Law, Governance and Technology Series 58

Henrique Sousa Antunes ·
Pedro Miguel Freitas · Arlindo L. Oliveira ·
Clara Martins Pereira · Elsa Vaz de Sequeira ·
Luís Barreto Xavier *Editors*

Multidisciplinary Perspectives on Artificial Intelligence and the Law

OPEN ACCESS

Springer

Law, Governance and Technology Series

Volume 58

The *Law, Governance and Technology Series* is intended to attract manuscripts arising from an interdisciplinary approach in law, artificial intelligence and information technologies. The idea is to bridge the gap between research in IT law and IT-applications for lawyers developing a unifying techno-legal perspective. The series will welcome proposals that have a fairly specific focus on problems or projects that will lead to innovative research charting the course for new interdisciplinary developments in law, legal theory, and law and society research as well as in computer technologies, artificial intelligence and cognitive sciences. In broad strokes, manuscripts for this series may be mainly located in the fields of the Internet law (data protection, intellectual property, Internet rights, etc.), Computational models of the legal contents and legal reasoning, Legal Information Retrieval, Electronic Data Discovery, Collaborative Tools (e.g. Online Dispute Resolution platforms), Metadata and XML Technologies (for Semantic Web Services), Technologies in Courtrooms and Judicial Offices (E-Court), Technologies for Governments and Administrations (E-Government), Legal Multimedia, and Legal Electronic Institutions (Multi-Agent Systems and Artificial Societies).

Henrique Sousa Antunes • Pedro Miguel Freitas •
Arlindo L. Oliveira • Clara Martins Pereira •
Elsa Vaz de Sequeira • Luís Barreto Xavier
Editors

Multidisciplinary Perspectives on Artificial Intelligence and the Law

Springer

CATÓLICA
RESEARCH CENTRE
FOR THE FUTURE OF LAW
LISBON

Plataforma
de Apoio à
Investigação
em Direito
na Católica

Editors
Henrique Sousa Antunes
Faculty of Law
Universidade Católica Portuguesa
Lisbon, Portugal

Arlindo L. Oliveira
Instituto Superior Técnico
University of Lisbon
Lisbon, Portugal

Elsa Vaz de Sequeira
Faculty of Law
Universidade Católica Portuguesa
Lisbon, Portugal

Pedro Miguel Freitas
Faculty of Law
Universidade Católica Portuguesa
Porto, Portugal

Clara Martins Pereira
Durham Law School
Durham, UK

Luís Barreto Xavier
Faculty of Law
Universidade Católica Portuguesa
Lisbon, Portugal

ISSN 2352-1902 ISSN 2352-1910 (electronic)
Law, Governance and Technology Series
ISBN 978-3-031-41263-9 ISBN 978-3-031-41264-6 (eBook)
https://doi.org/10.1007/978-3-031-41264-6

This work was supported by PAIDC - Plataforma de Apoio à Investigação em Direito na Católica

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

AI Instruments for Risk of Recidivism Prediction and the Possibility of Criminal Adjudication Deprived of Personal Moral Recognition Standards: *Sparse Notes from a Layman*

Pedro Garcia Marques

Abstract In what follows lies a recount of a concerned criminal lawyer, a layman, as he observes the change foreshadowed by AI in the field of individual risk recidivism assessment for the purposes of criminal penalty imposition on convicted felons. The text will therefore reflect upon the nature of that assessment when promoted by new AI programs based on actuarial-meaning statistically derived-information. It then proceeds to compare that risk recidivism assessment with the one undertaken within the current traditional human paradigm. Identifying the ensuing challenges set by the technological alternatives on the very survival of criminal law's principiological mainstays. A final note will be drawn on what is lacking in the technological proposal, for all its technical upsides and perceived advantages. The approach here changes. From literature one will bring to the fore the very human account that lies at the center of anything resembling judgment. Both the judgment of the individual being assessed and the one of the court doing the assessment. Human as they both are, one heeds the kind of humanity an entire science—that of law—and its specific approach must acknowledge. Exactly that humanity that seems to be lacking in the technological AI proposals.

