CADERNOS DE DEREITO ACTUAL
www.cadernosdedereitoactual.es

Cadernos de Dereito Actual Nº 23. Núm. Extraordinario (2024), pp. 25-44
ISSN 2340-860X - ·ISSNe 2386-5229

Sentencing in the age of artificial intelligence: possibilities, ethical issues, and the intrinsic human nature of the adjudication process

La sentencia penal en la era de la inteligencia artificial: posibilidades, cuestiones éticas y la naturaleza humana intrínseca del proceso de determinación judicial de las penas

Américo Bedê Jr.[1]
Gabriel Silveira de Queirós Campos[2]
Faculdade de Direito de Vitória – FDV

Summary: 1. Introduction; 2. Technological evolution and the transformations of the State and its Justice; 3. Artificial Intelligence and sentencing in an "age of uncertainty"; 4. Are judges going to be replaced by machines? 5. Is sentencing essentially human?; 6. Conclusions; 7. References.

Abstract: The world's technological advancement in times of big data, machine learning, the 'Internet of Things' and artificial intelligence (AI) has been remarkable in recent decades. In the 'digital era', technological innovations such as electronic petition systems (PJe, EPROC and PROJUD) are already part of the reality of the Judiciary branch in Brazil. Artificial intelligence systems are also beginning to be developed in many courts. The present study addresses the problem of using AI at sentencing, exploring the potential of algorithms and the main ethical issues involved: fairness, accountability, and transparency. The research problematizes the most common approach found in specialized literature on the subject, according to which the advantages and disadvantages of using AI in sentencing must be measured vis-à-vis decisions traditionally made by human judges. Despite acknowledging that computerized sentences can produce more precise, consistent, and predictable judgments (therefore, more objective), the study argues that sentencing is essentially a human task, involving sensitivity and interpretation. Transforming sentencing into a purely mechanical, even machine-automated task, is to remove its inherently human element.

Keywords: Artificial Intelligence; algorithms; Ethics; judicial decision-making; sentencing.

[1] Ph.D. and Master in Law (Faculdade de Direito de Vitória). Law professor (Faculdade de Direito de Vitória). Judge in Espírito Santo, Brazil (Federal court system). Email: bedejunior@hotmail.com. ORCID: 0000-0003-0128-8790.
[2] Ph.D. Candidate in Law (Faculdade de Direito de Vitória). MSc in Criminology and Criminal Justice (University of Oxford). Federal Prosecutor in Brazil. Email: gabrielsqcampos@gmail.com. ORCID: 0000-0002-9244-157X.

Recibido: 15/01/2024
Aceptado: 15/02/2024
DOI: 10.5281/zenodo.10898304

Resumen: El avance tecnológico mundial en tiempos de big data, aprendizaje automático, 'Internet de las cosas' e inteligencia artificial (IA) ha sido notable en las últimas décadas. En la "era digital", innovaciones tecnológicas como los sistemas de petición electrónica (PJe, EPROC y PROJUD) ya son parte de la realidad del Poder Judicial en Brasil. También se están empezando a desarrollar sistemas de inteligencia artificial en muchos tribunales. El presente estudio aborda el problema del uso de la IA en la sentencia, explorando el potencial de los algoritmos y las principales cuestiones éticas involucradas: equidad, rendición de cuentas y transparencia. La investigación problematiza el enfoque más común encontrado en la literatura especializada sobre el tema, según el cual las ventajas y desventajas del uso de la IA en la sentencia deben medirse frente a las decisiones tradicionalmente tomadas por jueces humanos. A pesar de reconocer que las sentencias computarizadas pueden producir juicios más precisos, consistentes y predecibles (por lo tanto, más objetivos), el estudio sostiene que la sentencia es esencialmente una tarea humana, que involucra sensibilidad e interpretación. Transformar la sentencia en una tarea puramente mecánica, incluso automatizada por una máquina, es eliminar su elemento inherentemente humano.

Palabras-clave: Inteligencia artificial; algoritmos; Ética; toma de decisiones judiciales; sentencia penal.

1. INTRODUCTION

Would you rather be judged by a human or a robot?

If we think about the reality of criminal justice, in Brazil and around the world, the question above is nothing more than a mere hypothesis. When we try to describe what a judge looks like, a circumspect, middle-aged man wearing a black robe and holding a gavel on his hand immediately comes to mind.

However, times are changing. As in the film '2001: A Space Odyssey', by Stanley Kubrick, released in the distant year of 1968, perhaps one day human judges will disappear and court decisions will be made by some form of Artificial Intelligence like HAL 9000, one which will be capable of listening, speaking, reasoning, feeling, and – perhaps not – deciding the fate of human lives within the scope of criminal justice.

Science fiction metaphors aside, the issue of deploying Artificial Intelligence (AI) within Criminal Justice Systems, particularly at sentencing, demands greater attention. This is a current and complex topic. It is necessary to be clear about the possibilities of using algorithms in criminal proceedings and their ethical limits. There are still many doubts and misunderstandings. With a great deal of exaggeration, people even defend the entire replacement of (human) judges by computers.

The research problem to be faced in the present study is the following: is the use of Artificial Intelligence in sentencing compatible with the human nature of decision-making at the sentencing phase of a criminal proceeding?

In the pursue of an answer to that question, the article will initially outline a brief historical retrospective of the evolution of technology, culminating in the so-called 'fourth Industrial Revolution' and the development of Artificial Intelligence. The transformations undergone by the State and its Justice as a result of new technologies will be synthetically explored in the initial part of the study, contextualizing the use of AI tools within the scope of the Brazilian Judiciary today.

Afterwards, the article will explore some possibilities for the use of computer algorithms in sentencing, seeking to dispel some common doubts at a historical moment in which there is still a lot of uncertainty surrounding the topic of Artificial Intelligence. The biggest concerns involve the eventual overcoming of human intelligence by machines and the question of whether machines will make decisions traditionally attributed to humans.

Subsequently, the study will address the main ethical issues raised by the possibility of using algorithms in sentencing, highlighting problems related to justice

(fairness), responsibility (accountability), and transparency. While such issues are in some ways equally present in any kind of human judgment, the focus of this part of the article will be to critically examine the problems of algorithmic discrimination and opacity.

Finally, the inherently human nature of sentencing will be discussed, presenting a comprehensive view of the complexity of sentencing – understood as a social and human process and interpretative activity – and the allocation of punishment.

2. TECHNOLOGICAL EVOLUTION AND THE TRANSFORMATIONS OF THE STATE AND ITS JUSTICE

The world's technological advancement in times of big data, machine learning, the 'Internet of Things' and Artificial Intelligence (AI) has been remarkable in recent decades. The history of technological revolutions, however, is much older, with references to at least four major 'industrial revolutions' recorded in a historical interval of just under four hundred years[3]. The first Industrial Revolution, often referred to simply as the 'Industrial Revolution', which occurred around 1760 to 1840, was characterized above all by the transition from manual production processes to mechanized factory systems. The use of water and steam as propellants for engines and machines for industrial use allowed for a huge leap in production. The second Industrial Revolution, from the second half of the 19th century, between 1850 and 1870, until the end of the Second World War (1939-1945), was basically an electrical revolution, which meant the end of dependence on coal as an energy source, allowing new leap in productivity for the industry. The third Industrial Revolution, which began in the post-World War II period, was marked by the creation of electronic equipment used to further increase production, by the development of transistors and processors and by the invention of computers. It became known, therefore, as the automation revolution or digital revolution. Today we are experiencing a fourth Industrial Revolution, the so-called 'industry 4.0', 'Internet of Things' (IoT), Artificial Intelligence, robotics, and cloud computing. In the 'digital age', we have surpassed the analog world of decades ago and started to live surrounded by computers and other electronic devices with inversely proportional architecture and processing capacity, that is, increasingly smaller and more powerful.

So many technological revolutions could not occur without corresponding and equally profound transformations in society and the State. From the point of view of power relations, the digital world is experiencing an unprecedented form of governance, concentrated in a few corporations (Big Techs), holding a huge amount of information (big data) from internet users around the world. Data acquired through constant surveillance or cyber patrolling, especially on social networks, is transformed into the raw material of the digital economy, therefore monetized, generating the so-called 'surveillance capitalism'[4].

The entire human experience, previously analog, is transformed into the hyper-connected world of artificial intelligence and omnipresent (omniscient and omnipotent?) algorithms. There are those who argue that the 'digitization of life' is changing human existence itself. For the philosopher of technology Éric Sadin, we would be granting machines a tacit delegation so that they, from the height of their wisdom, decide on the evolution of the world. We would thus live in a historical moment in which our condition would go from human to hybrid, as human and Artificial Intelligence would irreversibly combine, generating an *humanité augmentee*[5]*.*

---

[3]SCHWAB, K. *A quarta revolução industrial,* Edipro, São Paulo, 2016, pp. 15-16.
[4]ZUBOFF, S. *A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira do poder*, Intrínseca, Rio de Janeiro, 2020, p. 18.
[5]SADIN, É. *La humanidad aumentada: la administración digital del mundo*, Caja Negra, Buenos Aires, Argentina, 2022, p. 154.

