# Governing Artificial Intelligence and Algorithmic Decision Making: Human Rights and Beyond

Vasiliki Koniakou

# Governing Artificial Intelligence and Algorithmic Decision Making: Human Rights and Beyond

Vasiliki Koniakou

Doctoral Candidate of Law, University of Turku, Faculty of Law, Caloniankuja 3, FI-20014 Turku, Finland – Research Fellow at ELTRUN Research Center and ISTLab, Department of Management Science & Technology, Athens University of Economics and Business, Evelpidon 47A, 11362, Athens, Greece
vaskon@utu.fi, koniakou@eltrun.gr
(https://orcid.org/0000-0001-7316-7723)

**Abstract.** In the context of Artificial Intelligence Ethics, human rights have been commonly invoked as a promising basis for an ethical framework. They have been also promoted as guidelines for Artificial Intelligence and Automatic Decision-making governance, or as engineering principles that may be turned into design requirements. Since literature so far engages only partially with the relevance and suitability of the extension of human rights in the realm of proprietary algorithms and privately owned Artificial Intelligence systems, this paper offers the necessary background and justification, building upon international human rights law theory and the concept of radiance of human rights. It aims to contribute to the scholarship promoting the human rights not only as ethical values but also as governance principles for Artificial Intelligence and algorithms. It also stresses the significance of concretizing and implementing the values of transparency, accountability, and explicability. Moreover, it suggests that for the ethically sound and societally beneficial employment of Artificial Intelligence and algorithms, useful insights may be derived from the field of technology governance. Stemming from that, it emphasizes the necessity to embrace the role of designers, and the need of conscious democratic control.

**Keywords:** Algorithms, Algorithmic decision-making (ADM), Artificial Intelligence (AI), Human Rights, AI Ethics, AI governance, Science and Technology Studies (STS), Technology Theory

## 1 Introduction

The last two decades we have witnessed impressive advances in Artificial Intelligence (AI) and algorithmic decision-making (ADM). AIs and various forms of ADM are increasingly employed, permeating several aspects of contemporary society[1].Enabling data-driven, automatic decision-making, they have rapidly become integral for numerous sectors and industries, ranging from healthcare, taxation and policy-making to pricing, products, processes, and services innovation[2], [3]. Additionally, bearing the promise of effective and efficient decision-making, they are considered as providing

new opportunities for people, and the society at large, to improve and augment their capabilities and wellbeing[4], [5].They are also expected to contribute to global productivity[6], [7],the achievement of sustainable development goals[7], [8], and broader environmental objectives[9], [10].

However, instances of discrimination and bias[11]–[13], disinformation and opinion manipulation [14, Ch. 4], [15], private censorship [16], [17] pervasive monitoring or surveillance [18]–[20], as well as adverse job market effects [6], [21] have raised serious concerns, attracting attention to the negative implications of automated 'intelligent' processes. Moreover, as ADM and AIs are increasingly implemented to inform critical decisions in the legal system, or to define individuals' eligibility or entitlement to critical opportunities and/or benefits, it is apparent that they relate and may also interfere with human rights [2], [22]–[25]. Hence, whereas the significance and impactful role of algorithms and AIs is not in question, whether and to what extent their impact will be positive or negative is hotly contested.[4] Furthermore, as reliance on AIs and ADMs deepens, the ethical questions and concerns amplify. Numerous researchers have engaged with the ethical aspects of AIs and algorithms, seeking to offer insights and create a road map towards their socially beneficial and ethically sound employment [3], [26]–[30].

AI Ethics is a broad interdisciplinary field of research, reflecting a wide range of value-based and societal concerns related to AI applications and the extensive employment of algorithms.[22, p. 78]. In this discourse, human rights have been invoked from various aspects. They have been proposed by scholars, policy-makers and civil society organizations as offering a promising set of ethical standards for AIs [2], [22, Ch. 4]. Researchers have also suggested ways of translating human rights into design requirements through various methodologies[3]. Additionally, they have been suggested as governance principles for AIs, to *"underlie, guide and fortify"* an AIs governance model[31]. The distinction between human rights as ethical standards, and human rights as governance principles and legal requirements is a meaningful one, particularly regarding the binding effects of legal requirements and the actual reach human rights may have in each case. It is also particularly relevant as a significant portion of algorithms and AIs are proprietary, privately designed, owned, and operated, whereas human rights as legal obligations are in principle vertical in nature.[32]

