**Hungarian Journal of Legal Studies** 64 (2023) 2, 236–267
DOI: 10.1556/2052.2023.00481

AKADÉMIAI KIADÓ

# "Humanity's new frontier": Human rights implications of artificial intelligence and new technologies

NOÉMI NAGY[*]

Ludovika University of Public Service, Budapest, Hungary

ORIGINAL RESEARCH PAPER

Received: May 12, 2023 • Accepted: July 24, 2023
Published online: November 28, 2023

## ABSTRACT

New technologies based on digitalization, automation, and artificial intelligence have fundamentally transformed our lives and society as a whole, in just a few decades. These technologies support human well-being and prosperity by enhancing progress and innovation, however, they also have the potential to negatively impact human rights, democracy, and the rule of law. Discrimination, the violation of privacy, increasing surveillance, the weakening of personal autonomy, disinformation and electoral interference are but a few of the many concerns. This paper examines the specific human rights implications of AI-driven systems through the lens of the most important international instruments adopted by the UN and regional human rights mechanisms. The paper shows how AI can affect the exercise of all human rights, not only a most obvious few. In line with major international organizations, the author calls on decision-makers to take a precautionary approach by adopting AI regulations that are consistent with the standards of fundamental human rights, and that balance the realization of the opportunities with the potential risks which AI presents.

**KEYWORDS**

AI and human rights, civil, political, economic, social and cultural rights, discrimination

[*] Corresponding author. E-mail: nagynoemi@uni-nke.hu

## 1. INTRODUCTION

The rapid development of digital and artificial intelligence based technologies has profoundly transformed various aspects of our lives, from work to law enforcement, from healthcare to transportation, from socializing to education. AI systems play a role in deciding where a crime is likely to take place, how to allocate social security benefits, or whether someone is at a serious health risk. From facial recognition technology through self-driving cars to agricultural and nursing robots, AI is making our lives easier and more efficient. However, it also has the potential to negatively affect human rights, democracy, and the rule of law.[1] As the Director-General of UNESCO warned us, "AI is humanity's new frontier. Once this boundary is crossed, AI will lead to a new form of human civilization. The guiding principle of AI is not to become autonomous or replace human intelligence. But we must ensure that it is developed through a humanist approach, based on values and human rights. We are faced with a crucial question: what kind of society do we want for tomorrow? The AI revolution opens up exciting new prospects, but the anthropological and social upheaval it brings in its wake warrants careful consideration."[2]

When it comes to the concept of AI, definitions are plenty.[3] For example, the European Commission's High-Level Expert Group on AI views it as "software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions."[4] For the purposes of this article, I use the term AI in a broad sense: software (algorithmic model) that represents human intelligence, i.e. performs cognitive or perceptual functions. I do not focus on individual AI technologies but rather on AI systems as a conceptual object of analysis. Nevertheless, the clarification of particular concerns relies on distinct illustrations of AI technologies. In addition, many findings reported in this paper also apply to new technologies and digitalization in the widest possible context, so it is not indispensable to agree on a certain definition of AI in order to understand the challenges that these technologies present.

This paper sets out from a human rights perspective, and examines the possible implications of diverse AI systems on a wide range of first- and second-generation human rights, provided in the most important international instruments adopted under the auspices of the United Nations and regional intergovernmental organizations. Although research is abundant on the implications of AI on specific human rights – particularly the right to privacy, non-discrimination, and freedom of expression –, it is still difficult to get an overview of the full spectrum of human rights that can be affected. The aim of this article is to provide the reader with a comprehensive overview that enables them to truly appreciate the extent and overarching nature of the issue.

[1] CAHAI (2020a), (2020b); Leslie et al. (2021).

[2] Azoulay (2018).

[3] For a thorough overview of these, see Hárs (2022) 2–8.

[4] AI HLEG (2019) 36.

A basic premise of any such research is that "the same rights that people have offline must also be protected online".[5]

This paper will focus on those human rights which are protected by the Universal Declaration of Human Rights (UDHR) – the cornerstone of the international system of human rights protection as we know it today – and the following international treaties: International Covenant on Civil and Political Rights (ICCPR), International Covenant on Economic, Social and Cultural Rights (ICESCR), European Convention on Human Rights (ECHR), American Convention on Human Rights (ACHR), African Charter on Human and Peoples' Rights (ACHPR). These legal instruments protect first- and second-generation human rights, that is civil and political rights (classical freedoms and liberties) as well as economic, social, and cultural rights. Some third-generation (solidarity) rights are included in the African Charter, such as the right of all peoples to equality, free disposal of wealth and natural resources, the right to development, the right to peace and security, and the right to "a general satisfactory" environment.[6] Although assessing the impact of AI on third-generation human rights[7] would certainly be instructive and useful in gaining a clearer view of the overall situation, it will not be explored in this paper. Beyond space constraints and the scarcity of relevant high-quality research, another reason for this is that the international community has not yet come to an agreement on the specific range, scope, and nature of third-generation human rights.[8] It is disputed which rights fall under this category. Consider, for example, the right of peoples to self-determination, the right to share in the benefits of the common heritage of mankind, or the right to humanitarian assistance. Is the right to water a third-generation human right or an aspect of the right to an adequate standard of living? Even though the author is aware of the weaknesses of the three generations theory,[9] this classification of human rights is actually quite useful from a didactic point of view and will satisfy the purposes of this paper.

After this introduction, the paper is structured as follows: Sections 2–5 provide a detailed analysis of the manifold effects of AI on individual human rights, starting with human dignity as a foundational value of all human rights, based on an extensive literature review. Relevant articles were selected from high-quality journals dedicated to either artificial intelligence or human rights, and from non-specific journal databases searched with the joint application of these two keywords. Reports of non-governmental and intergovernmental human rights organizations, online news, monographs and edited books were also used. After the analysis, Section 6 puts forward possible directions for future research, whereas Section 7 makes concluding remarks and recommendations for legislators and decision-makers.

[5] UN Human Rights Council (2017). That said, I do agree that certain adjustments to human rights norms and implementation strategies in the cyberspace may be necessary. For a critical view on the paradigm of 'normative equivalency' between offline and online spaces, see Dror-Shpoliansky and Shany (2021).

[6] ACHPR, Arts. 19–24.

[7] The concept of the right to development was introduced by Kéba M'Baye in 1972, which Karel Vasak later classified, together with other rights, as a third generation of rights or solidarity rights. Mubangizi (2004) 96.

[8] Algan (2004).

[9] Algan (2004) 126–128.; Mubangizi (2004) 96–100.

## 2. IT IS ALL ABOUT HUMAN DIGNITY

The edifice of international human rights law is premised upon human dignity.[10] Human dignity is understood as the foundation of all human rights, and at the same time a stand-alone human right, and an aspirational standard to reach for in relation to other rights.[11] It means that every individual has an intrinsic worth as a human being, whereby they need to be treated as an end in themselves, not as a means for our own purposes.[12] Teo refers to this as the non-instrumentalization of persons, where even individual consent can be displaced as dignity "goes beyond purely individualistic notions to encompass a wider communitarian concern".[13] In addition, human dignity has at least three other conceptions: the protection of vulnerable groups, the recognition and expression of self-worth (including autonomy, which is also interpreted as an element of privacy), and the protection of humanity as a species concern.[14]

Given its pervasiveness and the profound impact it may have on every aspect of life, AI can significantly affect all these aspects of human dignity. For instance, the use of lethal autonomous weapons systems engages the 'protection of humanity as a species' and the 'non-instrumentalization of persons' concerns. Bias and discrimination within AI systems relate to the 'non-instrumentalization' and the 'autonomy' conception. Content moderation and curation as well as constant monitoring and surveillance of persons by AI systems involve the 'recognition and expression of self-worth' conception of human dignity, notably through respect for informational privacy.[15]

Furthermore, the increasing use of AI in the medical, social, and care systems raise potential dignity issues for vulnerable groups, especially for the elderly and persons with disabilities. Therapy, care, and nursing robots that exist today are only capable of performing simple tasks, such as giving information, navigating, bringing and serving medicine or food, helping patients stand up, sit up, or move. More complex tasks such as feeding, washing, intimate hygiene, dressing, or undressing are beyond the capabilities of most robots currently in use. However, Pfeifer-Chomiczewska warns of the possibility of such a development in the near future, and urges us to ask ourselves whether we should allow a "soulless machine" to replace human personnel in performing the above activities, leading to the dehumanization of the health and care system.[16] Of course, much depends on robot design and the culture where these tools are applied.[17]

[10] UDHR, Art. 1.: "All human beings are born free and equal in dignity and rights."; ICCPR/ICESCR, Preamble: "[human] rights derive from the inherent dignity of the human person". Cf. UDHR, Preamble; ICCPR, Art. 10; Protocol No. 13 to ECHR (concerning the abolition of the death penalty in all circumstances), Preamble; ACHR, Art. 11(1); ACHPR, Art. 5.

[11] Teo (2023) 243.

[12] Corea et al. (2023) 524.

[13] Teo (2023) 247.

[14] Teo (2023) 245–253.

[15] Teo (2023) 255–259.

[16] Pfeifer-Chomiczewska (2023) 789, 796.

[17] Van Est, Gerritsen and Kool (2017) 28.

The various applications of AI in the social, economic, and cultural fields may contribute to the conditions of a dignified human life, but they can also lead to manipulation, decreasing human agency, threatening moral and physical integrity, and dissolving the uniqueness of each individual in the generality of statistical models and Big Data.[18] This should be kept in mind when considering the impact that AI can have on individual human rights, as those always have an underlying human dignity component.

# 3. AI, DISCRIMINATION AND VULNERABLE GROUPS

## 3.1. The prohibition of discrimination

Discrimination means the prejudiced or unfair treatment of an individual based on his or her membership in a certain group or category, such as race, color, sex, language, religion, political or other opinion, national or social origin, property or birth. The *prohibition of discrimination* entails that everyone is equal before the law, and is entitled to the equal protection of the law.[19] Yet, not all differentiated treatments are prohibited. When assessing the legality of a measure, most human rights courts and committees use the same test that the European Court of Human Rights (ECtHR) elaborated in the landmark Belgian Linguistic Case back in the 1960s. Accordingly, “the principle of equality of treatment is violated if the distinction has no objective and reasonable justification”, and when “there is no reasonable relationship of proportionality between the means employed and the aim sought to be realised”.[20]

Discrimination in the context of AI can be caused by the unequal access of certain (usually marginalized) groups to these technologies, bias in the data, and algorithmic bias. As for bias of the data, automated decision-making systems – via machine learning and deep learning – learn from large databases (Big Data). However, the available data is often not representative of the population or phenomenon of study, does not include variables that properly capture the phenomenon we want to predict, or includes content produced by humans which may contain biases against certain groups.[21] In general, this bias is not intentional: “Data represent our world: if they are biased, it is most often because our world itself is biased.”[22] In turn, algorithmic bias in the context of online content consumption means that “search algorithms and search engines by definition do not treat all information equally. While processes used to select and index information may be applied consistently, the search results will typically be ranked according to perceived relevance. Accordingly, different items of information will receive different degrees of visibility depending on which factors are taken into account by the ranking algorithm.”[23]

[18] Corea et al. (2023) 524; Mantelero and Esposito (2021) 11–13.

