 Etica & Politica / Ethics & Politics, XXIV, 2022, 3, pp. 451-463
ISSN: 1825-5167

AI AND THE GROUNDS FOR HUMAN RIGHTS

NIR EISIKOVITS
Philosophy Faculty
University of Massachusets Boston
Nir.Eisikovits@umb.edu

ABSTRACT
The second edition of Claudio Corradetti's Relativism and Human Rights[1] updates his influential account of the theory and practice of human rights and further deepens what was already a major contribution to the philosophical literature in this field. In Chapter 3 of the book Corradetti offers a detailed discussion and reinterpretation of several attempts to ground human rights. This paper will offer a new layer to the discussion of how human rights are grounded. I will focus on the interplay between technology and the human rights agenda and, in particular, on the relationship between the rise of Artificial Intelligence and the project of grounding human rights. Corradetti and other key thinkers on rights have not paid much attention to the impact of AI on the traditional grounds of human rights, and I hope this paper encourages them to take a second look. Part I provides a brief overview of statistical machine learning (the main variant of what the popular media calls AI) and its current uses. Part II considers the implications of this type of technology on classical and contemporary justifications of human rights. Part III takes up the relationship between algorithmic governance and human rights. The conclusion places the discussion in the context of the broader interplay between technological developments, self-perceptions, and political institutions.

KEYWORDS
Artificial intelligence, human rights, judgement, Kant, Mill, Gewirth

The second edition of Claudio Corradetti's *Relativism and Human Rights*[1] updates his influential account of the theory and practice of human rights and further deepens what was already a major contribution to the philosophical literature in this field. In Chapter 3 of the book Corradetti offers a detailed discussion and reinterpretation of several attempts to ground human rights – from Alan Gewirth's conception of Rational Purposive Agency and the Principle of Generic Consistency it yields, to the various permutations the idea of dignity has

[1] Corradetti, C. 2022. Relativism and Human Rights: A Theory of Pluralist Universalism (Springer, 2nd edition).

undergone in the history of philosophy and in recent political documents such as the Universal Declaration of Human Rights.

This paper will offer a new layer to the discussion of how human rights are grounded. I will focus on the interplay between technology and the human rights agenda and, in particular, on the relationship between the rise of Artificial Intelligence and the project of grounding human rights. Corradetti and other key thinkers on rights have not paid much attention to the impact of AI on the traditional grounds of human rights, and I hope this paper encourages them to take a second look. Part I provides a brief overview of statistical machine learning (the main variant of what the popular media calls AI) and its current uses. Part II considers the implications of this type of technology on classical and contemporary justifications of human rights. Part III takes up the relationship between algorithmic governance and human rights. The conclusion places the discussion in the context of the broader interplay between technological developments, self-perceptions, and political institutions.

PART I: STATISTICAL MACHINE LEARNING AND ITS USES

Artificial Intelligence (AI) is a popular term for algorithmic technology that is capable of making predictions based on historical data. Most so called AI's are statistical machine learning technologies that are very good at recognizing patterns in large data sets and making predictions or recommendations based on those patterns[2]. The technology is being deployed in a rapidly expanding set of contexts: it is used to predict which television shows or movies one would like to watch based on past preferences, to make decisions about who can get credit or be approved for a loan based on past performance (and other proxies for likelihood of repayment), for the detection of fraudulent credit transactions, for the identification of malignant tumors, for hiring and firing decisions in large chain stores, hiring and firing in public school districts, and for a variety of purposes in law enforcement – from assessing the chances of recidivism, to police force allocation, to the identification of criminal suspects[3]. The term AI conjures up some more speculative applications as well – in particular the prospect of AI being used to create robots who could outperform humans on tasks across the board. This view of the technology,