So far, the literature has not addressed the question whether human rights are more relevant and appropriate as ethical standards, as formal obligations and governance principles for AIs and ADM, or both. Additionally, it has only partially engaged with the suitability of extending the application of human rights to the private sphere. The argumentation is mostly premised on their relevance for AIs and ADMs as ethical principles, and the effects AIs and algorithms may have on human rights[22], [31]. Moreover, the discussion regarding the steering of new and disruptive technologies towards ethical and societally beneficial ends is hardly new nor unique to AIs and ADM. On the contrary, it is part of a broader discourse on the relationship between technology and society, centered around human-centric design and the necessity to humanize technology governance and to allow the development, employment and governance of technologies in a socially beneficial way[33]–[35]. In that context, human rights are an

essential part of a broader strategy towards technology governance that involves additional elements, values, and principles.

Stemming from these observations, this article wishes to offer additional argumentation on the relevance and suitability of human rights as governance principles for AIs and ADM based on international human rights theory. It argues that human rights, apart from ethical standards, ought to be applied as guiding governance principles, and their respect in terms of AIs and algorithms should be legally required. Furthermore, it stresses the significance of concretizing and implementing the values of transparency, accountability, and explicability and argues for the need to examine AI governance within the broader context of technology governance, under the light of Technology Theory and Science and Technology Studies (STS). More specifically, it emphasizes the necessity to embrace the decisive and ethically important role of designers and invest in ethics education, as well as the need of conscious democratic governance of AIs.

## 2 Human Rights, AIs and ADM

### 2.1 Human rights as guidelines for AI Ethics

Human rights constitute a rare set of values recognized internationally by the majority of societies[36, pp. 53–54], [37]. Even though they are not uncontroversial[38], nor universally applied[39], they represent a sum of principles and norms that are widely shared and institutionalized globally. Serving as the basic moral entitlements of every human being, they are deeply rooted in contemporary politics and law, recognized in political practice and legal institutions globally[40, pp. 2–3]. Hence, human rights, both in their strictly legal sense, and as norms encapsulating and reflecting moral and social values, are considerably comprehensive and widespread. Furthermore, the international human rights system includes a well-established institutional framework comprised by dedicated monitoring bodies and agencies, as well as conflict and tensions resolution mechanisms. It also involves a rich theoretical background and ample discursive tools aimed to protect and promote human rights, as well as monitor, and ensure compliance with human rights principles globally.

Contrary to human rights, currently in AI Ethics there is no commonly agreed upon set of ethical standards that may serve as governance principles[22, p. 80], [41]. The industry-driven self-governance model is largely premised on a variety of voluntarily adopted codes and self-commitments. Such codes are usually rather abstract and largely vague, while they often lack the necessary mechanisms and frameworks to ensure the enforcement of the norms and handle disputes, conflict and tensions[42]. Thus, the lack of binding effects, and their questionable enforceability combined with the absence of conflict resolution mechanisms hamper their effectiveness and normative function. Additionally, such self-commitments may be in fact proclamatory, invoked for ‘ethics washing’[22, p. 84], [43], [44] or simply to avoid direct regulatory interference, in the form of binding legislation[42], [45] . Therefore, considering the ineffectiveness of self-commitments, literature suggests that human rights offer a substantially better

alternative, promoting human rights as a more rich and elaborated set of principles that can serve as ethical standards for AIs and ADM.

### 2.2 Human rights as principles for AI and ADM governance and the extension of human rights to the private realm

Going a step further, some scholars promote human rights not merely as ethical guidelines, but as governance principles, suggesting a governance approach anchored in human rights [22, p. 85]. Essentially, they argue that instead of simply internalizing human rights in AI Ethics, human rights-premised obligations should be turned into concrete legal requirements in the field of AIs and ADM, and human rights should inform and shape AI and ADM governance [22, Ch. 4]. From a similar point of view, the High-Level Expert Group on Artificial Intelligence (AI HLEG) that the EU Commission tasked to offer input for the development and deployment of AI, stressed that *"respect for fundamental rights, […], provides the most promising foundations for identifying abstract ethical principles and values"* [46]. In its recommendations, human rights are identified as the foundational principles for a normative framework that may safeguard the development and deployment of AIs in a societally beneficial way. This suggestion progressively gains momentum in literature and policy discourse for various reasons, mainly related to the merits of the international human rights system and its potentials to prevent socially harmful uses of technology.