1 Introduction

Sparse notes, no more, will be found below. Reflections of a criminal lawyer, no doubt sensing the uneasy feeling of redundancy. A layman lead by curiosity and scholarly interest. One that for all his investment is in no doubt that he can only tag along the incessant technological revolution that here as elsewhere creates new

P. G. Marques (✉)
Law School, School of Lisbon, Catholic University of Portugal, Lisbon, Portugal
e-mail: p.marques@ucp.pt

H. Sousa Antunes et al. (eds.), *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, Law, Governance and Technology Series 58,
https://doi.org/10.1007/978-3-031-41264-6_18

possibilities and all-encompassing challenges. Challenges of a kind that in order to be properly met will require the one thing that cannot be provided: time for pause.

The following lines are not those of an expert. If expertise is what you, dear reader, are searching, please do look elsewhere. They are, on the other hand, the result of (a small amount of) time to pause and reflect. If, dear reader, you are willing to settle with that then, please, do come in and join me. What I offer are no more than mere reflections. Who knows, maybe your time will in end be well spent.

The immediate object of our attention is be centered around machines and computer programs endowed with artificial intelligence, with particular emphasis on those in which the degree of autonomy is determined by an independent decision-making capacity with the ability to learn, at a level of *deep learning.*[1]

The consideration of predictive tools for the level of risk of recidivism will gain particular importance, namely computer tools for anticipating the risk of recidivism of convicted felons.

Endowed with autonomy in considering the nature and measure of the individual penalties, COMPAS, LSI-R, VRAG and ORAS will be taken as a reference, as they are acronyms of currently available computer programs for "prediction" of *recidivism risk*, each of them dependent on *actuarial-meaning statistically derived-information.*[2]

All of them concern computer programs that, at the present time, calculate, through the attribution of points and ranking definition, the recidivism risk of any certain convicted felon in the commission of crimes in the future.

Offered today as auxiliary tools to judicial decisions, these are not yet programs endowed with autonomous decision-making attributes. Albeit that technological capacity may already be well within reach (Eaglin 2017, pp. 59–122).

At stake in any of these instruments of prediction is, therefore, at the present moment, its claim of capacity to predict future individual behavior. And not just that, but to do it accurately. And always, inevitably, the further claim that *that* judgment of probability, precisely because of its undoubted accuracy, should be able to serve as the criterion and the basis, not only for limiting the individual physical freedom, but also for deciding its severity and nature.

The possibility of criminal adjudication by machines and programs, i.e., by the artifact endowed with artificial intelligence, to any given individual based on any given set of facts of a specific penalty, proposed in a specific form and at a certain level of severity, based on the prediction of his/her future criminal behavior is now

[1] On this, please see Eaglin (2017), pp. 59–122.

[2] Acronyms for Correctional Offender Management Profiling for Alternative Sanction (COMPAS), Level of Service Inventory -Revised (LSI-R), Violence Risk Appraisal Guide—Revised (Vrag) and Ohio Risk Assessment System (ORAS), respectively. For the detailed description of these systems, among others, please *see* Eaglin (2017), pp. 59–122. For detailed discussion on risk recidivism assment tools, please see Eaglin (2017), pp. 59–122, Selbst (2017), pp. 96–98, Citron and Pasquale (2014), pp. 1–34, Notes (2017), pp. 1530–1537, Tillers (2002), pp. 1365–1380, and Vervaele (2014), pp. 115–128. In portuguese language, please *see* Rodrigues (2020), pp. 11–58. In Italian language, please *see* Gialuz (2019), pp. 1–23.

upon us. And with it, for the first time, the central role played in the decision-making process of adjudication of criminal guilt and of the assignment of blame by machines. That is, by objects utterly devoid of capability for moral judgment and reflection.

The question will inevitably arise over the possibility and legal justification of individual criminal adjudication devoid of moral recognition.