[2] For overviews of the field and the place of statistical machine learning in it see Sugiyama, M. 2015. Introduction to Statistical Machine Learning (Morgan Kaufmann) and Russel, S. and Norvig, P. 2020. Artificial Intelligence: A Modern Approach (Pearson). My discussion in part 1 draws on my co-written Eisikovits, N. and Feldman, D. (Forthcoming) "AI and Phronesis," Moral Philosophy and Politics

[3] For excellent overviews of current uses of the technology and the concerns it raises see O'Neil, C. 2016. Weapons of Math Destruction (Crown), Eubanks, V. 2018. Automating Inequality (St. Martin's Press) and Zuboff, S. 2019. The Age of Surveillance Capitalism (Public Affairs).

sometimes referred to as Artificial General Intelligence or Superintelligence[4], captures the imagination, stokes anxieties, and suggests associations of HAL, the rogue artificial intelligence that refused to be shut down in Stanley Kubrick's *2001: A Space Odyssey,* or the deadlier Skynet system that achieves world domination in the *Terminator* franchise. Public figures from Elon Musk to Stephen Hawking[5] have raised alarms about the risks such an AGI might pose. But, at the moment, the technology is simply nowhere near these worrisome capacities. Instead, statistical machine learning systems are being trained to replace specific and narrow human tasks as detailed earlier.

These narrow applications of AI have given rise to less cinematic but still very serious concerns about the entrenchment of biases and inequities in the distribution of resources. Could AI make existing social injustice worse? The worry is of the "garbage in-garbage out" variety: if the algorithms used for loan approval, facial recognition and hiring are trained on biased data, or if they assess the data based on biased models and proxies, wouldn't we end up with technology that simply perpetuates existing prejudices and inequalities? And wouldn't the opaque and mathematically complex algorithms used for these purposes give the new modes of distribution the patina of objective science and make them harder to dispute[6]?

These worries are well grounded. Algorithmic bias is already having an impact on the real world: police force allocation algorithms target neighborhoods based on historical arrest data and set in motion self-fulfilling prophecies about who is likely to commit crimes. Facial recognition technologies proved better at recognizing the faces of white men than any other demographic. Loan repayment predictors disadvantage applicants from economically depressed neighborhoods, and so on. But it's worth pointing out the conditional nature of such concerns. *If* the training data and models are discriminatory *then* the AI predictions based on them will be also be discriminatory and they will perpetuate a history of injustice. The trouble, then, is not with the technology, per se, but with the people who write the code or provide the training data. The possibility that AI could make fair predictions, indeed fairer and less biased than those made by humans is left open, provided we solve the "garbage in" problem. Cleaned up data and more rigorous models would go a long way towards eliminating algorithmic bias. And algorithmic decision

[4] For good discussions of AGI and Superintelligence see chapters 8,9,12 of Muller,V.ed. 2015. Risks of Artificial Intelligence (CRC Press), Bostrom, N. 2016. Superintelligence (Oxford), Tegmark, M. 2018. Life 3.0 (Vintage) and Gunkel, D. 2018. Robot Rights (MIT).

[5] See for example Dylan Love, "Stephen Hawking is Worried About Artificial Intelligence Wiping Out Humanity", in Business Insider, May 25, 2014 (https://www.businessinsider.com/stephen-hawking-on-artificial-intelligence-2014-5) and Camila Domonoske, "Elon Musk Warns Governors: Artificial Intelligence Poses 'Existential Risk'", in The Two-Way: Breaking News from NPR, July 17, 2017 (https://www.npr.org/sections/thetwo-way/2017/07/17/537686649/elon-musk-warns-governors-artificial-intelligence-poses-existential-risk)

[6] For the best accounts of these worries see O'Neil and Eubanks, Supra note 3.

making may have advantages over its human counterpart. Consider this: if you lived in a neighborhood targeted regularly by the police, would you prefer that decisions about search warrants or arrests be issued by an algorithm or by a human official? It's far from clear where one's interests would lie in such a scenario; an algorithm's biases are, at least in principle, incidental and fixable. Political and commercial pressures will often result in such corrections. Microsoft, for example, was quick to address racial bias in its facial recognition software[7]. A prejudiced person's biases are both easier to deny and more difficult to correct.