The rights enshrined in the Universal Declaration of Human Rights (UDHR), the Charter of Fundamental Rights of the European Union (EU Charter), as well as in the European Convention on Human Rights (ECHR) are arguably the most broadly embraced set of values and ethical principles, closely related to rule of law and the democratic polity. Hence, human rights, both in their strictly legal sense, and as norms encapsulating and reflecting moral and social values, are considerably comprehensive and widespread. Instead of fragmentary, abstract, or conflicting principles, premised on various views or aspirations, human rights can serve as common framework to address the majority of not only ethical but also normative concerns related to AIs and algorithmic decision-making. Simultaneously, the international human rights law system can provide guidance also in terms of the procedural aspects, offering a solid and tested tension and dispute resolution mechanisms and the necessary theoretical and discursive tools [22, Ch. 4], [31].

Finally as AIs, ADM and algorithms increasingly define opportunities and risks[47], having an often mediating role regarding human rights, the international human rights law system is not only suitable but also highly relevant. As AI applications become ubiquitous and pervasive, routinely relied upon to carry a wide range of tasks, they increasingly affect a wide variety of human rights, from freedom of expression and privacy to access to health care. From this angle, the extension of human rights to the private realm, in the form of concrete principles and specific obligations, and their integration to AI and ADM governance mechanisms is critical *"to maintain the character of our political communities as constitutional democratic orders."*[22, p. 81] Thus, human right should serve both as ethical guidelines, and as governance principles, while

they should be extended to regulate the development and deployment of AIs and ADM, informing and shaping their processes and procedures also from the design and technology-in-the-making point of view[3]. Nevertheless, a significant portion of algorithms and AIs are privately designed, owned, and operated, while human rights are in principle vertical in nature, [32] which makes their extension to the private sphere far from self-evident or uncontroversial.

### 2.3 The challenge of applying vertical rights in the private realm

The turn to human rights as a source of governance principles, and the necessity of extending human rights obligations to the private realm to achieve socially important ends are not new within the technology governance discourse[37], [48]–[50]. However, according to international human rights law it is the states and not private entities that are bound by them[40]. This means that while the application of human rights as ethical guidelines is largely unproblematic, their adoption as governance principles, as well as the extension of human rights-premised obligations to private actors, such as the owners and operators of AI systems and proprietary algorithms, is relatively challenging. More specifically, the suitability and appropriateness of the extension of human rights-related obligations to the private realm is a hotly contested topic. Practically, the scope and application of international human rights law in the private sphere constitutes one of the most topical issues in constitutional law and human rights discourse[51].

The horizontal application of human rights, namely the extension of human rights to relationships otherwise regulated under private law, is challenging both theoretically and practically. Enforcing the same duties as public bodies to private actors could affect the very core of private law and liberal autonomy having adverse effects for both private law and international human rights law. Nevertheless, the vertical nature of human rights is premised upon the *"far greater imbalance of power between the state and individuals,"*[52, p. 16] which is rapidly challenged in the context of the modern society, in which individuals' rights commonly depend on private entities' actions and decisions, business and revenue models, corporate policies and rules. The indubitable power of private actors to negatively affect human rights brings to the forefront the change in the global balance of power between state and non-state actors. It also highlights the distance between the human rights doctrine and the reality of several almost omnipotent non-state actors in contemporary society.[53, p. 192] Thus, there is an ever-growing volume of literature exploring the ways to protect human rights from non-state actors, through the extension of human rights to private relationships, allowing them to have *horizontal effects.* [54] In that context, the question over the so-called horizontal application of human rights is of considerable practical importance and political relevance.[52, p. 3]

### 2.4 Horizontality, the radiance of human rights and the human rights gap

As it is progressively becoming apparent that individuals' rights and freedoms as well as a wide array of societal and constitutional principles are threatened or restrained more frequently or severely in terms of private relationships[52, p. 20], the discussion

regarding the positive duties of private actors comes to the forefront. The sharp distinction between the public and private spheres seems increasingly obsolete[55]. Moreover, as we are rapidly moving from technologies of pervasive effects towards technologies that are themselves pervasive, or as Susan Brenner puts it, from '*dumb*' to '*smart*' technologies[56], the majority of which are privately owned, designed and/or operated, the role and placement of human rights is a critical discussion.

