


Research Brief – April 2023



# Protecting and Realizing Human Rights in the Context of Artificial Intelligence: A Problem Statement

By Alexander Kriebitz

As the transformation driven by AI is characterized by the replacement and augmentation of human intelligence, its growing influence in society has the potential to create adverse impacts for individuals and minorities, but also to undermine established procedures in society such as informed consent, rule of law or democratic accountability. Consequently, the discourse on AI ethics has identified several ethical concerns that might accompany the dissemination of AI. However, the implementation of AI ethics has been complicated by their perceived vagueness and the difficulty in enforcing non-binding ethical standards (Mittelstadt, 2019). As a result, scholars have explored the integration of human rights into the AI ethics discourse as a further avenue to hold companies accountable for violating norms and principles. In this Brief, the relevance of integrating human rights in the AI lifecycle is analyzed in order to shed light on the designing of normative frameworks.

Recent advances in artificial intelligence (AI) have created major repercussions in ethical debates and legal discourses, culminating in public statements to pause certain directions of AI development (*futureoflife*, 2023). This is owed primarily to the ascribed potential of this “family of technologies” throughout different business sectors and various areas of life, as well as the impact of AI solutions on the life of individuals (Council of the European Union, 2021). As the transformation driven by AI is characterized by the replacement and augmentation of human intelligence, its growing influence in society has the potential to create adverse impacts for individuals and minorities but also to undermine established procedures in society such as informed consent, rule of law or democratic accountability (Altman, 2023). While AI does not always determine the final outcome of a process, it exerts a strong influence on society by delivering content, predictions, recommendations, or decisions at an amplified pace (Council of Europe, 2021).

Consequently, the discourse on AI ethics has identified several ethical concerns that might accompany the dissemination of AI, including, but not limited to, the loss of human autonomy, exclusion, marginalization or discrimination of minorities and vulnerable groups, violation of privacy or lack of accountability (Rodrigues, 2020). In order to mitigate such threats, frameworks outlining the responsible use of AI have called for the integration, or at least observation, of ethical principles into the development, deployment, and use of AI (Floridi, 2018).

However, the implementation of AI ethics has been complicated by their perceived vagueness and the difficulty in enforcing non-binding ethical standards (Mittelstadt, 2019). As a result, scholars have explored the integration of human rights into the AI ethics discourse as a further avenue to hold companies accountable for violating norms and principles (Fukuda-Parr & Gibbons, 2021). Human rights constitute fundamental rights and freedoms that are inherent to all human beings, regardless of race, gender, nationality, religion, or any other status. Thus, they matter for the interpretation of existing legal obligations of United Nations (UN) member states and international organizations, most prominently through the Universal Declaration of Human Rights, the International Covenant on Political and Civil Rights, and finally, the International Covenant of Economic, Social and Cultural Rights (D’Amato, 1982). This applies to corporate actors as well, as companies developing and deploying AI. These are increasingly understood as accountable for human rights through the UN Guiding Principles on Business and Human Rights (Ruggie, 2007), which explicitly state that companies “should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved” (United Nations, 2011).

While the discussion of human rights in the AI context is still ongoing, there are some important aspects to reflect on, namely, what the realization and protection of human rights in the context of AI actually entail for companies developing and deploying AI solutions and societies exposed to these corporate activities. The following analysis sheds light on this issue by focusing on the relevance of integrating human rights in the AI lifecycle. Firstly, the article introduces human rights as a comprehensive - and sometimes ambiguous - normative concept. Following an analysis of the structural challenge AI poses to human rights, the article formulates the problem statement of realizing but also protecting human rights in the wider AI context and derives further conclusions for the design of normative frameworks.

### Human Rights as a comprehensive normative concept

*The discussion of human rights in the AI context is ongoing, important aspects to reflect on are what the realization and protection of human rights in the context of AI actually entail for companies developing and deploying AI solutions and societies exposed to these corporate activities*

Based on the philosophical tradition of enlightenment, human rights are regarded as the supreme norm of law and justice (UDHR, 1948). Philosophical and legal theories, as discussed by the likes of Francisco Suarez, John Locke, Immanuel Kant, John Rawls, Carole Pateman or Martha Nussbaum, have referred to the existence of a natural law, according to which individuals are entitled to basic freedom and rights including, but not

limited to, human dignity, the right to physical or mental integrity, the right to vote, equality before the law or freedom of opinion. Thus, these rights form major pillars of universal justice and are protected from breaches committed by states and other powerful organizations (Donnelly, 1982).