The absence of which becomes a fact when, in that process of ascription, the machine *calculae* departs from a mere (even if already significant) instrumental role and achieves full autonomy to the point of becoming sole arbiter in the decision-making process of individual adjudication of a criminal penalty. A well-advertised fact in a not too distant future by their respective system's providers (Eaglin 2017, pp. 59–122), as the ongoing development of these instruments and the improvement of their learning capacity will allegedly afford them the kind of algorithms capable of prediction calculations endowed with increasingly smaller margins of error (Eaglin 2017, pp. 59–122).

The moment may come when these instruments for calculating recidivism will become fully capable of replacing the court, the Public Prosecutor's Office and generally the social welfare apparatus in its court's auxiliary capacity, deciding by themselves, by *their own lights*, unaccompanied and unchecked, the nature of the penalty, determining its measure, its duration and the conditions for its completion.

On that day, the decision making process of penalty adjudication will become, not a machine based, not even a machine driven, but rather a machine decision-making process.

The combined consideration of these instruments and their adoption (in their current status and in the near future) will challenge the very existence of an array of traditional principles of criminal responsibility, such as, the personal nature of each individual criminal penalty, the *in dubio pro reo* or the acceptance of free will as the basis for contemporary philosophical justification of criminal individual adjudication.

2 The Predictability of Future Behavior

2.1 The Acceptance of a Judgment of Probability as a Criterion and Basis for Limiting Physical Freedom: Its Implications and Its Consequences

Let us take on each of the challenges one by one. And let us do it *through the lens* of the type of assessment that any of these programs is capable of asserting: one that is undoubtedly based, not on certainty, but only on probability.

Probability of recidivism does not mean certainty of recidivism and risk assessment calculation does not equate to certainty of danger. Therefore, a risk probability

judgment, even if rigorously determined in relation to an agent, does not imply that he is actually dangerous.

Thus, the imposition of a penalty based on that calculation may imply:

a. Its enforcement on an agent that, despite the calculated statistical probability, does not actually pose any of the risks calculated as probable;
b. That this negative margin of inefficiency is accepted, meaning that it is admitted as a *cost* of the system that a non-determinable number (since the penalty is applied to all according to a calculated risk calculation validated by the system) of convict felons will be subject to a type of penalty and to a level of severity that, in individual cases, are not justified, due to the non-coincidence between the risk probability calculation and the individual risk that the convicted person *actually* represents.

The penalty or penalties chosen, the measure of their severity, that is, the sentencing to which the accused is subject, subject as it is to a judgment based on a mere probability-based risk calculation, will apply to the individual regardless of the risk that, as mentioned above, *that convicted person actually* represents.

Thus the problem: if the applicable penalty is based on the pursuit of preventive needs, then the risk of justifying a penalty on the basis of a judgment of probability, surely will imply that the deprivation of individual liberty of *that* individual agent may pursue no foreseeable end, legitimate or otherwise. Hence corresponding to a penalty whose imposition no legitimate end can justify.

The abandonment is thus dictated, albeit without conscious realization, of one of the fundamental principles of contemporary criminal adjudication: that of the irrevocable inherently personal nature of each individual penalty.

Probability, risk, that is, doubt—even if supported by calculation—will do. It will suffice for the deprivation of liberty and above all for the imposition of a criminal penalty with the particular kind of acumen that any criminal ascription implies (and is supposed to imply). The convicted felon is blamed, singled out and labeled as dangerous, not because she/he actually is, but because everything *technological*—the machines, a spreadsheet, an algorithm—suggests that she/he is. Even if and even when she/he is not.

The commitment may very well be on the constant refinement of the mechanisms involved in the calculation of probability of risk as the basis for the proposal of—or decision on—the nature and the quantification of severity of the penalty to be applied to each convicted felon.

But that convicted felon will always be regarded by the system as the object of calculation. The object to be targeted with those measures.