[19] UDHR, Art. 7; ICCPR, Arts. 2 and 26; ICESCR, Art. 2; ECHR, Art. 14. and Protocol No. 12 to the ECHR; ACHR, Arts. 1 and 24; ACHPR, Art. 2.

[20] Case “relating to certain aspects of the laws on the use of languages in education in Belgium”, App nos 1474/62 and others (ECtHR, 23 July 1968), 31. Cf. UN Human Rights Committee (1989) para. 13.

[21] Greenstein (2022) 311.

[22] Devillers, Fogelman-Soulié and Baeza-Yates (2021) 79.

[23] Council of Europe (2018) 26.

Algorithms are influenced by the conscious and unconscious biases of their developers – such as gender identity, socioeconomic class, culture and upbringing – through direct programming or by the data used to train the algorithms.[24] Abundant research show that "some classes of information have a propensity to be used in a discriminatory way", such as a person's gender, sexual orientation, ethnicity, religion, political affiliation, or economic situation.[25] The use of historical discriminatory decision-making data to train AI can reinforce or perpetuate structural discrimination and power imbalances in societies, even if the protected attributes are not directly present in the data.[26]

As will be further discussed in Section 4.2, machine learning tools are widely applied in criminal justice. Predictive policing, for instance, is based on the use of historical crime data and statistical methods to forecast areas where crime is likely to take place and the individuals who may commit a crime. Research has shown that police decisions about where to patrol and whom to detain or search are highly motivated by race and ethnicity. "When historical racism and class discrimination are encoded in the outputs of an algorithm, minority and low-income communities might fall victim to a feedback loop of ever greater police attention".[27] O'Neil explains how this works in practice. PredPol (a predictive policing software used in the U.S.) examines a crime in one area, integrates it into historical patterns, and calculates, hour by hour, where it is most likely to occur next. Since the main inputs into the model are the type, location, and time of the commission of each crime, PredPol seems to be neutral to race and ethnicity. However, geography can be an accurate proxy for ethnicity because poor neighborhoods are primarily inhabited by ethnic minorities. This becomes especially problematic when "nuisance" crimes (vagrancy, aggressive panhandling, selling and consuming small quantities of drugs) are incorporated into the program, which are endemic to many impoverished areas. As such, the system amplifies racially biased policing.[28] In January 2020, the New York Times reported the first case of wrongful arrest (of a black man) due to racial bias in AI-based facial recognition technologies.[29] Risk assessment tools in the US criminal justice system have also been criticized as inherently unfair because they disproportionately target minority individuals and communities. One study has shown that COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) was twice as likely to label black offenders as high-risk than whites.[30]

Further examples of discrimination from other areas are abundant: Google's face recognition system identified black people as gorillas, Microsoft's chatbot Tay became a racist neo-Nazi in

[24] Ashraf (2020) 171.

[25] Affonso et al. (2021) 161.

[26] Alikhademi et al. (2022); Mancuhan and Clifton (2014).

[27] Alikhademi et al. (2022) 2.

[28] O'Neil (2016) 84–90.

[29] Hill (2020b).

[30] In fact, in these cases there was no element of specific targeting involved. The questionnaire informing the algorithmic assessment does not even mention race. However, it does include items that can be correlated with race, such as poverty, joblessness and social marginalization. These items work as reliable proxies for minority membership. Angwin et al. (2016) For more information on proxy discrimination see, Barocas and Andrew (2016) 691–92.

one day, AI systems categorize white men with a 1% error but show a 35% error rate in the case of dark-skinned women.[31] Racial discrimination was found when searching for people's names in Google. Names that are associated with persons of black ancestry were most likely to show arrest-related contents compared to other races, regardless of whether or not they were actually involved in police arrest.[32] Studies show that AI content moderation tools for hate speech detection were 1.5 times more likely to flag tweets written by African-Americans as offensive, and also discriminated against other marginalized groups who use non-Caucasian forms of speech.[33]

In China, more extensive surveillance based on facial recognition and other biometrics is being deployed in relation to ethnic minorities such as the Uighurs.[34] Even those who argue that algorithms can in fact help detect hidden forms of discrimination admit that "epistemic opacity may undermine a sound ethical examination of complex algorithms and this, in turn, can intensify issues of discrimination".[35]

## 3.2. Women

The above examples revealed how algorithms can discriminate against racial and ethnic minorities. Yet another group of society is disproportionately affected by the use of new technologies and especially AI: women.[36] In addition to the prevailing gender gap in the use of mobile internet and unbalanced access to digital devices in general, it has been found that job search engines systematically present women ads for lower-paying jobs, thus helping to create a glass ceiling. Classic examples include Amazon's former hiring tool that downgraded the resumes of women,[37] but LinkedIn and Google were also reported because they did not display high-paying jobs as frequently for searches by females as for males.[38]

Difficulties of representing the gender dimension arise in personalized medicine, too. In the screening and definition phase of clinical trials, some choices are consistent, and women's accounts about their discomfort or pain are taken less seriously than men's.[39] The main reason for gender bias is that most of the input data come from men. Women continue to be underrepresented in medical research, because these studies often involve long periods under clinical supervision that conflict with caring and other traditional responsibilities of women. Furthermore, "women (and more specifically those from a lower socioeconomic status and/or those belonging to ethnic minority groups) show high levels of ethical concern regarding the

[31] Devillers, Fogelman-Soulié and Baeza-Yates (2021) 76–77.

[32] Anshari et al. (2023) 711.

[33] Ashraf (2022) 773.

[34] Smith and Miller (2022) 168.

[35] Heinrichs (2022) 153.

[36] The equal rights of men and women are affirmed in the preamble of the UDHR; Art. 3 of ICCPR; Art. 3 and Art. 7. (a) (i) of ICESCR; and Art. 18(3) of ACHPR. In addition, the Convention on the Elimination of All Forms of Discrimination against Women has special relevance for the rights of women.

[37] Devillers, Fogelman-Soulié and Baeza-Yates (2021) 79.

[38] Stanila (2018).

[39] Carnevale et al. (2023) 835.

participation in medical research due to the history of harmful medical experiments, often targeting their reproductive health”.[40]

### 3.3. Children

Childhood enjoys special protection in the international system of human rights.[41] The ICESCR especially calls on State parties to provide for the healthy development of the child, including mental health. Yet, the ramifications of AI on minors’ rights receive inadequate attention in policy-making,[42] especially if we look beyond the context of education which is an obvious concern for children. Fosch-Villaronga et al. call attention to the fact that more and more smart connected toys – reactive and environmental-adaptable devices connected to the internet – use machine-learning, such as facial recognition. The negative side-effects of these toys include violations of privacy, the commercialization of play along with the commodification of children’s identities, the reinforcement and exacerbation of gender stereotypes and inherent biases in favor of certain ethnicities or social classes, dependency, social isolation, weakening children’s ability to differentiate between tangible and intangible reality, etc.[43] Privacy issues and the danger of abuse (sexual harassment, bullying, etc.) arise also when photos, videos, or other personal data about children are shared on the internet by their parents, relatives, teachers, peers, or the children themselves.[44]

## 4. THE IMPACT OF AI ON CIVIL AND POLITICAL RIGHTS

### 4.1. Privacy

The core of the right to *privacy*[45] or the right to *respect for private life*[46] is the right to be left alone, and is essentially meant to protect individuals against arbitrary interference by public authorities. However, international courts tend to interpret the concept in broad terms, entailing both negative and positive obligations relevant for building one’s own personality and relationship with other individuals and to the world.[47] Privacy is strongly connected to human dignity, physical, psychological and moral integrity, personal autonomy, self-determination, as well as personal and social identity.[48] The right to privacy is of paramount importance as it enables

[40] Carnevale et al. (2023) 836.

[41] UDHR, Art. 25(2); ICCPR, Art. 24; ICESCR, Art. 10, Art. 12(2) a); ACHR, Art. 19, Art. 13(4), Art. 17(5); ACHPR, Art. 18(3). In addition, the UN Convention on the Rights of the Child provides for a thorough protection.

[42] Fosch-Villaronga et al. (2023) 133.

[43] Fosch-Villaronga et al. (2023) 133–39.

[44] Pavlovic, Randelovic and Ivanovic (2018); Tilovska-Kechedji and Rakitovan (2018).

[45] UDHR, Art. 12: “No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.” Cf. ICCPR, Art. 17. and ACHR, Art. 11.

[46] ECHR, Art. 8. (The ACHPR does not contain the right to privacy or personal life.)

[47] Affonso et al. (2021).

[48] Çınar (2021); Mantelero and Esposito (2021).

individuals to enjoy other rights and to participate in political, economic, social, and cultural life. Thus, violations or abuses of privacy might affect the enjoyment of other human rights, including freedom of expression, assembly, and association.[49]

The internet, digitalization, and smart technologies have fundamentally changed the way in which human beings relate to each other. Amazon, Facebook, Google and other companies collect an unprecedented amount of data from their users on a daily basis. In fact, "[o]ver the last decade, humans have produced each year as much data as were produced throughout the entire history of humankind".[50] Not only the information of users is being digitized, but also the environment around them, with houses and domestic appliances connected to the internet and having the capacity to make services online (Internet of Things). The public space is also being digitalized with ubiquitous CCTV cameras, facial and vehicular recognition software, traffic monitoring, and surveillance technologies.[51] All this has reframed the traditional meaning of privacy, and even those human rights systems that did not previously recognize data protection as an autonomous right had to adapt themselves to face the new challenges in connection with the right to privacy.[52] These challenges include "non-consensual data collection by consumer products, using AI to identify individuals, AI profiling[53] of individuals based on population-level data, AI-generated inferences of information and identity based on non-sensitive data, and AI decision making".[54]

AI applications have also been widely used in combating the Covid-19 epidemic, including surveillance technology to track the propagation of the coronavirus, disaster prevention, rapid reaction, and improved communication between the government, public, business associations, and other stakeholders.[55] De Almendra Freitas, Pamplona and de Oliveira call attention to the possible damage caused by profiling and contact tracing techniques when personal data was gathered online without the individual's knowledge or consent, or when legislation was approved to make consent mandatory.[56]

Privacy issues frequently arise in psychiatric contexts where behavioral monitoring systems are applied,[57] or in the workplace, with companies increasingly utilizing data analytics to monitor the actions and performance of their employees, sometimes even extending to

[49] Human Rights Council (2017).

[50] Carnevale et al. (2023) 829.

[51] Continuous and invasive monitoring in public spaces, schools and workplaces, or via wearable devices, mobile applications, GPS, etc. also adversely affects the *freedom of movement*. Mantelero and Esposito (2021) 15. Cf. UDHR, Art. 13; ICCPR, Art. 12; Protocol No. 4. to the ECHR, Art. 2; ACHR, Art. 22; ACHPR, Art. 12.