Owing to their relevance in international law and constitutional discourses, human rights have been codified in the form of constitutional rights, including the American Declaration of Independence, the Charter of Fundamental Rights of the European Union or the fundamental rights of the German Basic Law and anchored in international law by international human rights treaties including the Universal Declaration of Human Rights, the International Covenant of Civic and Political Rights or the International Covenant on Economic, Social and Cultural Rights (D'Amato, 1982). Through their integration into constitutional and international law, they supersede other legal considerations and define how states interact with their citizens, but also with human beings in general (Besson, 2011). Doing so, human rights encompass different dimensions of human rights realization and human rights protection:

- *Human rights protection:* Human rights constitute not merely separate rights but rather a self-contained regime of indivisible rights. Thus, human rights, in their interpretation as negative liberties or as claim-rights, regard freedom and equality before the law as the underlying purpose of human rights, and any restrictions to this must be based on the observation of certain procedures such as adherence to the rule of law principle, democratic legitimacy or the right to a fair trial (Ripstein, 2006; Wenar, 2013; Joseph, 1999).
- *Human rights realization:* According to many legal scholars and international frameworks, human rights extend beyond a purely defensive character and specify the objectives of social organizations in realizing certain entitlements such as the right to health, the right to adequate food, or the right to education (Daniels et al., 2015; Fagundes et al., 2022; Binder, 2022).

The reconciliation of both perspectives on human rights proves difficult in practice and is, therefore,

subject to norm derogation. The term norm derogations describes the act of reconciling different aspects of human rights (Zysset, 2022). According to the harm principle, which forms the basis of human rights as claim-rights, 'the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others (Turner, 2014). Nevertheless, the question of where harm starts and how to balance conflicting human rights implications tend to be controversial when applied in practice. For example:

- The realization of economic rights through the redistribution of wealth and income illustrates a balancing act between the right of a decent standard of living and the human right to own and use private property (Müller, 2009).
- The question of whether to enact compulsory vaccination portrays a conflict between the duty to prevent a massive health crisis and the duty of the state from respecting the physical integrity and autonomy of individuals (Archard et al., 2021).

Such derogations are situated in a predefined normative space: The principle of proportionality dictates that state interference must be proportional to the damage averted. Nevertheless, there are cases where the principle of proportionality cannot be applied, as certain types of human rights cannot be derogated, including the prevention of torture, slavery, or dehumanizing actions (Baig et al., 2022).

Considering human rights as a wider concept, human rights can be divided into freedoms and entitlements, moreover, there is a further distinction between derogable and non-derogable human rights. The discussed aspects reveal the complexity of human rights as a normative concept but also underpin the relevance of cultural and contextual differences when applying proportionality and balancing between different aspects of human rights.

### AI as a Challenge to Human Rights

Human rights present a complex normative cosmos, rendering their interpretation, which is traditionally the domain of courts and governmental bodies, difficult for corporate actors. Although companies are increasingly understood as actors accountable to

human rights, the question of how to interpret normative implications in case of trade-offs between different aspects of human rights remains complicated, particularly in the case of AI (Stahl et al., 2022; Almeida et al., 2022; Kosta, 2022).

***Although companies are actors accountable to human rights, the question of how to interpret normative implications in case of trade-offs between different aspects of human rights remains complicated.***

First and foremost, AI itself depicts a complex technology, which transcends an easy definition. The Merriam-Webster Dictionary defines AI as the capability of a machine to imitate intelligent human behavior, which reflects a plethora of potential use cases (Merriam-Webster). In order to perform actions regarded as intelligent, AI requires a specific data input but also a methodology that converts data into a specific outcome (European Commission, 2019). This conversion typically involves approaches and techniques such as machine learning, machine reasoning, or robotics (European Commission, 2021). The key issue of this methodology is that aspects outside of the specified data input are not considered in the decision-making progress. This constitutes a major epistemological drawback of AI when compared to human beings that can rely on a moral compass, intuitions, and gut feelings when making complex decisions (Durán, 2021). Furthermore, depending on the model, the decisions AI arrives at can be deterministic or non-deterministic. Deterministic means that the methodology of an AI model is clear *ex-ante*, whereas non-deterministic models can be characterized as “black boxes” (Tan Ming, 2022). As many non-deterministic models are based on machine learning technologies, the consequences cannot be foreseen *ex-ante*, as the model is capable of adapting its behavior dynamically when receiving new information.