Artificial intelligence has not turned into the nightmare scenario that entrepreneurs like Musk and scientists like Hawking have been warning us about. Systems rivaling human intelligence are far off - so far off that the industry has not yet been able to create an AI that can competently cross a busy room, let alone achieve world domination. We do not have technology capable of unified, flexible judgment and we may never have it. But we do have systems that are increasingly good, and sometimes outstanding at making predictions (and judgments based on those predictions) in well delineated contexts – from figuring out exactly what our taste in movies is to figuring out who is a good candidate for a mortgage. Such predictions are, of course, put to various political and commercial uses (by displaying political ads, by recommending products we may want to purchase, or by introducing efficiencies into the operation of companies). The machines are not *making* practical judgments but they are certainly *replacing* the need for us to use our own judgments in a variety of areas. Decisions that were, until recently, made by middle management human executives or officials (who to hire, who to give credit to, where to send police forces, is that anomaly on the MRI a tumor) are now being made algorithmically. Sometimes this process of automation frees up humans to make higher level decisions: the radiologist who is not reading as many scans can decide which hospital and which patients need them most. But the economic logic of AI is to save on the costs of labor; the idea is, to use Marx's terminology, to eventually replace the means of production. For every radiologist making more strategic decisions, a few radiologists will lose their jobs. Statistical Machine Learning is gradually and often usefully encroaching on our need to make everyday judgements[8]. But if Aristotle was right that we only get better at making practical

[7]"Microsoft Improves Biased Facial Recognition Technology" Fortune, 6/27/2018 available athttps://fortune.com/2018/06/27/microsoft-biased-facial-recognition/

[8] By many predictions these developments will yield a 30% contraction in labor markets in the coming years. Job loss rates range between 14%-47% percent in the next few decades. For a useful overview of key studies see: https://www.brookings.edu/blog/techtank/2018/04/18/will-robots-and-ai-take-your-job-the-economic-and-political-consequences-of-automation/ The mean job loss rate reported is 38%. The World Economic Forum recently released a report that predicted that from the 1.37 million workers who will lose their jobs to automation in the next decade, only a quarter can benefit from programs that will teach them new skills. Three quarters will likely become unemployed. The report is available here: http://www3.weforum.org/docs/WEF_Towards_a_Reskilling_Revolution.pdf

judgments by making them, we are going to lose muscle tone in this area. As we make fewer practical everyday judgments at work, our capacity to make them will diminish[9].

PART II: AI AND THE GROUNDS FOR HUMAN RIGHTS

What does the advent of such predictive technology mean for our understanding of human rights? Why would the rise of movie recommendation engines, algorithmic hiring, or AI- based sentencing recommendations impact the status of our human rights? One reason, which I have touched on already, is that the technology might directly violate rights to due process or equality before the law. If failures of facial recognition technology lead to false arrests, if algorithmic credit decisions systematically disadvantage minorities, and if AI sentencing guidelines regularly disadvantage certain groups, then the technology contributes to and facilitates human rights violations.

In fact, these abuses are already taking place. But they are not our greatest concern. As I have noted, these problems, while very real, are (at least in principle) fixable. Just like AI powered cars will ultimately reduce the number of traffic accidents, and just like AI weaponry has the potential to economize on the collateral damage of war, it is quite possible that AI based judgment making will ultimately become more equitable than its human counterpart. Algorithms can be quality controlled. They are already being quality controlled based on commercial considerations and there is nothing to say that they can't be controlled on social grounds. Indeed, let us assume, with careful optimism about the future and for the sake of crystalizing the argument, that algorithmic predictions and judgments will eventually display less rather than more bias than human judgments. Does the worry about AI's impact on human rights then dissipate?