[52] Affonso et al. (2021). For the relevant case-law of the ECtHR, see European Court of Human Rights (2022), especially the sections on "Modern-day challenges of data protection", "Storage of personal data for the purposes of combating crime", and "Data collection by the authorities via covert surveillance". For a recent analysis of AI-related cases, see Szappanyos (2023).

[53] Profiling is the act of suspecting or targeting a person on the basis of observed characteristics or behavior.

[54] Ashraf (2022) 775.

[55] Trew (2020); Yuan (2020); Anshari et al. (2023).

[56] de Almendra Freitas, Pamplona and de Oliveira (2022) 1313.

[57] Ramli and Zakaria (2014).

non-job related behavior.[58] As the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights emphasized in his 2016 report, "the growing processing power of computers exacerbates the challenge as information can be harvested from multiple sources, processed and re-processed and then sold on. In fact, the entire business model of the most successful companies directly impinges upon the right to privacy".[59] Unfortunately, data protection protocols – even ones as advanced as the European Union's 2016 GDPR – do not fully address the imbalance of power between the data controllers and the data subjects.[60]

Mass surveillance and bulk interception of communications and metadata are also increasingly used in policing, and are mainly justified by the protection of national security and in particular the fight against terrorism and other serious crime, such as cybercrime, drug trafficking, human trafficking, and the sexual exploitation of children.[61] One of the most important and rapidly developing AI technologies currently available for this purpose is biometric facial recognition, involving the automated comparison of facial features to identify suspects from photos and closed circuit television. The technology is widely used in the UK, the U.S. and Australia, also at international airports and border control systems, integrating facial images from passports, driver's licenses and even social media into a national database for use by law enforcement and other government agencies.[62] Guo and Kennedy explain how automatic facial recognition is susceptible to misuse, leading to various individual (dignity, privacy, autonomy) and collective (trust, transparency) harms.[63] The danger lies in using the biometric information as a key to gather more information about the same person from other databases, with the capacity of integrating automatic facial recognition data and other data (phone metadata, internet history, financial, medical and tax records, etc.). This tendency of continually expanding the scope of collection and use of personal information ("data creep") is

> "clearly against the principle of purpose coherency in data processing. It not only directly violates data privacy [but] fundamentally endangers society as a whole – nurturing a culture of authoritarian control of objects and eroding the culture of policing by consent of autonomous, dignified subjects".[64] Furthermore, the pervasiveness and routinizing (normalization) of surveillance may lead people to consider it acceptable, even favorable, also in contexts where surveillance is used for establishing, maintaining, and expanding power. Normalization dynamics can contribute to a 'slippery slope trajectory' that weakens privacy and civil liberties.[65]

[58] Ebert, Wildhaber and Adams-Prassl (2021).

[59] Office of the Special Rapporteur for Freedom of Expression of the IACHR, 'Standards for a Free, Open and Inclusive Internet' (2016) para. 199. Cited by Affonso et al. (2021) 152.

[60] Vanberg (2021).

[61] Rusinova (2022).

[62] Smith and Miller (2022).

[63] Guo and Kennedy (2023). That is why the Artificial Intelligence Act proposed by the European Commission prohibits "the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary" for specifically given objectives, usually involving the prevention or punishment of serious crime. European Commission (2021) Article 5.1.*d*.)

[64] Guo and Kennedy (2023) 411.

[65] Selinger and Ree (2021).

It is not only authoritarian states that are prone to the normalization of surveillance, yet China shows an alarming example.[66] It has established a social credit system that rewards or punishes citizens on the basis of social norm compliance or non-compliance, facilitated by an extensive biometric surveillance network.[67] Elsewhere too, governments and law enforcement agencies work closely with AI companies, such as Clearview AI, to develop advanced surveillance systems to match photos of unknown individuals to their online images.[68] Smith and Miller correctly underline that privacy violations "on a large scale can lead to a power imbalance between the state and the citizenry and, thereby, undermine liberal democracy itself".[69]

Revisiting the traditional trade-off between privacy and security in the context of mass surveillance and bulk interception of communications, Rusinova warns against the danger of securitization, that is "the transformation of the otherwise exceptional state of emergency into something commonplace".[70] She further challenges the presumption that there is a direct relationship between the level of protection and the amount of the data collected from individuals. The general shortcomings of predictive analytics (which uses historical data to predict future events) include the inaccuracy of the raw data, human error, the reflection of the biases and values of programmers, the opacity of results (the "black box" phenomenon), etc., which may not only violate the right to privacy, but may also be incompatible with due process, the prohibition of discrimination, the presumption of innocence, and the reasonability of suspicion.[71]

## 4.2. Criminal justice and procedural guarantees

As anticipated in the previous section, AI systems are increasingly used in practically all phases of criminal justice, in both crime prevention via predictive and automated policing (profiling people and areas and to predict supposed future criminal behavior or occurrence of crime) and criminal proceedings, from investigation through predictive sentencing and risk assessment to the execution of penalties.[72] AI-based tools do not only raise privacy concerns but have been demonstrated to infringe on *liberty and security*,[73] *fair trial* (or *due process*) rights,[74] and disproportionately target marginalized groups.[75]

In the context of policing, AI is used to deduce crime correlations from countless data categories, leading to the construction of a profile based on which individuals may be identified

[66] Cataleta and Cataleta (2020) 56–58.

[67] Smith and Miller (2022).

[68] Hill (2020a).

[69] Smith and Miller (2022) 172.

[70] Rusinova (2022) 750.

[71] Rusinova (2022).

[72] For a very thorough analysis, see Quattrocolo (2020).

[73] ICCPR, Article 9(1): "Everyone has the right to liberty and security of person. No one shall be subjected to arbitrary arrest or detention. No one shall be deprived of his liberty except on such grounds and in accordance with such procedure as are established by law." Cf. UDHR, Art. 3; ECHR, Art. 5; ACHR, Art. 7; ACHPR, Art. 6.

[74] UDHR, Arts. 9-11; ICCPR, Art. 14; ECHR, Art. 6; ACHR, Art. 8; ACHPR, Art. 7.

[75] Alikhademi et al. (2022).

as potential suspects.[76] In addition to the adverse effect of profiling on the prohibition of *arbitrary arrest* (UDHR, Art. 9) and the *right to be presumed innocent until proved guilty* (UDHR, Art. 11.1), over-reliance on relative probability, lack of profile specificity, and potential for inaccurate data to infiltrate predictions are identified in research as possible risks. Furthermore, the employment of a predictive profile lessens objectivity in officer discretion, since "information available to the officer accordingly adjusts his/her perception of context and affects the application of the reasonable suspicion standard".[77] Predictive policing algorithms also provide statistics about the prevalence of certain crimes (such as burglaries or thefts) in certain areas, which can lead to increased patrolling of these hot-spot areas. Biasing effects can arise from incorrect, partial, or non-representative data.[78]

Regarding criminal proceedings, due process or fairness does not only cover the right of the accused to "a fair and public hearing by a competent, independent and impartial tribunal established by law" (ICCPR, Art. 14). In a broad sense, it also includes the presumption of innocence, the right to defense (including the right to access a lawyer and the principle of equality of arms between the prosecution and the defense), the prohibition of admitting evidence collected in violation of procedural guarantees ("the fruit of the poisonous tree"), etc. Many of these principles can be affected by AI technologies. Hacking techniques, for instance, have the advantage of more effective collection of information which can be used either in the investigation phase or/and as evidence at the trial stage. However, they may violate the right to privacy (as discussed in the previous section) and the principle of *equality of arms* when the accused has no chance to challenge the correctness or the selection of the automatedly generated evidence used against him.[79]

AI is also used in forensic investigations to help record witness and victim statements after an incident,[80] and to assist judges in making judicial decisions. Risk assessments tools are now widely used in the U.S., Canada and the UK to inform decisions about pre-trial detention, setting bail, the duration of prison sentences, and parole.[81] In addition to the concerns regarding traditional legal guarantees mentioned above, Greenstein also questions "the extent to which the

[76] Blount (2022).

[77] Blount (2022) 1.

[78] Degeling and Berendt (2018).

[79] Quattrocolo (2020) 90. This was one of the issues raised in the Einarsson case before the ECtHR, where the prosecution used an e-Discovery system (Clearwell) to sort out evidence. The applicants complained that their defense had not been given access to the vast amount of data collected during the investigation phase and were unable to have a say in the prosecution's electronic sifting of that data in order to gather relevant information for inclusion in the investigation file. *Sigurður Einarsson and Others v. Iceland*, App no 39757/15 (ECtHR, 4 June 2019).

[80] Minhas, Elphick and Shaw (2022).

[81] Recidivism risk assessments are increasingly commonplace in the U.S. in presentencing investigation reports (PSIs). In a 2016 case, the Wisconsin Supreme Court held that a trial court's use of COMPAS (an algorithmic risk assessment tool, already mentioned in Section 3.1) in sentencing did not violate the defendant's due process rights even though the methodology used to produce the assessment was a trade secret and as such unknown to both the court and the defendant. The state Supreme Court admitted that COMPAS provides only aggregate data on recidivism risk for groups akin to the offender, but they emphasized that since the PSI would not serve as the sole foundation for a decision, a sentencing process incorporating a COMPAS assessment would still retain enough individualization. In any case, the Court warned that judges must proceed with caution when using such tools. *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), reported by Harvard Law Review (2017).

judiciary, relying on AI developed by private corporations, can be deemed independent" (*independence of the judiciary* being an essential part of the right to a fair trial).[82]

Both within and outside the field of criminal justice, *the right to an effective remedy*[83] may be violated by AI. An effective remedy does not only mean judicial protection, but can also be provided by administrative or legislative authorities, or by any other competent authority. The point is that the decision must be individual and reasoned, based on a careful analysis of the specific context. However, automated decision-making processes and algorithmic data processing techniques might lack just these elements. Challenges include "the opaqueness of the decision itself, its basis, and whether the individuals have consented to the use of their data in making this decision, or are even aware of the decision affecting them. The difficulty in assigning responsibility for the decision also complicates individuals' understanding of whom to turn to address the decision".[84]

## 4.3. Political liberties

Political liberties are understood here as rights that enable an individual to participate in a democratic society, acknowledging of course that they are also critical for personal self-development and autonomy. In the following paragraphs, the impact of AI on the freedom of expression, the freedom of thought, conscience and religion, the freedom of assembly, and the freedom of association will be discussed.