Considering these aspects, human rights violations could originate through the following three channels:

1. **Purpose:** The action AI aims to imitate presents a human rights violation in itself.
2. **Process:** The very fact that a specific decision cannot be foreseen *ex-ante* but also not fully analyzed *ex-post* presents a violation of specific human rights, particularly when it comes to the involvement and participation of individuals or groups in relevant decisions.
3. **Outcomes:** The results produced by an AI solution conflict with human rights, in particular, with the right to non-discrimination.

While there are further human rights issues pertinent to the development and deployment of AI in society, the focus of the following elaboration lies in the three highlighted areas that constitute specific features of non-deterministic AI systems.

### ***Purposes of AI development, deployment, and use that conflict with human rights***

The entry point of considering human rights in an AI solution lies in its purpose and its guiding assumptions, which encapsulate the intentions of the designers behind the development of a specific model. In different scenarios, AI can be used as an instrument for creating but also reinforcing adverse human rights impacts.

Two examples underpin the relevance of such considerations:

- Amplifying human rights violations through the use of AI: A well-documented case is the treatment of ethnic minorities in the Xinjiang Uyghur Autonomous Region (XUAR) by the Chinese government and the regional administration of the XUAR (Stark, 2021). Reports of international organizations, including the United Nations, have criticized the mass deployment of face recognition tools aiming at identifying ethnic Uyghurs and AI solutions that aim at disenfranchising ethnic minorities in the region (Harwell & Dou, 2020). The human rights issue of the very development of such AI solutions lies in the assumption that ethnic Uyghurs, irrespective of their individual preferences, are perceived as harboring a tendency towards religious extremism (Byler, 2019).

Racial profiling of this kind is regarded as an inherent human rights violation and as a practice dehumanizing individuals (Giwa et al., 2020). Similar concerns apply to the general human rights and privacy-specific implications of AI solutions based on real-time biometric data and used in the context of public surveillance, but also to the development of social scoring schemes that could be used for law enforcement purposes (Jaiswal & Tarar, 2020).

- Using AI to circumvent human rights obligations: Intended human rights violations can occur in the context of seemingly harmless business operations such as human resources (Gratch & Fast, 2022). A high-risk area is AI solutions that seek to circumvent legal standards or internationally guaranteed labor rights. Discrimination on the basis of unionization - trade union membership - is regarded as a human rights violation, in particular, a labor rights violation (Baert & Omey, 2015). AI solutions that aim to find out whether a likely candidate for a job is a member of a trade union through the evaluation of social media profiles constitutes therefore a human rights violation due to the very use case. Similar issues relate to AI evaluating information related to employees that is regarded as confidential or which is protected by law, including pregnancy (Williams, 2005).

***The EU AI Act ignores the possibility that seemingly harmless AI solutions might be deployed within a wider context of human rights violations (...) and does not include actionable guiding principles or measures that could be taken to mitigate incentives adverse to human rights***

Some controversial purposes behind AI development have already been addressed by legislation approaches (Ebers et al., 2021). The current EU AI Act proposal, for example, aims to ban certain AI use cases, including the usage of real-time biometric data in public surveillance or deceptive AI solutions based on human rights considerations (Raposo, 2022). The

EU AI Act, however, ignores the possibility that even seemingly harmless AI solutions might be deployed within a wider context of human rights violations, such as in conflict-affected regions. Moreover, the EU Act does not include actionable guiding principles or measures that could be taken to mitigate incentives adverse to human rights.

To fill this gap, the application of the UN Guiding Principles on Business and Human Rights to AI development has much potential, as it calls on enterprises to scrutinize their business operations for human rights risks (UN Guiding Principles, Principle 18) (UDHR, 1948): Business enterprises, but also governments may face strong power or profit maximization incentives to use AI as a means to circumvent existing legal obligations. The identification of such toxic incentives is important to consider from a policy point of view when closing legal loopholes, but also when examining the risks of business relationships. The task of regulators under this approach would be to connect both the EU AI Act and the UN Guiding Principles in order to integrate traditional business and human rights issues to AI development and deployment.

### ***Designs of AI that conflict with procedural rights***

The non-deterministic nature of AI, but also the general transfer of human decision-making to a machine or software solution without human oversight, can cause systematic human rights violations. Indeed, many procedures related to consent and human control in society are relevant from a human rights perspective, especially in “high-stakes environments”, where it is critical to ensure human consent and understanding (De Hert & Gutwirth, 2006; Percy, 2018). This applies particularly to decisions that have a strong impact on the physical or mental well-being, opportunities, chances or social and political participation of an individual (Daten Ethik Kommission, 2019).