Reduced to the condition of object of calculation, the *calculae* on the *rei extensae* involved in the risk assessment necessary for penalty determination will inevitably fail to grasp the absence from the equation of the very person, the individual herself/himself, that forms that *object* of measurement.

Again, one can the swear by the commitment to allow for autonomy in calculation only when the best possible technical conditions are achieved. But with doubt hanging over each risk assessment exercise and with an object rather than

a person in view, the perspective adopted is necessarily one of damage limitation in the face of miscalculation by the part of implements that will dispense with any consideration of the person, replacing it by a mere object under measurement.

But, if this is true than one can always argue that probability is precisely at the heart of every assessment the future possibility of recidivist action by convicted felons. Be that the traditional judicial evaluation subject to open court discussion, bearing on the pertinent evidence over the personality of the defendant, as well as her/his family, social and economic standing and conduct before and after the perpetration of the crime; be it the technological cybernetic activation of algorithmically organized parameters that, in all honesty, will bear into the calculus much the same factors and parameters the judge and jury will comb through in order to arrive at a substantiated decision on the sentencing. All of which allowing no more than an estimate. Probability in the end underlies any prognostic judgment on recidivism and sustains any decision on the imposition and enforcement of criminal penalties in each particular case, on each respective convicted felon.

So one can argue that, at least in its technological form these new found methods of criminal recidivism assessment, the issue or rigor or lack of in the calculous of probability becomes expressly thematized. Along with the pledge towards its successful achievement in a measured and verifiable way.

Whereas the traditional judicial form of risk assessment will foster no such ambition as it will not allow for any equivalent capability. For any such close scrutiny will require the kind of mass data treatment and simultaneous differentiated parameter control that will fall quite beneath the scope of capability afforded by the traditional judicial method, human as it is in its nature and form.

Moreover, the treatment of the data and the care taken in its selection in order to properly support the recidivism risk assessment is, one can further argue, no less rigorous, as the traditional *human* method and the new technological *alternative* are set side by side.

This being true it is possible to defend that these new technological possibilities provide the platform for the much-needed criticism and denunciation of traditional systems of risk assessment and ascription—with the inherent blame assignment and individual criminal liability adjudication—for not acknowledging the fact that they, just as much as their cybernetic counterparts, are solely dependent and reliant upon probability. Probability stands for, on one and the other, still as the paradigm.

And by not coming to terms with that realization one can contend that the traditional *human* method will effectively conceal the requirement for the utmost rigor in its calculation of individual risk assessment. Dispensing with the very implements that allow for a thorough and convincing review as well as scrutiny of each individual decision on the penalty. Thus opening the door, without even realizing it, to judicial arbitrariness.

If the criticism is true. And it is. And if we can credit its particular thrust on the very existence today of technological alternatives with their unmatched capabilities; these technological proposals are none the less lacking. And they are lacking where the traditional approaches are not. At the core of their fundamental epistemic option they do fail. And they do fail where the criticized traditional approach,

notwithstanding the pertinence of the criticism, does not. The technological models fail in and because of their *epistemic self-sufficiency*.

Self-sufficiency as their epistemic central feature lie, no so much on the fact that the entire calculation model afforded by the ever-adaptable given algorithm is the epilogue of an exercise in probability. Rather, their stapled epistemic self-sufficiency resides on the assumption that probability suffices. No more is needed.

Probability is the exercise on which the cybernetic *calculator* grounds its entire risk assessment. *Doubt* therefore becomes the very thing on which the—every—effort of calculation is based. Henceforth, *doubt*, namely the one that under the persuasion of extraordinary technical capability is reduced to the fringes of statistical irrelevance, will not suffice to merit the reversal, let alone the putting into question of whatever level of risk assessment that has been set regarding *that* individual convicted felon. Despite the doubt, the assertion of statistical probability will suffice.

Ubiquitous as doubt may be on each exercise of calculation, none of that will hinder the imposition of penalty on the convicted felon to the exact measure of the risk that, despite the doubt, he is statistically deemed to represent.