It does not. The remaining concerns are not with how the technology violates human rights but with its impact on the grounding of human rights, with the philosophical and political justifications we can give for extending rights in the first place. As Corradetti shows, the project of grounding rights has a long history. Human rights make significant demands on citizens and governments (an effective rights regime will impose duties on citizens and require governments and international institutions to spend resources on enforcement) and such demands need justification for both normative and pragmatic reasons. Let's take a look at some key philosophical attempts to ground and justify human rights. I don't intend what follows as a comprehensive history but rather as an overview that will suggest how predictive AI challenges and undermines some of the main assumptions

[9] For a full account of this deskilling argument see Eisikovits, N. and Feldman, D. (forthcoming) supra note 2.

anchoring our views of human rights. Kant's account of personhood provides us with one classical grounding for human rights. Since rational beings can self-legislate moral directives and follow categorical imperatives, since they alone can act against inclinations and narrow perceptions of self-interest, they are free and must be afforded respect rather than merely "conditioned value[10]." "Everything in nature works according to Law," Kant writes famously in the *Groundwork for the Metaphysics of Morals*. "Only a rational being has the power to act according to his conception of laws...[11]" Because they are free to follow reason's categorical commands, rational beings must be understood as "ends in themselves". It is this potential split between inclination and will, this freedom to follow (or fail to follow) reason's commands that grounds the ideas of human dignity and personhood on a Kantian view, as well as the rights needed to protect it. The rise of statistical machine learning presents some straightforward philosophical challenges to the Kantian grounding of human rights. At the most fundamental level, there is the question of whether Kantian freedom (to go against inclinations and follow categorical imperatives) is consistent with an increasingly powerful predictive technology. This is the old philosophical puzzle about the consistency of free will and predetermination updated to the age of machine learning. Put a bit more subtly, how plausible and convincing it is to maintain our view of persons as having infinite rather than conditional value based on their capacity to make choices, when their choices are regularly predicted, aggregated, packaged and exploited for commercial purposes? The credibility and, ultimately, the majesty of free and open agency which any Kantian picture of human rights rests on is called into question and destabilized by a predictive technology that collects and parses myriad data about us and successfully replaces our professional decision making, predicts what we like, what we are interested in and what we are about to do. Further, the technology increasingly not only predicts what we would prefer but also, based on historical data, is capable of moving us towards some choices rather than others[12]. Commercial companies as well as political actors are increasingly engaging in AI based analysis of our preferences, sympathies and psychological dispositions so as to nudge us to act according to their interests. This latter development is perhaps the most explicit challenge that AI poses for the Kantian conception of agency. Corradetti's genealogy of the idea of dignity as a ground for human rights goes beyond the Kantian account of the term. In chapter 3 Corradetti highlights the tension between status-based theories of dignity and accounts that turn on actual moral capacities. The former hold that dignity is inherent to the class of being

[10] Kant, I. 1993. Grounding for the Metaphysics of Morals. (Hackett) Section 2, p.35

[11] Ibid, p. 23

[12] In section 7 of The Age of Surveillance Capitalism Zuboff (supra, note 3) describes Facebook's contagion experiment which was meant to alter the political behavior of users by showing them various forms of content. Facebook data scientists and their co-authors report the results of the experiment here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3834737/

