Losing Freedom by Du Feichen

ARTIFICIAL INTELLIGENCE HARM AND HUMAN RIGHTS:

A HIGH LEVEL EXPLORATION OF THE INTERACTION OF AI HARMS

ICAAD
Human Rights Innovation

KING&WOOD MALLESONS
金杜律师事务所

CONTENTS

1. INTRODUCTION

Over the last few years, the use of Artificial Intelligence (**AI**) has exponentially grown. With the International Monetary Fund predicting that almost 40% of global employment is exposed to AI, and Goldman Sachs predicting that generative AI alone could drive a 7% (or almost $7 trillion) increase in global GDP and lift productivity growth by 1.5% over a 10-year period,[1] AI has the potential to reshape the global economy. However, this potential must be balanced with recognition of the potential harms that AI presents. As best summarised by the drafters of the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law:[2]

> *'... artificial intelligence systems offer unprecedented opportunities to protect and promote human rights, democracy and the rule of law. At the same time, they also wished to acknowledge that there are serious risks and perils arising from certain activities within the lifecycle of artificial intelligence such as, for instance, discrimination in a variety of contexts, gender inequality, the undermining of democratic processes, impairing human dignity or individual autonomy, or the misuses of artificial intelligence systems by some States for repressive purposes, in violation of international human rights law'.*

Although there is significant academic literature and a growing body of international discourse on the harms presented by AI, there has been relatively little attention given to mapping the various methods by which harm may result from AI systems in the context of fundamental human rights.

This paper proposes a simple (and easily expandable) table approach to mapping the interaction of potential AI harms in the context of human rights that can be used by public and private actors when considering how AI systems interact with human rights and whether an AI system could have potential human rights implications (subject to further consideration of the applicable AI system, the context in which it is used and the nuances of human rights law).

1. Mauro Cazzaniga et al, 'Gen-AI: Artificial Intelligence and the Future of Work' (Staff Discussion Note SDN2024/001, International Monetary Fund, January 2024) (available at: https://doi.org/10.5089/9798400262548.006); 'Generative AI could raise global GDP by 7%' Goldman Sachs (5 April 2023, Web Page) <https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent>.
2. Paragraph 10 of the Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS 225) (available at: https://rm.coe.int/1680afae67). See also Preamble to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (available at: https://rm.coe.int/1680afae3c).

2. AI AND HUMAN RIGHTS

WHAT IS AI?

Despite the term being first coined in the 1950s, AI does not have a universally agreed definition as it does not refer to one easily definable concept: rather, it is used to describe the capabilities of computer systems and algorithms to imitate human intelligence, and captures a wide group of technologies. The OECD has adopted the following definition of AI systems:

> *'... a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predications, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment'.*

In approaching AI, it is important to distinguish between the use of 'Narrow AI' (also commonly referred to as 'Traditional AI' or 'Predictive AI') and 'Generative AI'. Both types of AI use machine learning combined with big data but have different objectives. For example:

- narrow AI focuses on machine learning models that use predetermined algorithms and rules to analyse data and make predictions, recommendations or decisions; while
- generative AI focuses on machine learning models, particularly neural networks, to create new content (including text, code, images, sounds and videos) based on the data contained in its training datasets.

Although generative AI (such as ChatGPT, Gemini, Claude and DeepSeek) has captured the bulk of the public attention in the last three years, narrow AI systems have been increasingly used by governments and companies around the world over the past two decades in a wide variety of industries from civil, healthcare, education, justice, employment, housing, and more. As such, AI should not be considered a new concept or even a new technology. What is new is the almost exponential rate of growth in technological development and popularity of AI (both narrow AI and generative AI).

Furthermore, AI systems are relatively unique compared to other technological developments over the last century as they (depending on the AI system in question):[3]

- are dependent on the use of large amounts of data throughout the AI system lifecycle (this includes both the inputs to an AI system, such as its training data and testing data, and its outputs);
- are often opaque or lack transparency/explainability as to how a particular output is reached (or even how the algorithm works);
- can interact with a range of interfaces (including IoT devices, infrastructure and robotic devices);
- can be easily replicated (and in some cases can even self-replicate); and
- have the potential to be either autonomous or semi-autonomous. This includes, but is not limited to, learning to perform tasks without being explicitly pre-programmed by its developer or deployer.

As best summarised in the Explanatory Memorandum for Europe's AI Act, the unique features of AI that present the greatest risk are '*opacity, complexity, bias, a certain degree of unpredictability and partially autonomous behaviour of certain AI systems … '*.[4]

3. Yonathan Arbel, Matthew Tokson and Albert Lin, 'Systemic Regulation of Artificial Intelligence' (2023) *56 Arizona State Law Journal* 545 at 551-552. Available at: https://ssrn.com/abstract=4666854.
4. Explanatory Memorandum, *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*, 2021/0106(COD). Available at: eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206.

HUMAN RIGHTS AND AI

In recent years, the interaction of human rights and AI is receiving increasing focus.

In September 2024, the first international human rights treaty specific to AI (the *Council of Europe's Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*) (**AI Convention**) was opened for signature. The AI Convention:

- presents a technology-neutral approach to regulating AI with a focus on ensuring that the various activities within the lifecycle of AI systems are fully consistent with human rights, democracy and the rule of law;
- obligates states to ensure that AI systems incorporate individual privacy protections, transparency and auditability requirements, and safety and security requirements; and
- is practically complemented by the HUDERIA Methodology (and soon to be released HUDERIA Model). HUDERIA is a tool that can be used by both public and private actors to identify and address the risks of AI to human rights, democracy and the rule of law.[5] HUDERIA is non-legally binding guidance.

As of 1 March 2025, only 13 countries (plus the European Union) have signed the AI Convention, and it will not enter into force until it has received 5 ratifications including at least 3 member States of the Council of Europe.[6]

Although the AI Convention is currently not in force, the core international human rights instruments (although developed without reference to AI) have equal relevance when they are breached by virtue of actions linked to AI systems. As a result, there is a (small) number of AI focused instruments (both binding and non-binding) that acknowledge the risk that AI can pose to human rights. This includes:

- the OECD's AI Principles that state *'AI actors should respect the rule of law, human rights, democratic and human-centred values throughout the AI system lifecycle. These include non-discrimination and equality, freedom, dignity, autonomy of individuals, privacy and data protection, diversity, fairness, social justice, and internationally recognised labour rights'*;[7] and
- Europe's AI Act (that regulates the use, deployment and development of AI systems within the European Union) which is designed to promote the uptake of human-centric and trustworthy AI while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems.[8]

However, subject to exceptions such as those listed above, human rights are often not central to AI governance or AI regulation. As summarised by Kate Jones (in her then role as Associate Fellow of Chatham House) in 2023:

- many AI governance principles (be they produced by companies, governments, civil society or international organizations) fail to explicitly mention human rights;
- most national AI strategies do not engage with human rights in depth;
- many in the AI industry do not engage with those in the human rights community when approaching responsible AI; and
- many businesses consider that human rights are not applicable to them.