A person is convicted, deprived of he/his freedom—and as mentioned *supra* blamed, singled out and labeled as dangerous—in the face of a justification that may very well be wanting. Since, despite the assessment, in that particular case, that particular person may have, in fact, posed no risk at all.

Of course, one may always turn to the traditional *human way* of going about that risk assessment. Based just as much on probability as its technological counterpart and *craving*, as it were, for the kind of certainty that will provide for its adequate juridical validity. But the *turn* in the analysis will not afford any conceivable answer to the question at hand. *Two wrongs will not make a right*, one would say and the misgivings of the traditional human approach will fall short on providing the kind of improvements the technological one is in dire demand.

But also and above all, albeit underlying a fact, it does not, on the other hand, describe the reality. And the reality of *that* human endeavor is its firm rootedness on the quest for certainty. One that provides an explanation for a penalty grounded on the conviction of that particular judge and through him of the State (and the Commonwealth for those in *those* whereabouts) regarding the specific risk that that individual person *actually* poses. And on the responsibility of the State for those risk assessments provided in error.

In this *human* traditional approach doubt is valued, subject to appreciation, challenged in open court, subject to argument by the parties, contested, reviewed over and over and challenged again. Doubt, even at the fringe of statistical irrelevance, is, even then, doubt. In fact one may say that in this human approach there is no such thing as irrelevant doubt, statistical or otherwise.

The difference in how doubt is handled in each of the *systems* is striking.

And perhaps therein lies the other kind of sufficiency that the technological approach boasts and takes upon itself: the procedural self-sufficiency. Apparently dispensing with discussion of any shape or form, on open court or *in camera* or any other, it calculates and assesses and proposes the end result. It may decide

one day by itself on it. In the meanwhile, as it aides the court on delivering its decision, what, may one ask, will the court have in its hands to counteract the *self-sufficiency* of a readymade decision that decides on what is *negligible* doubt, that equates it to sufficient certainty, without ever partaking its rationale, nor allowing for any contradictory discussion on what has already become the machine's sole *actuarial* (using the term on which the machine decides against itself) domain?

Doubt and probability thus fulfill, in the traditional human approach, the part of *costs* of the system They are evidence of its fallibility, the very sources of its injustice. The awareness of what therein lies in palpable threat of unfairness manifest traditional approach's commitment to the principle of presumption of innocence and serves as the backbone of its claim to constitutional validity.

This claim entails the acceptance, as an unquestionable moral assertion of lesser evil, that it is best to acquit a guilty person than to convict the innocent one. And to bring it to bare on the whole exercise of calculation and assessment of recidivism risk of any given convicted person.

Doubt will therefore be brought into consideration in favor of the one being evaluated, accepting risk of error by default. Insufficient weighing of risk assessment serves, in the face of doubt, as the *cost* of a traditional system that, by proclaiming the above mentioned principle of lesser evil, does so as part of its duty towards equal respect and consideration of each individual, recognizing in each and everyone her/his inherent humanity. Such a proclamation is endowed with immediate practical consequences, bringing to the fore the exact principle of which that proclamation and the *traditional* method are their all too human practical translation: that of the *in dubio pro reo*. Precisely the one that the technologic alternative fails to consecrate and to integrate in itself. Precisely the one that avows the constitutional validity of the former to detriment of the latter.

3 The Risk of Technological *Bias*

Furthermore, the linear understanding of prospective behavior at the heart of any technologically autonomous risk assessment will be afford the understanding of human behavior precisely as linear. The feasibility of technological risk assessment becomes possible only insofar as future human behavior is determinable. A determination that, let us remind ourselves, is only made possible through the transformation of doubt over actual risk, when statistically improbable, into *certainty* of risk.

At this level, no room will be afforded to those *rare* moments of regret, desistance, repentance, introspection or epiphany. Back tracking, changing of minds, since statistically highly improbable will equate to nonexistence in the *brave new world* of all too certain technological decisions on risk. Those precise cybernetic decisions that will perform the function, if and only if, the human element is extricated from the equation, utterly ignored from the calculous. Those moments, rare as they are, are as good as nonexistent. Yet human nature and human condition,

on those rare moments, stubbornly challenge the machine-like certainty, comforting as it may appear in its alleged accuracy.