Looking beyond the question of vertical nature or horizontal effects, stemming from the German constitutional tradition, and the concept of *"Drittwirkung"*[1] we may perceive human rights as "radiating" over the legal order, serving as a *"fundamental and objective system of values, which provides a blueprint for society as a whole."*[52, p. 129] From that angle, they have both an interpretative and guiding effect towards private law, without necessarily applying directly. They may inform the work of legislators and the decisions of judiciary, forming an indivisible, interdependent and interrelated whole which unites the legal order. This view, acknowledging that human rights and private law *"no longer exist in isolation from each other"*[57] is valuable for approaching and framing the role of human rights in the context of AIs and ADM governance. It allows them to be employed in AIs and ADM governance, enabling the extension of human rights-premised duties to private actors, as well as the employment of human rights and human rights due-diligence as a basis of assessment for private policies and governing structures in the field of AIs. Simultaneously, it does not absolve the states from their positive obligations to protect human rights, nor allows them to outsource this duty to private actors. Furthermore, the extension of human rights to the private sphere via the concept of radiance allows us to interpret the existing framework under the light of human rights, offering a much-needed time window to prepare the rules without the risk of a normative vacuum.

Finally, such an extension is also necessary to prevent a *"human rights gap"* in the governance of AIs and ADM. Given the increasingly relevant role of AIs and algorithms for human rights, keeping human rights strictly public (in the sense that only the states are obligation holders) and not extending them to the governance of AIs and ADM may result into a *"human rights gap."* This 'gap' is essentially the void created by the fact that although human rights are impactfully affected, mediated or even governed by non-state actors, these actors and technologies remain shielded from human rights obligations, leading to a vacuum of human rights protection. However, specific technologies, particularly those that penetrate the *"lifeworld"* producing consequential impacts that shape and affect individuals' options and choices, rights, and freedoms, should not be left outside the human rights discourse and system[53, p. 71].

[1] BVerfGE 7, 198 ff of 15 January 1958.

# 3 Looking beyond human rights

## 3.1 Transparency, accountability, explicability

As mentioned in the introduction, for the ethically sound and socially beneficial development and deployment of AIs and ADM, human rights ought to be part of a larger governance strategy. In this context, values and principles derived from the self-adopted ethical codes in private sector, along with insights from various Recommendations, Declarations and Ethical Principles suggested by several organizations, think tanks and institutions [42], [58]–[60], can also have a role. Particularly the commonly recurring values of "transparency", "accountability" and "intelligibility" or "explicability", shared among most of these recommendations [4], [42], should be concretized into rules and turned into specific and viable governance guidelines. Jointly these three values are essential for the meaningful scrutiny, and integral for good governance in the field of technology. Thus, they are also particularly relevant for AIs and ADM governance, especially as they constitute new and powerful forms of smart agency.

Transparency about the input and outcomes of algorithmic decision-making criteria is crucial, given that algorithms, as forms of automatic decision-making, control or significantly influence key aspects of daily life, affecting eligibility to life-changing opportunities, defining access to goods and services [22, Ch. 4]. Simultaneously, they increasingly penetrate the judicial system and law enforcement [47], [61]. However, ADM is *"essentially concealed behind a veil of code"*[62] often protected by intellectual property rights (IPR). This means that although algorithms may reach decisions with major impact for individuals' lives, the way they reached upon these decisions and the data they acted upon is opaque to the affected individual.[44] In turn, the lack of transparency significantly obscures both explicability and accountability. AI systems and algorithms are largely presented as back boxes, too complex and difficult to be explained and/or understood. Yet the lack of explicability rises serious about due process, and the possibility of meaningful human control and scrutiny [63], [64]. Simultaneously, the question "who is responsible for the way it works?" is close to impossible to be answered if transparency is absent and no one can answer "how it works?" for reasons of allegedly complexity or IPR protection.

Nonetheless, if we are indeed entering an era of omnipresent smart agents, wherein algorithms largely determine and shape the exercise of power, affecting public policy, and human rights, we need to find meaningful ways to ensure transparency, accountability, and explicability, rejecting the black box approach and realigning private rights with public interest [62], [65]. Law and regulatory intervention have here a significant part to play. The EU has taken a number of regulatory initiatives towards this direction, most prominently through the General Data Protection Regulation (GDPR), that emphasizes the principles of transparency and accountability, while stresses the need of explainability in case of automated processes, such as automated profiling. Yet, this also entails finding new ways to balance private interests and IPRs with the requirements of transparency, accountability, and explicability, without risking the malicious exploitation of algorithmic transparency, or hampering innovation.

### 3.2 From AI and ADM governance to technology governance

These challenges are not unique for AIs and ADM governance [13]. Some of the key questions for the future of smart agents governance are inherent in the field of technology governance and have been thoroughly discussed in terms of Technology Theory[66], [67] and STS[68]. From that angle, it may be insightful to examine AIs and ADM governance under the light of Technology Theory and STS, building upon the rich literature of technology governance. In that context, a necessary first step towards establishing a governance model that will contribute to the ethically sound and socially beneficial employment of AIs and ADM would be demystifying them. Regardless their opacity and the "veil of mystery" that covers their processes, they are both human constructs, in the sense that they are designed, programmed, applied by human beings. Hence, those creating them have both considerable control over how they function[2], and the responsibility to ensure that they are employed within a sound ethical framework. However, responsibility here is not to be perceived narrowly, in terms of liability or accountability in the legal sense, but as the moral and social virtue of steering intellectual creations towards the public good.

