Lusíada

Repositório das Universidades Lusíada

Universidades Lusíada

Meiréis, Manuel Augusto Alves

Artificial intelligence in the criminal court : the recidivism algorithm

http://hdl.handle.net/11067/7196
https://doi.org/10.34628/yvyj-3c15

Metadados

Data de Publicação	2023
Palavras Chave	Inteligência artificial, Direitos humanos, Discriminação racial
Tipo	article
Revisão de Pares	Não
Coleções	[ULL-FD] LD, s. 2, n. 29 (2023)

Esta página foi gerada automaticamente em 2024-09-21T12:21:49Z com informação proveniente do Repositório

ARTIFICIAL INTELLIGENCE IN THE CRIMINAL COURT: THE RECIDIVISM ALGORITHM

(Conference Paper) [58]

INTELIGÊNCIA ARTIFICIAL NO TRIBUNAL PENAL: O ALGORITMO DA CONTINUIDADE CRIMINOSA

(Notas de Conferencia)

Augusto Meireis[59]

DOI: https://doi.org/10.34628/yvyj-3c15

Abstract: The period of the Jim Crow legislation, a segregationally one, was rooted with the so-called "*white supremacy*" legislation mainly of the Southern States of the US after the *Reconstruction*. It only ended with the beginning of the civil rights movement in the 1950s and 1960s. After the approval of laws separating the *colored* from the white people the U. S. Supreme Court, in an infamous decision, legitimized racial segregation. A denial of equality. Its 18th May 1896 and the Supreme Court decision of the *Plessy v. Ferguson* legitimized the "*separate but equal*" doctrine. The American decision that marked the end of the nineteenth century with the acceptance racial oppression and a recognized hierarchy based on the color of the skin with discriminatory consequences, our goal is to analyze the decision and the inheritance that, nowadays, still disseminates in racial inequities and ostracism with severe consequences, including, allegedly, in AI and "racist" algorithms.

Keywords: Racial Inequities; Equality; Human Rights; Human Dignity, AI.

[58] Paper presented at LSA 2022, in Lisbon. It should be understood as a research document and a work in progress aimed exclusively to serve as a guideline paper to an oral presentation at the conference. The paper was not revised, and the quotes and citations are merely indicative, with no formal revision, and some might be missing.

[59] Assistant Professor at the Faculty of Law, of Lusiada University at Porto; Researcher at CEJEA – Centre for Legal, Economic, and Environmental Studies of the said University.

Resumo: O período "Jim Crow", com raiz em princípios da "supremacia branca" do Sul (dos EUA) depois da *Reconstruction, só terminaria com o movimento dos direitos civis nas décadas de 50 e 60 do século XX*. Foi a 18 de maio de 1896 que o Supremo Tribunal, na decisão que ficou conhecida como *Plessy vs. Fergunson*, legitimaria a doutrina, com repercussões além-fronteiras, "*iguais mas separados*". A decisão americana que marcou o fim do século XIX com a aceitação da opressão e segregação racial e o reconhecimento de uma hierarquia entre as pessoas baseada na cor da pele com evidentes consequências discriminatórias, o nosso objetivo é analisar esta vil decisão e a pesada herança que, ainda hoje, se vai disseminando em discriminação racial e marginalização, com consequências nefastas, incluindo, alegadamente, na IA e em algoritmos "racistas".

Palavras-chave: Discriminação Racial; Igualdade; Direitos Humanos; Dignidade Humana; IA.

Presentation:

Artificial intelligence (AI) tools are being used (including in the criminal justice system) and becoming increasingly popular. These digital tools are performing risk assessments, the likelihood of the defendants become a recidivist.

The use of these technologies is not peaceful, and some critic's voices arise stating that machines are replacing the judges and we are relying on, secret, recidivism algorithm to predict the future of criminals.

The potential problems arise, including ethical and legal issues, but the question is if, for that reason, we should reject the AI tools in the justice system.