In May 2025 the United Nations Human Rights Council Working Group on the issue of human rights and transnational corporations and other business enterprises released its report on 'Artificial intelligence procurement and deployment: ensuring alignment with the Guiding Principles on Business and Human Rights'.[9] As part of this report, the working group noted that:

> *'although the protection of human rights is increasingly emphasized in regulatory developments...evidence from the Working Group's consultations and submissions has shown that there are still significant gaps when it comes to legislative frameworks on rights-respecting procurement and deployment of AI systems by States and businesses. Further, for many businesses, understanding of the human rights implications of the deployment of AI systems remains in its early stages. Thus, the rapid, mainly unregulated, uptake of AI systems by States and businesses is creating situations with high potential for adverse impacts across a variety of human rights, in a context where existing access-to-remedy mechanisms are struggling to keep up.'*[10]

5. 'HUDERIA: New tool to assess the impact of AI systems on human rights', Council of Europe (Web Page, 2 December 2024) <https://www.coe.int/en/web/portal/-/huderia-new-tool-to-assess-the-impact-of-ai-systems-on-human-rights>.
6. 'Chart of signatures and ratifications of Treaty 225', Council of Europe (Web Page) <https://www.coe.int/en/web/conventions/full%20list?module=signatures-by-treaty&treatynum=225>. The web page was accessed on 1 March 2025.
7. 'AI principles', OECD (Web Page) <https://www.oecd.org/en/topics/ai-principles.html>.
8. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 2024/1689, art 1. Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.
9. United Nations Human Rights Council, 'A/HRC/59/53: Artificial intelligence procurement and deployment: ensuring alignment with the Guiding Principles on Business and Human Rights - Report of the Working Group on the issue of human rights and transnational corporations and other business enterprises.' Available at: https://www.ohchr.org/en/documents/thematic-reports/ahrc5953-artificial-intelligence-procurement-and-deployment-ensuring
10. Ibid [4]

3. CAN AI HARM BE MAPPED AGAINST HUMAN RIGHTS?

WHAT IS AI HARM?

In many cases an AI system itself is unlikely to cause harm. This is particularly the case where the AI system does not operate in an autonomous manner and does not produce or influence a decision or undertake an action that will impact an individual or have real-world implications. However, it cannot be presumed that the use of AI systems will not result in harm. To the contrary, AI systems have the potential to cause significant harm to individuals, society or the environment depending on how they are used and their level of autonomy. As stated in the 2023 Bletchley Declaration, this risk is particularly acute for highly capable general purpose AI models with:

> *'... [the] potential for serious, even catastrophic, harm, either deliberate or unintentional, stemming from the most significant capabilities of these AI models.'* [11]

However - given the wide breadth of the types of AI systems, the wide breadth of sectors in which it can be used and the wide breadth of potential use cases for which it can be employed, there is no simple method for determining whether an AI system poses harm and whether it has caused harm. To fill this gap, there is currently an increasing number of:

- AI harm taxonomies that seek to produce a methodology by which the harms of AI can be identified. Depending on the taxonomy, the methodology may include consideration of whether the harm is tangible or intangible; the timing of the harm; the entity responsible for the harm (i.e. the AI system or humans); the intent of the harm (i.e. is it intentional, unintentional or unknown); types of impacted individuals (e.g. older persons, adult or child, men and women, Indigenous, LGBTIQ+, disabled); geographies (e.g. global north or south); industries and sectors (e.g. healthcare, finance, criminal justice), and dimensions (e.g. recurrence and reversibility). Examples include the MIT AI Risk Repository,[12] the Centre for Security and Emerging Technology AI Harm Taxonomy,[13] the AI, Algorithmic and Automation Incidents and Controversies Database (**AIAAIC**) harm taxonomy,[14] TASRA (a Taxonomy and Analysis of Societal-Scale Risks from AI)[15] and the OECD AI Incident definition;[16] and
- AI incident databases that (using harm taxonomies) seek to track real-world incidents where AI has caused harm. Examples include the Artificial Intelligence Incident Database (**AIID**),[17] the AIAAIC Repository,[18] the Atlas of AI Risk[19] and OECD's AI Incidents Monitor (AIM).[20]

11. The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023' *GOV.UK* (Web Page) <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>.
12. See https://airisk.mit.edu.
13. Mia Hoffmann et al, *CSET AI Harm Taxonomy for AIID and Annotation Guide* (25 July 2023). Available at: https://github.com/georgetown-cset/CSET-AIID-harm-taxonomy/blob/main/CSET%20V1%20AI%20Annotation%20Guide%20(with%20Schema%20and%20Field%20Descriptions)%2025Jul2023.pdf
14. See https://www.aiaaic.org/projects/ai-algorithmic-risks-harms-taxonomy.
15. See https://arxiv.org/pdf/2306.06924.
16. *Stocktaking for the Development of an AI Incident Definition* (OECD Artificial Intelligence Papers No 4, OECD, October 2023) at 8. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/10/stocktaking-for-the-development-of-an-ai-incident-definition_64c69a10/c323ac71-en.pdf.
17. See https://incidentdatabase.ai.
18. See https://www.aiaaic.org/home.
19. See https://social-dynamics.net/atlas.
20. See https://oecd.ai/en/incidents.

Together, harm taxonomies and AI incident databases provide both public and private actors with a means of identifying how a potential use for an AI system (or an AI system itself) could result in harm or, alternatively, how an AI system has already resulted in harm.

However - with the notable exception of the CSET AI Harm Taxonomy, the OECD AI Incident definition and the AIAAIC harm taxonomy – many harm taxonomies do not expressly consider harms from a human rights perspective. Although it is acknowledged that less tangible or intangible harms (such as those that arise from human rights infringements) are harder to evidence than tangible harms, this presents a potential gap when public and private actors approach AI. In 2023, Volker Türk (the UN High Commissioner for Human Rights), in calling for urgent action by governments and by companies regulating AI, stated that:

> *' The starting point should be the harms that people experience and will likely experience. This requires listening to those who are affected, as well as to those who have already spent many years identifying and responding to harms … [and any AI] … regulations need to require assessment of the human rights risks and impacts of AI systems before, during, and after their use … AI technologies that cannot be operated in compliance with international human rights law must be banned or suspended until such adequate safeguards are in place.'*[21]

WHAT NOW?