If one accepts this *technological* state of affairs, along with the above-mentioned methodological conundrums, a *turn* will be imposed on *the* philosophical premise that sustains and forms the basis for any exercise in criminal adjudication and blame ascription. That of its reliance on the understanding of each person as a free agent. That, in the end, of a commitment to *indeterminism*.

At stake is the abandonment of an indeterministic conception of human action. And at the statistical fringes, where improbability lurks in, the irrelevance of free will is for all intents and purposes accepted. Replaced with a kind of self-appointed statistically plausible determinism, albeit unchallenged and uncritically accepted.

Then there is the risk of perpetuation and promotion of the existing model of social reality in its prejudices and asymmetries. And, with it, the conjunctural set of problems that persist in challenging the credibility of all technological attempts at risk assessment.

Since any exercise will need to take into account data on social environment and provenance of the convicted felon, such an exercise may imply the risk of considering the defendant in terms of a model, and, as such, as a member of a paradigmatic community for statistical purposes. And the risk of bias creeps in.[3]

In addition to the inherent risk of miscalculation, such a model may tend to perpetuate a dominant majority perception of a social reality with the kind asymmetries that feed its underlying prejudices. When lacking critical instruments able to denounce them, the algorithmic model of risk assessment posed by the defendant, instead of being able to ascertain the actual danger that she/he represents as an individual, may undergo the risk of overburdening the exact same persons who, as members of disfavored communities, already suffer the effects of those exact same asymmetries and prejudices.

Even worse. By providing an arsenal of data and methods of calculation that seem to order a necessary conclusion in a logical, plausible and linear way, it will tend to validate and, with that, to legitimize ways of collecting and processing data and consequent proposals for penalties that, underneath that arsenal, replicate—in a fashion that lends itself to be hidden and camouflaged—the precise *bias* that will label those individuals whose risk is being assessed with the same criminogenic factors that affect in a marked way the members of the communities they come from. And for no other reason than the fact that they come from those communities.

Hence, those who, in fact, most often find themselves grappling with justice may, in the logical plausible linear assessment of the machine, find themselves labeled as the personification of danger because mechanical logic demands it. Despite of who she/he individually are and in spite of what a human consideration, under the light

[3] The risk of bias in risk predictive assessment tools has a long-standing issue that has merited significant discussion. For detailed analysis on the issue, please *see*, among others, Angwin et al. (2016), Chander (2017), pp. 1023–1045, Flores et al. (2016), pp. 38–45, Spielkamp (2017), pp. 96–98, Temming (2017), pp. 26–29, and Zarsky (2014), pp. 1375–1412.

(a no less) human reasoning, might be able to do in challenging that (machine-like calculated) logic.

4 Conclusion

What is then left out?

Most definitely the consideration of the improbable, although always possible, *change of mind*.

But how so?

We will try to provide an answer in much in the same way as in previous occasions (Marques 2016, pp. 505, 506; Marques 2021, p. 103). The path will be unorthodox. We start with literature.

The kind of *change* that the brazilian writer JORGE AMADO, in his novel *Terras do Sem Fim,* lets us in *in the mind* of Damião, a hired killer (*jagunço*), as he prepares to ambush and kill yet another victim. A change of mind, an epiphany even, that leads him, for the first time, to miss the shot (Amado 1942, pp. 75, 76).

Damião is tormented by the questions he had heard from Sinhô Badaró addressed to another killer (*jagunço*), "Do you think it's good to kill people? Don't you feel anything? Nothing inside?". And now in the ambush, as he was preparing to kill Firmo, an idea emerges. At first, as a mere conjecture—"And suddenly, the terrifying idea cut his head: what if Dona Teresa were pregnant, a child in her belly? (...) He would have been born without a father, the father would have been under the aim" of Damião.