human (De Mirandola's *Oration on the Dignity of Man* or Jefferson's *Declaration of Independence*, with its emphasis on "self-evident rights," are examples). Moral capacity- based accounts are predicated on our ability to be free in the Kantian sense (that is to follow the rational requirements of morality even against instinct and inclination); such dignity is awarded if and to the extent that we realize this capacity. It's important to clarify that the two views differ in emphasis but do not conflict. In fact, they are mutually reinforcing. Part of why it makes sense to say that some rights inhere in us, or that we self-evidently possess them, is because we have capacities worthy of such recognition. And the motivation to cultivate these capacities, the sense of legitimacy and indeed urgency in developing these moral abilities, is buttressed by the status view. It is unsurprising, therefore, to find "capacity" language in theorists focused on status and vice versa. Thus, for example, Mirandola, a "status theorist" if we are to adhere to Corradetti's distinction, can also be found describing humans along these "moral capacity" lines: ``We have given you, O Adam, no visage proper to yourself, nor endowment properly your own, in order that whatever place, whatever form, whatever gifts you may, with premeditation, select, these same you may have and possess through your own judgment and decision...We have made you a creature neither of heaven nor of earth, neither mortal nor immortal, in order that you may, as the free and proud shaper of your own being, fashion yourself in the form you may prefer. It will be in your power to descend to the lower, brutish forms of life; you will be able, through your own decision, to rise again to the superior orders whose life is divine[13]."

Be that as it may, AI challenges both status and capacity conceptions of dignity: it pushes against the capacity view insofar as it curtails the exercise of the capacity to make judgements and choices, as I have just argued. But as such capacities are reduced, a diminution of status may follow. To choose a telling caricature, anyone who watched the engorged, bored humans in WALL-E, Disney's dystopia about a world in which humans have outsourced most of their work to machines and have stopped making substantial decisions, would be hard pressed to see them as majestic creatures worthy of specials status.

On the other side of the philosophical spectrum, John Stuart Mill, in his famous account of what makes a life valuable to the person living it in *On Liberty*, argues for the freedom to conduct experiments (including failed experiments) with one's life: "as it is useful that while mankind are imperfect there should be different opinions, so it is that there should be different experiments of living; that free scope should be given to varieties of character, short of injury to others... it is desirable in short that in things which do not primarily concern others individuality should assert

[13] Mirandola, G.P. D (2016), Oration on the Dignity of Man, Borghesi F., Papio M., Riva M. (eds.), (Cambridge).

itself.[14]" The key to human flourishing for Mill is being allowed the space for one's individuality and "spontaneity" to assert themselves – "free development of individuality," as he puts it, "is one of the leading essentials of well-being.[15]" Mill complains about the smothering weight of tradition on the development of our personality: "spontaneity is looked on with jealousy, as a troublesome and perhaps rebellious obstruction...[16]" He reminds us that we need protection from such conservative forces if we are to feel that we are actually living *our* lives: "it is the proper privilege and proper condition of a human being arrived at the maturity of his faculties to use and interpret experience in his own way. It is for him to find out what part of recorded experience is properly applicable to his own circumstances and character[17]..."

To arrive at our preferences and make our decisions guided primarily by what others have preferred and decided in the past risks alienating people from their experience – risks making them lose ownership of their lives. "The human faculties of perception, judgment, discriminative feeling, mental activity and even moral preference are exercised only in making a choice," Mill writes. "He who does anything because it is the custom makes no choice... the mental and moral, like the muscular powers, are improved only by being used[18]."

For Mill, like for Aristotle, it is the exercise of choice and judgment that lead us gradually to develop the capacity for making choices and judgments and to feel that those choices and judgment are ours. Now AI raises a variety of difficulties if this is our conception of what makes a life valuable: not only is it a choice-replacing technology, which weakens the human "choice-making" muscle, but the choices and recommendations the technology makes are fundamentally *conservative*. If the data set guiding them is going to be helpful, it is going to draw on as many as possible previous choices made by others in similar situations as well as previous preferences by the agent herself. Such recommendations will likely yield satisfying or at least non-offensive results, on average, but they are emptied of the quality of spontaneity. Finally, the interaction with predictive technology – especially with successful predictive technology that is good at anticipating the books you like, the movies you enjoy, and the professional decisions you would take - cannot fail to make us feel that we are predictable (because we are). And yet, it is the sense of being unpredictable – whether grounded in reality or not - that is essential, according to Mill, for us to feel committed to our lives and to occasionally perform groundbreaking, unpredictable feats. To summarize, Statistical Machine Learning technology is increasingly recommending choices and experiences to us. That