Without clear guidance on how an AI system may result in human rights breaches, it is difficult (especially for private actors) to easily engage with how the design, deployment or use of an AI system could result in harms that may result in human rights breaches.

Internationally, there is increasing recognition that this gap is best fixed by the use of human rights impact assessments (or a fundamental rights impact assessment) before AI systems are deployed or when they are substantially modified. Although fundamental rights impact assessments for high-risk AI systems will be required under Europe's AI Act from 2 August 2026,[22] they are still relatively rare. Notable exceptions include:

- the Australian Human Rights Commission's human rights impact assessment tool for artificial intelligence-informed decision-making systems in banking;[23] and
- Ontario's Human Rights Commission's human rights impact assessment tool.[24]

However, these human rights impact assessment templates are contextually specific and will not assist in bridging the gap identified in section 2 above that many harm taxonomies do not consider human rights. Accordingly, Table 2 below presents a simple method that:

- existing harm taxonomies and incident databases can take into consideration when approaching the intersection of human rights and AI harm;
- private companies can have reference to when considering whether an AI system could have potential human rights implications (subject to further consideration of the applicable AI system, the context in which it is used and the nuances of human rights law); and
- public entities can have reference to when considering how AI systems could interact with human rights.

Please note that Table 2 is not intended to be used solely to determine whether an AI system could potentially result in human rights breaches. Rather it is designed to be easily updatable, complement other tools and is illustrative only. Furthermore, when considering the harms of an AI system, it will also be important to consider the severity of the impacts (including the scope, gravity and irremediability of that impact).

21. ''Artificial intelligence must be grounded in human rights, says High Commissioner' *United Nations Human Rights Office of the High Commissioner* (Web Page, 12 July 2023) <https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner>.
22. See https://artificialintelligenceact.eu/article/27/.
23. See https://humanrights.gov.au/our-work/technology-and-human-rights/publications/hria-tool-ai-banking.
24. See https://www.ohrc.on.ca/sites/default/files/Human%20Rights%20Impact%20Assessment%20for%20AI.pdf.

METHODOLOGY

In order to produce Table 2, the following approach was taken:

- nine global human rights instruments and the Sustainable Development Goals (**SDGs**) were reviewed in order to produce a list of 10 fundamental human rights that are most likely to result in AI harms. The global human rights instruments and the in-scope fundamental human rights are listed below;
- the existing harm taxonomies and AI incident databases were reviewed to establish how (or if) they approach human rights and common features of incidents that were identified as having a human rights component;
- existing literature and core human rights documents on the interaction of human rights and AI was reviewed to determine what the common features of AI systems (or how they are used) that could potentially result in interference with human rights (subject to further consideration of the applicable AI system, the context in which it is used and the nuances of human rights law); and
- (where applicable) news articles and court cases were selected to provide illustrations of how the harms discussed in the table have arisen in practice.

Table 1: Summary of human rights instruments and the in scope fundamental human rights considered

Human rights Instruments[25]	• Universal Declaration of Human Rights (**UDHR**) • International Convention on the Elimination of All Forms of Racial Discrimination (CERD) • International Covenant on Civil and Political Rights (ICCPR) • International Covenant on Economic, Social and Cultural Rights (ICESCR) • Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) • Convention on the Rights of the Child (CRC) • International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families (CMW) • Convention on the Rights of Persons with Disabilities (CRPD)
Fundamental rights most applicable to AI	• Rights to equality and non-discrimination Examples: ICESCR 2(2), ICCPR 2 and 3, UDHR 2, 7, 23(2), CRC 2, CRPD 5, CEDAW 2, CERD 2, CMW 7 • Right to enjoyment of scientific progress Examples: ICESCR 15(1)(b) • Procedural fairness Examples: ICCPR 14, UDHR 10, CRC 37 and 40, CRPD 12 and 13, CERD 5, CMW 16 and 18 • Right to privacy Examples: UDHR 12, ICCPR 17, CRC 16, CRPD 22, CMW 14 • Right to meaningful employment Examples: ICESCR 7, UDHR 23(1), CRPD 27, CEDAW 11 • Freedom from physical and psychological harm and interference Examples: UDHR 3, ICCPR 9, ICCPR 16, CRPD 10, CMW 16 • Freedom of religion, opinion, expression and access to information Examples: UDHR 18 and 19, ICCPR 18 and 19, CRC 13 and 17, CRPD 21, CMW 13 • Prohibition of advocacy of national, racial or religious hatred Examples: ICCPR 20, CERD 4 • Right to freedom of assembly and the freedom of association Examples: UDHR 20, ICCPR 21 - 22, CRC 15 • Right to take part in public affairs Examples: UDHR 21, ICCPR 25, CRPD 29, CEDAW 7 and 8

25. In addition to the nine instruments listed, the Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (CAT) and the International Convention for the Protection of All Persons from Enforced Disappearance (CPED) were also considered. However, they were not included on the basis they had a looser nexus to the most common AI uses at this time. In addition, this report does not include any human rights instruments not yet in force (such as the AI Convention) or regional human rights instruments (such as the African Charter on Human and Peoples' Rights (ACHPR), the American Convention on Human Rights (ACHR), and the European Convention on Human Rights (ECHR)). The Sustainable Development Goals (SDGs) were also considered given their vision for sustainable development grounded in international standards for human rights. Some references to the SDGs are included but they are not the focus of this report given they are not legally binding.

Table 2: Mapping AI harm and human rights

Important Notes:

(a) human rights are not absolutist. Rather, human rights law balances rights and interests to reach a conclusion. Accordingly, whether a particular AI system, or even a particular harm listed below, could give rise to an interference of a human right will always be subject to further consideration of the applicable AI system, the context in which it is used and the nuances of human rights law; and

(b) the use of examples are illustrative only. Although in some cases the examples have been found to result in human rights breaches, in many cases the situations are illustrative of the harms discussed in the table (rather than a suggestion that the example amounts to a potential breach of human rights).