> ≪And it shudders all over, its huge giant body.
>
> You can see Dona Teresa's face (...) before there was the moonlight, white as milk, spilling over the ground. (...) She is asking him not to kill Firmo, for God's sake he does not kill ... On the moonlit floor the black man sees Teresa's face perfectly≫.
>
> What if I didn't kill Firmo? (Amado 1942, pp. 75, 76) [4]

Unpredictable as it may be, therein lies humanity.

And Arendt's words spring to mind: "the fact that man is capable of action means that the unexpected can be expected from him, that he is able to perform what is infinitely improbable" (Arendt 1998, p. 178). And so because man "is unique", since, with each birth, "something uniquely new comes into the world" (Arendt 1998, p. 178).

Every beginning, thus Arendt, every origin bares with it the element of unpredictability. Each moment of *difference* is, therefore, an unexpected and unpredictable moment of novelty, of surprise, confirming both that the "new" "always appears in the guise of a miracle" (Arendt 1998, p. 178).

[4] Translation by the author from the original in portuguese.

Here, thought is born as the political condition for action, as the very possibility of language and discourse and, therefore, of a life founded on, because marked by, humanity. And, with that, the word, the thought that reflects it and the judgment that, then, through one and the other becomes possible.

Now, what is at stake here is precisely, with ARENDT, the return of thought to action (Arendt 2005, pp. 188, 189). Of action considered as freedom. Since, when action, that which begets human creativity, is left aside, thought itself, as the "activity of sharing in the difference in which equals are recognized" (Arendt 2005, pp. 188, 189), will be forgotten. And, on that moment, judgment, that on which the very possibility of freedom relies, will disappear from human existence.

Hence, what is definitively left out?

Again, literature.

We turn to PRIMO LEVI in his *Se Questo è un Uomo* (Levi 1989). While captive in Auschwitz he recounts that one day while being escorted by a guard, named Alex, an oiled steel pipe crossed their way. As the guard passed, he leaned on the pipe and accidentally got oil on his hands. "Without hate and without derision", Levi tells us, "Alex wipes his hand on my shoulder, the palm and the back of your hand, to clean it". And concludes Levi, "(. . .) he would be very surprised, the poor and brutish Alex, if someone told him that today I judge him for this act, he and Pannwitz, and the countless people who were like him, big and small, in Auschwitz and elsewhere. " (Levi 1989, pp. 96, 97). [5]

This—human judgment as the essence of *human* thought and the mark of human dignity—becomes forgotten and, as forgetfulness sets in, P. F. TRAWSON's lament becomes our own. A lament that he expresses as follows: "it is a pity that the talk about moral sentiments has fallen out of favor" (Strawson 2008, p. 26). Fallen into disfavor is not only the doubt regarding the restriction of moral concepts into a single, rational origin, but also the perception towards the insufficiency of mere reason as the prime motivational impetus for action.

A lament that serves as the point of departure for the unravelling of his proposed *reactive attitudes of participation*. As those that in the face of each behavior of each one of us trigger common reactive attitudes (ordinary reactive attitudes). The kind of attitudes—participatory in nature—that the Author calls "essentially natural and human reactions", without which "it is doubtful that we have something that we can consider intelligible as a system of human relations, as human society" (Strawson 2008, p. 26).

At stake is the need to take into consideration the network of attitudes and feelings that form an essential part of moral life as we know it, as a perceptible source of the meaning enclosed in the "language of morals", and, as such, of everything we intend to say when we speak of worthiness, responsibility, guilt, condemnation, and justice. Each of them resonates in those attitudes and feelings that constitute the cement of the specific kind of relationship that forms a community. Of the kind of relationship that equate to *life* itself, again with ARENDT, in its ontologically

[5] Translation by the author from the original in portuguese.

characterizing *togetherness*. A relationship that, in the end lies, at the very he very foundation of moral responsibility.