[14] Mill, J.S. 1998 On Liberty and Other Essays (Oxford) p. 63
[15] Ibid
[16] Ibid, p.65
[17] Ibid
[18] Ibid

means that we are increasingly not making our own (often wrong and outrageous) choices and don't stumble into (often absurd and pointless) experiences. That is perhaps an efficient and pleasant state of affairs. But with the efficiency and pleasantness comes a loss of spontaneity and a degradation of the sense that one owns her experience. On a Millian account, our basic procedural rights, the basic defenses we have against encroachment, are justified by the need to protect and maintain such spontaneity – the freedom to live our own lives in our own way. But it is exactly this freedom that algorithmic recommendation technology erodes.

An important and more recent attempt to ground human rights which Corradetti takes up is the one offered by Alan Gewirth. Gewirth begins with the assumption that human action is typically both voluntary and purposive. For action to have this character, citizens need the state to guarantee their freedom and well-being. Freedom is understood as the lack of impediments and well-being "consists in having the various substantive conditions and abilities, ranging from life and physical integrity to self-esteem and education, that are required if a person is to act either at all or with general chances of success in achieving the purposes for which he acts.[19]" An agent must be able to choose to act and must enjoy certain background conditions if the action has a chance of being effective. In other words, from the observed fact of purposive agency flows a commitment to certain basic entitlements or rights – namely the negative and positive freedoms that underpin that kind of agency. Corradetti points out that such purposive agency might be too contingent a basis for grounding rights: "if rights are argued this way, then different degrees of allocation of rights must be provided to agents in accordance to the amount of their factual capacity of being purposive...[20]" The increasing prevalence of AI in our everyday lives highlights one surprising explanation for such contingency. An agent's purposive potential may be curtailed due to reasons of cognitive capacity, or mental health. But it may also be curtailed by technology that limits the contexts in which agents can develop purposes and make choices. As we have noted, algorithms are replacing human decisions in contexts as diverse as loan approval, credit worthiness and police force allocation. AI is emerging as an array of (narrowly focused) decision replacement technologies. If this results in people having fewer opportunities to practice their practical judgments or make decisions, the technology signals the retreat and narrowing of the very idea of purposive agency. My claim is, of course, predicated on the fact that we now practice much of our purposive agency at work. It may well be that as we work less we will shift to deploying purposive agency as part of our leisure activities (although recommendation engines are already replacing everyday judgments in that context as well). Whether or not our purposive agency will simply migrate from work to

[19] Gewirth, A. 1984. "The Epistemology of Human Rights" Social Philosophy and Policy 1:2

[20] Corradetti, 2022, p.104

leisure without much loss is an open question. At the very least we can propose the following conditional: if purposive agency is central to being human and is a key way to ground human rights, and if AI is beginning to narrow the zone in which we make choices and project purposes, then the technology raises serious questions about our understanding of valuable human activity and about why some basic human rights are needed. Purposive agency might migrate to other areas of our lives; we might find other ways, except purposive agency, of accounting for a meaningful life and grounding rights. But as long as the jury is out on these replacements, AI calls this important justification for human rights into question as well.

PART III. AI, HUMAN RIGHTS AND THE FORMIDABLE STATE

There is yet another way in which the rise of AI influences the human rights agenda. This impact is not direct (as are violations of privacy and due process rights involved in the application of facial recognition or algorithmic loan approvals). This impact is also not immediately corrosive to the grounds of human rights as discussed in the previous section. The influence at issue is indirect and concerns the way in which AI facilitates the operation of government.

