



## Evaluating the Risk of AI Systems to Human Rights from a Tier-based Approach

### I. *Evaluating AI risks to human rights begins with assessing the rights.*

The European Commission acknowledged in its White Paper on Artificial Intelligence that “The main risks related to the use of AI concern the application of rules designed to protect fundamental rights.”[1] And yet, it intends to regulate AI from a risk-based approach that would focus on operational risks and those related to external factors. In line with other civil society organizations,[2] **ECNL urges the Commission to consider risks within a rights-based framework, i.e. as risks to human rights, rule of law and democracy.**

Making this distinction is necessary, as multiple actors conceive risks from different perspectives. As the European Union Agency for Fundamental Rights (FRA) cautioned, “When asked about general risks when using AI, the interviewees did not always mention fundamental rights as the main risks, although some highlighted related topics.”[3] In their analysis, risk perceptions ranged from inaccuracy as a risk of using AI, to potential bias and business risks. The OECD, for example, considers risks to human rights and well-being, as well as risks to critical sectors or functions.[4] For these reasons, and because people’s fundamental rights must be prioritized over other interests, all AI regulation must be based on a human rights framework.

As expressed previously,[5] it is alarming that the European Commission plans to limit robust oversight and safeguards to AI systems determined as “high risk” only. This would lead to a legal vacuum for all other AI systems, and unduly put the burden of proof on the person subjected to the AI system, as opposed to those developing or deploying it. This would also create an uneven playing field and incentivize classifying AI systems as low risk to avoid further regulation. **There are compelling arguments that a thorough, inclusive and transparent human rights impact assessment (HRIA) must be the starting point for all subsequent regulatory actions of any AI system, which is in line with the EU’s proclaimed approach.**

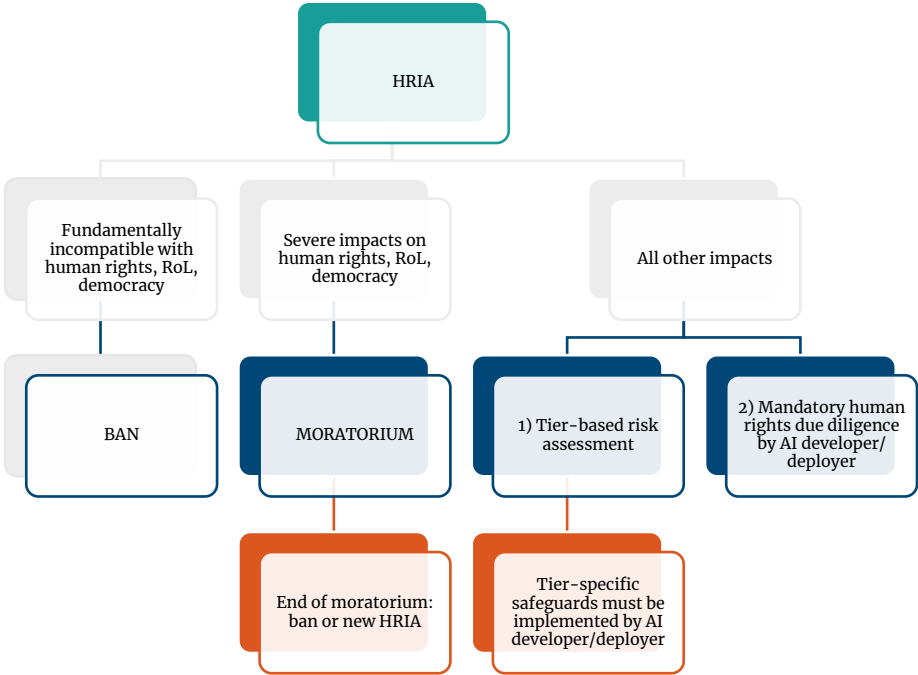
To ensure public interest, the HRIA should be conducted within a clear governance framework and oversight by a regulatory agency or independent expert, as opposed to the developers of the AI system themselves. Importantly, the ex-ante HRIA should be based on input from affected communities and stakeholder groups, including civil society. This should begin – and where necessary end – with the questions: **is the technology effective for [x] purpose and is there no other less intrusive way to achieve the same result? Is such a purpose even legitimate?**

The HRIA should be made available and accessible, as recommended by the Council of Europe[6] (CAHAI para. 118). Public registers should be established for AI systems being deployed in the public sector. The results of the HRIA should then guide future regulatory action, *not* the risk classification. Of note, any risk measurement should follow the precautionary principle: where it is difficult or impossible to determine a metric due to



uncertainty, the risk should be determined as higher rather than lower. What’s more, when the consequences are unacceptable because they are inherently incompatible with human rights, the precautionary principle would support a ban or moratorium.[7]

**II. A risk approach is only appropriate for AI systems that should not otherwise be banned.**



Based on the existing research, discussions and regulatory attempts, ECNL recommends three regulatory actions following the HRIA:

**a. Regulation prior to deployment.** To prevent adverse impacts on human rights, all AI systems that do not fall into the ban or moratorium categories are subject to further appropriate levels of regulation prior to deployment, including mandatory tier-based risks assessments and human rights due diligence.