Perhaps most obviously in the world of social media, the *freedom of expression*[85] comes to mind. In the digital space, freedom of expression and the right to privacy are closely interrelated.[86] We have a right to protect our personal data, whereas the public has a right to access information. We have the right to hold and impart opinions, but certain expressions (such as hate speech) are prohibited[87] because they violate human dignity and privacy. Strange as it may seem, AI also affects our freedom to hold opinions without interference. Social media platforms, news websites, video streaming services and search engines utilize algorithms (like Facebook's EdgeRank or Google's PageRank) to recommend content to users based on their preferences, behaviors, and historical interactions. The intention of this process – *content display* or *content curation* – is to enhance user engagement and provide a more tailored user experience. As the Special Rapporteur on Freedom of Expression explained, "AI applications determine how widely, when and with which audiences and individuals content is shared. […] AI in the field of content display is driving towards greater personalization of each individual's online

[82] Greenstein (2022) 314.

[83] UDHR, Art. 8; ICCPR, Art 2.3; ECHR, Art. 13; ACHR, Art. 25; ACHPR, Art. 7.1.

[84] Council of Europe (2018) 24.

[85] UDHR, Art. 19: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers." Cf. ICCPR, Art. 19; ECHR, Art. 10; ACHR, Art. 13; ACHPR, Art. 9.

[86] Çınar (2021).

[87] ICCPR, Art. 20: "1. Any propaganda for war shall be prohibited by law. 2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law." Cf. ACHR, Art. 13.5.

experience [which] may also minimise exposure to diverse views, interfering with individual agency to seek and share ideas and opinions across ideological, political or societal divisions".[88]

To put it simply, if a user selects online articles with a specific viewpoint, search algorithms will offer them other similar articles, leading to the filter bubble or echo chamber effect, which only amplifies our bias, limits the diversity of our opinions, and creates polarization.[89] Hoax, fake, misleading and unverified news, and deepfake videos[90] – which can spread all too easily through social media – also interfere with our freedom of expression.[91] "In an AI-governed system, the dissemination of information and ideas is governed by opaque forces with priorities that may be at odds with an enabling environment for media diversity and independent voices".[92] This leads us to another issue of online communication: *content moderation*. Social media platform providers apply AI to remove information which breaches their terms of services, including hate speech, abusive content or spam. However, content moderation technologies still fail to understand context, and are limited in their ability to take into consideration variations of language cues, meaning, cultural and linguistic specificities, thus posing risks to users' right to free speech and access to information.[93] To illustrate the extent to which algorithmic content display and content moderation can interfere with our lives, the term 'algorithmic censorship' has been coined.[94]

As for the *freedom of religion*,[95] Ashraf's comprehensive study shows how the internet has become central to the faith of millions: ranging from online forums and scripture discussion groups, community-building religious memes, blogging as a form of observance, and faith-promoting online rituals, to utilizing apps or social media for teaching and worship.[96] Religious radicalization and online religious hate speech that social media has facilitated are just two of the many concerns. As UN Special Rapporteur on Freedom of Religion or Belief Ahmed Shaheed stated in his 2019 report, "the emergence of 'digital authoritarianism' through increased surveillance, encroachment on privacy and broad restrictions on expression related to religion or belief has rendered cyberspace a perilous place for dissenters and religious minorities. […]

[88] Kaye (2018) 6–7.

[89] Although there are studies indicating that this may not necessarily be the case. See the collection of these in Haidt and Bail (ongoing) 34–51.

[90] Deepfake is an AI application that alters visual content by digitally inserting the face and voice of one person into a video of another person, creating highly realistic but entirely synthetic video content. Livingston and Risse (2019) 144. Maas warns that deepfake techniques may also adversely affect the probative value of evidence in investigations and judicial proceedings. Maas (2019) 14.

[91] Anshari et al. (2023).

[92] Kaye (2018) 12.

[93] Dias Oliva (2020); Kaye (2018) 8.

[94] Ashraf (2022) 770.

[95] UDHR, Art. 18: "Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief, and freedom, either alone or in community with others and in public or private, to manifest his religion or belief in teaching, practice, worship and observance." Cf. ICCPR, Art. 18; ECHR, Art. 9; ACHR, Art. 12; ACHPR, Art. 8.

[96] Ashraf (2022).

Digital applications [are] being used to report allegations of blasphemy, and digital footprints can be used to assess compliance with faith-related observances."[97] Furthermore, as mentioned in connection with the freedom of expression, algorithmically customized content display as a form of subconscious and personalized persuasion influence how we think and perceive the world, and thus constitutes a violation of freedom of thought.[98]

Freedom of religion also includes the right not to disclose one's religion or belief to anyone, including state authorities. However, AI can lead to public disclosure of an individual's religious preferences by recommending them to certain friends, advertisers, merchants, or other individuals. This is especially problematic when someone is member of a targeted religious minority, or in cultures where women's online participation is discouraged. These people may become less likely to use social media if they perceive a risk of being exposed. Furthermore, jurisdictional content moderation rules where certain religious content is considered blasphemous may also deter individuals from exercising their right to freedom of religion online. Via content moderation, AI can also influence how religion or belief manifests online, eliminating entire conversations, pages, videos, events, and other content from social media.[99]

The *right to peaceful assembly and association*[100] can also be exercised through new technologies. Each day, billions of people come together and/or form groups peacefully on the internet to discuss politics or daily life, organize protests, gather signatures for petitions, raise funds, or meet on forums. The use of online spaces to organize protests "has been well-known since the 2009 Iranian Green Movement popularized the phrase 'Twitter Revolution' and became the first major world event broadcast worldwide almost entirely via social media".[101] The #MeToo movement brought millions of women together online to share their stories of sexual harassment, and over 6 million people assembled to sign a petition to annul Brexit – apparently the largest petition ever delivered to parliament.[102]

Threats to the freedom of assembly/association in cyberspace include shutdowns, internet censorship, and excluding individuals from the internet, especially at key political moments. Of all the possible threats, AI is currently used in content display, content moderation, and surveillance. In addition to what has already been said above in relation to freedom of expression and freedom of religion, AI-driven personalization may also minimize how and where individuals assemble online, and what types of associations can be formed. Via content moderation, AI can influence assembly by eliminating conversations or events from social media. AI-based surveillance, even if used for legitimate purposes such as to anticipate epidemics or identify potential terrorist threats, makes online spaces vulnerable, and as a result, some individuals – especially those belonging to high-risk groups – will become less likely to

[97] Cited by Ashraf (2022) 769.

[98] Ashraf (2022) 769. As Alegre (2017) and Aswad (2020) explain, there are three key elements to the right to freedom of thought and the related right to freedom of opinion: the right to keep our thoughts and opinions private; the right not to have our thoughts and opinions manipulated; and the right not to be penalized for our thoughts and opinions. The current business model for big tech ("surveillance capitalism") has implications for all three aspects.

[99] Ashraf (2022) 772–73.

[100] UDHR, Art. 20; ICCPR Arts. 21-22; ICESCR Art. 8; ECHR Art. 11; ACHR, Arts. 15-16; ACHPR, Arts. 10-11.

[101] Ashraf (2020) 167.

[102] Ashraf (2020) 167.

attend gatherings.[103] In the offline space, the use of drones by law enforcement authorities to monitor public demonstrations can have a similar deterrent effect.[104]

## 4.4. The right to free elections

The above-mentioned mechanism of algorithms and automated recommender systems creating filter bubbles “in which individuals only see pieces of information that confirm their own opinions” also impacts democracy at its core: the process of elections.[105] The *right to vote* freely at genuine periodic elections is one of the participatory rights that ensure that everyone can take part in the public affairs of their own country.[106] In this context, AI and other forms of technology can be used in various forms, from voting online to the – as yet – hypothetical idea of fully replacing the legislature’s human representatives with algorithms. Although there is some positive reception of these possibilities in the scholarship,[107] most authors are worried about the negative impacts that the use AI may bring about in the context of elections.

Mainz, Sønderholm, and Uhrenfeldt argue that if one has access to massive amounts of data about specific electors, AI makes it possible to infer with high levels of accuracy individual electors’ past voting behavior, therefore “the secret ballot is now much less effective at protecting individual voters against social ostracism and social punishment”.[108] Google and Facebook, for instance, “routinely employ AI models specifically designed to predict individual people’s voting behavior. Links we click on online, news we read, posts we ‘like’ on social media sites etc., generate an accurate digital picture of our political orientation”.[109]

After predicting how individual electors are going to vote in a forthcoming election, AI can be used in microtargeting to persuade them to vote in a particular way. A recent example involves Cambridge Analytica, a London-based data-mining firm with connections to former US President Trump, which lifted the Facebook profiles of tens of millions of users without their permission in order to manipulate elections.[110]

Algorithms can also be deployed to spread fake news which can undermine the ability of voters to reach informed conclusions on political issues, as occurred in the U.S. during the 2016 elections, when Russia’s Internet Research Agency manipulated Facebook’s algorithms to promote dis/misinformation, anti-immigrant sentiment, and hate speech.[111] In fact, it seems warranted to believe that “elections may be won not by the candidates with the best political argument, but by those who use the most efficient technology to manipulate voters, sometimes emotionally and irrationally”.[112]

[103] Ashraf (2020).

[104] Mantelero and Esposito (2021) 17.

[105] Council of Europe (2018) 30.

[106] UDHR, Art. 21; ICCPR, Art. 25; Protocol No. 1 to the ECHR, Art. 3; ACHR, Art. 23; ACHPR, Art. 13.

[107] Burgess (2021).

[108] Mainz, Sønderholm and Uhrenfeldt (2022) 3.

[109] Mainz, Sønderholm and Uhrenfeldt (2022) 4.

[110] Anderson (2018).

[111] Frenkel and Benner (2018).

[112] Council of Europe (2018) 31.

# 5. THE IMPACT OF AI ON SECOND-GENERATION HUMAN RIGHTS

## 5.1. Labor rights

Labor rights relate to labor relations between workers and employers, including the right to work, the enjoyment of just and favorable conditions of work (such as fair wages, equal remuneration for equal work, safe and healthy working conditions, reasonable working hours), and the right to form and join trade unions (a special aspect of the freedom of association).[113] Science-fiction writers have long fantasized about what our "robotic future" might look like, ranging from humans being liberated from performing tedious or repetitive tasks (such as providing sanitation services, monitoring security feeds, or working on an assembly line) to being subjugated by killer robots.[114] Instead of idealistic or dystopian visions, it is more reasonable to see AI as a potentially disruptive technology with a multi-layered and comprehensive impact on the structures and institutions of labor markets, including on jobs, working conditions, organization of work, and social dialogue. The benefits of creating new productive activities and eliminating tiresome tasks must be counterbalanced by the challenges of unemployment, inequality, unfair competition, and unbalanced distribution of value.[115]

Most authors agree that the general presumption of AI replacing human labor with robots, leading to massive unemployment, is simply not warranted. AI is already widely used in a variety of industries, in legal work, healthcare, agriculture, etc.[116] It is estimated that around 70% of businesses will be using at least one type of AI technology by 2030, while less than half of large businesses will be deploying the full scale of AI technologies.[117] This will require a large amount of workforce with information technology skills. As a result, employment patterns in terms of age, education, and income will profoundly change. According to Borenstein, "in general, the jobs resulting from technological innovations typically draw from a different skill set than those that are lost", which means that certain groups of the population may be disproportionately affected, including older workers (who may be difficult to re-train) and workers without advanced educational degrees.[118]

Another field of applying AI in the workplace is algorithmic management. This can be present in multiple stages of employment decision-making: job advertisements, matching tools, shortlisting, conducting pre-employment tests and interviews (even analyzing facial expressions, voice and word choices, eye contact), hiring (from CV selection to automation of the full hiring process), career coaching, optimization of the labor process (through the tracking of worker movements), evaluation of employees (through rating systems), automated scheduling of shifts,

[113] UDHR, Arts. 23-24; ICESCR, Arts. 6-8; ACHPR, Art. 15. The ECHR does not contain second-generation human rights, with the only exception of the right to education. The ACHR has one article (Art. 26) on economic, social and cultural rights, whereby states parties undertake to achieve progressively the full realization of these rights. In both the European and the Inter-American systems of human rights protection there are separate treaties dedicated to economic, social and cultural rights.