AI has the potential to make the operation of government far more efficient than what we are used to. The Chinese use of AI provides examples that illustrate both the peril and promise of technologically mediated governance. In the peril column one thinks of China's Social Credit system – a relatively new form of social control which combines wide spread facial recognition cameras, geo location technologies, social media tracking, and centralized payment technologies to create a vast network of state surveillance and influence of the citizenry. The social credit system efficiently aggregates citizens' activities across different areas of their lives and determines how well they conform and are likely to conform to various objectives and values set by the Chinese Communist Party. Some scholars have described the social credit system as a powerful way to graft the consciousness of the masses onto that of the party – an old authoritarian dream technologically realized[21]. Chinas' sophisticated facial recognition apparatus as well as other aspects of the Social Credit system have also been put to ominous use in tracking and suppressing the Uighur minority in Xinjiang province.

On the other hand, the same technological apparatus, which allows the Chinese to deploy a highly potent form of social control, has also yielded real public health benefits for the country. Consider, for example, China's relative success in subduing its Covid 19 outbreak. That success has been facilitated by the widespread use of

[21] Singer, P. and Brooking, E.2018. Like War: The Weaponization of Social Media (Houghton Mifflin) Chapter 4.

geolocation, facial recognition, temperature scanning and streamlined payment and identification systems. These (largely AI) technologies make it simple to locate where an infected person has been, who she has been close to, and where all of those people have been in a given time frame. This information has been used for effective and targeted contact tracing efforts and has been very useful in quickly cutting short infection chains.

The upshot of the Chinese example is that AI has the potential to usher in what we may call the *formidable state*: a governmental apparatus that challenges the stereotype of bloated, inefficient, bumbling bureaucracy and that functions with impressive and sometimes frightening efficiency. How is this development related to our understanding of human rights? Much of what human rights are meant to guarantee is a defense from state power. And, on the face of things, if we need human rights to protect ourselves from the bumbling state, we need them all the more to protect ourselves from the formidable one. AI governance helps the state apply power to us more efficiently and seamlessly, which gives it more power over us, which means we need more robust human rights.

The problem is that it is easier to argue for limiting the bumbling state than for limiting the formidable state. Said differently, curtailing the bumbling state puts the breaks on something with limited potential. Curtailing the formidable state puts the brakes not only on abuse but on a level of benefit and protection that we are historically not used to receiving from the state. The theory of human rights as shields from state abuse was developed in the shadow of an inefficient state causing as much harm by its incompetence as by its malice (and often through the combination of both). But it may become harder to maintain our constitutional protections when what is at issue is a state that is not only better at oppressing us but also better, perhaps even incomparably better, at orchestrating important public health and public safety interventions that could save many thousands of lives.

I want to suggest that if AI governance fulfils some of its promise, if it ushers in anything close to the "formidable state," the political cost-benefit calculations involved in the extension and defense of human rights may well change. This is a descriptive rather than a normative statement. A prediction rather than a wish. If AI can successfully foretell a coming public health crisis on the basis of scanning our google searches or finding an uptick in anti-histamine purchases; if AI governance can deliver absolute success in contact tracing by means of wide spread surveillance; if AI governance can efficiently coordinate between national security agencies and help prevent the miscommunications and blind spots that allowed the Sep 11th attacks to happen in the United States, it is going to be harder and harder – even for democratic polities – to make a case for rights that curtail the technology.

One could, of course, object that what allows the Chinese to technologically oppress their people and efficiently control the spread of Covid is the lack of western constitutional guarantees and traditions (and the rights they typically

involve). That is certainly true, but my point is that the causality may also come to work in the other direction. How we think of rights and how much importance we attach to them may change with our view of government and what it can do. Part of why we have so many due process rights is because government gets things wrong as often as it does. If that track record improves, as it very well may with the help of AI, and the rights serve not only to protect us from potential mistakes but also to limit the operation of a vastly improved machinery – will the original protections (and the traditions and institutions underpinning them) survive?