Specific matters for human rights are, for instance, policing, law enforcement, or decisions by public authorities, which require the consideration of proportionality before making a final judgment. The Universal Declaration of Human Rights explicitly anchors the right to an effective remedy for acts violating fundamental rights (Art. 8), the prevention of “arbitrary interference with his [their] privacy, family, home or correspondence” (Art. 12), the rights of “free

choice of employment” (Art. 23) and the “right to education” (Art. 26). These rights set a high bar for using AI systems, particularly non-deterministic ones, in environments such as judicial systems including courts, policing, human resources, health and the education system, where understanding but also justifying why a certain decision has been adopted is critical for an individual (Barth & Arnold, 1999; Young et al., 2019).

Therefore, the criticality of such a system rises if the consequences made by AI decisions are completely or partly irreversible. This concerns the deployment of AI in public health, where informed consent of patients determines further steps taken by medical personnel in treatment (Beauchamp, 2011). Informed consent can be undermined by a lack of transparency in an AI system. Consequently, the existence of a case-dependent level of human oversight and a sufficient degree of explainability is critical for the realization of human rights.

One takeaway for the human rights responsibilities of AI developers and users is the consideration of procedural rights such as informed consent, accountability, administrative discretion, or rule of law in “high-stakes environments”, including the realization of an effective right to correct decisions that are erroneous.

### ***AI-generated outcomes that violate human rights***

Apart from the overall purpose and the safeguarding of specific processes, the consequences AI generates, such as recommendations, decisions, or any other outcomes, are relevant for the dimension of human rights as claim-rights. Human rights violations can originate in breaches of the right of physical integrity - as a treatment method does not consider the specific conditions of a vulnerable population but also relate to the collective rights of minorities and vulnerable groups. Human rights treaties have codified such rights in the form of rights that protect females, children, individuals with disabilities, or racial and ethnic minorities such as indigenous peoples (Hendriks, 2007).

Such specifications matter in the context of biased AI: The input an AI receives for solving a specific task, such as prescribing a specific treatment, can miss important information and subsequently lead to a

wrong conclusion. This happened, for example, in the context of AI solutions dealing with the allocation of health resources to individuals deployed in the US (Obermeyer et al., 2019). The assumption of the model was that higher spending on health products indicates a worsening health condition. While this assumption might be plausible on first sight, the model did not factor in that individuals have varying financial means to spend on health products. Consequently, the AI model did not give an accurate picture of the health risks of, for example, the African American community, which has lesser financial means than other parts of the population thus leading to a general racial bias in the tool.

***One takeaway for the human rights responsibilities of AI developers and users is the consideration of procedural rights such as informed consent, accountability, administrative discretion, or rule of law in “high-stakes environments”, including the realization of an effective right to correct decisions that are erroneous.***

The aforementioned case exemplifies how biases can result in decisions of public authorities at the expense of specific parts of the society - understood as a minority group - but also individuals that are disadvantaged or penalized for belonging to a specific group. Such consequences - however unintended they might be - are particularly relevant in areas where the progressive realization of human rights is very urgent, for example, in the context of the right to equitable health (Samorani et al., 2022). Here, biases conflict with two relevant human rights aspects: On the one side, it impedes the realization of the right to health for the affected individuals. On the other side, it presents a structural breach of the right to equal treatment and furthermore relates to the rights of minorities protected by specific legal frameworks (United Nations, 1948).

The issue of discrimination and exclusion being amplified by AI, however, transcends questions relating to data input and bias as it pertains to technical designs in general (Keats & Clarkson, 2003). Certain hardware components of AI solutions that are inaccessible for individuals with physical impairments would be one example. Embedding diversity within such design choices is critical for a

human rights perspective on AI, as the structural disregard of such might not only impede the realization of human rights but also reinforce existing patterns of exclusion addressed by group-specific conventions (United Nations, 2006).

***Embedding diversity is critical for a human rights perspective on AI, as the structural disregard of such might not only impede the realization of human rights, but also reinforce existing patterns of exclusion.***

### Final Thoughts

This Brief has given a structural view of the implications of AI for companies that are understood to be accountable for human rights realization and protection. The analysis of the human rights and AI context drives two important conclusions:

The first observation we can make is that human rights accountability of enterprises in the AI context presents an enormous task for companies. This owes to the nature of both concepts: Human rights and AI governance constitute two complex systems that cannot be easily integrated.