[114] Borenstein (2011).

[115] Harayama, Milano, Baldwin et al. (2021) 62.

[116] Deranty and Corbin (2022).

[117] Arifin (2021) 99.

[118] Borenstein (2011) 90.

coordinating customer demand with service providers, monitoring of worker behavior, algorithmic incentivization (through algorithm-based "nudges" and penalties), and even firing workers with algorithmically determined low productivity scores.[119]

Algorithmic management raises obvious issues of privacy, not just at the workplace, but also at home, especially given the shift towards remote working which resulted from the Covid-19 pandemic.[120] Furthermore, it might reproduce discriminatory practices and perpetuate societal biases at work. For instance, through social media platforms, AI can provide job advertisements to targeted audiences and enable businesses to personalize recruitment. However, in these advertisements, "search engines may deliver job postings on well-paying technical jobs that are targeted at men only, possibly discriminating against women job-seekers".[121] Potential sources of bias are manifold. At the input stage, bias may be caused by the inaccuracies or non-representativeness of the training data from the organization itself or from external sources such as LinkedIn. Data may be mislabeled "based on the employer's prejudiced interests in favour of certain groups of candidates which may in turn influence the recommendations offered by the AI system… [T]he selection of people in the training data may be biased (e.g., the data does not include certain marginalised groups) or the selection of the attributes of the people are incomplete (e.g., where it is difficult to collect data such as [persons] suitable for specific jobs)".[122] The collation of data from certain online sites might also be biased, skewing the data in favor of individuals using such sites.[123]

## 5.2. Social rights

One of the most important social rights is the *right to health*, including medical care.[124] AI is already widely present in health care services, for example in automatic acute care triaging and chronic illness management, including remote monitoring, preventative treatment, patient intake, referral help via AI-enabled Telehealth,[125] and personalized and precision medical practices. In the definition of Carnevale et al., personalized medicine means "the consideration of the genotypical and phenotypical (environment, lifestyle, social relationships, etc.) characteristics of each individual receiving health care".[126] Although it seems that technological developments in healthcare have improved the overall quality of medicine, as in other fields, the use of AI brings about dangers in this context as well. Risks include obsolescence for human theorization in medical diagnoses, losing a holistic vision of society's health problems, data determinism based

[119] Chan (2022); Deranty and Corbin (2022); Ebert, Wildhaber and Adams-Prassl (2021).

[120] Deranty and Corbin (2022) 9.

[121] Chan (2022) 2.

[122] Chan (2022) 3.

[123] Chan (2022) 3.

[124] The ICESCR (Art. 12) and the ACHPR (Art. 16) provide for the enjoyment of the highest/best attainable standard of physical and mental health, whereas the UDHR (Art. 25.1) considers food, clothing, housing, medical care and social services as preconditions for a standard of living adequate for everyone's health and well-being.

[125] Anshari et al. (2023) 707–8.

[126] Carnevale et al. (2023) 831.

on provisory and incomplete inference, and a totalitarian digital society based on biomedical data control.[127]

Impacts on the right of patients to autonomy are complex. On the one hand, AI-driven systems can empower patients through self-monitoring and self-management of health, including exercise promotion, medication adherence, chronic disease self-care management and daily diabetes routines.[128] On the other hand, patient autonomy might be affected by undue influence, e.g. by "refusing to cover certain digital medicine costs or by encouraging patient to use digital devices through financial incentives, insurance companies might constrain and limit patient's choices".[129] Furthermore, the extensive use of AI in making decisions regarding diagnosis and therapy may lead to errors and concerns about the right of patients to make decisions about their own treatments.[130]

A further risk is presented by prejudice, which can also appear in health algorithms.[131] For example, Obermeyer et al. found evidence of racial bias in one widely used algorithm in the US health care system. It turned out that black patients who were assigned the same level of risk by the algorithm were actually sicker than white patients. This bias occurs because the algorithm uses health costs as a proxy for health needs: since less money is spent on black patients who have the same level of need, the algorithm falsely concludes that black patients are healthier than their equally sick white counterparts. The authors estimated that this racial bias reduces the number of black patients identified for extra care by more than half.[132]

The Covid-19 pandemic serves with a current and convenient topic for exploring the impact of AI on the right to health. Wakunuma, Jiya and Aliyu presented the use of AI-based robots in Africa to fight Covid-19 infection rates of health staff while treating Covid-19 patients. The robots screen temperature and read other vital signs, deliver video messages to health care practitioners, detect people not wearing masks and instruct them on how to wear them. However, they may offer solutions based on incomplete data or decisions based on unclear patterns, so misdiagnosis is a possibility. Furthermore, as the authors explain, especially in the least developed countries, there are often no guidelines or provisions to address the issues of responsibility, and AI-based technologies are used without consent from most of the population due to poor literacy, lack of awareness of rights and weak enforcement, which again raises issues of privacy.[133]

The recent study of Galetsi, Katsaliaki and Kumar[134] presents a wide range of measures of how Big Data analytics and AI have been involved in the management of Covid-19, including the identification of Covid-19 positive patients, predicting and monitoring the spread of the virus in the population, suggesting policy decisions, predicting mortality risk, optimizing Covid-19 patient management, creating warning systems for society, detecting the probability

[127] Carnevale et al. (2023) 833–35.

[128] Mirbabaie et al. (2022).

[129] Žaliauskaitė (2021) 579.

[130] Mirbabaie et al. (2022).

[131] Parfett, Townley and Allerfeldt (2021).

[132] Obermeyer et al. (2019).

[133] Wakunuma, Jiya and Aliyu (2020).

[134] Galetsi, Katsaliaki and Kumar (2022).

of false negative or false positive Covid-19 cases, tracking and communicating individuals' vital signs to their doctors, measuring the spread of misinformation to verify the credibility of data from social media, etc. The authors identified privacy concerns, the bias of output and the spread of false information as the main challenges. Specifically, genetic information and the DNA sequences of the virus are included in electronic databases without any control over who is allowed to use the information and for what purpose. Biased outputs may result from hastily collected "dirty" data, the circumvention of validation model checks, and the composition of teams developing AI-based Covid-19 applications which may not adequately reflect the diversity of the population.[135]

In the context of privacy and Big Data, the ECtHR noted that disclosure of a person's medical data may endanger the health of a person or community.[136] Moreover, AI is not able to give hope and comfort to a patient or communicate emotionally, thus cannot replace human contact and empathy.[137]

The *right to social security*, including social insurance[138] is also affected by the use of digital technologies, not necessarily for the better. Philip Alston, the former UN Special Rapporteur for Extreme Poverty and Human Rights warned in his 2019 report, "as humankind moves, perhaps inexorably, towards the digital welfare future, it needs to alter course significantly and rapidly to avoid stumbling, zombie-like, into a digital welfare dystopia".[139] Alston drew attention to concerns with the rise of automated eligibility assessments, calculation of welfare benefits, fraud detection, and risk scoring. In his opinion, the right to human dignity is at particular risk: coupled with the rigid and robotic application of rules, "the way in which determinations are framed and communicated may be dehumanized and allow no room for meaningful questioning or clarification".[140] He further mentioned the lack of accuracy and the fact that technologies overlook structural disadvantages based on inequality, poverty, and racism.

Racial discrimination in the practice of insurance companies goes back to the late nineteenth century, when a renowned statistician named Frederick Hoffmann published his study claiming

[135] Galetsi, Katsaliaki and Kumar (2022) 7–8.

[136] "The disclosure of [data about a person's HIV infection] may dramatically affect his or her private and family life, as well as social and employment situation, by exposing him or her to opprobrium and the risk of ostracism. For this reason it may also discourage persons from seeking diagnosis or treatment and thus undermine any preventive efforts by the community to contain the pandemic". *Z v. Finland*, App no 22009/93 (ECtHR, 25 February 1997), para. 96. The case was referred to in Çınar (2021) 44.

[137] Žaliauskaitė (2021)

[138] UDHR, Art. 25(1): "Everyone has […] the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control." ICESR, Art. 9: "The States Parties to the present Covenant recognize the right of everyone to social security, including social insurance."

[139] Alston (2019) 21.

[140] Alston (2019) 17.

that “the lives of black Americans were so precarious that the entire race was uninsurable”.[141] Other authors worry that in the digital welfare state “the good governance triad of transparency, accountability and participation may be restricted [...], especially through the loss of reason-giving and discretion”, and that digital illiteracy can hinder access to social services.[142] *Housing rights* (explicitly mentioned in Article 25 of the UDHR and Article 11 of the ICESCR as a *sine qua non* of an adequate standard of living), just like the right to work, can also be adversely affected by automated decision-making processes, including credit scoring.[143]

In relation to the *right to food* (also explicitly mentioned in Article 25 of the UDHR and Article 11 of the ICESCR as part of an adequate standard of living), the various usages of AI in the agri-food sector from digital farming to agricultural robots may have significant benefits. In fact, State parties to the ICESCR are obliged to “improve methods of production, conservation and distribution of food by making full use of technical and scientific knowledge”, and thus ensure that everyone is “free from hunger”.[144] AI software can retrieve plenty of data from the farm, local climate and machinery, based on which they can provide farmers with forecasts and recommendations about when to seed, harvest, and sell their crops, about the health and behavior patterns of their livestock, they are able to detect plant disease, etc. AI robots and drones work relatively autonomously on farms such as picking fruit and vegetables, cleaning manure from stalls, weeding and hoeing, spraying herbicides and pesticides, measuring crop health, pruning trees, milking cows, etc.[145] Whereas the benefits of applying AI in this area seems to overshadow the risks, issues of responsibility, transparency, inequality, the weakening of human agency, and the depreciation of human labor must not be overlooked.