Consider the rather typical dynamic of public opinion in times of crisis in the west – in the wake of national security or public health disasters, the public is often sympathetic to the curtailment of political rights (recall the initial support expressed by Americans for the Patriot Act, or, more recently, the initial acceptance by many Israelis, early during the Covid outbreak, of cell phone surveillance for contact-tracing purposes). The initial public support usually levels off and declines as people both get used to the crisis and government fails to deliver the benefits that were promised as rewards for restricted freedoms. But what if the rewards were actually and regularly delivered? Wouldn't the public indifference about political rights deepen and become more entrenched[22]?

IV CONCLUSION: TECHNOLOGY AND OUR SELF-UNDERSTANDING

Philosophers from Rousseau to Heidegger to Carl Schmitt have argued that technology is never a neutral tool for achieving human ends. Technological innovations reshape or, if we take Rousseau at his word, misshape us as we use them to control our environment. Our technological capacities and inventions don't just make life easier; they shift our perspective and provide a new framework for seeing the world. As Heidegger argues in "The Questions Concerning Technology," with the advent of more and more tools we learn to see the natural world around us as a reservoir of resource to make use of – as potentially useful to us[23]. Once you can chop wood the naïve view of the forest is layered or replaced by a tendency to see trees as a source of heat. Once you can build a dam the river is experienced not primarily as an awesome object of beauty or a formidable natural barrier (indeed the word awesome has lost its original meaning denoting fear and wonder) but as a source of hydro electrical energy. Or, to shift to Schmitt's language[24], with the

[22] For some recent evidence of such trends see: Ziller, C. and Helbling, M. 2020. "Public Support for State Surveillance" Available at SSRN: https://ssrn.com/abstract=3556953 or http://dx.doi.org/10.2139/ssrn.3556953

[23] Heidegger, M. 1977. "The Question Concerning Technology" In The Question Concerning Technology and Other Essays (Garland).

[24] Schmitt, C.1929. "The Age of Neutralizations and Depoliticizations" (Trans. M. Konzett and J. P. McCormick). Telos, No. 96, 130-142.

development of ever more advanced tools we become beholden to an almost magical "technicity" – a mode of understanding in which all problems have solutions and experience is described in increasingly neutral terms such that it may be carefully quantified, calibrated and ultimately controlled.

Put differently, the history of our use of technology has always been a history of co-evolution: the tools we master shape how we think and talk about ourselves. The rise of Newtonian mechanics famously made us view ourselves as intricate clocks, the rapid development of computing power after World War II led to us think about the mind as "hardwired" for various functions, and then in computational terms. These shifts of metaphor[25] are, of course, evidence that our self-conception is contingent. How we understand ourselves and the language we use to conceive of our bodies are in constant flux. How will the rise of AI impact these dynamics? The argument of this paper has been that powerful predictive technologies will eventually challenge our self-conception as agentic beings defined by free, spontaneous choices. The advent of efficient, fair machine learning technology capable of offering us good suggestions and replacing an array of every day judgments will slowly dislodge our self-understanding as judgment making creatures who value stumbling into their own tastes and preferences. And to the extent that this (traditional) self-understanding underpins and justified our idea of human rights, that idea may become destabilized as well.

This does not have to happen. Perhaps the automation of everyday judgments will free us up to make more rarefied ones about our hobbies and artistic inclinations. Perhaps organic serendipity and "different experiments in living" will be replaced by randomly generated algorithmic correlates. Maybe we will identify other rationales for rights that have nothing to do with the Kantian and Millian foundations we have detailed. This paper is not a luddite plea to hang on to the old grounds for human rights. But it is a warning against dismissing these grounds thoughtlessly, a warning against the blithe acceptance of wholesale philosophical and psychological change as the inevitable "collateral damage" that accompanies progress.

[25] Some important studies on the relationship between technology and self-conception include Marshall, J.1977. "Minds, Machines and Metaphors." Social Studies of Science, vol. 7, no. 4, pp. 475–488; Turkle, S. 2005 The Second Self: Computers and the Human; Spirit (MIT); Kurt, D. 2008. Marking the Mind: A History of Memory (Cambridge).