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?
1. Rights to equality and non-discrimination ***Summary:*** *Everyone has the right to equality and freedom from discrimination on protected grounds. These grounds include race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth, or other status. Discrimination can include any distinction, exclusion, restriction, or preference, which can impair people's enjoyment of rights and freedoms.* **Example Sources: ICESCR 2(2), ICCPR 2 and 3, UDHR 2, 7, 23(2), CRC 2, SDG 5 and 10, CRPD 5, CEDAW 2, CERD 2, CMW 7**	Harm may arise when the output of an AI system results in differential treatment, exclusion, restriction, or preference based on protected grounds. This harm is particularly pronounced where AI systems are used to apply rules en masse rather than assessing the merits of individual situations. Such harm (which includes where the AI system output itself is discriminatory and when the AI system output is used to influence a human decision-maker to make a discriminatory decision) may be the result of: • direct discriminatory programming (e.g. the AI system treats an individual or group differently based on protected grounds) • indirect discriminatory programming (e.g. the AI system treats an individual or group the same but in a way that results in a disadvantage to people from a protected group) • the design of the underlying training data (e.g. datasets that only record gender as binary) • inaccurate, irrelevant or outdated training data (e.g. datasets with historical bias will be replicated and exaggerated in an AI system)	• Predictive policing tools[26] • Recidivism algorithms in sentencing[27] • Biased access to health care and social services (e.g. health care systems trained on patients from higher socioeconomic backgrounds)[28] • Employment assessment/ support tools (e.g. government employment tools that negatively rate women, disabled people and those over 30)[29]	**Most at risk sectors** • Administrative decision making • Education • Core services (including banking and health care) • Employment

26. See, for example, Will Douglas Heaven 'Predictive policing algorithms are racist. They need to be dismantled.' *MIT Technology Review* (Web Page, 17 July 2020) < https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice>; Tzu-Wei Hung and Chun-Ping Yen 'Predictive policing and algorithmic fairness' (2023) 201 *Synthese* 206. Available at: https://link.springer.com/article/10.1007/s11229-023-04189-0.
27. See, for example, *State v Loomis* 881 NW 2d 749 (Wis, 2016); see also, for example, https://harvardlawreview.org/print/vol-130/state-v-loomis.
28. See, for example, https://incidentdatabase.ai/cite/124; Richard Chen et al, 'Algorithm Fairness in Artificial Intelligence for Medicine and Healthcare' (2023) 7(6) *National Biomedical Engineering* 719 (available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10632090/pdf/nihms-1940941.pdf); Ziad Obermeyer et al, 'Dissecting racial bias in an algorithm used to manage the health of populations' (2019) 366 *Science* 447 (available at: https://doi.org/10.1126/science.aax2342).
29. See, for example, https://algorithmwatch.org/en/austrias-employment-agency-ams-rolls-out-discriminatory-algorithm/ and https://incidentdatabase.ai/cite/95; Prasanna Tambe, Peter Cappelli, and Valery Yakubovich, 'Artificial Intelligence in Human Resources Management: Challenges and a Path Forward' (2019) 61(4) *California Management Review* 15.

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?
	• automatic content moderation based on datasets that incorporate discriminatory assumptions • the design of the algorithm itself (e.g. algorithms designed to draw inferences from disparate data to assess future behaviours based on inferences, predictions or correlations or that fail to take into consideration key information; the use of coarse variables as predictors; blurred boundaries on algorithmic categories) • indirect influence from AI developers (e.g. non-diverse AI developers may inadvertently introduce bias or discrimination into the design of the AI system) • probabilistic outputs (e.g. an output may be based on the most probable answer but does not take into account particular factors) • accuracy measures (e.g. users are shown false positives but no information about false negatives) • inappropriate AI systems (e.g. an AI system that is not appropriate for the use it is being put to or does not consider the social conditions in how the output will be used) • The impact of discriminatory harm from AI systems is amplified where individuals do not know an AI system was used to either make, or was a factor in the making of, a decision that significantly impacts them.	• Biased fraud detection on racialised communities (e.g. automated claims processing that disproportionately delays claims of African American homeowners)[30] • Car insurance premiums directly influenced by gender and birthplace[31] • Racially biased facial recognition technology[32]	

30. See, for example, Huskey v State Farm Fire & Cas Co, 22 C 7014 (ND Ill, 2023); 'A suit filed by the Center for Race, Inequality, and the Law takes a new approach to proving racial bias in the insurance industry', NYU Law (Web Page, 22 December 2022) https://www.law.nyu.edu/news/deborah-archer-cril-alexander-rose-state-farm.
31. See, for example, http://www.dei.unipd.it/~silvello/papers/2021_aies2021.pdf.
32. See, for example, https://www.scientificamerican.com/article/police-facial-recognition-technology-cant-tell-black-people-apart/; Marcus Smith and Monique Mann Facial Recognition Technology and Potential for Bias and Discrimination (Cambridge University Press, 2024) at 87-95 (available at: https://www.cambridge.org/core/books/cambridge-handbook-of-facial-recognition-in-the-modern-state/facial-recognition-technology-and-potential-for-bias-and-discrimination/B1C4A7F38AE00781EC8A559EFE48B3DF).

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?
2. Right to privacy *Summary: No one shall be subjected to arbitrary or unlawful interference with his or her privacy, family, home or correspondence, nor to unlawful attacks on his or her honour and reputation.* *Everyone is entitled to protection by law to address this.* Example Sources: UDHR 12, ICCPR 17, CRC 16, CRPD 22, CMW 14	Harm may arise when an AI system is used in a way that contributes to, or causes, a breach of an individuals' privacy. This harm is particularly pronounced where AI systems: • involve the collection of sensitive data (including health data); • are used to intrude into the seclusion of an individual; or • are used to ground decisions that have a legal or other similarly significant impact on an individual. Such harm (which can arise at any point through the AI lifecycle) may be the result of: • the collection of personal data to train an AI system without individuals' knowledge or (if required by national laws) consent • the collection of personal data (either as training data or input data) that is unnecessary and/or disproportionate for the purpose which it is being collected (i.e. to train an AI system or use the AI systems) • the generation of personal data (either correct or incorrect) about individuals based on inferences, predictions and correlations found in other data (this includes developing profiles to inform decisions, such as about health care, social benefits, insurance and employment) • the amendment of personal data without the knowledge of the organisation deploying the AI system or the relevant individual	• The Netherlands SyRI (System Risk Indication) algorithm system applied by the Dutch Government for digital welfare fraud that was found to interfere with Article 8 ECHR (Right to private life)[33] • Automated facial recognition tools used by the South Wales Police without consent and without clear limits on its use that was found to interfere with Article 8 ECHR[34] • The training of ChatGPT on personal data of Italians' without an appropriate legal basis and in violation of the transparency principles in the GDPR[35] • The scraping of photos and biometric information from the internet for use in facial recognition services provided to law enforcement and intelligence agencies[36]	**Most at risk technology types** • Any technology that uses personal data **Most at risk sectors** • Any sector that collects, or uses, personal data

