms. More important than knowing how an AI agent gets its results is assuring it has enough explanatory power, for instance, through the use of unconditional counterfactuals as a mean to provide explanations and surpass the opacity of the black box [81]. Regardless of the technique employed or the uses of the AI agent, it should always be guaranteed, as the ECtHR strongly points out, the adversarial process in the judicial decision-making, in order to assure transparency and reinforce people's confidence in the rule of law.

Another ethical concern is related to a possible anchoring effect. If the AI decision is evidence-based, the judge will tend to follow it, relinquishing his own decision. And the more he trusts his AI assistant's expertise, the more the judge will be depending on the machine for his rulings [82]. Nonetheless, if the AI agent is «trustworthy», in the sense meant by the AI-HLEG, this is actually good news. With this powerful ally, the judge would make better decisions, faster, and more fairly, provided that the dialectical nature of the procedure would be assured [83] and that the AI's assistance could always be challenged by the parties.

5.   Final remarks

In his 2018 book, **_Enlightenment Now_**, Steven Pinker noted that "_intelligence is a contraption of gadgets: software modules that acquire, or are programmed with, knowledge of how to pursue various goals in various domains_". When defined as the "ability to deploy novel means to attain a goal", intelligence is a common property of machines and humans alike, and those two very different forms of its manifestation, artificial and human, will hopefully allow for a fruitful coexistence and cooperation in the Judiciary.

Volition, values and affection play a significant role in human decision-making. AI does not have intentionality or a real attitude, but only set tasks and goals; it does not make real judgements based on principles, rules, priorities or values. Even if the algorithm learns some principles, values and rules, the range would be limited to those which are significant to the model in order to accomplish its goal. On the other hand, human intelligence goes beyond this strictly cognitive domain, because it is connected to actions and rests

---

[81] SANDRA WACHTER, BRENT MITTELSTADT, CHRIS RUSSELL, _Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR_, available at https://ssrn.com/abstract=3063289 (last access 09 June 2019). A counterfactual (or contrary-to-fact) is a conditional sentence in the subjunctive mood, such as 'if you had broken the bone, the X-ray would have looked different', or 'if the reactor were to fail, this mechanism would click in'. It carries the suggestion that the antecedent of such a conditional is false. Since counterfactuals could be related to all kind of possible worlds, it's important that the world we are using is close to the real world, that is, it should be the closest possible world, see SIMON BLACKBURN, _The Oxford Dictionary of Philosophy_, Oxford University Press, p. 85-86.

[82] IAN KERR, CARISSIMA MATHEN, _Chief Justice John Robert is a Robot_, p. 8, paper presented on 5th April, at the WeRobot 2014 Conference, available at http://robots.law.miami.edu/2014/wp-content/uploads/2013/06/Chief-Justice-John-Roberts-is-a-Robot-March-13-.pdf (last access 13 June 2019).

[83] JOÃO MARQUES MARTINS, «A system of communication rules for justifying and explaining beliefs about facts in civil trials», _Artificial Intelligence and Law_, 05 March, 2019.

on a large collection of values [84]. Applying the law is more than a simple logical syllogism, as Justice Oliver Wendell Holmes once implied. Judging is a mix of skills, including research, language, logic, creative problem solving and social skills [85]. Nonetheless, interpretation and application of law necessarily imply argumentation, oral or written, and explanatory capacities in which logic analysis play an important role [86]. AI systems could be helpful devices to the judicial ruling, above all in preventing biases or transient emotional instability affecting the decision. Judges are subject to personal and work-related stress and burnout, which can naturally shake the decision-making objectivity [87], where AI are less prone to these flaws [88].

We should encourage the use of AI agents that are less susceptible to inspire mistrust as a way of incrementing the judiciary's productivity [89]. One should not bet on a "legal singularity", in which AI assistance to the judicial ruling will get it right all the time, thus eradicating any legal uncertainty. However, all the help in trying to achieve this purpose should be *prudently* welcome.

***

---

[84] Wenceslao J. Gonzalez, «From Intelligence…», p. 10 (see note 15).

[85] Richard A. Posner, *Cómo Deciden los Juices*, Marcial Pons, 2011, pp. 16 e ss. (Victoria Roca Pérez, Spanish translation of *How Judges Think*, Harvard University Press, 2008).

[86] E. Bulygin, «What can one expect from Logic in the Law? (Not everything, but more than something)», *Ratio Juris*, 2008, p. 21. As Bulygin points out, "*that logic cannot give a full account of any legal system is obvious; I wonder who (…) could expect it to. I know of no legal philosopher who would raise such a claim. What logic, or rather logical analysis, can do, however, is to clarify legal concepts and thus introduce greater order, thereby deepening our understanding of legal phenomena.*"

[87] Take, for instance, a study which concluded that judges were more likely to accept prisoner's requests for parole at the beginning of the day than at the end, Shai Danzigera, Jonathan Levavb, and Liora Avnaim-Pesso, «Extraneous factors in judicial decisions», PNAS, vol. 108, n.º 17, 2011, available at https://www.pnas.org/content/pnas/108/17/6889.full.pdf (last visit 11.06.2019).

[88] Past experiences had shown that AI decision-making could reveal structural biases, such as the Correctional Offender Management Profiling for Alternative Sanctions – COMPAS, that was used for predicting the likelihood of defendants committing a future crime, which was considered racially biased against African American defendants. However, these shortcomings can always be corrected.

[89] Thomas Julius Buocz, « Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary», *Retskraft – Copenhagen Journal of Legal Studies*, vol. 2, n.º 1, 2018, p. 50.