Even without the pretension of any sociological and anthropological analysis, the fact is that society has changed in the digital age. At the political-legal level, major transformations are also underway. In reality, the State itself, as a historical institution, has always been in constant change. Jose Luis Bolzan de Morais describes the dynamics of the State – a preferable term to 'evolution' – and its transformations based on the issues that confronted it (and still confront it) in its trajectory. First, the individual issue, that is, the protection of the individual against the State, which marked the 18th century, in the transition from monarchical absolutism to the Liberal State. Then, with the advance of capitalism and its ills, the social issue emerged and in response to it the state institution took the form of the Social State, with the incorporation of social rights, especially those related to health and work. Later, in the middle of the 20th century, the environmental issue challenged the idea of territoriality of the State, crumbling political borders and weakening modern statehood itself. New global arrangements needed to be implemented, mitigating the notion of sovereignty in favor of greater solidarity and cooperative state action. Finally, the advent of cybernetics, the expansion and massification of microcomputers and the private development of the internet have placed the State before a digital issue, which also challenges boundaries and exposes the smallness of the state institution in dealing with a progressively complex reality[6].

The digital age, therefore, has had a huge impact on the State itself. Probably the biggest challenge involves the regulation of digital life, which confronts traditional state heteroregulation with new forms of self-regulation, sometimes referred to as 'regulated self-regulation'[7]. From another perspective, in the daily life of the State and its institutions, there is a strong process of virtualization or digitalization, which is also the effect of the unstoppable advance of new technologies. In the limited space available for this study, therefore, our greatest interest will be in the incorporation of technological innovations such as artificial intelligence by the Judiciary, particularly within the Criminal Justice System.

When it comes to the use of new technologies by the Brazilian Judiciary, it is legitimate to ask where exactly we stand. The digitalization of processes is a reality, with a multitude of electronic petition systems available to lawyers, members of the Prosecution Service (*Ministério Público*) and judges, such as PJe, Eproc and PROJUD, for instance. Regarding Artificial Intelligence, there are some systems in production and even in operation in several courts[8]. In the Federal Supreme Court (*Supremo Tribunal Federal* or STF), e.g., RAFA 2030 (Artificial Networks Focused on Agenda 2030) is a machine learning algorithm whose objective is to classify processes by sustainable development objective (SDG) of the UN Agenda 2030. The STF also develops an IA system named VICTOR, which supports the analysis by judges of the admissibility of sentence appeals by indicating topics of general repercussion, with an AI based on machine learning and natural language (text). At the Superior Court of Justice (*Superior Tribunal de Justiça* or STJ), ATHOS identifies repetitive themes, triangulates jurisprudence and searches for similar findings. The STJ also has a robot to identify reasons for inadmissibility of the REsp and to index procedural documents in original processes. In turn, the Superior Labor Court (*Tribunal Superior do Trabalho* or TST) has the Bem-Te-Vi system, an artificial intelligence that assists offices in screening cases. Similar intelligent computational tools are under development/production in other courts, such as TRF1 (ALEI – Intelligent Legal Analysis) and TJDFT (AMON, TOTH, SAREF, ARTIU and Horus).

AI tools are already part of the reality of the activities carried out by the

[6]BOLZAN DE MORAIS, J. L. *As crises do estado e da constituição e a transformação espaço-temporal dos direitos humanos*, 2nd edn., Livraria do Advogado, Porto Alegre, 2011, p. 18.
[7]ABBOUD, G. & CAMPOS, R. "A autorregulação regulada como modelo do Direito proceduralizado: regulação de redes sociais e proceduralização", *in:* ABBOUD, G., NERY Jr., N. & CAMPOS, R. (eds.). *Fake news e regulação*, Thomson Reuters, São Paulo, 2018, p. 19.
[8]SALOMÃO, L. F. (ed.). *Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário*, 3ª ed., CIAPJ/FGV, Rio de Janeiro, 2022, available in: https://bdjur.stj.jus.br/jspui/handle/2011/156490, access: Aug. 14, 2023.

Brazilian Judiciary. They assist judges, most of the time, in simple tasks such as screening cases, classifying events according to certain criteria and identifying common themes in petitions presented by the disputing parties. This is the current state of the art as described in the specialized literature[9]. There is, however, enormous potential to explore in the development of computational solutions to aid decision-making. So, we need to question not just where we are, but where we can (and want) to go. There is no simple answer.

3. ARTIFICIAL INTELLIGENCE AND SENTENCING IN AN "AGE OF UNCERTAINTY"

AI is the greatest product of the current technological revolution, with practical applications present in almost all of our daily activities. There is Artificial Intelligence in virtual assistants such as Siri and Alexa, available on televisions and other electronic devices, including cars; in route software such as Google Maps and Waze, currently used by any app driver (e.g. Uber); in increasingly powerful internet search engines, such as Google; on streaming platforms such as Netflix, Deezer and Spotify, where the product catalog is presented to us according to our tastes and preferences; in more complex facial recognition services, existing in airports, but also simple ones, such as Apple Face ID, used to unlock the screen on iPhone cell phones; in online translation services, such as Google Translate, whose accuracy rate exceeds 90% in translations between English and Spanish[10]; in chatbots, such as ChatGPT, capable of processing natural language and producing texts that appear to be written by real people; and, of course, on social networks such as Facebook, Instagram and Twitter, where the offering of paid products and services caters exactly to consumers' digital profiles (microtargeting).

As it is widespread in practically all fields of digital life, the great extent of the hype, including academic, surrounding Artificial Intelligence is understandable. A quick search on Google Scholar, using the English terms 'Artificial Intelligence', returns 5,740,000 results, including books and scientific articles from different areas. At Amazon's online store for Brazil, if someone types the words 'Artificial Intelligence' and 'Law' in the search bar, they will find almost 1,000 books published in Portuguese. The relevance and timeliness of the topic, therefore, are indisputable.

Despite the relative newness of AI, whose origins date back to the 1950s[11], the relationship between man and technique/technology, that is, human-machine interaction, has been an object of study for a much longer time. In a world undergoing a growing process of technology, the perspectives to adopt have already been described as antagonistic, pessimistic and optimistic[12]. The confrontation of such perspectives results in some fundamental questions for discussion, such as the problem of technological 'determinism': who makes decisions, the man (programmer) or the machine (algorithm)?

From a pessimistic standpoint, Oswald Spengler wrote, in 1931, therefore still in the pre-digital era, a book named 'Man and Technique', showing a negative vision about the future of humanity. Men would become slaves to their own creation (the machine) and, ultimately, technology would destroy civilization itself, at the stage in which we know it. In an almost apocalyptic outcome, '(...) only dreamers

---

[9]TAUK, C. & SALOMÃO, L. F. "Inteligência artificial no Judiciário brasileiro: estudo empírico sobre algoritmos e discriminação", *Diké (UESC)*, vol. 22, nº 23 (2023), pp. 2-32.

[10]TAIRA, B. R.; KREGER, V.; ORUE, A. & DIAMOND, L. C. "A pragmatic assessment of Google Translate for emergency department instructions", *Journal of General Internal Medicine*, vol. 36, nº 11 (2021), pp. 3361-3365, available in: https://doi.org/10.1007/s11606-021-06666-z.

[11]BODEN, M. A. *Inteligência artificial: uma brevíssima introdução,* Editora Unesp, São Paulo, 2020, pp. 21-22.

[12]PIMENTEL, A. F.; BOLZAN DE MORAIS, J. L. & SALDANHA, P. M. "Estado de Direito e tecnopoder", *Justiça do Direito*, vol. 35, nº 3 (2021), pp. 6-43, available in: https://doi.org/10.5335/rjd.v35i3.13241.

believe there is a way out. Optimism is cowardice'[13].

Already in the digital era, Jacques Ellul also expressed, in 1950, an apocalyptic expectation about the risks of man's domination by technology. We would lose our decision-making power in front of the machine. For Ellul, the most frightening feature of modern technology is its autonomy. Man participates less and less in the creation of the technique, being reduced to the level of mere catalyst, resembling 'a bullet inserted into a slot machine: he starts the operation without participating in it'[14].

Herbert Marcuse was another thinker who observed man's submission to the machine. For him, we should be concerned about the unbridled development of technology, the dangers of 'technological rationality'[15] and the devaluation of reason in favor of technique.