## 5.3. Cultural rights

In addition to the digital divide in the population caused by unequal access to modern technologies, the *right to education*[146] can be more directly affected by the application of AI. AI tools have been successfully applied to overcome barriers to learning, also in special education.[147] Regrettably, however, undesired negative effects have been observed in the context of education as well. For instance, US universities are using algorithmic systems to recommend applicants for admission. These are often customized to meet institutional preferences, and use historical data of previously admitted students. Since many elite universities have historically been attended by

[141] O’Neil (2016) 161. Other cases of discriminatory or faulty welfare algorithms from the Netherlands include the welfare freud system used by the city of Rotterdam (reported by Constantaras et al. 2023), the child care benefit scandal (“Toeslagenaffaire”, reported by Peeters and Widlak 2023) and System Risk Indication (SyRI) used by the Dutch government. In the famous *NCJM* et al. *and FNV v The State of the Netherlands (‘SyRI’)* case, the District Court of the Hague ruled that neither the legislation governing SyRI nor its use met the requirements laid down in Article 8(2) of the ECHR for an interference with the right to private life. Curiously, although the right to social security lies at the core of SyRI’s functioning, the plaintiffs did not refer to that in their claim. Rachovitsa and Johann (2022).

[142] Langford (2020) 143.

[143] McGregor, Murray and Ng (2019).

[144] ICESCR, Art. 11.2. (a)

[145] Ryan (2022) 1–3.

[146] UDHR, Art. 25; ICESCR, Arts. 13-14; Protocol No. 1. to the ECHR, Art. 2; ACHPR, Art. 17.1.

[147] Drigas and Ioannidou (2013).

prosperous white men, any model that uses these data may lead to discrimination and perpetuating past trends.[148] Furthermore, AI is increasingly being used in grading and essay scoring, sometimes in high-stakes standardized testing environments. These tools depend on the collection, storage, and analysis of a vast amount of written material, which raises the usual privacy-related concerns related to most AI systems.[149] Similarly to human beings, AI is not infallible, as evidenced by the 2020 fiasco around secondary school exams in the United Kingdom, where a controversial model downgraded students' results as opposed to their earlier, teacher-assessed grades.[150]

Experts warn that "given the growth of research into early childhood predictors of success, it is likely that such a system could be used to restrict the opportunities of students at increasingly younger ages, resulting in significant discrimination, with students coming from underprivileged backgrounds ultimately being denied opportunities because people from that background tend to have more negative outcomes. Such a system would ignore the students that overcome adversity to achieve academic and professional success, and would entrench existing educational inequalities".[151] Furthermore, since educational achievement is the precondition of social mobility, restrictions on the right to education may negatively affect the right of individuals to participate in economic, social, and public life.

The impact of AI on the *right to take part in cultural life* and *enjoy benefits of scientific progress*[152] is less direct but no less significant. The report of Access Now calls attention to the risk that AI could be used to "criminalize" certain cultures: "When members of a particular culture are disproportionately arrested or otherwise targeted by law enforcement, the behaviours and customs associated with these cultures could become linked with criminal activities. For example, a [machine learning] system analysing video or photographic footage could learn to associate certain types of dress, manners of speaking, or gestures with criminal activity, and could be used to justify the targeting of these groups under the guise of preventing crime."[153] Thus, AI technologies and surveillance may inspire "fear of being identified or suffering reprisals for cultural identity, leading people to avoid cultural expressions altogether".[154]

As for enjoying the benefits of scientific progress, it is worth recalling General Comment No. 25 of the UN Committee on Economic, Social and Cultural Rights that this expression is "not restricted to the material benefits or products of scientific advancement, but includes the development of the critical mind".[155] This brings us back to what has been said about the impact of AI on the freedom of thought and expression, offering another example for the interrelatedness of human rights. Furthermore, in an era where a vast amount of scientific knowledge is only available online, the population gap in accessing new technologies not only influences who is

[148] O'Neil (2016) 50–67.

[149] Raso et al. (2018) 49–50.

[150] Ferguson and Savage (2020).

[151] Access Now (2018) 28.

[152] UDHR, Art. 27; ICCPR, Art. 15; ACHPR, Art. 17.2.

[153] Access Now (2018) 28.

[154] Access Now (2018) 28.

[155] Cited by Shaheed and Mazibrada (2021) 113.

able to participate in the production of knowledge, but also what we consider to be "knowledge".[156]

## 6. FURTHER AVENUES FOR RESEARCH

As explained in the Introduction, this paper did not investigate the impact of AI on third-generation rights, such as the right of peoples to social, economic, and cultural development or the right to a healthy environment. In these areas, AI admittedly offers promising opportunities. Big Data gathered from a wide variety of observation points, including satellites, can be analyzed to map environmental threats and challenges to climate, ocean and marine resources, forests, land, water, air, and biodiversity, among others.[157] AI tools can automate agricultural practices and increase agricultural supply, enhance the efficiency and predictability of renewable energy, streamline energy-usage and waste-management, improve the management of water quality, quantity, and access, etc.[158] Therefore, the importance of AI in addressing poverty, climate change, and challenges affecting sustainable development cannot be underestimated. However, equal access and fair distribution must be ensured, or AI will only lead to further polarization and division in societies and globally. For instance, the research of Galetsi, Katsaliaki and Kumar reveals that less developed countries have been excluded from many AI-based solutions fighting Covid-19 because their healthcare systems were not able to adapt to modern digital technologies.[159] Future research could map the possible advantages and drawbacks of AI in the context of third-generation human rights, and clarify their relationship to sustainability, and more specifically, the Sustainable Development Goals of the United Nations.

Another topic worth exploring is the impact of artificial general intelligence (AGI), also called superintelligence, on human rights. Based on the theory of singularity, one day – perhaps in the not-too-distant future – machines may be clever enough to program and improve themselves until they become self-aware, independent from their human creators, and may even outsmart human beings.[160] AGI is a technology that does not yet exist,[161] but captures the imaginations of many.[162] Should this vision come true, our whole human-centered world as we know it could cease to exist, perhaps together with the very concept of human rights.[163] This is a scenario almost impossible to properly comprehend; nonetheless, it cannot hurt to prepare ourselves by thinking about possible implications.

[156] Shaheed and Mazibrada (2021) 113–114.

[157] Wakunuma, Jiya and Aliyu (2020).

[158] Lee (2021).

[159] Galetsi, Katsaliaki and Kumar (2022).

[160] Cataleta and Cataleta (2020) 58–59.

[161] However, Bubeck et al. (2023) believe that given the breadth and depth of GPT-4's capabilities, it could be viewed as an early (yet still incomplete) version of an AGI system.

[162] Livingston and Risse (2019).

[163] Maas mentions legal obsolescence as a possible consequence of AI technology, for example because the given rule can no longer be justified, as could be the case with the right to work. Maas (2019) 13–14.

## 7. CONCLUSIONS

The results of this paper confirm that the widespread use of digitalization, modern technologies, and especially AI does not only affect the exercise of certain human rights, but possibly has ramifications on all human rights.[164] In the indivisible, interdependent, and interconnected arena of human rights, the application of AI presents a general challenge, therefore it is not sufficient to focus only on specific rights (privacy, freedom of expression, etc.), even if these are more evidently or easily violated by AI technologies. My analysis shows that we are facing a general problem permeating practically all aspects of individual and social life, whereby all human rights and fundamental freedoms are potentially impacted.[165]

The term "weapons of math destruction", coined by Cathy O'Neil,[166] might seem sensationalist, but it accurately emphasizes the dramatic consequences of AI on society. Empirical studies referred to in this paper have shown that the positive effects of modern technology are often overrated, and along with exaggerated expectations and reliance on AI, built-in racial, ethnic and gender biases continue to go unnoticed. Algorithmic censorship determines "what we can or cannot see online and the extent to which we can interact with content and shape our online environments".[167] Experts warn that "[t]he traditional asymmetry of power and information between state structures and human beings is shifting towards an asymmetry of power and information between operators of algorithms (who may be public or private) and those who are acted upon and governed".[168] As AI is becoming more sophisticated and embedded in society, the rule of law – based on principles such as accessibility (predictability), transparency, fairness, explainability – is being diminished. This is so because most AI-based decision-making systems use exceptionally complex technology that is beyond the reach of human cognition.

The law also contributes to the "black box" phenomenon by inhibiting transparency, for instance by protecting trade secrets and other intellectual property rights.[169] We have seen how often the right to private life has been sacrificed on the altar of national security, and how fragile the equilibrium is between human rights and security considerations. "Digital services risk eliminating, almost entirely, much of the human interaction and compassion"[170] that are indispensable in many contexts from health care to social services. The list of AI-related dangers and unintended negative effects is long. Against this backdrop, decision-makers and legislators cannot sit idly by.

[164] And beyond, since this paper did not examine for example autonomous weapon systems and other military AI tools, since these are covered primarily by international humanitarian law – a closely connected but still different field of international law. Similarly, AI-enabled surveillance technology can play a role in war crime investigations. Maas (2019) 16–17.

[165] For a similar view, see Council of Europe (2018) 32.

[166] O'Neil (2016).

[167] Ashraf (2022) 770.

[168] Council of Europe (2018) 33.

[169] Greenstein (2022).

[170] Alston (2019) 17.

Although some authors fear that AI technology is developing faster than policy-makers and the law can react to it[171] or propose the conclusion of a new, AI-specific international treaty,[172] I agree with those who claim that the existing universal human rights framework is appropriate to face the challenges emerging as a result of new technologies.[173] Using this framework in the governance of AI has the advantages of both global reach and normative force – none of which can be claimed by the ever-growing ethical frameworks and initiatives developed by civil society organizations, private companies, groups of experts commissioned by IGOs, academics or other individuals/groups.[174] It does not seem convenient or even possible to react to every new technology on a case-by-case basis as they emerge. Instead, human rights must be embedded (coded) into the design and deployment of all AI applications, like a sort of filter. Adherence to human rights must be represented in every stage of Al development, similar to what Isaac Asimov called the Three Laws of Robotics in his 1942 short story "Runaround" (although these laws are criticized for being too unspecific[175]). Donahoe and Metzger underline that with the advances of AI, "the status of human beings as the focal point of AI-reliant governance decisions cannot be assumed – it will have to be ensured".[176] This means that AI must be trained in line with human dignity and specific human rights, and to use *ex ante* AI-focused human rights impact assessment mechanisms (HRIA).[177]

Furthermore, international human rights treaties have been constantly interpreted by their relevant monitoring bodies (courts, committees, etc.) in a progressive, evolutionary manner to meet the needs of contemporary society.[178] A prime example is how the ECtHR developed the concept of privacy, taking into consideration social and technological developments.[179] This dynamic interpretation should make the available international treaties suitable to address new

[171] Anshari et al. (2023).

[172] Gervais (2023).

[173] Donahoe and Metzger (2019); McGregor, Murray and Ng (2019); Nonnecke and Dawson (2021).

[174] As Mantelero and Esposito put it, "the point is not to cut off the ethical roots, but to recognise that rights and freedoms flourish on the basis of the shape given them by law provisions and case law. There is no conflict between ethical values and human rights, but the latter represent a specific crystallisation of these values that are circumscribed and contextualised by legal provisions and judicial decisions". Mantelero and Esposito (2021) 4. For a detailed account of the benefits of the international human rights approach, see Douek (2021) 44–49.