**b. Moratorium.** AI systems that have severe impacts on actual and potential human rights should be temporarily banned. At the end of the moratorium period, a new HRIA should be conducted, and further actions re-evaluated. In order for the moratorium to be legitimate, a ban must be a genuine potential outcome of the second HRIA, if not the default one.

**c. Ban.** AI systems that are inherently incompatible with fundamental rights, rule of law, and democracy should be permanently banned. This includes facial recognition and remote biometric recognition systems whose indiscriminate or arbitrarily targeted use can lead to unlawful mass surveillance and profiling, and risk assessment tools for criminal justice, among others.<sup>1</sup>

<sup>1</sup> For a broader list, see the joint civil society letter on AI red lines at <https://edri.org/our-work/civil-society-call-for-ai-red-lines-in-the-european-unions-artificial-intelligence-proposal/>.

Indeed, it is only at the level of ‘further regulation’ that we believe a risk-based approach could be useful. The risk analysis would be conducted to determine the risk level of the AI system. Each risk level should then determine what additional safeguards must be taken to prevent or mitigate human rights risks before deploying the AI system. When these additional safeguards are defined, it is up to the developer and/or end-user of the AI system to implement them. Only once the safeguards are implemented through targeted measures can the AI system be deployed. Of note, this includes monitoring and maintenance responsibilities at various stages of the AI’s life cycle.

In parallel, mandatory human rights due diligence laws should be enacted, including binding Directors’ duties.<sup>[6][8] [9]</sup> As consistent with the UN Guiding Principles on Business and Human rights (UNGPs), all AI developers and end-users should conduct human rights due diligence. Company-specific human rights impact assessments can play an important role in carrying out this process. <sup>[10][11]</sup> Depending on the risk level that the AI system is assigned to, and the specific needs raised during the human rights due diligence process, AI developers and end-users should take appropriate measures to mitigate and prevent risk.

### **III. *A tiered model is more appropriate than a “low vs high” approach for determining the human rights risks of AI.***

The EU should reject an over-simplified low-high risk structure and consider a tier-based human rights risk approach. **We propose a framework and criteria for determining the level of risk to human rights based on a tiered approach.**<sup>[12]</sup>(CAHAI para. 125<sup>2</sup>)

**We recommend that at least five levels be considered to determine the level of risk, from 1) low to 2) medium low, 3) medium, 4) medium high to 5) high.** The level of risk to human rights should determine specific regulatory and governance measures that should be implemented before deploying the AI system. While merely illustrative and not exhaustive, we propose in section IV key criteria that should be considered when determining the risk level of each AI system. In any case, a tiered approach should be fluid, i.e., remain subject to continuous assessment and potential variations of levels depending on context and circumstances.

**Beyond looking at individual criteria, it’s important to consider the process for conducting the risk assessment.** While this is not the purpose of this analysis, we caution that *who* conducts the assessment and *when* this is done will highly influence the outcomes and legitimacy of the assessment. Future analysis should also carefully consider the following issues: first, how much weight is given to each factor?<sup>3</sup> Second, on what scale/range of responses are criteria evaluated? This requires moving away from binary “yes/no” type responses and offering multiple options on a numerical or qualitative scale.<sup>4</sup> Third, who has

---

<sup>2</sup> “When member States take measures to safeguard the listed principles, rights and requirements in the context of AI, a risk-based approach – complemented with a precautionary approach where needed – is recommended. Such approach acknowledges that not all AI systems pose an equally high level of risk, and that regulatory measures should take this into account.” (CAHAI para 125).

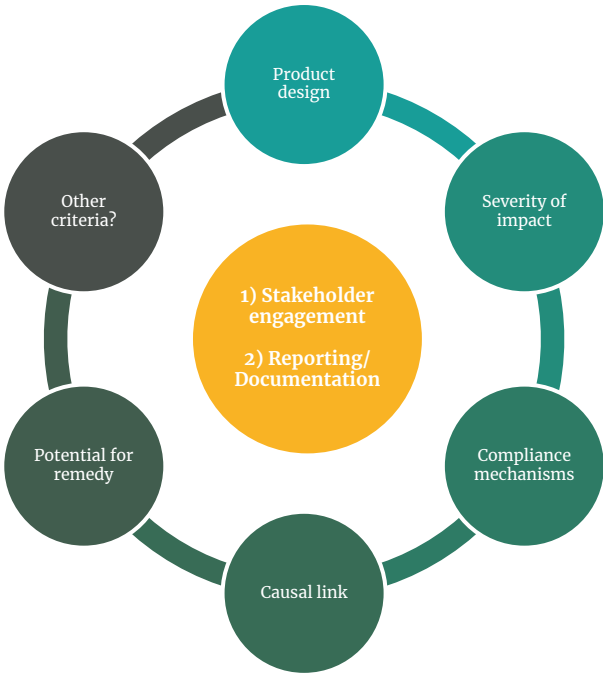
<sup>3</sup> Example: For category A (product design), the motive could be counted as 1x, the intended user as 3x, and the matching threshold as 2x.