AI decisions, based on algorithms, develop information based in coded information and the machine can develop knowledge trough the "machine learning" that enables the constructions of models from data, including a large number of variables. Parameters are set using data to find patterns and classify them.

Although, the many questions that these super-beings pose is rooted in the problematic between rationality and morality.

From a human rights point of view, regardless of the futuristic scenario described, AI already raises some questions and problems.

Recognizing AI can be used to help societies to overcome challenges and improve people's lives, it has also a negative side and it can affect human rights, democracy, and the rule of law. (Teixeira & Almeida Araújo, to be published)

In the report of Thorbjørn Jagland to the 129th Session of the Committee of Ministers, new challenges to humankind were emphasised

> «[...] *for which Council of Europe legal standards are required. Three immediate challenges stand out: how to harness the benefits of the artificial intelligence revolution, while identifying and mitigating its threat to human rights, democracy and the rule of law»*[60].

These risks are not a price to pay, and Michelle Bachelet calls for an urgent action to assess the risks, and concluded that «[…] ***until compliance with human rights standards can be guaranteed, governments should implement a moratorium on the sale and transfer of surveillance technology.***» (Bachelet, 2021)

One of these scenarios is the use of AI in justice.

For instance, the use of "recidivism algorithms".

In the process known as *Loomis v. In Wisconsin*, one defendant argued that the algorithm that calculated its likelihood of recurrence discriminated against defendants by gender and, furthermore, that the algorithm was commercial in nature and protected by intellectual property secrecy rules, which made it impossible for it to syndicate the decision of the machine.

The ProPublica analysed a commercial AI tool made by Northpointe, Inc. and tested whether the recidivism algorithm, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), was predisposed against certain groups.

The analysis, reported on May 23, 2016 - "How we analysed the COMPAS Recidivism Algorithm" -, found that black defendants were more likely, than white ones, to be incorrectly judged to be at higher risk of re-offense (Larson, Mattu, Kirchner, & Angwin, 2016).

A significant case was discussed in the USA. In the case Wisconsin v. Loomis[61] the Wisconsin Supreme Court analyzed for the first time the use of algorithms and the right to a due process[62].

Loomis defense argued that, among others, it was not possible to know and understand how the algorithms used predicted the defendant's recidivism since it was protected by IP rights. The defendant was unable to challenge the validity of the risk assessment produced by COMPAS tools. (Teixeira & Almeida Araújo, to be published)

Can we conclude that the algorithm is racist?

The report analysed that:

---

[60] Jagland, T. (2019, May 16-17). Ready for future challenges – reinforcing the Council of Europe. Helsinki. Retrieved February 14, 2020, from https://rm.coe.int/168093af03, p. 9.

[61] State of Wisconsin v. Loomis. Retrieved March 21, 2022, from https://casetext.com/pdf-email?slug=state-v-loomis-22

[62] Freeman, K. (2016, 12). Algorithmic Injustice: How The Wisconsin Supreme Court Failed To Protect Due Process Rights in State v. Loomis. North Carolina Journal of Law & Technology, 18, pp. 75-106, pp. 89-89.

"In forecasting who would re-offend, the algorithm made mistakes with black and white defendants at roughly the same rate but in very different ways.

The formula was particularly likely to falsely flag back defendants as future criminals, wrongly labelling them this way at almost twice the rate as white defendants.

White defendants were mislabelled as low risk more often than black defendants." (Angwin, Larson, Mattu, & Kirchner, 2016)

Whatever our conclusion is, truth is that skin colour is, still, an issue and a source of discrimination.