33. *NJCM v the Netherlands (SyRI)*, District Court of The Hague 5/02/20, ECLI:NL:RBDHA:2020:865.
34. *R (on the application of Edward Bridges) v The Chief Constable of South Wales* [2020] EWCA Civ 1058.
35. See https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10085432#english.
36. See, for example, https://www.autoriteitpersoonsgegevens.nl/en/current/dutch-dpa-imposes-a-fine-on-clearview-because-of-illegal-data-collection-for-facial-recognition;

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?
	• the disclosure (including sale) of personal data to third parties without the knowledge (or if required by state laws) consent of the relevant individual • inadequate security of AI systems that utilise personal data (this can, for example, expose both personal data and individuals using AI enabled devices to cyber-criminal attacks) • a failure to delete or de-identify personal data that is out-of-date and/or no longer necessary for the purpose for which it was collected *Note: Although the above risks are not new – the risk to privacy posed by AI is amplified due to the amount of data that is utilised by AI systems and the reduced level of human input*	• Indiscriminate mass surveillance by real-time facial recognition systems[37] • Collection of data on social media without users' consent or knowledge[38] • Use of facial recognition technology to display personalised advertisements on subway car doors[39]	

37. See, for example, https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9575877.
38. See, for example, https://www.law.com/therecorder/2020/01/29/facebook-agrees-to-550m-deal-to-settle-biometric-suit-over-tag-suggestions/;
39. See, for example, https://globalfreedomofexpression.columbia.edu/cases/the-case-of-sao-paulo-subway-facial-recognition-cameras/#:~:text=The%20Court%20held%20that%20the,to%20cease%20using%20the%20technology.

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?
3. Procedural fairness *Summary: Everyone is entitled to procedural fairness, including fair and public trial by an independent and impartial decision maker for civil, criminal and administrative matters.* Example Sources: ICCPR 14, UDHR 10, CRC 37 and 40, CRPD 12 and 13, CERD 5, CMW 16 and 18	Harm may arise when the output of an AI system is used to restrict, or deprive, individuals of their rights to a fair trial (where that decision would have a significant or serious impact on civil rights). This harm is particularly pronounced where AI systems are used to make the decision without human input. Such harm (which includes where the AI system output itself makes a decision and when the AI system output is used to influence a human decision-maker) may be the result of decisions (either by law enforcement, judiciary or other decision making bodies) made, or influenced by: • an AI system that is biased or discriminatory (see row 1 above) • the level of accuracy of an AI system (inaccuracy and errors may arise from data quality, coding errors, misinterpretation of the underlying laws, model errors, including a failure to account for particular variables) • an AI system that is not appropriate for a particular use case but is used nonetheless • decisions made by an AI system without human involvement or oversight • a decision maker relying (either in full or part) on the output of the AI system without understanding how the output has been produced and/or key limitations in that output • individuals subject to decision-making by an AI system not being made aware that an AI system has been involved or where they are made aware, being unable to challenge that decision[40] *Note: Not all decisions by public bodies are subject to the right to fair trial. However, the potential harms that AI presents to the decision-making process apply equally to all decisions.*	• Predictive policing tools[41] • Recidivism algorithms in sentencing[42] (such as the Correctional Service of Canada's use of psychological and actuarial risk assessment tools to assess inmates' risk of recidivism)[43]	**Most at risk technology types** • Automated decision-making • Facial recognition **Most at risk sectors** • Law enforcement • Judicial system • Administrative decision making

40. Arun Rai, 'Explainable AI: From Black Box to Glass Box' (2020) 48 *Journal of the Academy of Marketing Science* 137. Available at: https://link.springer.com/article/10.1007/s11747-019-00710-5.
41. See, for example, https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/; Tzu-Wei Hung and Chun-Ping Yen 'Predictive policing and algorithmic fairness' (2023) 201 *Synthese* 206. Available at: https://link.springer.com/article/10.1007/s11229-023-04189-0;
42. See, for example, *State v Loomis* 881 NW 2d 749 (Wis, 2016); see also, for example, https://harvardlawreview.org/print/vol-130/state-v-loomis/.
43. *Ewert v Canada* [2018] 2 SCR 165. Available at: https://decisions.scc-csc.ca/scc-csc/scc-csc/en/item/17133/index.do.

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?
4. Freedom from physical and psychological harm and interference *Summary: Everyone has the right to life, liberty, and security of person and the right to enjoy the best attainable state of physical and mental health.* Example Sources: UDHR 3, ICCPR 9, ICCPR 16, CRPD 10, CMW 16	Harm may arise when either the use of AI itself, or the output of an AI system, results in: • threats to a person's life, liberty or security (e.g. incitement of violence by deepfakes) • restrictions on, or loss of, a person's liberty (e.g. wrongful arrest or imprisonment based on inaccurate output of an AI system) • harms to a person's mental health (e.g. from the creation and distribution of material by an AI system) • threats to a person's physical or mental health (e.g. the use of AI to threaten another person). Such harm may be the result of: • the use of AI systems to develop material based on particular individuals (e.g. the use of generative AI to develop sexually explicit deepfakes that cause serious harm to the mental or physical health of a depicted individual)[44] • the use of AI systems to create or amplify threats to life or security of particular individuals or groups (e.g. the use of AI powered voice and video cloning by cyber criminals to impersonate individuals or to incite hatred or racial discrimination)[45] • reliance by the police or other government agency on inaccurate, or discriminatory, output of AI systems for arrests (e.g. reliance on facial recognition technology or sound detection software)	• Police reliance on sound detection technology (with known limitations) to detain or imprison individuals[46] • Predictive policing tools[47] • Recidivism algorithms in sentencing.[48] See in particular Canada's psychological and actuarial risk assessment tools used for determining the risk of recidivism[49] • Chatbots that provide advice about committing suicide or advice about eating disorders[50]	**Most at risk technology types** • Facial recognition • Chatbots • Deepfakes/Cloning • Automated decision-making **Most at risk sectors** • Government • Law enforcement • Health care • Social Media