- Firstly, human rights encompass different dimensions of human rights realization and human rights protection. In cases of conflict, trade-offs within human rights are subject to the principle of proportionality, which is usually applied by courts or public

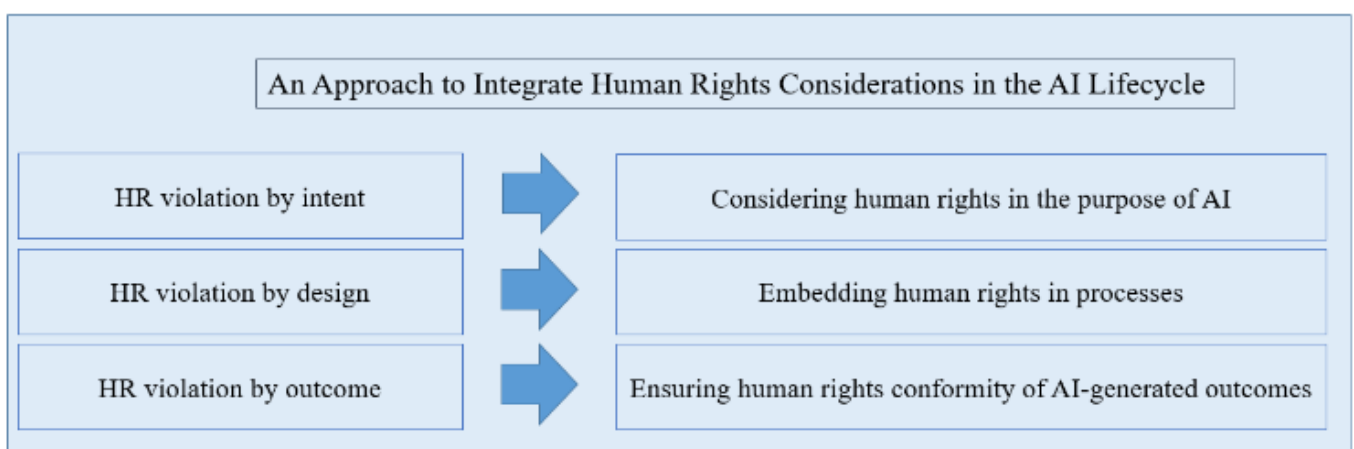
administration. Thus, it is not always clear ex/ante whether a specific action can be regarded as a human rights issue or not.

- Secondly, AI presents a complex technology itself: It refers to the performance of fundamentally different and unrelated tasks such as visual perception, speech recognition, decision-making, and language translation. AI is achieved through the use of algorithms, machine learning, and neural networks that enable computers to learn from data and improve their performance over time. Thus, governing its use in a general way is a complex task.

Consequently, the challenges that go along with integrating both concepts are enormous, as both AI and human rights affect virtually all walks of life and are both difficult to handle.

The second observation is that establishing conceptual clarity is generally possible in spite of the stated challenge. In order to create conceptual clarity, the best way to navigate this challenge is, therefore, to look at the different stages which come up in the development of AI and where human rights considerations can be integrated. In this effort, a life cycle analysis can be very beneficial:

- The first stage concerns the very purpose of the deployment case and the intentions behind its use. AI presents a human rights violation by intent if its purpose is to violate principles which are protected by human rights. This might sound obvious, but several cases underpin the relevance of this point. AI is used to persecute ethnic minorities, and



policymakers have been contemplating AI-based social scoring solutions. Moreover, AI can also be used in order to create deep fakes used for deception and to violate the privacy rights of specific individuals.

- The second question concerns the non-deterministic nature of many AI solutions. Certain processes in society necessitate a high level of transparency and explainability. The use of non-deterministic AI in such processes is therefore problematic, particularly from a human rights perspective. The use of AI in courts, however well-intended this might be, can conflict with the right to remedy and can constitute a human rights violation by design.
- The third question concerns the outcomes of AI, which are dependent on the model of AI. Research has highlighted many cases of biases leading to the structural exclusion of individuals based on criteria that are regarded as discriminatory. Human Rights violations of this sort constitute violations by the outcome. Discrimination of individuals, particularly when it comes to social services or access to rights is critical, and can be mitigated by measures and policies adopted by corporate actors developing and using AI.

Both observations reinforce one particular conclusion, namely that the question of human rights needs, therefore to be integrated into different steps of the AI lifecycle, including the very early stage of product conceptualization. The pressing issue is the thorough analysis of the use case itself, particularly considering the incentives potential users are exposed to. Risk assessments based on human rights must therefore consider the possible consequences if a particular AI solution is deployed by a company with commercial interests or by law enforcement in an authoritarian or politically unstable country.