"The matter of bias and discrimination is specifically addressed in the Executive Office of the President's Report[63] that grouped the challenges to promote fairness and overcoming the discriminatory effects of data: challenges related to the data used as inputs and challenges related to the inner workings of the algorithm itself. In the first group, the report the decision to use certain data inputs can result in discriminatory outputs. Such as poorly selected data, incomplete, incorrect, or outdated data, selection bias, and unintentional perpetuation and promotion of historical biases". (Teixeira & Almeida Araújo, to be published)

"On the second group, the report highlights the flaws related to poorly designed matching systems, personalization and recommendation services that narrow instead of expand user options, decision-making systems that assume correlation necessarily implies causation, data sets that lack information or disproportionally represent certain populations[64], and that algorithms encode discrimination and bias outputs related to the different participation in the digital ecosystem, «[…] *due to economic, linguistic, structural or socioeconomical barriers, among others.*»[65]" (Teixeira & Almeida Araújo, to be published)

"It is undeniable that AI entrenches bias, and, for that reason, outputs can not only replicate but also amplify it and, even, be weaponized against certain groups of people. In another aspect, AI is mainly a product of the Global North and, if so, «*if privileged white men are designing the technology and business models of AI, how they design for the south?*»[66]" (Teixeira & Almeida Araújo, to be published)

"Therefore, not only the amount of data is crucial to mitigate discriminatory outputs but also, guarantee a diversity of data to cover most spectrums. And, besides the substantial and diverse data, as stated by the AI Act, it is an obligation of States for *ex ante* test, manage risk and guarantee human oversight and,

[63] Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights from the Executive Office of the President, 2016.

[64] Executive Office of the President, 2016, p. 77.

[65] Ibid, p. 9.

[66] Arun, C. (2020). AI and the Global South: Designing for Other Worlds. In M. D. Dubber, F. Pasquale, & S. Das, The Oxford Handbook of Ethics of AI (pp. 588-607). Oxford: Oxford University Press, p. 591.

also, to guarantee that AI systems maintain «[…] *its level of performance under any circumstances*»[67]. These obligations will facilitate the respect of other fundamental rights by minimising the risk of erroneous or biased AI-assisted decisions." (Teixeira & Almeida Araújo, to be published)

Promoting a fair and impartial AI embraces, as an imperative, diversity and inclusion.

Almost a century later the founding fathers of the UDHR assumed the formulation of *universal standards*, "(…) associated with equality and therefore intrinsic and inalienable to man, for the *simple* fact of being a man and, the fact that the concept of human dignity "anchors different worldviews"(…)" (Caulfield & Chapman, 2005). And, as the French Philosopher Maritain, commenting the drafting of the UDHR, stated that "(…) at one of the meetings of a UNESCO National Commission where Human Rights were being discussed, someone expressed astonishment that certain champions of violently opposed ideologies had agreed on a list of those rights. "Yes", they said, "we agree about the rights but on the condition that no one asks us why". The "why" is where the argument begins" (1948, p. I). Human dignity gained the formal strength of being the catalyst for the discussion and assumed as the guiding principle. The French Philosopher, Maritain commenting the drafting of the UDHR, stated that "(…) at one of the meetings of a UNESCO National Commission where Human Rights were being discussed, someone expressed astonishment that certain champions of violently opposed ideologies had agreed on a list of those rights. "Yes", they said, "we agree about the rights but on the condition that no one asks us why". The "why" is where the argument begins" (1948, p. I).

The racial identity discourse in *Plessy v. Ferguson* "(…) left tragically to later generations of Americans the problem of sorting out the source of identity in law" (Davis, 2004, p. 41).

But one thing is for sure, "(…) there is no superior, dominant, ruling class of citizens", considering Justice Harlan words. And, as he said law (he referred specifically to the US Constitution, but we use his words in a broader context) is "(…) color-blind, and neither knows or tolerates classes among citizens" (Plessy v. Ferguson, 1896).

Even though the echo of these magnificent words, racial inequities are, still, a social and legal problem around the world and, like Duwell "(…) in line with Ricoeur or Lévinas, one could say: the phenomenology of moral experiences confronts us with the worth of the other, the other has the same authority over us that is the origin of the respect we owe to him" (Duwell, 2014, p. 43).

Funding: *This work is financed by national funds by FCT - Foundation for Science and Technology, under the Project UIDB/04053/2020*

[67] Ammanath, op. cit., p. 44, quoting the definition of «AI robustness» of The International Organization for Standardization.