More recently, a more optimistic view of man's relationship with new technologies began to be disseminated by Pierre Lévy. The enormous expansion of the internet and the ubiquity of information in the algorithmic world facilitate the expansion of collective human intelligence. Furthermore, machines will never replace men. The leading role will never fall to the computer, because the person who decides is – and always will be – men. 'Techniques do not determine anything'[16]. The image of an 'autonomous, separate, fatal, all-powerful, evil-causing technoscience or privileged instrument of progress' is false[17]. Artificial intelligence must be at the service of human intelligence.

Pessimists and optimists, each in their own way, may have legitimate justifications for their expectations regarding the use of new technologies by man, especially artificial intelligence. There is no right or wrong in this regard. The biggest mistake is to think of the man-machine relationship as a fierce fight between humans and robots and therefore, with extreme pessimism, envision a dystopian future, in which men will have been defeated and subjugated by machines. Another mistake is to see the algorithmic world of artificial intelligence through a naively optimistic lens, believing in the utopia of technological neutrality and ignoring the strength of 'technopower'[18].

Outside of these two extremes, the important thing is to seek greater clarity about the possibilities brought by AI and the limits that should be imposed on it. Important issues that cause some misunderstanding, such as, for example. e.g., knowing whether Artificial Intelligence is truly intelligent, and whether machines decide or just execute, must be faced and minimally answered.

Regarding the 'intelligence' of Artificial Intelligence, amid all the hype surrounding the topic, the underlying concern is that human intelligence will, sooner or later, be surpassed by computers. There is, however, a huge dose of wonderment, even romanticization, surrounding such a prediction. When, in 1997, the Deep Blue supercomputer, created by IBM, finally defeated world chess champion Garry Kasparov, in a revenge for the defeat suffered the previous year, the machine's formidable capacity was revered, as if, from that moment on, Artificial Intelligence was finally supplanting human intelligence. What is truly remarkable about this episode is not the machine's capabilities, but Kasparov's brilliant intellect and fantastic brain. After all, it is estimated that Deep Blue could analyze 100 million

---

[13] SPENGLER, H. *Man and technics: a contribution to a philosophy of life*, Allen & Unwin, London, 1932, p. 104, available in: https://archive.org/details/oswaldspenglermanandtechnicsgeorgeallenunwin1932/page/n5/mode/2up, access: Aug. 8, 2023.

[14] ELLUL, J. *The technological society*, Vintage Books, New York, 1964, p. 135.

[15] MARCUSE, H. *O homem unidimensional: estudos da ideologia da sociedade industrial avançada*, Edipro, São Paulo, 2015, p. 153.

[16] LÉVY, P. *As tecnologias da inteligência: o futuro do pensamento na era da informática,* 2ª ed., Editora 34, São Paulo, 2010, p. 188.

[17] LÉVY, P. *As tecnologias da inteligência* (...), Ob. cit., p. 198.

[18] PIMENTEL, A. F.; BOLZAN DE MORAIS, J. L. & SALDANHA, P. M. "Estado de Direito (...)", Ob. cit., p. 8.

moves per second in the 1996 match, and even so, the Soviet chess player managed to win. IBM needed to improve its computer, increasing its capacity to 200 to 250 million moves per second in 1997, to finally defeat Kasparov in a rematch[19].

The question of whether AI is truly intelligent is quite complex and controversial[20] [21] [22] and any answer to it would go beyond the limits of the present study. In any case, arguments aside, it cannot be denied that human beings are especially endowed with capabilities such as intuition, creativity, sensitivity, feelings and emotions, and that such attributes are not extendable to machines. Human intelligence encompasses a vast array of cognitive and emotional capabilities, which computers can, at best, try to identify and simulate. Therefore, regarding the concern about the possibility of machines eventually surpassing the intelligence of men, we prefer to be in the company of scholars who misunderstand '(...) the idea that a system, whatever it may be – say, a human brain – can create something more complex than itself'[23].

Furthermore, it also seems clear to us that machines create nothing and equally decide nothing. They are programmed to perform tasks faster and more efficiently than humans. However, decision-making power remains in human hands, as Pierre Lévy has always argued. There is no dispute between man and machine regarding decision-making protagonism. The dystopian future imagined by some pessimists has already been interpreted accurately: 'The fear of domination of humanity by robots perhaps resides in the sapiens subconscious'[24]. There is no reason to fear technological determinism.

Despite these clarifications, the truth is that we know little about the potential of algorithms, especially machine learning, in ordinary life and within the criminal Justice System. In this sense, we live in an 'age of uncertainty'[25], characterized by such an astonishing development of technology that we are not even able to predict/understand the possibilities and limits of using Artificial Intelligence. In the words of Kai-Fu Lee, '(...) when it comes to understanding our future with AI, we are all children in kindergarten. We are full of unanswered questions, trying to peer into the future with a mixture of childish wonder and adult concerns'[26]. It is possible that not even the researchers and developers of Artificial Intelligence systems themselves would be able to predict what the most advanced computer programs will be capable of doing in a few years.

In such a scenario of uncertainty, it is reasonable to think that computers will be capable of drafting criminal sentences, including applying penalties, deciding the best punishment for each defendant. As long as they are correctly programmed and receive, in each specific case, the input of process data, the machines will have the technological aptitude to evaluate the set of evidence, identify whether the case involves a decision of acquittal or conviction and, in this case, choose the most fit sentence, both in qualitative and quantitative terms. This can basically occur in two different ways: through an expert system programmed 'manually', which, based on the rules established in the Penal Code, teaches the machine how to sentence offenders; or by machine learning, a method that does not directly teach the

---

[19]ASSUNÇÃO, J. C. "Kasparov faz 2: defesa da humanidade", available in: https://www1.folha.uol.com.br/fsp/esporte/fk020522.htm, access: Feb. 29, 2024.
[20]BODEN, M. A. *Inteligência artificial* (...), Ob. cit., p. 165.
[21]NICOLELIS, M. *O verdadeiro criador de tudo: como o cérebro humano esculpiu o universo como nós o conhecemos*, Planeta, São Paulo, 2020, pp. 137-165.
[22]SANTAELLA, L. *A inteligência artificial é inteligente?* Edições 70, São Paulo, 2022, pp. 143-166.
[23]NICOLELIS, M. *O verdadeiro criador de tudo* (...), Ob. cit., p. 139.
[24]PIMENTEL, A. F.; BOLZAN DE MORAIS, J. L. & SALDANHA, P. M. "Estado de Direito (...)", Ob. cit., p. 18.
[25]KENNETH, J. G. *The age of uncertainty,* Houghton Mifflin Harcourt, Boston, 1979.
[26]LEE, K. *Inteligência artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos*, Globo Livros, Rio de Janeiro, 2019, pp. 10-11.

computer how to sentence, but which teaches it to learn how to do it[27]. In this case, the system would be built to receive certain input data and, applying statistical models, make predictions of results (outputs), e.g., the most appropriate penalty for a given case.

Obviously, the possibilities and intensity of the use of technological tools in sentencing would depend, in theory, on the way each legal system regulates the sentencing stage of criminal proceedings. There is certainly wide variety of sentencing systems throughout the world. For purely didactic purposes, we could establish the following distinction: in countries with a Romano-Germanic tradition, such as Brazil, sentencing regimes tend to be centered on the idea of legality, which means that statutes determine a methodology to be followed by judges in sentencing, such as the *'dosimetria penal'* defined by the Brazilian Penal Code; on the other hand, in common law jurisdictions, such as the United States of America and England, for example, in the absence or insufficiency of positive legal rules on sentencing, the task of guiding judges has been performed, in the last three decades, by sentencing guidelines, developed and published by commissions charged with this specific mandate, imbued with a (not always) disguised objective of mathematizing the operation of sentencing[28] [29].

Before the beginning of the sentencing guidelines movement in the Anglo-Saxon world, sentencing was highly discretionary. A famous North American federal judge, Marvin E. Frankel, went so far as to state that there was an absolute absence of law ('lawlessness') in the US sentencing system[30]. It is interesting to speculate that, in essentially discretionary systems, such as the historical case of the USA, algorithms built to operate through machine learning would be capable of, from a database of criminal sentences, identifying a type of 'average sentence', starting from based on the premise that the penalties actually applied by judges would reflect the presumably 'correct' sentence. In other words, machine learning algorithms used to prescribe penalties would do without legal rules establishing any methodology for applying penalties. They would learn from real sentencing practices (training data), that is, they would learn to sentence from the experience of judges. But of course, at the time of Frankel's critique of the American legal system, Artificial Intelligence technology was not available.