[175] Risse (2018) 9.

[176] Donahoe and Metzger (2019) 117.

[177] Tzimas (2021) 138–45. Nonnecke and Dawson point out that the design and implementation of impact assessments in the realm of AI (AIA) is still in its early stage. Consequently, there is no consensus regarding the appropriate methodology or application of constitutive components (Nonnecke and Dawson 2021, 6). HRIA presents specific challenges, because its scope tends to be broader, more complex and more forward-looking than that of an AIA by default, where technical aspects (the potential for bias, fairness, explainability, etc.) and the immediately foreseeable and measurable risks are assessed (Ibid., 7–8.) For a convincing model of HRIA, see Mantelero and Esposito (2021).

[178] Cf. the living-instrument doctrine of the ECtHR (Grover [2020] 191–231.) or the preamble of the ECHR stating the purpose of the treaty as not only the maintenance, but also the "further realisation" of human rights.

[179] Çınar (2021).

challenges brought about by AI technologies.[180] In fact, to determine the lawfulness of any deployment and to protect against arbitrariness, decision-makers should conduct the usual test that courts apply to human rights restrictions.[181] If an AI-related measure is seen to interfere with a human right, it must be examined whether the measure in question is being *prescribed by law* and has a *legitimate aim*. The measure must be *suitable*, that is reasonably likely to realize its objectives; *necessary in a democratic society*, that is there should be no other less intrusive means capable of achieving the desired result; and *proportional*, that is a balance of interests must exist between the objectives pursued and the restriction of a given human right.

To conclude, AI is in fact humanity's new frontier. In light of the possibility of widespread human rights abuses and risks, it is imperative that decision-makers practice caution in the adoption and regulation of AI. Their approach should be in harmony with the universally acknowledged standards of fundamental human rights, with a view to create a proper balance between realizing the opportunities and averting the dangers that AI present to humanity.

## ACKNOWLEDGEMENTS

This article was written in the framework of Project no. TKP2021-NKTA-51, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NKTA funding scheme. I am indebted to the two anonymous reviewers for their valuable comments, which greatly improved the quality of the article.

## LITERATURE

Access Now, *Human rights in the Age of Artificial Intelligence* (2018) <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf> accessed 12 May 2023.

Affonso, S. C., César de Oliveira, C., Perrone, C. and Carneiro, G., 'From privacy to data protection: the road ahead for the Inter-American System of human rights' (2021) 25 The International Journal of Human Rights 147–77.

African Charter on Human and Peoples' Rights, adopted 27 June 1981, O.A.U. Doc. CAB/LEG/67/3/Rev. 5.

AI HLEG, *Ethics guidelines for trustworthy AI* (European Commission 2019) <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed 12 May 2023.

Alegre, S., 'Rethinking Freedom of Thought for the 21st Century' (2017) 3 European Human Rights Law Review 221–33.

[180] Of course, questions of practical implementation are quite relevant. For example, how can courts restore lawfulness if the data is already collected, stored and transferred, when it cannot be permanently deleted? Human rights violations by AI occur almost immediately, but the exhaustion of domestic remedies – as required in the proceedings of international human rights bodies – can take years. Yet, in the long run, the deterrent effect of these mechanisms and their capability to induce change within the national legal systems remain viable reasons for relying on them, even if only as a last resort.

[181] Murray (2020); Degeling and Berendt (2018). For the constraints of this approach, see Douek (2021) 50–66.

Algan, B., ‘Rethinking “Third Generation” Human Rights’ (2004) 1 Ankara Law Review 121–55.

Alikhademi, K., Drobina, E., Prioleau, D., Richardson, B., Purves, D. and Gilbert, J. E., ‘A review of predictive policing from the perspective of fairness’ (2022) 30 Artif Intell Law 1–17.

de Almendra Freitas, C. O., Pamplona, D. A. and de Oliveira, D. H. Z., ‘Duty to protect and responsibility to respect: data privacy violations in pandemic times’ (2022) 26 The International Journal of Human Rights 1313–32 .

Alston, P., ‘Report of the Special Rapporteur on extreme poverty and human rights’, 11 October 2019, A/74/493 <https://digitallibrary.un.org/record/3834146> accessed 12 May 2023.

American Convention on Human Rights, adopted 22 November 1969, O.A.S. T.S. No. 36.

Anderson, M., ‘Facebook privacy scandal explained’, The Associated Press (6 April 2018) <https://www.ctvnews.ca/sci-tech/facebook-privacy-scandal-explained-1.3874533> accessed 12 May 2023.

Angwin, J., Larson, J., Mattu, S. and Kirchner, L., ‘Machine Bias. There’s software used across the country to predict future criminals. And it’s biased against blacks’, *ProPublica* (23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 21 July 2023.

Anshari, M., Hamdan, M., Ahmad, N. et al., ‘COVID-19, artificial intelligence, ethical challenges and policy implications’ (2023) 38 AI & Society 707–20.

Arifin, S., ‘Artificial Intelligence in the Workplace – How Should Moral and Legal Issues Be Addressed?’ (2021) 9 Pro Publico Bono – Public Administration 94–109.

Ashraf, C., ‘Artificial Intelligence and the Rights to Assembly and Association’ (2020) 5 Journal of Cyber Policy 163–79.

Ashraf, C., ‘Exploring the impacts of artificial intelligence on freedom of religion or belief online’ (2022) 26 The International Journal of Human Rights 757–91.

Aswad, E., ‘Losing the Freedom to Be Human’ (2020) 52 Columbia Human Rights Law Review 306–71.

Azoulay, A., *Towards an Ethics of Artificial Intelligence* (2018) <https://www.un.org/en/chronicle/article/towards-ethics-artificial-intelligence> accessed 12 May 2023.

Barocas, S. and Andrew, D. S., ‘Big Data’s Disparate Impact’ (2016) 104 California Law Review 671–732.

Blount, K., ‘Using artificial intelligence to prevent crime: implications for due process and criminal justice’, AI & Society (2022) <https://doi.org/10.1007/s00146-022-01513-z> accessed 12 May 2023.

Borenstein, J., ‘Robots and the changing workforce’ (2011) 26 AI & Society 87–93.

Bubeck et al., ‘*Sparks of Artificial General Intelligence: Early experiments with GPT-4*’ (2023) <https://doi.org/10.48550/arXiv.2303.12712> accessed 20 July 2023.

Burgess, P., ‘Algorithmic augmentation of democracy: considering whether technology can enhance the concepts of democracy and the rule of law through four hypotheticals’ (2021) 37 AI & Society 97–112.

CAHAI (Ad hoc Committee on Artificial Intelligence), ‘The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law’ (2020a) Strasbourg, The Council of Europe, 24 June 2020, CAHAI(2020)06-fin.

CAHAI, ‘Feasibility Study’ (2020b) Strasbourg, The Council of Europe, 17 December 2020, CAHAI(2020)23.

Carnevale, A., Tangari, E., Iannone, A. et al., ‘Will Big Data and personalized medicine do the gender dimension justice?’ (2023) 38 AI & Society 829–41.

Cataleta, M. S. and Cataleta, A. ‘Artificial Intelligence and Human Rights, an Unequal Struggle’ (2020) 1 CIFILE Journal of International Law 41–63.

Chan, G. K. Y., ‘AI employment decision-making: integrating the equal opportunity merit principle and explainable AI’, *AI & Society* (2022) <https://doi.org/10.1007/s00146-022-01532-w> accessed 12 May 2023.

Çınar, Ö. H., 'The current case law of the European Court of Human Rights on privacy: challenges in the digital age' (2021) 25 The International Journal of Human Rights 26–51.

Constantaras, E., Geiger, G., Braun, J-C., Mehrotra, D. and Aung, H. 'Inside the Suspicion Machine', *WIRED* (6 March 2023) <https://www.wired.com/story/welfare-state-algorithms/> accessed 20 July 2023.

Corea, F., Fossa, F., Loreggia, A., Quintarelli, S. and Sapienza, S. 'A principle-based approach to AI: the case for European Union and Italy' (2023) 38 AI & Society 521–35.

Council of Europe, *Algorithms and Human Rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications* (2018) DGI (2 017)12 <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5> accessed 12 May 2023.

Douek, E., 'The Limits of International Law in Content Moderation' (2021) 6 UC Irvine Journal of International, Transnational, and Comparative Law 37–75.

Dror-Shpoliansky, D. and Shany, Y., 'It's the End of the (Offline) World as We Know It: From Human Rights to Digital Human Rights – A Proposed Typology' (2021) 32 European Journal of International Law 1249–82.

Degeling, M. and Berendt, B., 'What is wrong about Robocops as consultants? A technology-centric critique of predictive policing' (2018) 33 AI & Society 347–56.

Deranty, J. P. and Corbin, T., 'Artificial intelligence and work: a critical review of recent research from the social sciences', AI & Society (2022) <https://doi.org/10.1007/s00146-022-01496-x> accessed 12 May 2023.

Devillers, L., Fogelman-Soulié, F. and Baeza-Yates, R., 'AI and Human Values. Inequalities, Biases, Fairness, Nudge and Feedback Loops' in B Braunschweig and M Ghallab M (eds), *Reflections on Artificial Intelligence for Humanity* (Springer 2021) 76–89.

Dias Oliva, T., 'Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression' (2020) 20 Human Rights Law Review 607–40.

Donahoe, E. and Metzger M. M., 'Artificial Intelligence and Human Rights' (2019) 30 Journal of Democracy 115–26.

Drigas, A. S. and Ioannidou, R. E., 'A review on artificial intelligence in special education' in MD Lytras and others (eds), *Information systems, e-learning, and knowledge management research*. WSKS 2011. Communications in computer and information science, vol. 278. (Springer 2013) 385–91.

Ebert, I., Wildhaber, I. and Adams-Prassl, J., 'Big Data in the workplace: Privacy Due Diligence as a human rights-based approach to employee privacy protection' (2021) 8 Big Data & Society 1–14.

ECtHR: *Case "relating to certain aspects of the laws on the use of languages in education in Belgium"* App nos 1474/62 and others (ECtHR, 23 July 1968).

ECtHR: *Z v. Finland* App no 22009/93 (ECtHR, 25 February 1997).

ECtHR: *Sigurður Einarsson and Others v. Iceland* App no 39757/15 (ECtHR, 4 June 2019).

ECtHR, 'Guide to the Case-Law of the European Court of Human Rights, Data protection', updated on 31 August 2022 <https://www.echr.coe.int/Documents/Guide_Data_protection_ENG.pdf> accessed 12 May 2023.

European Commission, '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', 21 April 2021, COM/2021/206 final.

European Convention for the Protection of Human Rights and Fundamental Freedoms(ECHR), adopted 4 November 1950, ETS No. 005.

Ferguson, D. and Savage, M., 'Controversial exams algorithm to set 97% of GCSE results', *The Guardian* (15 August 2020) <https://www.theguardian.com/education/2020/aug/15/controversial-exams-algorithm-to-set-97-of-gcse-results> accessed 12 May 2023.