Furthermore, AI development and deployment must consider the relevance of human oversight in the field of operation. This is particularly relevant if the right to remedy, but also autonomy and the right to make informed decisions are at stake. High-stakes

environments that have a close link to particular human rights are more vital here than in other deployment scenarios.

Finally, human rights entail implications for the overall quality of an output generated by AI: Depending on the data, the model, but also the expertise of human developers involved, an AI system might be better or worse. Furthermore, quality variances can overlap with criteria deemed as discriminatory. Companies have therefore a due diligence to control for potential human rights violations and to use the resources they have at their disposal to minimize the potential risk of human rights violations through the entire AI lifecycle.



**References:**

- Almeida, D., Shmarko, K., & Lomas, E. (2021). The ethics of Facial Recognition Technologies, surveillance, and accountability in an age of Artificial Intelligence: A comparative analysis of US, EU, and UK Regulatory frameworks. *AI and Ethics*, 2(3), 377–387. <https://doi.org/10.1007/s43681-021-00077-w>
- Altman, S. (2023, February 24). Planning for AGI and beyond. Retrieved April 21, 2023, from <https://openai.com/blog/planning-for-agi-and-beyond>
- Archard, D., Brierley, J., & Cave, E. (2021). Compulsory childhood vaccination: Human rights, solidarity, and best interests. *Medical Law Review*, 29(4), 716–727. <https://doi.org/10.1093/medlaw/fwab024>
- Baert, S., & Omeij, E. (2015). Hiring discrimination against pro-union applicants: The Role of Union density and firm size. *De Economist*, 163(3), 263–280. <https://doi.org/10.1007/s10645-015-9252-1>
- Baig, K., Kiran, I., & Ahamad, W. (2022). The implementation of non-derogable rights under the ICCPR in the context of the defined role of the Treaty Bodies. *Review of Education, Administration & Law*, 5(4), 657–664. <https://doi.org/10.47067/real.v5i4.296>
- Barth, T. J., & Arnold, E. (1999). Artificial Intelligence and administrative discretion. *The American Review of Public Administration*, 29(4), 332–351. <https://doi.org/10.1177/02750749922064463>
- Beauchamp, T. O. M. L. (2011). Informed consent: Its history, meaning, and present challenges. *Cambridge Quarterly of Healthcare Ethics*, 20(4), 515–523. <https://doi.org/10.1017/s0963180111000259>
- Besson, S. (2011). Human rights and democracy in a global context: Decoupling and Recoupling. *Ethics & Global Politics*, 4(1), 19–50. <https://doi.org/10.3402/egp.v4i1.6348>
- Binder, C. (2022). The Committee on Economic, social and cultural rights—the power of subjective rights: Comment with special focus on the case law of the Inter-american and the European Courts of Human Rights. *Journal of Human Rights Practice*, 14(1), 75–83. <https://doi.org/10.1093/jhuman/huac021>
- Byler, D. (2019). Preventative policing as community detention in Northwest China. *Made in China Journal*, 4(3). <https://doi.org/10.22459/mic.04.03.2019.13>
- D'Amato, A. (2017). The concept of human rights in international law\*. *International Law of Human Rights*, 21–70. <https://doi.org/10.4324/9781315092492-3>
- Daniels, N., Charvel, S., Gelpi, A. H., Porteny, T., & Urrutia, J. (2015). Role of the courts in the progressive realization of the right to health: Between the threat and the promise of judicialization in Mexico. *Health Systems & Reform*, 1(3), 229–234. <https://doi.org/10.1080/23288604.2014.1002705>
- Daten Ethik Kommission. (2019, October). *Opinion of the data ethics commission - executive summary*. Bundesministerium der Justiz. Retrieved April 22, 2023, from [https://www.bmj.de/SharedDocs/Downloads/DE/The men/Fokusthemen/Gutachten\\_DEK\\_EN.pdf?\\_\\_blob=publicationFile&v=2](https://www.bmj.de/SharedDocs/Downloads/DE/The men/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&v=2)
- Donnelly, J. (1982). Human rights as natural rights. *Human Rights Quarterly*, 4(3), 391. <https://doi.org/10.2307/762225>
- Durán, J. M., & Jongsma, K. R. (2021). Who is afraid of Black Box Algorithms? on the epistemological and ethical basis of trust in Medical Ai. *Journal of Medical Ethics*. <https://doi.org/10.1136/medethics-2020-106820>
- Ebers, M., Hoch, V. R., Rosenkranz, F., Ruschemeier, H., & Steinrötter, B. (2021). The European Commission's proposal for an artificial intelligence act—a critical assessment by members of the Robotics and AI Law Society (Rails). *J*, 4(4), 589–603. <https://doi.org/10.3390/j4040043>
- European Commission. (2019). *Ethics guidelines for Trustworthy Ai*. Shaping Europe's digital future. Retrieved April 22, 2023, from <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- European Commission. (2021, April 21). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (ARTIFICIAL INTELLIGENCE ACT) and amending certain UNION LEGISLATIVE ACTS*.