This historical digression serves to better understand the types of algorithms that may be used in sentencing. In fact, it is no coincidence that the development of sentencing guidelines in common law countries, where, until then, there were no legal rules guiding the sentencing judges, was already considered a form of algorithm, obviously not in a computational sense, but in terms to reflect the real sentencing practices in force at that historical moment[31]. Therefore, the use of Artificial Intelligence to define sentences, especially if we think about machine learning models, currently seems to be a reality closer to common law than to countries with a tradition of legality, such as Brazil.

In civil law, on the other hand, the existence of positive rules, albeit imperfect, would allow the development of traditional programming, which, using logical operators ('if...then'), would be capable of recommending appropriate sentences, in a typical expert system model.

---

[27]KAUFMAN, D. "Inteligência artificial: repensando a mediação", *Brazilian Journal of Development*, vol. 6, nº 9 (2020), pp. 67621-67639, available in: https://doi.org/10.34117/bjdv6n9-264

[28]QUIRÓS, D. Z. *Castigo e determinação da pena nos E.U.A.: um estudo sobre as United States Sentencing Guidelines*, Empório do Direito, Florianópolis, 2017, pp. 100-116.

[29]CAMPOS, G. S. Q. *Aplicação da pena e o problema da discricionariedade judicial: breve estudo comparativo entre a dosimetria penal brasileira e o modelo de sentencing guidelines norte-americano e inglês*, JusPodivm, Salvador, 2021, pp. 45-76.

[30]FRANKEL, M. "Lawlessness in sentencing", *University of Cincinnati Law Review*, vol. 1, nº 41, 1972.

[31]DONOHUE, M. "A replacement for Justitia's scales?: machine learning's role in sentencing", *Harvard Journal of Law & Technology*, vol. 32, nº 2, 2019, pp. 657-678.

As with other applications, the potential for using computer algorithms in sentencing is virtually limitless. Technological evolution offers criminal justice the use of machines as tools to optimize the work of lawyers, police officers, members of the *Ministério Público* and judges, in the same way that, one day, Hewlett-Packard scientific calculators facilitated the activities of a large number of professionals, such as engineers, mathematicians and accountants. The question, finally, is not one of technological capacity, which is enormous; The crucial point for ethical and philosophical reflection is to be clear about what we actually want Artificial Intelligence to do on our behalf.

4. ARE JUDGES GOING TO BE REPLACED BY MACHINES?

It seems to be taken for granted that, in some way, Artificial Intelligence will be incorporated into Criminal Justice Systems, including at the sentencing stage. Although we are still far from such a reality in Brazil, it is not unreasonable to imagine a computerized criminal sentence, algorithmically decided, with reduced (or no) participation of a human judge. In specialized literature, two different approaches to this issue have been proposed: for the first, machines should completely replace judges in sentencing; a second approach argues that Artificial Intelligence should only assist judicial decision-making, establishing a kind of division of tasks between computer and judge[32]. There is also a third approach, radically contrary to the so-called 'robot judge', but such a position is far from typical[33].

The main objections made against the use of AI in sentencing involve ethical issues that can be roughly grouped around three terms: fairness, accountability, and transparency[34]. Problems related to the Anglo-Saxon concept of fairness arise when algorithms used in criminal proceedings have a discriminatory effect on certain individuals or groups of individuals. In fact, it has already been proven that algorithms can have implicit biases, undermining the desired equality of treatment within the Criminal Justice System. The best-known example of this is COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), developed in 1998 by the North American company Equivant (formerly Northpointe) and widely used by US state jurisdictions. COMPAS is a computational tool which analyses recidivism risk, therefore used to predict, based on machine learning and statistical models, future criminal behavior, calculating the risk score of individuals[35] [36]. The major criticism directed at the use of population-level actuarial data as input in sentencing is that the algorithm would exacerbate existing racial and socioeconomic disparities in society in general and particularly within the prison system[37].

---

[32]SCHWARTZE, M. & ROBERTS, J. V. "Reconciling artificial and human intelligence: supplementing not supplanting the sentencing judge", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence*, Oxford University Press, New York, 2022, pp. 206-229, available in: https://doi.org/10.1093/oso/9780197539538.003.0011.

[33]GRECO, L. *Poder de julgar sem responsabilidade de julgador: a impossibilidade jurídica do juiz-robô*, Marcial Pons, São Paulo, 2020, p. 17.

[34]CHIAO, V. "Fairness, accountability and transparency notes on algorithmic decision-making in criminal justice", *International Journal of Law in Context*, vol. 15 (2019), pp. 126-139, available in: https://doi.org/10.1017/S1744552319000077.

[35]In the current stage of American criminal justice, COMPAS integrates a new model that has been called, by criminologists, Actuarial Criminal Justice, *in:* FEELEY, M. M. & SIMON, J. "The new penology", *in:* McLAUGHLIN, E; MUNCIE, J. & HUGHER, G. (eds.). *Criminological Perspectives: Essential Readings*, 2ª ed., SAGE, London, 2005, pp. 434-446.

[36]In Brazil, although there is no such movement towards the administration of criminal justice under an actuarial logic, it has already been said that it represents 'the criminology of the end of history', *in:* DIETER, M. S. *Política criminal atuarial: a criminologia do fim da história*, Revan, Rio de Janeiro, 2013, p. 195.

[37]DONOHUE, M. E. "A replacement for Justitia's scales? (…)", Ob. cit., pp. 662-663.

Building the programming of an 'ethical algorithm'[38] is a task that will require great effort to identify models that minimize injustices. Software developed by biased (human) programmers will inevitably be biased. Algorithms trained on biased data will inevitably produce discriminatory results. Greater human diversity (ethnic, geographic, socioeconomic, among others) will be necessary in the development of programming codes and machine training data sets. Furthermore, computer scientists, legal professionals, philosophers, and ethicists will need to participate in the construction of algorithms in the near future, if not immediately.

Aside from the issue of fairness, problems relating to accountability for the use of AI would tend to become more acute in private decision-making processes, but not within the Criminal Justice System, where, at least in principle, there will always be human judges responsible for making decisions. However, there are those who deny the (legal) possibility of the robot judge precisely because it means 'power to judge without the responsibility of a judge'[39]. This is an important issue on an abstract level; concretely, however, it seems unlikely that one day we will have machines deciding in an entirely autonomous way, without being subjected to some human agency, the latter being responsible, after all. In any case, the possibility of establishing a specialized regulatory body, with technical expertise to ensure the reliability of algorithmic instruments, has already been raised[40].

More serious are the problems arising from the lack of transparency of algorithms. The issue of transparency involves an important concern about the risks that the introduction or expansion of the use of algorithms in sentencing may bring to the scrutability of the computerized decision-making process[41]. In this sense, 'algorithmic opacity'[42] is a serious potential threat.

Two observations are crucial so that the ethical issue of algorithmic transparency can be correctly understood. First, it is necessary to keep in mind that the level of transparency of the algorithms essentially depends on the type of artificial intelligence used. 'Manually' programmed algorithms (expert systems), which follow a simple flowchart logic ('if…then'), are quite simple to understand and therefore transparent. Algorithms based on machine learning, especially deep learning, tend to be opaque, especially because it is not always easy to understand the internal mechanisms that lead to certain results (decisions)[43].

Second, algorithmic transparency is generally a function of two factors: access and complexity[44]. Access/availability of the algorithm's source code is often highlighted as an essential element for transparency. At the current stage, some computer programs used by the Judiciary of different countries enjoy the privilege of secrecy assured by intellectual property, with no legal obligation on the part of their developers/owners to make their source codes available[45] [46]. In Brazil, the legal protection of software arises from Laws Nos. 9,609/98 and 9,610/98. The program license can choose to keep the programming code 'closed' (secret), which is more

[38]KEARNS, M. & ROTH, A. *The ethical algorithm: the science of socially aware algorithm design*, Oxford University Press, New York, 2020, p. 4.
[39]GRECO, L. *Poder de julgar* (...), Ob. cit., p. 47.
[40]CHIAO, V. "Fairness, accountability and transparency (…)", Ob. cit., pp. 137-138.
[41]RYBERG, J. "Sentencing and algorithmic transparency", *in:* RYBERG, J. & ROBERTS, J. V. (eds.), *Sentencing and artificial intelligence* (…), Ob. cit., p. 13.
[42]RYBERG, J. "Sentencing and algorithmic transparency", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence* (…), Ob. cit., p. 13.
[43]VAN WINGERDEN, S. & PLESNIČAR, M. M. "Artificial intelligence and sentencing: humans against the machines", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence* (…), Ob. cit., p. 242.
[44]THOMSEN, F. K. "Iudicium ex Machinae: the ethical challenges of ADM at sentencing", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence* (…), Ob. cit., p. 261.
[45]RYBERG, J. "Sentencing and algorithmic transparency", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence* (…), Ob. cit., pp. 14-15.
[46]McKAY, C. "Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making", *Current Issues in Criminal Justice*, vol. 32, nº 5 (2019), pp. 1-18, available in: https://doi.org/10.1080/10345329.2019.1658694.

usual, or 'open' (open source). Private and closed algorithms can be true black boxes[47]. In the USA, the lack of transparency of COMPAS, caused by the legal protection of the right to property, was the constitutional argument raised in the State v. Loomis before the Wisconsin Supreme Court. The court, however, ruled that the use of risk prediction algorithms by judges does not violate defendants' fundamental rights.