Opening the back box and perceiving AIs and algorithms as malleable, human creations, sheds light on the dilemmas, social processes, institutions and arrangements that affect the development of technology[69, p. 568]. From that angle, the design of technologies and technological artifacts involves more than technical skills or creative insight, as the final outcome reflects also the character, views, values and ethics of the designers and developers [70], [71]. Acknowledging that engineering practice involves choice, value straggles and value-informed decisions, highlights the fact that algorithms and AI design choices are not neutral[72]. Embracing the key role of the designers[69, p. 573], and the ethically important aspects of engineering, brings to the forefront the necessity to include ethics modules and courses in Higher Education Institutions, at least in fields of engineers and computer science[73], [74]. Thus, improving access to ethics modules and stand-alone courses related to ethical considerations in design, and responsible engineering,[58], [75], [76] which still remain relatively low[41], may be a vital to steer AIs and ADM towards socially beneficial and ethically sound ends.

Similarly, it is equally important to place AI governance withing a framework of democratic scrutiny, and conscious democratic control, allowing policy and decision-making about such impactful and consequential technologies to reflect and adequately represent the views, considerations, values, fears, hopes and expectations of the citizens. Whereas in modern constitutional democracies such a request sounds self-evident or presumed already satisfied, in fact technology governance is commonly a non-democratic procedure[66], [77]. More specifically, as a reductionist way of thinking about the relationship between technology and society, technological determinism remains deeply rooted in our casual way of thinking about technologies[78]. As such, it has informed several socio-economic configurations [79], promoting non-democratic, technocratic arrangements, and preventing the conscious democratic control of technologies[77], or allowing for non-democratic practices to be accepted as inevitable[80].

From that aspect, identifying and rejecting technological determinism and its entailments from AIs and ADM governance may constitute a necessary and relatively

demanding step to ensure that their governance will not be an exception of democratic control. Considering the expanding role of AIs and ADM in contemporary society, as well as their far-ranging implications for individuals and human rights, it is of at most importance to premise their governance upon a democratic framework. To put it differently, although AIs may be privately owned, while algorithms are in their majority proprietary, their governance, how they are regulated and the larger policy framework about them should be subject to democratic steering. In turn, this is closely related with ensuring transparency, accountability, and explicability,[81], [82] as well as with rejecting technological determinism that leads to the decoupling of technology governance and democratic decision-making.

## 4 Concluding Thoughts

Algorithms and AI are not simply *"another utility that needs to be regulated once it is mature."*[4] They comprise a powerful and disruptive new form of smart agency, that bears significant promises as well as risks. This paper argued that to steer this force towards the benefit of the society it is necessary to introduce human rights not only as guidelines for AI Ethics, but also as governance principles for AIs and ADM. Whereas the literature has already argued for the need to extend human rights obligations to AI governance, it largely tends to avoid engaging with the question of horizontality. Yet, without clearly articulating the relevance and the suitability of human rights as governance principles for AI, the proposed models may seem ill-grounded from an international human rights law point of view. Building on this observation, the paper sought to offer background and justification regarding the relevance of human rights and the suitability of extending them into the private sphere building upon the theory of *Drittwirkung* and the concept of human rights' radiance. It also highlighted the risk of a *"human rights gab"* in case private actors are left to act outside the scope of human rights. Looking beyond human rights, it emphasized the need to concretize the values of transparency, accountability, and explicability and turn them to pillars of AI governance. Finally, it sought to bring to the forefront the valuable insights AI and ADM governance may derive from Technology Theory, technology governance and STS. Embracing the key role of engineers and developers it is critical to invest in their ethics education and take specific legislative and normative initiatives to address the black box approaches towards technology. Additionally, it is vital to ensure that governance of AIs will be a democratic procedure, rejecting technological determinism and exploring meaningful ways to align private interests with the public good.

## References

[1] C. Cath, "Governing artificial intelligence: Ethical, legal and technical opportunities and challenges," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2133. Royal Society Publishing, 28-Nov-2018.

[2] J. Gerards, "The fundamental rights challenges of algorithms," *Netherlands Quarterly*

*of Human Rights*, vol. 37, no. 3, pp. 205–209, Sep. 2019.