EUR-Lex. Retrieved April 24, 2023, from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

European Parliament. (2023). Regulation 2021/0106. *Official Journal of the European Union*, 66. Retrieved April 20, 2023, from [https://eur-lex.europa.eu/TodayOJ/fallbackOJ/l\\_10620230420en.pdf](https://eur-lex.europa.eu/TodayOJ/fallbackOJ/l_10620230420en.pdf).

Fagundes, A., de Cássia Lisboa Ribeiro, R., de Brito, E. R., Recine, E., & Rocha, C. (2022). Public infrastructure for food and nutrition security in Brazil: Fulfilling the constitutional commitment to the human right to adequate food. *Food Security*, 14(4), 897–905. <https://doi.org/10.1007/s12571-022-01272-1>

Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—an ethical framework for a good AI Society: Opportunities, Risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>

Fukuda-Parr, S., & Gibbons, E. (2021). Emerging consensus on 'ethical ai': Human rights critique of stakeholder guidelines. *Global Policy*, 12(S6), 32–44. <https://doi.org/10.1111/1758-5899.12965>

Future of Life Institute. (2023, April 13). *Pause giant AI experiments: An open letter*. Future of Life Institute. Retrieved April 20, 2023, from <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

Giwa, S., Mullings, D. V., Adjei, P. B., & Karki, K. K. (2020). Racial erasure: The silence of social work on police racial profiling in Canada. *Journal of Human Rights and Social Work*, 5(4), 224–235. <https://doi.org/10.1007/s41134-020-00136-y>

Gratch, J., & Fast, N. J. (2022). The power to harm: AI assistants pave the way to unethical behavior. *Current Opinion in Psychology*, 47, 101382. <https://doi.org/10.1016/j.copsyc.2022.101382>

Gutwirth, S., & De Hert, P. (2022). Privacy, Data Protection and law enforcement. opacity of the individual and transparency of power. *Direito Público*, 18(100). <https://doi.org/10.11117/rdp.v18i100.6200>

Harwell, D., & Dou, E. (2020, December 8). *Huawei tested AI software that could recognize Uighur minorities and alert police, report says*. The Washington Post. Retrieved April 22, 2023, from <https://www.washingtonpost.com/technology/2020/12/08/huawei-tested-ai-software-that-could-recognize-uighur-minorities-alert-police-report-says/>

Hendriks, A. (2007). UN Convention on the rights of persons with disabilities. *European Journal of Health Law*, 14(3), 273–298. <https://doi.org/10.1163/092902707x240620>

Jaiswal, A., & Tarar, S. (2020). Real-time biometric system for security and surveillance using face recognition. *Communications in Computer and Information Science*, 293–304. [https://doi.org/10.1007/978-981-15-6634-9\\_27](https://doi.org/10.1007/978-981-15-6634-9_27)

Joseph, S. (1999). A Rights Analysis of the Covenant on Civil and Political Rights. *J. Int'l Legal Stud*, 5(57).

Kosta, E. (2022). Algorithmic state surveillance: Challenging the notion of agency in human rights. *Regulation & Governance*, 16(1), 212–224. <https://doi.org/10.1111/rego.12331>

Merriam-Webster. (n.d.). *Artificial Intelligence Definition & meaning*. Merriam-Webster. Retrieved April 22, 2023, from <https://www.merriam-webster.com/dictionary/artificial%20intelligence>

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>

Müller, A. (2009). Limitations to and derogations from economic, social and cultural rights. *Human Rights Law Review*, 9(4), 557–601. <https://doi.org/10.4324/9781315257044-4>