Although complex, the problem of the absence of publicity could receive a theoretically simple solution: the removal of commercial secrecy from algorithms used in the sphere of the Judiciary and other public decision-making bodies, thus opening their entire operation to public control[48]. Another possibility would be for bodies within the Judiciary to hire only open source software for use in the judges' core (decision-making) activities, as has already happened with the performance of core activities, facilitated by programs such as the Linux operating system and data processors. text LibreOffice/OpenOffice.

Furthermore, it has already been observed that the greater the complexity of the algorithm programming, e.g. in the case of machine learning, the chances of the general public, and even some experts, being able to understand its internal workings will be lower[49]. That is to say, the process that leads an algorithm to extract its predictions from the data provided to it (input data), or, put another way, the correlations between inputs and outputs made by the program, can be extremely difficult to interpret[50]. The complexity factor, also known as intelligibility, emerges as a serious challenge even if we adopt the knowledge of experts as a parameter. As an AI expert once stated: 'We have built automated systems that we cannot fully understand, that are much faster and more powerful than the human mind, and that are trained on data sets too large for us to understand'[51].

Justice (fairness), responsibility (accountability), and transparency, therefore, are key concepts that summarize some of the main objections directed to the use of artificial intelligence in criminal justice, especially in sentencing[52]. But after all, will human judges be replaced by machines? Are there limits to the use of algorithms in criminal sentencing, and what are they?

There is a clear tendency to answer such questions through a comparison between human decision-making and algorithms, in terms of their advantages and disadvantages. Such an approach, which we could call 'comparative analysis', basically asks whether a computerized decision can turn out to be better than the

[47]McKAY, C. "Predicting risk in criminal procedure (…)", Ob. cit., p. 11.
[48]CHIAO, V. "Fairness, accountability and transparency (…)", Ob. cit., pp. 137-138.
[49]THOMSEN, F. K. "Iudicium ex Machinae: the ethical challenges of ADM at sentencing", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence* (…), Ob. cit., pp. 261-262.
[50]CHIAO, V. "Transparency at sentencing: are human judges more transparent than algorithms?", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). Sentencing and artificial intelligence, Oxford University Press, New York, 2022, pp. 34-56, https://doi.org/10.1093/oso/9780197539538.003.0003.
[51] WALMSLEY, J. "Artificial Intelligence and the value of transparency", *AI & Society*, vol. 36, 2020, pp. 585-595, available in: https://doi.org/10.1007/s00146-020-01066-z.
[52]The concern is such that, in Brazil, the National Council of Justice (*Conselho Nacional de Justiça* or CNJ) issued Resolution nº 332/2020, establishing ethical parameters for the use of Artificial Intelligence in the Brazilian Judiciary. The standard provides that the use of AI in decision-making processes must meet ethical criteria of transparency, predictability, auditability, impartiality, and substantial justice. Furthermore, judicial decisions supported by Artificial Intelligence must preserve equality, non-discrimination, plurality, solidarity, and fairness.

status quo in criminal justice, that is, human decisions[53] [54] [55].

The biggest comparative advantage of an algorithmic criminal sentencing model would probably be the gain in consistency/predictability in the application of sentences. Regardless of whether they are programmed manually, using logical operators such as 'if...then' based on legal rules on the application of sentences – what, in Brazil, we call *'dosimetria penal'* –, or constructed by machine learning, Based on the set of past criminal sentences, algorithms have greater potential to define sentences consistently and without so many disparities[56]. Therefore, ideally speaking, they would be able to treat similar cases in a similar way, fulfilling an important requirement of justice. There are a huge number of studies proving that (human) judges fail to give identical cases the same solution. The application of punishment is traditionally a discretionary and not very consistent/predictable activity. This is a problem with the judicial sentencing system that the use of computers would solve or, at the very least, help to mitigate.

Other distinctions are more complex.

In terms of fairness, concern about the phenomenon of algorithm bias is legitimate, as happened in the famous North American COMPAS case. Algorithmic bias can become a serious ethical problem, as it results in discriminatory effects against individuals from vulnerable groups, normally defined as 'preferred targets' of the criminal system's selectivity. Discriminatory bias can essentially arise through two distinct mechanisms, depending on the type of algorithm. First, programming may simply reflect biases found in training data in the case of machine learning. And second, bias may result from the inclusion in your code of factors that vary systematically between different population groups, e.g. e.g., socioeconomic status, educational level, employment history or criminal record in risk assessment algorithms, and thus, indirectly, result in discrimination. In this last hypothesis, the algorithms in question can be programmed based on logical rules, not using machine learning.

In any case, the algorithm bias is due to human bias. There is no 'digital-atavic bias'[57] that is part of the genetics of algorithms. Judicial decisions made by (human) judges are equally discriminatory, they just often do not appear to be so. In other words: bias is not an algorithmic creation, but the result of structural and systemic inequalities, especially in countries like Brazil, and legal rules for applying penalties that are insensitive to the peculiarities of individuals from marginalized groups. It is no surprise that one of the suggestions found in the specialized literature for the problem of algorithmic bias is the adoption of affirmative action, with the insertion of positive discrimination rules in favor of vulnerable minorities, for example, by considering race and ethnicity as mitigating factors[58].

On the other hand, questions related to algorithmic transparency can also be posed, with some nuances, to human decisions. Thus, the question of whether algorithms can be more transparent in their decisions than (human) judges requires a better understanding: a) of the judicial (human) decision-making process, and b) of the legal issue of the motivation of judicial decisions, which, in Brazil, it has constitutional status (article 93, IX, of the Federal Constitution).

On the one hand, it is known that every human decision-making process is

[53]RYBERG, J. "Sentencing and algorithmic transparency", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence* (…), Ob. cit., pp. 28-30.
[54]CHIAO, V. "Fairness, accountability and transparency (…)", Ob. cit., p. 127.
[55]BAGARIC, M. & WOLF, G. "Sentencing by computer: enhancing sentencing transparency and predictability, and (possibly) bridging the gap between sentencing knowledge and practice", *George Mason Law Review*, vol. 25, nº 3 (2018), pp. 653-709.
[56]BAGARIC, M. & WOLF, G. "Sentencing by computer (…)", Ob. cit., p. 656.
[57]PIMENTEL, A. F.; BOLZAN DE MORAIS, J. L. & SALDANHA, P. M. "Estado de Direito (...)", Ob. cit., p. 13.
[58]THOMSEN, F. K. "Iudicium ex Machinae: the ethical challenges of ADM at sentencing", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence* (…), Ob. cit., pp. 263-264.

complex, and, in the case of judicial sentencing, there are certainly decision-making factors that judges often use unconsciously. In criminological scholarship, researchers from Israel tested the caricatured image of Justice as being what the judge ate for breakfast, associated with 'legal realism'. They found that judicial decisions on provisional release/parole in Israel are decisively influenced by the time elapsed since the last meal eaten by the judges, so that the hungrier they felt, the greater the number of decisions denying release from prison[59]. Methodological criticisms aside[60], there seem to be clear signs that, like some algorithms, a judge's mind can also be a black box[61]. Decisions made by human judges can be as or more unintelligible than computerized decisions[62].

With good reason, the judicial activity of imposing sentences has already been critically described as 'a very open exercise of discretion by a human judge, often under time pressure, with limited capacity to learn from his own mistakes and little information about the general patterns of sentences throughout the system beyond what can be obtained from their own experience'.[63]

On the other hand, the requirement to motivate judicial decisions is not always effective in providing transparency at each stage of the judges' work. In the specific case of sentencing, for instance, practice reveals that the sentence usually does not individually examine each of the relevant circumstances of article 59 of the Brazilian Penal Code. That is to say, the sentencing judge often makes a global and comprehensive analysis of the case, setting the base penalty at the legal minimum or at some higher level, violating the requirement for detailed reasoning. This finding inspired some experts to argue that publicizing the source code of an algorithm would better meet the idea of transparency than the requirement of motivation for judicial decisions typical of some legal systems[64], such as the Brazilian one.

In comparative terms, therefore, bias and opacity are not problems exclusive to algorithms, affecting human decisions in the same way.