That, in the end, is missing. Left out. [6]

References

Amado J (1942) As terras do sem fim (from the 1942 original). Livros do Brasil, Lisbon

Angwin J, Larson J, Mattu J, Kirchner L (2016) Machine bias. There is software that is used across the country to predict future criminals. And it is biased against blacks. https://www.propublica.org/article/machine-bias-risk-assessmentsin-criminal-sentencing. Accessed 8 Sept 2021

Arendt H (1998) The human condition. Introduction by Margaret Canovan. University of Chicago Press, Chicago

Arendt H (2005) Thinking and moral considerations. In: Kohn J (ed) Responsibility and judgment, edition and introduction. Schocken Books, New York, pp 159–189

Chander A (2017) The racist algorithm? Mich Law Rev 115:1023–1045

Citron DK, Pasquale F (2014) The scored society: due process for automated predictions. Wash Law Rev 89:1–34

Eaglin JM (2017) Constructing recidivism risk. Emory Law J 67:59–122

Flores AW, Bechtel K, Lowenkamp CT (2016) False positives, false negatives, and false analyses: a rejoinder to machine bias: there's software used across the country to predict future criminals. and it's biased against blacks. Fed Probation 80:38–45

Gialuz M (2019) Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa. Direitto Penale Contemporaneo 2019:1–23

Levi P (1989) Se questo è un uomo. Einaudi, Turim

Marques P (2016) O juízo crítico da culpa, PhD. Dissertation, Universidade Católica Portuguesa

Marques P (2021) Inteligência artificial e humanismo – prolegomena. In: Sequeira EVD (ed) Católica talks – direito e tecnologia. Universidade Católica Editora, Lisbon, pp 95–114

Rodrigues AM (2020) Inteligência artifical – a justiça preditiva entre a Americanização e a Europeização. In: Rodrigues AM (ed) A inteligência artificial no direito penal. Almedina, Coimbra, pp 11–58

Selbst AD (2017) Disparate impact in big data policing. Georgia Law Rev 52:108–195

Spielkamp MA (2017) Inspecting algorithms for bias. MIT Technol Rev 120:96–98

State v. Loomis (2017) Criminal law-sentencing guidelines – Winsconsin Supreme Court requires warning before use of algorithmic risk assessments in sentencing. 88 1 NW.2d 749. Harv Law Rev 130:1530–1537

Strawson PF (2008) Freedom and resentment. In: Strawson PF (ed) freedom and resentment and other essays. Routledge, London, pp 1–28

[6] *See* also, on AI and judicial reasoning, in this book L M Pereira, F C Santos and A B Lopes - AI Modelling of Counterfactual Thinking for Judicial Reasoning and Governance of Law; W Gravett - Judicial Decision-making in the Age of Artificial Intelligence; D Durães, P M Freitas and P Novais - The Relevance of Deepfakes in the Administration of Criminal Justice; and J C Abreu - The "Artificial Intelligence Act" Proposal on European e-Justice Domains Through the Lens of User-focused, User-friendly and Effective Judicial Protection Principles. *See* also, on biases, in this book W Gravett - Judicial Decision-making in the Age of Artificial Intelligence; and D Durães, P M Freitas and P Novais - The Relevance of Deepfakes in the Administration of Criminal Justice. *See* finally, on the COMPAS system, in this book W Gravett - Judicial Decision-making in the Age of Artificial Intelligence.

Temming M (2017) Fair-minded machines–a new drive to revamp artificial intelligence may cut down on bias, Science News, September 16, 2017, pp 26–29

Tillers P (2002) Introduction: a personal perspective on'artificial intelligence and judicial proof'. Cardozo Law Rev 22:1365–1380

Vervaele JAE (2014) Surveillance and criminal investigation: Blurring of threshholds and boundaries in the criminal justice system? In: Gutwirth S, Leenes R, Hert PD (eds) Reloading data protection Multidisciplinary insights and contemporary challenges. Springer, Heidelberg, pp 115–128

Zarsky TZ (2014) Understanding discrimination in the scored society. Wash Law Rev 89:1375–1412