[3] E. Aizenberg and J. van den Hoven, "Designing for human rights in AI," *Big Data & Society*, vol. 7, no. 2, p. 205395172094956, Jul. 2020.

[4] L. Floridi *et al.*, "AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," vol. 28, pp. 689–707, 2018.

[5] M. Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf, 2017.

[6] A. Agrawal, J. Gans, and A. Goldfarb, "Artificial Intelligence, Automation, and Work," 2019.

[7] Vincent Pedemonte, "AI for Sustainability: An overview of AI and the SDGs to contribute to the European policy-making," 2020.

[8] R. Vinuesa *et al.*, "The role of artificial intelligence in achieving the Sustainable Development Goals," *Nature Communications*, vol. 11, no. 1. Nature Research, pp. 1–10, 01-Dec-2020.

[9] G. Elshafei and A. Negm, "AI Technologies in Green Architecture Field: Statistical Comparative Analysis," in *Procedia Engineering*, 2017, vol. 181, pp. 480–488.

[10] K. S. Mishra, Z. Polkowski, S. Borah, and R. Dash, *AI in Manufacturing and Green Technology: Methods and Applications*. Routledge, 2021.

[11] S. Murray, R. Wachter, R. C.-H. A. Blog, and U. 2020, "Discrimination By Artificial Intelligence in a Commercial Electronic Health Record—a Case Study," *Healthaffairs.Org*, 2020. [Online]. Available: https://www.healthaffairs.org/do/10.1377/hblog20200128.626576/. [Accessed: 07-Apr-2021].

[12] R. Gorwa, R. Binns, and C. Katzenbach, "Algorithmic content moderation: Technical and political challenges in the automation of platform governance," *Big Data and Society*, 2020.

[13] F. Z. Borgesius, "Discrimination, artificial intelligence, and algorithmic decision-making," 2018.

[14] J. R. Allen and G. Massolo, "AI in the Age of Cyber-Disorder | ISPI," 2020.

[15] C. Cadwalladr, "Fresh Cambridge Analytica leaks 'shows global manipulation is out of control," *The Guardian*, 2020. [Online]. Available: https://www.theguardian.com/uk-news/2020/jan/04/cambridge-analytica-data-leak-global-election-manipulation?CMP=Share_AndroidApp_Slack. [Accessed: 07-Apr-2021].

[16] T. Gillespie, "The Relevance of Algorithms," in *Media Technologies*, 2014.

[17] T. Gillespie, *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. 2018.

[18] M. Hildebrandt and B.-J. Koops, "The Challenges of Ambient Law and Legal Protection in the Profiling Era," *Modern Law Review*, vol. 73, no. 3, pp. 428–460, 2010.

[19] S. Feldstein, "The Global Expansion of AI Surveillance," 2019.

[20] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *Journal of Parallel and Distributed Computing*, vol. 74, no. 7, pp. 2561–2573, Jul. 2014.

[21] M. Vochozka, T. Kliestik, J. Kliestikova, and G. Sion, "Participating in a highly automated society: How artificial intelligence disrupts the job market," *Economics, Management, and Financial Markets*, vol. 13, no. 4, pp. 57–62, 2018.

[22] M. D. Dubber, F. Pasquale, and S. Das, *Oxford Handbook of Ethics of AI*. Oxford University Press, 2020.

[23] F. Raso, H. Hilligoss, V. Krishnamurthy, C. Bavitz, and L. Y. Kim, “Artificial Intelligence & Human Rights: Opportunities & Risks,” *SSRN Electronic Journal*, Oct. 2018.

[24] G. Buchholtz, “Artificial intelligence and legal tech: Challenges to the rule of law,” in *Regulating Artificial Intelligence*, Springer International Publishing, 2019, pp. 175–198.

[25] MSI-NET, “Algorithms and Human Rights : Study on the Human Rights Dimensions of Automated Data Processing Techniques (in particular Algorthims) and Possible Regulatory Implications,” *Council of Europe Study DGI*, 2017. [Online]. Available: https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html. [Accessed: 08-Apr-2021].

[26] A. Van Wynsberghe, “A method for integrating ethics into the design of robots,” *Industrial Robot*, 2013.

[27] S. Umbrello, “Atomically precise manufacturing and responsible innovation: A value sensitive design approach to explorative nanophilosophy,” *International Journal of Technoethics*, vol. 10, no. 2, pp. 1–21, 2019.