44. See, for example, https://www.bbc.com/news/world-europe-66877718; https://www.pbs.org/newshour/show/women-face-new-sexual-harassment-with-deepfake-pornography.
45. See, for example, https://www.fbi.gov/contact-us/field-offices/sanfrancisco/news/fbi-warns-of-increasing-threat-of-cyber-criminals-utilizing-artificial-intelligence.
46. See, for example, https://scholar.google.com.au/scholar_case?case=14253973677235046996&q=Williams+v.+City+of+Chicago+%2B+shotspotter&hl=en&as_sdt=2006&as_ylo=2021&as_vis=1 and https://www.chicagotribune.com/2021/08/20/how-ai-powered-tech-landed-a-chicago-grandfather-in-jail-for-nearly-a-year-with-scant-evidence/.
47. See, for example, https://worldcrunch.com/tech-science/ai-images-extremists-germany and https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/; Tzu-Wei Hung and Chun-Ping Yen 'Predictive policing and algorithmic fairness' (2023) 201 *Synthese* 206. Available at: https://link.springer.com/article/10.1007/s11229-023-04189-0.
48. See, for example, *State v Loomis* 881 NW 2d 749 (Wis, 2016); see also, for example, https://harvardlawreview.org/print/vol-130/state-v-loomis.
49. *Ewert v Canada* [2018] 2 SCR 165. Available at: https://decisions.scc-csc.ca/scc-csc/scc-csc/en/item/17133/index.do.
50. See, for example, https://www.vice.com/en/article/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says/ and https://www.psychiatrist.com/news/neda-suspends-ai-chatbot-for-giving-harmful-eating-disorder-advice/.

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?

5. Freedom of religion, opinion, expression and access to information
 	
Summary: Everyone has the right to have a religion or belief (and the freedom to manifest that religion) and to hold and express opinions without interference except where provided by law and necessary for respecting the right or reputation of others or for protecting national security, public order, health or morals.
 	
Example Sources: UDHR 18 and 19, ICCPR 18 and 19, CRC 13 and 17, CRPD 21, CMW 13
	Harm may arise when either the use of AI itself, or the output of an AI system, suppresses or restricts an individual's speech, information or ideas in a way that is considered unfair, arbitrary, or disproportionate.
	
	Such harm may be the result of:
	• the utilisation of AI systems (especially facial recognition and other surveillance systems) to identify and/or monitor individuals
	• the suppression of information by AI systems. This may either be due to:
	  • a deliberate design decision of the AI developer or deployer (e.g. through moderation techniques that block or restrict content);
	  • inadvertent result of the design, deployment or use of AI. This can arise in a number of ways – for example, information may blocked for certain individuals or groups due to an AI system being biased or discriminatory (see Row 1 above), due to errors in the AI system or due to over-moderation of content;
	  • systemic issues with the design of AI models or AI systems. This mainly arises in the context of large language models that are not trained on lower-resource languages[51]
	• the utilisation of AI systems to either:[52]
	  • censor individuals or groups (e.g. mass disinformation campaigns can be used to generate large amounts of false content about individuals or groups online and suppress genuine information);[53] or
	  • force self-censorship (e.g. an AI system can be weaponised to harass particular individuals into taking (or not taking) particular actions).
		• Alleged uses of facial recognition technology by governments targeting legitimate protests and/or political opponents[54]
		• Alleged censorship of content relating to Palestine on social media platforms[55]
		• Removal of cultural, historical or artistic content by automated moderation[56]
			Most at risk technology types
			• Facial recognition
			• Recommender systems
			• Chatbots
			• Social media
			• Deepfakes/Cloning
			Most at risk sectors
			• Media/News Industry
			• Tech Industry
			• Government
			• Law enforcement / intelligence agencies

51. See, for example, Alena Gorbacheva 'No Language Left Behind: How to Bridge the Rapidly Evolving AI Language Gap' *UNDP* (Web Page, 6 October 2023) https://www.undp.org/kazakhstan/blog/no-language-left-behind-how-bridge-rapidly-evolving-ai-language-gap; see also, for example, Karen Hao 'A new vision of artificial intelligence for the people' *MIT Technology Review* (Web Page, 22 April 2022) <https://www.technologyreview.com/2022/04/22/1050394/artificial-intelligence-for-the-people/>.
52. Niva Elkin-Koren, 'Contesting algorithms: Restoring the public interest in content filtering by artificial intelligence' (2020) *7 Big Data & Society* (available at: https://doi.org/10.1177/2053951720932296); Emma Llanso et al, 'No amount of "AI" in content moderation will solve filtering's prior-restraint problem' (2020) *7 Big Data & Society* (available at: https://doi.org/10.1177/2053951720920686).
53. Niva Elkin-Koren, 'Contesting algorithms: Restoring the public interest in content filtering by artificial intelligence' (2020) *7 Big Data & Society* (available at: https://doi.org/10.1177/2053951720932296); Emma Llanso et al, 'No amount of "AI" in content moderation will solve filtering's prior-restraint problem' (2020) *7 Big Data & Society* (available at: https://doi.org/10.1177/2053951720920686); see a range of publications from the Transatlantic Working Group on Content Moderation Online and Freedom of Expression (available at: https://www.ivir.nl/twg/); Transatlantic Working Group Final Report - *Freedom and Accountability: A Transatlantic Framework for Moderating Speech Online* (2020) (available at: https://www.annenbergpublicpolicycenter.org/wp-content/uploads/Freedom_and_Accountability_TWG_Final_Report.pdf).
54. See, for example, https://www.unwantedwitness.org/download/Surveillance-State-Parliament-Endorses-Unregulated-Surveillance.pdf and https://www.techspot.com/news/102148-russian-authorities-used-facial-recognition-tech-identify-arrest.html.
55. See, for example, https://www.hrw.org/report/2023/12/21/metas-broken-promises/systemic-censorship-palestine-content-instagram-and.
56. See https://incidentdatabase.ai/cite/275.