An analysis of the type detailed above can be useful to have greater clarity about the advantages and disadvantages of a computerized sentence, compared to decisions made by (human) judges. However, we believe that the introduction of algorithms in sentencing, as it represents such a significant change in forensic practice and involves such ethical challenges, should not be analyzed in comparison to the status quo, but rather in ideal terms, especially in the approach that advocates substitution tout court of judges by machines: is an algorithmic decision-making system in criminal justice ideal in aspects such as rationality, consistency, transparency and justice? We will return to such questions in the next topic, when we will investigate the very human nature of the act of judging.

First, it is worth discussing one last question about the ethical parameters of fairness, responsibility, and algorithmic transparency. They will often put themselves on a collision course with one of the objectives of using Artificial Intelligence in sentencing: the desire for greater accuracy[65]. In the case of risk prediction

---

[59]DANZIGER, S., LEVAV, J. & AVNAIM-PESSO, L. "Extraneous factors in judicial decisions", *PNAS,* vol. 108, nº 17 (2011), pp. 6889-6892.

[60]GLÖCKNER, A. "The irrational hungry judge effect revisited simulations reveal that the magnitude of the effect is overestimated", *Judgment and Decision Making*, vol. 11, nº 6 (2016), pp. 601-610, available in: https://doi.org/10.1017/S1930297500004812.

[61]VAN WINGERDEN, S. & PLESNIČAR, M. M. "Artificial intelligence and sentencing: humans against the machines", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence* (…), Ob. cit., p. 241.

[62]CHIAO, V. "Fairness, accountability and transparency (…)", Ob. cit., p. 136.

[63]CHIAO, V. "Transparency at sentencing (…)", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). Sentencing and artificial intelligence (…), Ob. cit., p. 35.

[64]CHIAO, V. "Transparency at sentencing (…)", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). Sentencing and artificial intelligence (…), Ob. cit., pp. 49-51.

[65]RYBERG, J. & PETERSEN, T. S. "Sentencing and the conflict between algorithmic accuracy and transparency", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial*

algorithms, such as COMPAS, it is self-evident that the objective is to obtain greater accuracy. On the other hand, in the case of software that makes sentence recommendations, it may be more difficult to establish that the algorithm 'hit' the most appropriate sentence. In discretionary sentencing systems like the Brazilian one, there are not even 'correct' sentences. However, the accuracy of the results (outcomes) of algorithms will always depend on the calibration of fairness and transparency, especially the latter: theoretically speaking, the more transparent the programming, e.g., when its complexity is reduced for better understanding by users, the less accurate its outputs will be. There is, therefore, a clear trade-off between precision and ethical considerations in the use of algorithms. Could we be facing a situation that could receive some kind of intermediate solution? Would there be a sweet spot between the values of accuracy and algorithmic ethics? Our position is that ethical requirements are urgent and non-negotiable, and, therefore, the loss of accuracy of algorithms is an inevitable consequence, compromising the very purpose of using artificial intelligence in criminal justice. This is a difficult to overcome paradox.

5. IS SENTENCING ESSENTIALLY HUMAN?

Sentencing cannot be conceptualized, in a reductionist way, as the formal adjudication of a criminal case. On the contrary, sentencing is a social process, resulting from social expectations, cultural factors and, obviously, the (human) interaction between procedural actors[66]. It is insufficient to examine sentencing from an exclusively technical perspective, as it is much more complex, expressing social meanings and feelings (e.g. what is a crime? what is punishment?) and communicating important messages to the defendant, the victim and to the community in general. There is also a democratic function of sentencing, as it allows the general public to discuss fundamental questions about the legitimacy of the State's use of violence and punishment.

Sentencing is also a human process, for at least two reasons. First, because it is the result of successive interactions between the various parties and procedural subjects (prosecution, defense, defendant, judge, witnesses, and so on). Even if one wants to conceive sentencing as a unilateral act that materializes the power of the State-judge, it is necessary to recognize the influence of the actions of these multiple actors in the formation of judicial conviction. Second, it is undeniable that sentencing has a communicative/expressive function[67] [68], transmitting, at the same time, a judgment of moral censure for the criminal behavior to the offender, a certain comfort to the victim due to the sentence and, also, a message of reaffirmation of the legal order and discouraging the practice of new offenses by the entire community. These are clear signs that the task of sentencing is essentially human. Artificial Intelligence would not be able to play such a role.

Evidently, recognizing that sentencing is a human task implies accepting an inherent characteristic of every decision-making process, including within the Criminal Justice System: discretion. Around the world, however, efforts have been made to reduce or limit judicial discretion in sentencing, with the aim of promoting greater consistency (predictability) of sentences, and avoiding or, at least, mitigating some of the undesirable effects of discretion, especially a) illegitimate disparities, which violate equality embodied in the principle that equal cases should receive

intelligence, Oxford University Press, New York, 2022, pp. 57-73, available in: https://doi.org/10.1093/oso/9780197539538.003.0004.

[66]TATA, C. "Resolute ambivalence: why Judiciaries do not institutionalize their decision support systems", *International Review of Law Computers & Technology*, vol. 14, nº 3 (2000), pp. 297-316, available in: https://doi.org/10.1080/713673373.

[67]TEIXEIRA, A. *Teoria da aplicação da pena: fundamentos de uma determinação judicial da pena proporcional ao fato*, Marcial Pons, São Paulo, 2015.

[68]GUIMARÃES NETO, S. L. *Uma teoria da pena baseada na vítima: a busca pela satisfação do indivíduo vitimado como finalidade da pena*, Marcial Pons, São Paulo, 2020.

similar sentences, and b) discrimination, which undermines the very notion of justice[69].

As seen previously, the path taken by common law countries has been marked by the development of sentencing guidelines, which seek to guide judges in their sentences, through an undisguised 'mathematization' of sentencing. Especially in the USA, e.g., the guidelines consist of a numerical table with two axes, representing the level of severity of the crime committed and the criminal history of the defendant. The 'correct penalty' would always be found at the intersection of the two axes, in a very simple operation[70] [71]. All other potential factors relevant to sentencing are left out in the North American guideline regime, thus 'impoverishing' the work of judges[72]. In England and Wales, on the other hand, the creation of the guidelines clearly sought to avoid such mathematization, having adopted a less numerical and more textual model to guide judges.

In civil law countries, which have never had sentencing guidelines, it is the law that defines rules guiding the judicial activity of determining sanctions. Undeniably, sentencing legal systems tend to suffer from problems related to the openness and vagueness of the language used in the statues, in addition to some vagueness or conceptual ambiguity, as is the case with the wording of article 59 of the Penal Code[73]. Discretionary ends up finding few limitations. It is then up to jurisprudence to give safer contours to the sentencing legal methodology, including through mathematical criteria. Therefore, in countries with a Romano-Germanic tradition, such as Brazil, there is a constant attempt to rationalize sentencing, making it more consistent.

In any case, sentencing cannot be seen as the result of a mechanical and artificial decision-making process. It is not limited to a simple arithmetic operation through which the judge would be able to identify the presence (or absence) of certain factors, normally defined by law, weigh them, and calculate the best possible result in terms of punishment. It is an illusion to imagine that sentences are the perfectly logical product of rationality. Sentencing is an essentially human activity, therefore imperfect.

Sentencing is, likewise, an interpretative task. A sentencing judge must not analyze, in an isolated and mechanical way, each of the circumstances of the specific case that influence the individualization of the sentence. Instead, he must look more broadly at the overall context of the case. For example, even when faced with a defendant who has a criminal record, which, by law, would result in automatic exasperation of the sentence, the judge may assess the relevance, seniority and time elapsed between previous criminal convictions, and thus, in some situations, stop aggravating the sanction. In other words, past criminal life needs to be interpreted by the judge, giving it the most appropriate weight given the circumstances of the case and the agent. Thus, in parallel with the general and abstract prediction, in the law, of the relevant factors for the application of punishment, it is in the operationalization of such circumstances in each specific case that each factor becomes 'unavoidably unstable and inescapably contingent'[74]. Even though a defendant's criminal history could be reduced to a mechanical computational formula, Artificial Intelligence would only be able to classify them in a crude way (e.g., 'good',

---

[69]CAMPOS, G. S. Q. *Aplicação da pena e o problema da discricionariedade judicial: breve estudo comparativo entre a dosimetria penal brasileira e o modelo de sentencing guidelines norte-americano e inglês*, JusPodivm, Salvador, 2021, p. 38.

[70]QUIRÓS, D. Z. *Castigo e determinação da pena* (...), Ob. cit., p. 103.

[71]CAMPOS, G. S. Q. *Aplicação da pena e o problema da discricionariedade judicial* (...), Ob. cit., p. 49.

[72]TONRY, M. *Sentencing matters*, Oxford University Press, New York, 1996, p. 20.