[28] J. Bryson and A. Winfield, “Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems,” *Computer*, vol. 50, no. 5, pp. 116–119, May 2017.

[29] A. A. Tubella, A. Theodorou, V. Dignum, and F. Dignum, “Governance by Glass-Box: Implementing Transparent Moral Bounds for AI Behaviour,” *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2019-August, pp. 5787–5793, Apr. 2019.

[30] M. Taddeo and L. Floridi, “How AI can be a force for good,” *Science*, vol. 361, no. 6404, pp. 751–752, 2018.

[31] N. A. Smuha, “Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea,” *Philosophy and Technology*, 2020.

[32] L. Lane, *The Horizontal Effect of International Human Rights Law in Practice*, vol. 5, no. 1. 2018.

[33] M. Bucchi, *Beyond technocracy: Science, politics and citizens*. 2009.

[34] J. Strobel and H. Tillberg-Webb, “Applying a Critical and Humanizing Framework of Instructional Technologies to Educational Practice,” in *Learning and Instructional Technologies for the 21st Century*, 2009.

[35] W. Benedek, M. C. Kettemann, and M. Senges, “The Humanization of Internet Governance: A Roadmap Towards a Comprehensive Global (Human) Rights Architecture for the Internet,” *SSRN Electronic Journal*, no. December, 2017.

[36] S. Walkila, *Horizontal Effect of Fundamental Rights in EU Law*. Europa Law Publishing, 2017.

[37] I. Brown, D. D. Clark, and D. Trossen, “Should specific values be embedded in the Internet architecture?,” in *Proceedings of the Re-Architecting the Internet (ReArch) Workshop, Held in Conjunction with CoNEXT 2010*, 2010.

[38] S. Hopgood, *The Endtimes of Human Rights*. Cornell University Press, 2018.

[39] S. Tharoor, “Are Human Rights Universal?,” *World Policy Journal*, vol. 16, no. 4, pp. 1–6, 2000.

[40] A. Etinson, *Human Rights*. Oxford University Press, 2018.

[41] D. Zhang *et al.*, “Artificial Intelligence Index Report 2021,” 2021.

[42] T. Hagendorff, “The Ethics of AI Ethics: An Evaluation of Guidelines,” *Minds and*

*Machines*, vol. 30, no. 1, pp. 99–120, Mar. 2020.

[43] E. Bietti, “From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy.” 01-Dec-2019.

[44] V. C. Muller, “Ethics of Artificial Intelligence and Robotics (Stanford Encyclopedia of Philosophy),” *Stanford Encyclopedia of Philosophy*, 2020. [Online]. Available: https://plato.stanford.edu/entries/ethics-ai/. [Accessed: 08-Apr-2021].

[45] B. WAGNER, “Ethics As an Escape From Regulation.,” 2019.

[46] M. Ala-Pietilä, Pekka; Bauer, WIihelm; Bergmann, Urs;Bielikova, “Ethics guidelines for trustworthy AI - Publications Office of the EU,” 2018. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1. [Accessed: 09-Apr-2021].

[47] F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2015.

[48] C. Cath and L. Floridi, “The Design of the Internet’s Architecture by the Internet Engineering Task Force (IETF) and Human Rights,” *Science and Engineering Ethics*, 2017.

[49] M. Zalnieriute and S. Milan, “Internet Architecture and Human Rights: Beyond the Human Rights Gap,” *Policy & Internet*, vol. 11, no. 1, pp. 6–15, Mar. 2019.

[50] M. L. Mueller and F. Badiei, “Requiem for a Dream: On Advancing Human Rights via Internet Architecture,” *Policy and Internet*, 2019.

[51] O. B. Hall, “Private Authority: Non-State Actors and Global Governance.,” *Harvard International Review*, vol. 27, no. 2, pp. 66–70, 2005.

[52] D. Oliver and J. Fedtke, “Human Rights and the Private Sphere,” *UCL Human Rights Review*, 2008.

[53] T. Mylly, *Intellectual Property and European Economic Constitutional Law : the Trouble with Private Informational Power*. Edward Elgar Publishing, 2009.

[54] J. H. Knox, “Horizontal human rights law,” *American Journal of International Law*, vol. 102, no. 1, pp. 1–47, 2008.

[55] L. Lane, *The horizontal effect of international human rights law in practice: A comparative analysis of the general comments and jurisprudence of selected united nations human rights treaty monitoring bodies*, vol. 5, no. 1. 2018.

[56] S. W. Brenner, *Law in an Era of “smart” Technology*. Oxford University Press, 2007.