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?
6. Right to meaningful employment *Summary: Everyone has the right to choose to engage in meaningful work and the labour market must remain open to those suitably qualified.* *Individuals must not be unjustly deprived of work, and reasonable accommodations must be made to support them.* **Example Sources: ICESCR 7, UDHR 23(1), SDG 8, CRPD 27, CEDAW 11**	Harm may arise when either the use of AI itself, or the output of an AI system, impacts an individual's ability to access employment or limits their workplace rights (such as the right to strike or join a trade union). Such harm (which can arise both as a result of reliance on, or influence by, AI systems within the recruitment process or in relation to managing workers) may be the result of: • (for the recruitment process) the use of indirect discriminatory programming or biased datasets/algorithms. This can affect who job advertisements are shown to, whether a candidate progresses to interview based on a review of the CV by an AI system and whether a candidate is offered a position based on how they performed in AI-assisted interviews • (for the recruitment process) employers placing undue reliance on 'objective' results produced from the analysis of CVs or interviews by emotion recognition or sentiment analysis[57] • (for the recruitment process) the retention of large amounts of historical data by employers which is then used by AI systems to sort and determine prospective candidates • (for managing workers) the use of AI systems to monitor or predict employee performance, or determine promotions or workplace opportunities, (including via surveillance technology or chatbots which could introduce errors or bias into the process, e.g. favouring employees who have worked longer hours, potentially disadvantaging those with caregiving responsibilities).	• AI algorithms that make female job seekers less likely to be shown adverts for highly paid jobs than males[58] • Automated recruitment platform built on employees' CVs that discriminates against women[59] • Booking system based on workers' reliability and participation that does not distinguish between reasons for absences (e.g. absence for sickness treated the same as unauthorised absences (no-shows, lateness))[60]	**Most at risk technology types** • Automatic skill assessment • Facial Recognition Technology • Automated decision-making **Most at risk sectors** • Government • Human resources

57. See, for example, https://interaktiv.br.de/ki-bewerbung/en/.
58. Samuel Gibbs 'Women less likely to be shown ads for high-paid jobs on Google, study shows' *The Guardian* (Web Page, 8 July 2015) <https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study>.
59. Jeffrey Dastin 'Amazon scraps secret AI recruiting tool that showed bias against women', *Reuters* (Web Page, 11 October 2018) <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.
60. See, for example, *Filcams CGIL Cologna, Nidil CGIL Bologna, and Filt CGIL Bologna v Deliveroo Italia S.R.L.* Trib. Bologna, Ord. no 2949/2019 (2020).

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?
7. Right to enjoyment of scientific progress *Summary: Everyone has the right to enjoy the benefits of scientific progress and its applications and to the protection of the moral and material interests resulting from any scientific production which they author. This right is countervailing to the other rights in this table, as it implies that there is an interest in minimising undue regulation of AI.* **Example Sources: ICESCR 15(1)(b), SDG 9**	Harm may arise when individuals, communities, or nations are denied the ability to engage with, benefit from, or contribute to developments in AI and the benefits that it can bring. This can occur due to various factors, including socioeconomic status, geographical location, political restrictions, educational barriers, or systemic inequalities.[61] Such harm may be the result of wide number of factors including: • unequal access to the benefits of AI (e.g. countries in the global north are disproportionately benefiting from AI productivity gains, and AI research is currently dominated by China and the United States)[62] • structural limitations (e.g. the global south currently has a lower ability to adopt AI on a scaled level, including in relation to the disparities in the availability of talent and capability, data, models and technical infrastructure)[63]	• Not available	**Most at risk technology types** • All **Most at risk sectors** • All

61. Tanima Bag, 'Socio-economic Impacts of Scientific-Technological Advancements' (2023) 12 *International Journal of Multidisciplinary Educational Research* 70. Available at: https://ijmer.s3.amazonaws.com/pdf/volume12/volume12-issue8(4)/13.pdf.
62. See, for example, https://www.nature.com/articles/s41598-024-79863-5.
63. See, for example, https://www.weforum.org/stories/2023/01/davos23-ai-divide-global-north-global-south/ and https://www.un.org/digital-emerging-technologies/sites/www.un.org.techenvoy/files/MindtheAIDivide.pdf.

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?

8. Prohibition of advocacy of national, racial or religious hatred
 	
Summary: Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.
 	
Example Sources: ICCPR 20, CERD 4
	Harm may arise when either AI itself, or the output of an AI system, is used to incite discrimination, hostility or violence. This is particularly pronounced in generative AI, as it has the ability to generate content (including content that is inaccurate or biased).
	
	Such harm may be the result of:
	
	• AI systems that:[64]
	  • are biased or discriminatory (see row 1);
	  • are designed to promote particular outcomes (e.g. deliberate programming choices of developers);
	  • are modified to create a particular outcome (e.g. through the modification or removal of moderation processes); or
	  • are used in such a way that inadvertently results in harmful outcomes (e.g. the prioritisation of content that incites hatred and violence)
	• the utilisation of AI systems to generate and disseminate content that incites hatred and violence (e.g. using generative AI to generate automatic text for recruitment purposes or spreading customized fake news and terrorism related conspiracy theories)
	• the utilisation of AI systems to target content at particular individuals (e.g. using AI to target messages at individuals that have repeatedly searched for violent content online).
		• Concerns of international bodies and intelligence agencies that violent extremists will develop deepfakes and AI-powered fake news sites to instrument for propaganda, radicalization or as a call for action[65]
		• Alleged manipulation of Facebook algorithms by Russia's Internet Research Agency to promote anti-immigrant rhetoric and hate speech (which resulted in physical gatherings in Houston, Texas)[66]
			Most at risk technology types
			• Generative AI
			• Deepfakes/Cloning
			• Social Media
			• Recommender systems
			
			Most at risk sectors
			• Social media
			• Media and Entertainment Industry

64. Jane Bailey et al, 'AI and Technology-Facilitated Violence and Abuse' (2020) *Artificial Intelligence and the law in Canada* (available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3734663); Thomas King et al, 'Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions', (2019) 26 *Science and Engineering Ethics* 89 (available at: https://link.springer.com/article/10.1007/s11948-018-00081-0); Tais Fernanda Blauth et al, 'Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI', (2022) 10 *IEEE Access* 77110 (available at: https://ieeexplore.ieee.org/abstract/document/9831441); Mark Latonero, *Governing Artificial Intelligence: Upholding Human Rights & Dignity* (Report, 10 October 2018) (available at: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf).
65. See, for example, https://unicri.org/sites/default/files/2021-06/Malicious%20Use%20of%20AI%20-%20UNCCT-UNICRI%20Report_Web.pdf and https://www.dni.gov/files/NCTC/documents/jcat/firstresponderstoolbox/151s_First_Responders_Toolbox-Violent_Extremists_Use_of_Generative_Artificial_Intelligence.pdf.
66. See, for example, https://www.nytimes.com/2018/02/17/technology/indictment-russian-tech-facebook.html and https://www.tandfonline.com/doi/full/10.1080/23738871.2020.1778760.

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?