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>

- Percy, S. L. (2018). *Disability, civil rights, and public policy: The politics of implementation*. University of Alabama Press.
- Raposo, V. L. (2022). Ex machina: Preliminary critical assessment of the european draft act on artificial intelligence. *International Journal of Law and Information Technology*, 30(1), 88–109. <https://doi.org/10.1093/ijlit/eaac007>
- Ripstein, A. R. T. H. U. R. (2006). Beyond the harm principle. *Philosophy Public Affairs*, 34(3), 215–245. <https://doi.org/10.1111/j.1088-4963.2006.00066.x>
- Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, 4, 100005. <https://doi.org/10.1016/j.jrt.2020.100005>
- Ruggie, J. G. (2007). Business and human rights: The evolving international agenda. *American Journal of International Law*, 101(4), 819–840. <https://doi.org/10.1017/s0002930000037738>
- Samorani, M., Harris, S. L., Blount, L. G., Lu, H., & Santoro, M. A. (2022). Overbooked and overlooked: Machine learning and racial bias in medical appointment scheduling. *Manufacturing & Service Operations Management*, 24(6), 2825–2842. <https://doi.org/10.1287/msom.2021.0999>
- Stahl, B. C., Rodrigues, R., Santiago, N., & Macnish, K. (2022). A european agency for artificial intelligence: Protecting fundamental rights and ethical values. *Computer Law & Security Review*, 45, 105661. <https://doi.org/10.1016/j.clsr.2022.105661>
- StarkTu, R. (2021). China's use of artificial intelligence in their war against Xinjiang. *Tul. J. Int'l & Comp. L*, 29(153).
- Tan Ming, J. (2022). Non-deterministic artificial intelligence systems and the future of the law on unilateral mistakes in Singapore. *Singapore Academy of Law Journal*, 34(1), 91–124.
- Turner, P. N. (2014). "Harm" and Mill's harm principle. *Ethics*, 124(2), 299–326. ISO 690
- United Nations Human Rights, Office of the High Commissioner. (2011). Implementing the United Nations "Protect, Respect and Remedy" Framework. *Guiding Principles on Business and Human Rights* .
- United Nations. (1948). *Universal declaration of human rights*. United Nations. Retrieved April 21, 2023, from <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
- United Nations. (n.d.). *Convention on the rights of the child*. OHCHR. Retrieved April 23, 2023, from <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child>
- Wenar, L. (2013). The nature of claim-rights. *Ethics*, 123(2), 202–229. <https://doi.org/10.1086/668707>
- Williams, N. (2005). Pre-hire pregnancy screening in Mexico's maquiladoras: Is it discrimination. *Duke J. Gender L. & Pol'y*, 12, 131.
- Young, M. M., Bullock, J. B., & Lecy, J. D. (2019). Artificial discretion as a tool of governance: A framework for understanding the impact of artificial intelligence on Public Administration. *Perspectives on Public Management and Governance*. <https://doi.org/10.1093/ppmgov/gvz014>
- Zysset, A. (2022). To derogate or to restrict? the COVID-19 pandemic, proportionality and the Justificatory Gap in european human rights law. *Jus Cogens*, 4(3), 285–301. <https://doi.org/10.1007/s42439-022-00065-6>
- Spahn, A. (2018) "The First Generation to End Poverty and the Last to Save the Planet?"—Western Individualism, Human Rights and the Value of Nature in the Ethics of Global Sustainable Development', *Sustainability*, 10(6), p. 1853. Available at: <https://doi.org/10.3390/su10061853>.
- WHO (2021) *The impact of COVID-19 on health and care workers: a closer look at deaths*. WHO/HWF/WorkingPaper/2021.1. Geneva.
- Ujomudike, P.O. (2015) 'Ubuntu Ethics', in *Encyclopedia of Global Bioethics*. Cham: Springer International Publishing, pp. 1–14. Available at: [https://doi.org/10.1007/978-3-319-05544-2\\_428-1](https://doi.org/10.1007/978-3-319-05544-2_428-1).
- Viviers, H. and Mzondi, A. (2016) 'The end of essentialist gods and Ubuntu: a feministcritical investigation', *Pharos Journal of Theology*, 97, pp. 1–17.

Wiener, N. (1960) 'Some Moral and Technical Consequences of Automation', *Science*, 131(3410), pp. 1355–1358. Available at: <https://doi.org/10.1126/science.131.3410.1355>.

Wiens, J., Creary, M. and Sjoding, M.W. (2022) 'AI models in health care are not colour blind and we should not be either', *The Lancet Digital Health*, 4(6), pp. e399–e400. Available at: [https://doi.org/10.1016/S2589-7500\(22\)00092-9](https://doi.org/10.1016/S2589-7500(22)00092-9).