[73]CARVALHO, S. "O controle do substancialismo e do decisionismo na aplicação da pena", *Revista Direito em Debate*, vol. 29, nº 54 (2020), pp. 311-317, available in: https://doi.org/10.21527/2176-6622.2020.54.311-317.

[74]TATA, C. *Sentencing: a social process – rethinking research and policy*, Palgrave Macmillan, London, 2020, p. 6.

'bad' or 'very bad' history). Human qualitative judgments would remain essential[75]. Interpretation may often be necessary for the sentence to do justice to the specific case.

Only human judges enjoy an essential characteristic so that the capacity for interpretation can be exercised: sensitivity. It is a relational attribute, and not an absolute one, that is, sensitivity, in addition to being a fundamentally human characteristic, is always established in intersubjective relationships. It is in human contact, in the case between judge and defendant, that an empathetic relationship can be developed, which is important for the correct application of the sentence. A simple example can help clarify this point. If a defendant, in addition to admitting to committing the crime, also appears genuinely remorseful, the sentencing judge may find a shorter sentence to be more fit not only due to the confession, but because he considers that the accused has already started an individual rehabilitation process, justifying a shorter sentence. Although controversial, the topic of remorse in sentencing is relevant and has occupied the attention of specialists[76] [77]. What we are interested in understanding is that remorse is a legitimate human feeling that, once manifested during the trial by someone accused of committing crimes, can only be correctly assessed by a sensitive and empathetic, therefore human, judge. Empathy and compassion are human values that algorithms cannot capture at all[78].

There is also another interesting fact to be considered. Some empirical studies have already proven that, in general, people prefer to be judged by human judges and not by robots. Such studies observed the reaction of individuals to decisions involving predictions in different fields, e.g., the performance of university students, which requires the elaboration of prognoses of academic success, or the granting/revocation of provisional release, which involves analyzing the risk of recidivism. Evidently, predictions made by algorithms based on statistical models and predictions made by human methods can fail. One of the empirical findings involves the so-called 'algorithmic aversion' as a possible explanation for the preference for human judgment: there is greater intolerance to algorithmic error than human error, even when it is proven that robots can make less mistakes than humans[79].

In short: sentencing, like human reason itself, is not purely logical and mechanical. It can be understood as a social and human process. Sentencing involves social expectations and meanings, cultural factors, intersubjective interactions, interpretation, feelings, and human values. Furthermore, sentencing has a strong moral element. Artificial Intelligence, at its current stage of development, is not capable of encompassing all this complexity. It will probably never be able to do so. Judges should not be replaced by machines. Statistical models, computational solutions, and AI algorithms should be used as auxiliary tools and devoid of decisional autonomy; they can, without a doubt, be useful in sentencing, as long as they are developed based on ethical parameters and subject to human supervision. The last word should always be that of a judge.

6. CONCLUSIONS

David, the robot-boy from the film `Artificial Intelligence` (2001), by the

[75]SCHWARTZE, M. & ROBERTS, J. V. "Reconciling artificial and human intelligence (…)", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence* (…), Ob. cit., p. 212.
[76]PROEVE, M. & TUDOR, S. *Remorse: psychological and jurisprudential perspectives*, Ashgate, New York, 2010.
[77]SMITH, N. *Justice through apologies: remorse, reform, and punishment*, Cambridge University Press, Cambridge, 2014.
[78]VAN WINGERDEN, S. & PLESNIČAR, M. M. "Artificial intelligence and sentencing: humans against the machines", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence* (…), Ob. cit., p. 238.
[79]DIETVORST, B. J., SIMMONS, J. P. & MASSEY, C. "Algorithm aversion: people erroneously avoid algorithms after seeing them err", *Journal of Experimental Psychology: General*, vol. 144, nº 1 (2015), pp. 114-126, available in: https://doi.org/10.1037/xge0000033.

duo Stanley Kubrick and Steven Spielberg, perfectly represents the idea of technology that prevailed for many decades. A kind of Pinocchio 4.0, the robotic character played by Haley Joel Osment looked every bit like a human being, including his ability to feel love. In the real world, until recently we considered, romantically and even naively, that artificial intelligence would be synonymous with humanoid robots. What has changed is that, today, AI no longer intends to physically resemble humans, although it continues to seek to emulate fundamental capabilities of the human species, especially reasoning. The current stage of the technological revolution is no longer marked, therefore, by android robots as in another famous science fiction film, 'Blade Runner' (1982), but by Artificial Intelligence tools that have become ubiquitous. We need to learn to live with them.

The profound impact of technologies such as Artificial Intelligence makes us reflect on the possibility of `algorithmic justice' and its ethical limitations. The possibility of human judges being replaced by machines (computers) is, to say the least, unsettling. Currently, an optimistic view seems to prevail regarding the use of AI in sentencing. It is argued with a supposed comparative advantage of algorithms in relation to human judges in aspects of the decision-making process such as accuracy, consistency, and predictability. In short, computerized criminal sentences would produce more objective (and technical?) judgments, as if the elimination of subjectivity were the main aim of traditional judicial tasks such as sentencing. It is not.

In times of great fascination with the almost unlimited potential of Artificial Intelligence, we dare to defend a humanistic vision on the use of algorithms in sentencing. Deciding a criminal case and applying the most appropriate sentence to a guilty defendant is a highly complex task. It is not just about receiving information about the crime committed and the criminal agent, processing it and calculating the most appropriate sanction. The justice of the specific case requires interpretation. Transforming sentencing into a purely mechanical, even machine-automated task, is to remove its inherently human element.

There is no doubt about the technological potential available to Criminal Justice Systems all around the world. If properly developed, algorithms would certainly be able to learn how to sentence, probably more consistently than (human) judges. Ethical questions related to Artificial Intelligence, although complex, could also be resolved. The question, basically, is what we actually want to do with algorithms: use them or not, and to what extent we want to use them. This will always be a human decision.

7. REFERENCES

ABBOUD, G. & CAMPOS, R. "A autorregulação regulada como modelo do Direito proceduralizado: regulação de redes sociais e proceduralização", *in:* ABBOUD, G., NERY Jr., N. & CAMPOS, R. (eds.). *Fake news e regulação.* Thomson Reuters, São Paulo, 2018.

BAGARIC, M. & WOLF, G. "Sentencing by computer: enhancing sentencing transparency and predictability, and (possibly) bridging the gap between sentencing knowledge and practice", *George Mason Law Review*, vol. 25, nº 3, 2018.

BODEN, M. A. *Inteligência artificial: uma brevíssima introdução,* Editora Unesp, São Paulo, 2020.

BOLZAN DE MORAIS, J. L. *As crises do estado e da constituição e a transformação espaço-temporal dos direitos humanos,* 2nd edn, Livraria do Advogado, Porto Alegre, 2011.

BRENNAN, T., DIETRICH, W. & EHRET, B. "Evaluating the predictive validity of the COMPAS Risk and Needs Assessment System", *Criminal Justice and Behavior*, vol. 36, nº 1 (2009), pp. 21-40, available in: https://doi.org/10.1177/0093854808326545,

BROCHADO, M. *Inteligência artificial no horizonte da filosofia da tecnologia: técnica, ética e direito na era cybernética,* Editora Dialética, São Paulo, 2023, available in: https://doi.org/10.48021/978-65-252-7454-6.

CAMPOS, G. S. Q. *Aplicação da pena e o problema da discricionariedade judicial: breve estudo comparativo entre a dosimetria penal brasileira e o modelo de sentencing guidelines norte-americano e inglês,* JusPodivm, Salvador, 2021.

CARVALHO, S. "O controle do substancialismo e do decisionismo na aplicação da pena", *Revista Direito em Debate,* vol. 29, nº 54 (2020), pp. 311-317, available in: https://doi.org/10.21527/2176-6622.2020.54.311-317.

CHIAO, V. "Fairness, accountability and transparency notes on algorithmic decision-making in criminal justice", *International Journal of Law in Context,* vol. 15 (2019), pp. 126-139, available in: https://doi.org/10.1017/S1744552319000077.

CHIAO, V. "Predicting proportionality: the case for algorithmic sentencing", *Criminal Justice Ethics,* vol. 37, nº 3 (2018), pp. 238-261, available in: https://doi.org/10.1080/0731129X.2018.1552359.

CHIAO, V. "Transparency at sentencing: are human judges more transparent than algorithms?", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence*, Oxford University Press, New York, 2022, pp. 34-56, available in: https://doi.org/10.1093/oso/9780197539538.003.0003.

CONSELHO NACIONAL DE JUSTIÇA. *Resolução nº 332,* de 21 de agosto de 2020. Available in: https://atos.cnj.jus.br/atos/detalhar/3429. Access: August 14, 2023.