Fosch-Villaronga, E., van der Hof, S., Lutz, C. et al., 'Toy story or children story? Putting children and their rights at the forefront of the artificial intelligence revolution' (2023) 38 AI & Society 133–52.

Frenkel, S. and Benner, K., 'To Stir Discord in 2016, Russians Turned Most Often to Facebook', *The New York Times* (16 July 2018) <https://www.nytimes.com/2018/02/17/technology/indictment-russian-tech-facebook.html> accessed 12 May 2023.

Galetsi, P., Katsaliaki, K. and Kumar, S., 'The medical and societal impact of big data analytics and artificial intelligence applications in combating pandemics: A review focused on Covid-19' (2022) 301 Social Science & Medicine 114973.

Gervais, D. J., 'Towards an effective transnational regulation of AI' (2023) 38 AI & Society 391–410.

Greenstein, S., 'Preserving the rule of law in the era of artificial intelligence (AI)' (2022) 30 Artificial Intelligence and Law 291–323.

Grover, S. C., *Judicial Activism and the Democratic Rule of Law* (Springer 2020).

Guo, Z. and Kennedy, L., 'Policing based on automatic facial recognition' (2023) 31 Artificial Intelligence and Law 397–443.

Haidt, J. and Bail, C., *Social media and political dysfunction: A collaborative review* (New York University, ongoing) <https://tinyurl.com/PoliticalDysfunctionReview≥accessed 20 July 2023.

Harayama, Y., Milano, M., Baldwin, R. et al., 'Artificial Intelligence and the Future of Work' in B Braunschweig and M Ghallab (eds) *Reflections on Artificial Intelligence for Humanity* (Springer 2021) 53–67.

Hárs, A., 'AI and international law – Legal personality and avenues for regulation' (2022) 62 Hungarian Journal of Legal Studies 320–44.

Heinrichs, B., 'Discrimination in the age of artificial intelligence' (2022) 37 AI & Society 143–154.

Hill, K., 'The Secretive Company That Might End Privacy as We Know It', *The New York Times* (18 January 2020a) <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> accessed 12 May 2023.

Hill, K., 'Wrongfully Accused by an Algorithm', *The New York Times* (24 June 2020b) <https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html> accessed 12 May 2023.

International Covenant on Civil and Political Rights (ICCPR), adopted 16 December 1966, United Nations Treaty Series No. 14668.

International Covenant on Economic, Social and Cultural Rights (ICESCR), adopted 16 December 1966, United Nations Treaty Series No. 14531.

Kaye, D., 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression', 29 August 2018 <https://digitallibrary.un.org/record/1643488> accessed 12 May 2023.

Langford, M., 'Taming the Digital Leviathan: Automated Decision-Making and International Human Rights' (2020) 114 AJIL Unbound 141–46.

Lee, J., 'Artificial Intelligence and Human Rights: Four Realms of Discussion, Research, and Annotated Bibliography' (2021) 1 Rutgers International Law and Human Rights Journal <https://www.rutgers-ilhr-journal.org/spring-2021> accessed 12 May 2023.

Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M. and Briggs, M., *Artificial intelligence, human rights, democracy, and the rule of law: a primer* (The Council of Europe 2021).

Livingston, S. and Risse, M., 'The Future Impact of Artificial Intelligence on Humans and Human Rights' (2019) 33 Ethics and International Affairs 141–58.

Maas, M. M., 'International Law Does Not Compute: Artificial Intelligence and the Development, Displacement or Destruction of the Global Legal Order' (2019) 20 Melbourne Journal of International Law 29–56.

Mainz, J. T., Sønderholm, J. and Uhrenfeldt, R., 'Artificial intelligence and the secret ballot', *AI & Society* (2022) <https://doi.org/10.1007/s00146-022-01551-7> accessed 12 May 2023.

Mancuhan, K. and Clifton, C., 'Combating discrimination using Bayesian networks' (2014) 22 Artif Intell Law 211–38.

Mantelero, A. and Esposito, M. S., 'An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems' (2021) 41 Computer Law & Security Review 105561 <https://doi.org/10.1016/j.clsr.2021.105561> accessed 12 May 2023.

McGregor, L., Murray, D. and Ng, V., 'International Human Rights Law as a Framework for Algorithmic Accountability' (2019) 68 International and Comparative Law Quarterly 309–43.

Minhas, R., Elphick, C. and Shaw, J., 'Protecting victim and witness statement: examining the effectiveness of a chatbot that uses artificial intelligence and a cognitive interview' (2022) 37 AI & Society 265–81.

Mirbabaie, M., Hofeditz, L., Frick, N. R. J. and Stieglitz, S., 'Artificial intelligence in hospitals: providing a status quo of ethical considerations in academia to guide future research' (2022) 37 AI & Society 1361–1382.

Mubangizi, J. C., 'Towards a new approach to the classification of human rights with specific reference to the African context' (2004) 4 African Human Rights Journal 93–107.

Murray, D., 'Using Human Rights Law to Inform States' Decisions to Deploy AI' (2020) 114 AJIL Unbound 158–162.

Nonnecke, B. and Dawson, P., '*Human Rights Implications of Algorithmic Impact Assessments: Priority Considerations to Guide Effective Development and Use*' (2021) Carr Center for Human Rights Policy. <https://carrcenter.hks.harvard.edu/files/cchr/files/nonnecke_and_dawson_human_rights_implications.pdf> accessed 20 July 2023.

Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S., 'Dissecting racial bias in an algorithm used to manage the health of populations' (2019) 366 Science 447–453.

O'Neil, C., *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Penguin Books 2016).

Parfett, A., Townley, S. and Allerfeldt, K., 'AI-based healthcare: a new dawn or apartheid revisited?' (2021) 36 AI & Society 983–999.

Pavlovic, Z., Randelovic, D. and Ivanovic, A. R., 'The right to privacy of children in the digital world and the responsibility of adults with special focus on social networks as modern media communication' (2018) 1 Journal of Eastern-European Criminal Law 7–18.

Peeters, R. and Widlak, A. C., 'Administrative exclusion in the infrastructure-level bureaucracy: The case of the Dutch daycare benefit scandal' (2023) 83 Public Administration Review 863–77.

Pfeifer-Chomiczewska, K., 'Intelligent service robots for elderly or disabled people and human dignity: legal point of view' (2023) 38 AI & Society 789–800.

Quattrocolo, S., *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for A European Legal Discussion* (Springer 2020).

Rachovitsa, A. and Johann, N., 'The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case' (2022) 22 Human Rights Law Review 1–15.

Ramli, R. and Zakaria, N., 'Privacy issues in a psychiatric context: applying the ISD privacy framework to a psychiatric behavioural monitoring system' (2014) 29 AI & Society 203–13.

Raso, A., Hilligoss, H., Krishnamurthy, V., Bavitz, C. and Kim, L., 'Artificial Intelligence & Human Rights: Opportunities & Risks' (2018) Berkman Klein Center Research Publication 6 <http://dx.doi.org/10.2139/ssrn.3259344> accessed 12 May 2023.

Risse, M., 'Human Rights and Artificial Intelligence: An Urgently Needed Agenda' *HKS Faculty Research Working Paper Series* (RWP18-015, May 2018) <https://www.hks.harvard.edu/publications/human-rights-and-artificial-intelligence-urgently-needed-agenda> accessed 12 May 2023.

Rusinova, V., 'Privacy and the legalisation of mass surveillance: in search of a second wind for international human rights law' (2022) 26 The International Journal of Human Rights 740–56.

Ryan, M., 'The social and ethical impacts of artificial intelligence in agriculture: mapping the agricultural AI literature' *AI & Society* (2022) <https://doi.org/10.1007/s00146-021-01377-9>, accessed 12 May 2023.

Selinger, E. and Rhee, H. J., 'Normalizing Surveillance' (2021) 22 SATS 49–74.

Shaheed, F. and Mazibrada, M. 'On the Right to Science As a Cultural Human Right' in H Porsdam and S Porsdam Mann (eds), *The Right to Science: Then and Now* (Cambridge University Press 2021) 107–23.

Smith, M. and Miller, S., 'The ethical application of biometric facial recognition technology' (2022) 37 AI & Society 167–75.

Stanila, L., 'Artificial intelligence and human rights: challenging approach on the issue of equality' (2018) 2 Journal of Eastern-European Criminal Law 19–30.

State v. Loomis. Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing' (2017) 130 Harvard Law Review 1530–1537.

Szappanyos, M., 'Artificial Intelligence: is the European Court of Human Rights prepared?' (2023) 11 Acta Humana – Emberi Jogi Közlemények 93–110.

Teo, S. A., 'Human dignity and AI: mapping the contours and utility of human dignity in addressing challenges presented by AI' (2023) 15 Law, Innovation and Technology 241–279.

Tilovska-Kechedji, E. and Rakitovan, D., 'The digital world affecting children rights and the affects of internet governance' (2018) 1 Journal of Eastern-European Criminal Law 140–147.

Trew, B., 'Coronavirus: Controversial Israeli Spyware Firm NSO Builds Software Tracking Mobile Data to Map Covid-19' *The Independent* (18 March 2020) <https://www.independent.co.uk/news/world/middle-east/coronavirus-israel-cases-tracking-mobile-phone-nso-spyware-covid-19-a9410011.html> accessed 12 May 2023.

Tzimas, T., *Legal and Ethical Challenges of Artificial Intelligence from an International Law Perspective* (Springer 2021).

UN Human Rights Committee, 'General Comment No. 18 – Non-discrimination', HRI/GEN/1/Rev.9 (Vol. I), 10 November 1989.

UN Human Rights Council, 'The Right to Privacy in the Digital Age' Resolution adopted on 23 March 2017, A/HRC/RES/34/7.

Universal Declaration of Human Rights, adopted 10 December 1948 by UN General Assembly Resolution 217.

Vanberg, A. D., 'Informational privacy post GDPR – end of the road or the start of a long journey?' (2021) 25 The International Journal of Human Rights 52–78.

Van Est, R., Gerritsen, J. B.A. and Kool, L., 'Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality'. Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (Rathenau Instituut, 2017).

Wakunuma, K., Jiya, T. and Aliyu, S., 'Socio-ethical implications of using AI in accelerating SDG3 in Least Developed Countries' (2020) 4 Journal of Responsible Technology 100006 <https://doi.org/10.1016/j.jrt.2020.100006> accessed 12 May 2023.

Yuan, S., 'How China Is Using AI and Big Data to Fight the Coronavirus' *Al Jazeera* (1 March 2020) <https://www.aljazeera.com/news/2020/03/china-ai-big-data-combat-coronavirus-outbreak-200301063901951.html> accessed 12 May 2023.

Žaliauskaitė, M., 'Role of ruler or intruder? Patient's right to autonomy in the age of innovation and technologies' (2021) 36 AI & Society 573–83.