9. Right to freedom of assembly and the freedom of association
 	
Summary: Everyone has the right to freedom of peaceful assembly and association
 	
Example Sources: UDHR 20, ICCPR 21 - 22, CRC 15
	Harm may arise when either AI itself, or the output of an AI system, is used to prevent individuals from gathering for collective action (whether that be for social, political, economic, or other purposes).[67]
	
	Such harm may be the result of:
	
	• AI systems that:
	  • are biased or discriminatory (see row 1);
	  • are designed to promote particular outcomes (e.g. through the choices of human moderators who can design AI systems to delete events or conversations from social media);
	  • are modified to create a particular outcome (e.g. through the manipulation of AI rules to have content prioritized or deprioritized); or
	  • are used in such a way that inadvertently affects individuals' actions (e.g. AI systems designed to personalise the content viewed by individuals may minimise how and where individuals assemble online and what types of association can be formed)
	• the utilisation of AI to generate and disseminate content that impacts the freedom of assembly and association
	• the utilisation of AI systems to target content at particular individuals
	• the utilisation of AI systems to pre-emptively identify threats, monitor potential dissent and track particular individuals.
	
	Note: It is accepted that the freedom of peaceful assembly, of expression and of association may apply to both physical interactions and analogous interactions taking place online.[68]
		• Use of AI systems to monitor and control local minority groups
		• Alleged manipulation of Facebook algorithms by Russia's Internet Research Agency to promote anti-immigrant rhetoric and hate speech (which resulted in physical gathering in Houston, Texas)[69]
			Most at risk technology types
			• Facial recognition
			• Predictive policing
			
			Most at risk sectors
			• Police
			• Government
			• Education

67. Hamid Akin Unver, *Artificial Intelligence (AI) and Human Rights: Using AI as a Weapon of Repression and its Impact on Human Rights* (Technical Report, May 2024) (available at: https://www.researchgate.net/publication/381407889); Steven Feldstein, 'The Road to Digital Unfreedom: How Artificial Intelligence is Reshaping Repression', (2019) 30 *Journal of Democracy* 40 (available at: https://www.journalofdemocracy.org/articles/the-road-to-digital-unfreedom-how-artificial-intelligence-is-reshaping-repression/); Steven Feldstein, *The Rise of Digital Repression: How Technology is Reshaping Power, Politics, and Resistance* (Oxford University Press, 2021).
68. UN Human Rights Council *The Promotion and Protection of Human Rights in the Context of Peaceful Protests* A/HRC/RES/38/11 (16 July 2018).
69. See, for example, https://www.nytimes.com/2018/02/17/technology/indictment-russian-tech-facebook.html.

FUNDAMENTAL RIGHTS	HARM	EXAMPLES	WHERE MAY THIS ARISE?

10. Right to take part in public affairs
 	
Summary: Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.
 	
Example Sources: UDHR 21, ICCPR 25, CRPD 29, CEDAW 7 and 8
	Harm may arise when either AI itself, or the output of an AI system, is used to inhibit or restrict individuals or groups from engaging in democratic processes, such as voting, campaigning, or participating in public discourse.[70]
	
	Such harm may be the result of:
	
	• AI systems that:
	  • are biased or discriminatory (see row 1); or
	  • (similar to rows 8 and 9) are designed to promote particular outcomes or are modified to create a particular outcome or whose use inadvertently affect individuals' actions
	• the utilisation of AI to generate and disseminate content that influences candidates or voters (e.g. smear campaigns using deepfakes of candidates or fake press releases that are used to negatively influence how voters view a particular candidate or force candidates to withdraw)
	• the utilisation of AI to generate and disseminate content that suppresses voter turnout or otherwise impacts the demonstration process (e.g. fake content about how or where to vote or deepfakes suggesting changes to the election or candidates).
		• Deepfakes of independent candidates in the 2024 Bangladesh national elections announcing their withdrawal (which was incorrect)[71]
		• CETAS found deceptive AI-generated content did shape the 2024 US election discourse by amplifying other forms of disinformation and inflaming political debates. This included 24 instances of AI enabled smear campaigns, 14 instances AI-enabled voter targeting; 6 instances of AI-generated misattribution); 6 instances of AI-generated parody and satire disinformation); 4 instances of AI based information campaigns using fake US new sources) and 2 fabricated celebrity endorsements)[72]
			Most at risk technology types
			• Generative AI
			• Deepfakes/Cloning
			• Social Media
			• Recommender systems
			
			Most at risk sectors
			• Social media
			• Media and Entertainment Industry

70. Karl Manheim and Lyric Kaplan, 'Artificial Intelligence: Risks to Privacy and Democracy' (2019) 21 *Yale Journal of Law and Technology* 106 (available at: https://heinonline.org/HOL/Page?handle=hein.journals/yjolt21&div=4&g_sent=1&casa_token=&collection=journals); Celal Hakan Kan, 'Artificial Intelligence (AI) in the Age of Democracy and Human Rights: Normative Challenges and Regulatory Perspectives' (2024) 9(25) *International Journal of Eurasian Education and Culture* 145 (available at: https://www.ijoeec.com/Makaleler/1355005649_8.%20145-166%20Celal%20Hakan%20Kan.pdf); Chen Yu, 'How Will AI Steal Our Elections?' (2024) (available at: https://files.osf.io/v1/resources/un7ev/providers/osfstorage/65d879c8c3ab490b7846b045?direct=&mode=render); Masabah Bint E. Islan et al., 'AI Threats to Politics, Elections, and Democracy: A Blockchain-Based Deepfake Authenticity Verification Framework' (2024) 2 *Blockchains* 458 (available at: https://www.mdpi.com/2813-5288/2/4/20).
71. See https://www.boomlive.in/decode/deepfake-elections-disinformation-bangladesh-india-us-uk-indonesia-24087.
72. See https://cetas.turing.ac.uk/sites/default/files/2024-11/cetas_research_report_-_ai-enabled_influence_operations_-_safeguarding_future_elections.pdf.

ABOUT KING & WOOD MALLESONS

A firm born in Asia, underpinned by world class capability. With over 3,700 lawyers in 26 global locations, we draw from our Eastern and Western perspectives to deliver incisive counsel.

We help our clients manage their risk and enable their growth. Our full-service offering combines un-matched top tier local capability complemented with an international platform. We work with our clients to cut through the cultural, regulatory and technical barriers and get deals done in new markets.

ABOUT ICAAD

ICAAD is a international human rights center that empowers our network to become leaders in driving systemic change.

We build the capacity of marginalized communities as systems-level changemakers, offer a creative space for innovation, and strategically connect those with greater influence to data and insights to increase their capacity for informed and meaningful action.

www.kwm.com www.icaad.ngo