[57] O. O. Cherednychenko, “Fundamental Rights and Private Law: A Relationship of Subordination or Complementarity?,” *Utrecht Law Review*, vol. 3, no. 2, pp. 1–25, 2007.

[58] The IEEE Global Initiative, “Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2,” 2017.

[59] U. de Montréal, “The Montreal Declaration for the Responsible Development of Artificial Intelligence,” 2020. [Online]. Available: https://www.montrealdeclaration-responsibleai.com/. [Accessed: 09-Apr-2021].

[60] o.V., “AI Principles - Future of Life Institute,” *Future of Life Institute*, 2017. [Online]. Available: https://futureoflife.org/ai-principles/?submitted=1#confirmation. [Accessed: 09-Apr-2021].

[61] R. Kitchin, “Thinking critically about and researching algorithms,” 2016.

[62] M. Perel (Filmar) and N. Elkin-Koren, “BLACK BOX TINKERING: Beyond Transparency in Algorithmic Enforcement,” *SSRN Electronic Journal*, 2016.

[63] M. Whittaker *et al.*, "AI Now Report 2018," 2018.
[64] S. Robbins, "A Misdirected Principle with a Catch: Explicability for AI," *Minds and Machines*, vol. 29, no. 4, pp. 495–514, Dec. 2019.
[65] M. Steen, "Upon Opening the Black Box and Finding It Full: Exploring the Ethics in Design Practices," *Science Technology and Human Values*, 2015.
[66] A. Feenberg, "The technocracy thesis revisited: On the critique of power," *Inquiry (United Kingdom)*, vol. 37, no. 1, pp. 85–102, 1994.
[67] L. Winner, *Autonomous Technology: Technics-Out-Of-Control As A Theme In Political Thought*. MIT Press, 1977.
[68] D. J. Hess, "Engaging science, technology, and society.," *Engaging Science, Technology, and Society*, vol. 1, no. 0, pp. 121–125, 2015.
[69] W. M. Jameson and D. G. Johnson, "STS and Ethics: Implications for Engineering Ethics," in *The Handbook of Science and Technology Studies*, MIT Press, 2008.
[70] C. Whitbeck, "Ethics as Design: Doing Justice to Moral Problems," *The Hastings Center Report*, vol. 26, no. 3, p. 9, 1996.
[71] C. Whitbeck, "Ethics as Design: Doing Justice to Moral Problems," *Ethics in Engineering Practice and Research*, pp. 135–154, 2011.
[72] M. Kranzberg, "Technology and History: 'Kranzberg's Laws,'" *Technology and Culture*, vol. 27, no. 3, p. 544, Jul. 1986.
[73] M. S. Pritchard, "Responsible Engineering: The Importance of Character and Imagination," *Science and Engineering Ethics*, vol. 7, no. 3, pp. 391–402, 2001.
[74] U. Pesch, "Engineers and Active Responsibility," *Science and Engineering Ethics*, no. 4, pp. 925–939, 2015.
[75] K. Krippendorff and R. Butter, "Semantics: Meanings and contexts of artifacts," in *Product Experience*, 2008.
[76] W. T. Lynch and R. Kline, "Engineering practice and engineering ethics," *Science Technology and Human Values*, 2000.
[77] T. Dotson, "Technological Determinism and Permissionless Innovation as Technocratic Governing Mentalities: Psychocultural Barriers to the Democratization of Technology," *Engaging Science, Technology, and Society*, vol. 1, pp. 98–120, 2015.
[78] S. Wyatt, "Technological Determinism is Dead: Long Live Technological Determinsm," in *The Handbook of Science and Technology Studies*, no. October, E. J. Hackett, Ed. The MIT Press, 2008, pp. 165–180.
[79] S. Cole, S. Jasanoff, G. E. Markle, J. C. Peterson, and T. Pinch, "Handbook of Science and Technology Studies.," *Contemporary Sociology*, 1995.
[80] F. N. Laird, S. Science, H. Values, and N. Summer, "Participatory Analysis , Democracy , and Technological Decision Making Stable URL : http://www.jstor.org/stable/689725 Participatory Participatory Analysis , Analysis , Democracy , Democracy , and and Technological Technological Decision Decision Making M," vol. 18, no. 3, pp. 341–361, 2016.
[81] E. O. Eriksen, "Governance between expertise and democracy: The case of European Security," *Journal of European Public Policy*, vol. 18, no. 8, pp. 1169–1189, 2011.
[82] M. E. Williams, "Escaping the zero-sum scenario: Democracy versus technocracy in Latin America," *Political Science Quarterly*. 2006.