DANZIGER, S., LEVAV, J. & AVNAIM-PESSO, L. "Extraneous factors in judicial decisions", *PNAS,* vol. 108, nº 17 (2011), pp. 6889-6892.

DIETER, M. S. *Política criminal atuarial: a criminologia do fim da história,* Revan, Rio de Janeiro, 2013.

DIETVORST, B. J., SIMMONS, J. P. & MASSEY, C. "Algorithm aversion: people erroneously avoid algorithms after seeing them err", *Journal of Experimental Psychology: General,* vol. 144, nº 1, 2015, pp. 114-126, available in: https://doi.org/10.1037/xge0000033.

DONOHUE, M. E. "A replacement for Justitia's scales?: machine learning's role in sentencing", *Harvard Journal of Law & Technology*, vol. 32, nº 2 (2019), pp. 657-678.

DYMITRUK, M. "Artificial intelligence as a tool to improve the administration of justice?", *Acta Univ. Sapientiae, Legal Studies*, vol. 8, nº 2 (2019), pp. 179-189, available in: https://doi.org/10.47745/AUSLEG.2019.8.2.03.

ELLUL, J. *The technological society,* Vintage Books, New York, 1964.

FEELEY, M. M. & SIMON, J. "The new penology", *in:* McLAUGHLIN, E., MUNCIE, J. & HUGHER, G. (eds.). *Criminological Perspectives: Essential Readings*, 2nd edn, SAGE, London, 2005, pp. 434-446.

FRANKEL, M. "Lawlessness in sentencing", *University of Cincinnati Law Review*, vol. 1, nº 41, 1972.

GLÖCKNER, A. "The irrational hungry judge effect revisited simulations reveal that the magnitude of the effect is overestimated", *Judgment and Decision Making,* vol. 11, nº 6 (2016), pp. 601-610, available in: https://doi.org/10.1017/S1930297500004812.

GRECO, L. *Poder de julgar sem responsabilidade de julgador: a impossibilidade jurídica do juiz-robô*, Marcial Pons, São Paulo, 2020.

GUIMARÃES NETO, S. L. *Uma teoria da pena baseada na vítima: a busca pela satisfação do indivíduo vitimado como finalidade da pena,* Marcial Pons, São Paulo, 2020.

HUTTON, N. "Sentencing, rationality, and computer technology", *Journal of Law and Society,* vol. 22, nº 4, 1995, pp. 549-570, available in: https://doi.org/10.2307/1410614.

KAUFMAN, D. "Inteligência artificial: repensando a mediação", *Brazilian Journal of Development,* vol. 6, nº 9, 2020, pp. 67621-67639, available in:

https://doi.org/10.34117/bjdv6n9-264.
KAUFMAN, D. *Desmistificando a inteligência artificial,* Autêntica, Belo Horizonte, 2022.
KEARNS, M. & ROTH, A. *The ethical algorithm: the science of socially aware algorithm design,* Oxford University Press, New York, 2020.
KENNETH, J. G. *The age of uncertainty,* Houghton Mifflin Harcourt, Boston, 1979.
LEE, K. *Inteligência artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos,* Globo Livros, Rio de Janeiro, 2019.
LÉVY, P. *As tecnologias da inteligência: o futuro do pensamento na era da informática,* 2 ed., Editora 34, São Paulo, 2010.
MARCUSE, H. *O homem unidimensional: estudos da ideologia da sociedade industrial avançada*, Edipro, São Paulo, 2015.
McKAY, C. "Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making", *Current Issues in Criminal Justice,* vol. 32, nº 5 (2019), pp. 1-18, available in: https://doi.org/10.1080/10345329.2019.1658694.
NICOLELIS, M. *O verdadeiro criador de tudo: como o cérebro humano esculpiu o universo como nós o conhecemos,* Planeta, São Paulo, 2020.
PIMENTEL, A. F., BOLZAN DE MORAIS, J. L. & SALDANHA, P. M. "Estado de Direito e tecnopoder", *Justiça do Direito,* vol. 35, nº 3, 2021, pp. 6-43. https://doi.org/10.5335/rjd.v35i3.13241.
PROEVE, M. & TUDOR, S. *Remorse: psychological and jurisprudential perspectives,* Ashgate, New York, 2010.
QUIRÓS, D. Z. *Castigo e determinação da pena nos E.U.A.: um estudo sobre as United States Sentencing Guidelines,* Empório do Direito, Florianópolis, 2017.
RYBERG, J. "Sentencing and algorithmic transparency", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence,* Oxford University Press, New York, 2022, pp. 13-33, available in: https://doi.org/10.1093/oso/9780197539538.003.0002.
RYBERG, J. "Sentencing disparity and artificial intelligence", *The Journal of Value Inquiry*, vol. 57, nº 3 (2021), pp. 447-462, available in: https://doi.org/10.1007/s10790-021-09835-9.
RYBERG, J. & PETERSEN, T. S. "Sentencing and the conflict between algorithmic accuracy and transparency", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence,* Oxford University Press, New York, 2022, pp. 57-73, available in: https://doi.org/10.1093/oso/9780197539538.003.0004.
SADIN, É. *La humanidad aumentada: la administración digital del mundo*, Caja Negra, Buenos Aires, 2022.
SALOMÃO, L. F. (ed.). *Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário*, 3rd edn, CIAPJ/FGV, Rio de Janeiro, 2022, available in: https://bdjur.stj.jus.br/jspui/handle/2011/156490, access: Aug, 14, 2023.
SANTAELLA, L. *A inteligência artificial é inteligente?,* Edições 70, São Paulo, 2022.
SCHWAB, K. *A quarta revolução industrial,* Edipro, São Paulo, 2016.
SCHWARTZE, M. & ROBERTS, J. V. "Reconciling artificial and human intelligence: supplementing not supplanting the sentencing judge", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence,* Oxford University Press, New York, 2022, pp. 206-229, available in: https://doi.org/10.1093/oso/9780197539538.003.0011.
SMITH, N. *Justice through apologies: remorse, reform, and punishment,* Cambridge University Press, Cambridge, 2014.
SPENGLER, H. *Man and technics: a contribution to a philosophy of life,* Allen & Unwin, London, 1932, available in: https://archive.org/details/oswaldspenglermanandtechnicsgeorgeallenunwin1932/page/n5/mode/2up, access: Aug. 8, 2023.

TAIRA, B. R., KREGER, V., ORUE, A. & DIAMOND, L. C. "A pragmatic assessment of Google Translate for emergency department instructions", *Journal of General Internal Medicine,* vol. 36, nº 11 (2021), pp. 3361-3365, available in: https://doi.org/10.1007/s11606-021-06666-z.

TATA, C. "Resolute ambivalence: why Judiciaries do not institutionalize their decision support systems", *International Review of Law Computers & Technology,* vol. 14, nº 3 (2000), pp. 297-316, available in: https://doi.org/10.1080/713673373.

TATA, C. *Sentencing: a social process – rethinking research and policy,* Palgrave Macmillan, London, 2020.

TAUK, C. & SALOMÃO, L. F. "Inteligência artificial no Judiciário brasileiro: estudo empírico sobre algoritmos e discriminação", *Diké (UESC),* vol. 22, nº 23 (2023), pp. 2-32.

TEIXEIRA, A. *Teoria da aplicação da pena: fundamentos de uma determinação judicial da pena proporcional ao fato*, Marcial Pons, São Paulo, 2015.

THOMSEN, F. K. "Iudicium ex Machinae: the ethical challenges of ADM at sentencing", *in:* RYBERG, J. & ROBERTS, J. V. (eds.). *Sentencing and artificial intelligence,* Oxford University Press, New York, 2022, p. 252-275, available in: https://doi.org/10.1093/oso/9780197539538.003.0013.

TONRY, M. *Sentencing matters,* Oxford University Press, New York, 1996.

TURING, A. "Computing machinery and intelligence", *Mind*, vol. 59, nº 236 (1950), pp. 433-460, available in: https://doi.org/10.1093/mind/LIX.236.433.

VAN WINGERDEN, S. & PLESNIČAR, M. M. "Artificial intelligence and sentencing: humans against the machines", *in:* RYBERG, J. & ROBERTS, J. V. (eds*.). Sentencing and artificial intelligence,* Oxford University Press, New York, 2022, pp. 230-251, available in: https://doi.org/10.1093/oso/9780197539538.003.0012.

WALMSLEY, J. "Artificial Intelligence and the value of transparency", *AI & Society,* vol. 36 (2020), pp. 585-595, available in: https://doi.org/10.1007/s00146-020-01066-z.

WIENER, N. *Cibernética ou controle e comunicação no animal e na máquina*. Perspectiva, São Paulo, 2017.

ZUBOFF, S. *A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira do poder*, Intrínseca, Rio de Janeiro, 2020.