<sup>4</sup> Example: For a question such as “would the use of an AI system lead to disparate impact?”, the available options would be “no; sometimes; often; never”.

the authority to evaluate each of the criteria, and who has access to the necessary data to make these evaluations?<sup>5</sup> Fourth, on what scale/range should the final score be determined?<sup>6</sup>

**IV. Risk levels should be based on product design, severity of impact, compliance mechanisms, causal link, and potential for remedy.**

We propose a criteria model for assessing the risk levels. Detailed examples of criteria or factors influencing the level of risk to human rights of an AI system are enclosed in Annex 1.



**1. Product design**

The AI system’s intricate features and design play a significant role in increasing or reducing human rights risks. For example, when data collection is widespread, when there are numerous end-users, or when the potential for misuse is high, the risk of adverse impacts increases. Conversely, where human control and oversight is strong, robust safeguards are put in place to prevent adversarial use of the product, or where the AI system is explainable for a wide audience, the risk of adverse impacts decreases. As many of the harms depend on the use of the AI system, the product developer can incorporate features to prevent misuse and abuse, thus lowering the system’s risk level.

---

5 Examples: Who determines the number of potentially impacted people? Who has access to match thresholds or datasets used for training?

6 Example: A 5-level risk base approach could consist of “low; medium-low; medium; medium-high; and high” levels.

## 2. Severity of impact on human rights, rule of law and democracy

Taking a rights-based approach implies that impact is considered on several domains: human rights, rule of law and democracy. As mentioned in the introduction, human rights impact assessments play a critical role in the overarching architecture of AI governance. This category doesn't look at *what* rights are impacted, but at the *severity* of the impact. Overall, the level of risk is proportionate to factors such as the range of affected people, the stakes of a decision, or the severity of rule of law or democratic concern. Where multiple rights are simultaneously at risk, a necessity and proportionality test must be conducted for all.

## 3. Causal link of harm and attribution

Access to remedy can only be ensured if the causal link of harm can be properly determined, and the cause of harm can be attributed to specific actors. AI developers or deployers also have different remediation responsibilities depending on whether they cause, contribute to, or are linked to adverse human rights impacts by their operations, products, services, or business relationships.<sup>[13]</sup>(UNGP 22)<sup>7</sup> Yet the more the AI developer or deployer is removed from direct impact, the harder it is to prevent, mitigate and remedy harm. This in turn influences the level of risk that an AI system would have on human rights. Other critical factors that elevate the risk of adverse impacts include the potential for misuse of an AI system, or weak internal mechanisms to assign individual accountability.

## 4. Compliance mechanisms

Core to the human rights based-approach is process. For AI systems, the more expansive and rigorous compliance mechanisms are, the lower the risk of adverse impacts is. Compliance mechanisms are internal processes and governance structures that enable AI developers and deployers to detect and prevent human rights risks.<sup>[14]</sup><sup>8</sup> Both AI developers and deployers should establish specific measures to reduce risk; the level and quality of implementation of such measures thus influence the risk level. Governance mechanisms include internal policies, risk management strategies, human rights due diligence programs, internal audits, testing and evaluation, and monitoring, among others. While the structure and scale of these mechanisms will depend on the size of the company, all companies *regardless of their size* should establish appropriate internal mechanisms.<sup>9</sup>

## 5. Potential for remedy

---

<sup>7</sup> UNGP 22: "Where business enterprises identify that they have caused or contributed to adverse impacts, they should provide for or cooperate in their remediation through legitimate processes.

UNGP 22 Commentary: "Where adverse impacts have occurred that the business enterprise has not caused or contributed to, but which are directly linked to its operations, products or services by a business relationship, the responsibility to respect human rights does not require that the enterprise itself provide for remediation, though it may take a role in doing so."

<sup>8</sup> For CAHAI's understanding of compliance mechanisms, see CAHAI para. 157.

<sup>9</sup> As noted by Human Rights Watch, "Concerns about whether smaller businesses would find the administrative costs of compliance with due diligence requirements onerous can be effectively addressed in the legislation. The UNGPs and the OECD MNE Guidelines state that though all businesses have a responsibility to respect and protect rights, their policies and practices should be commensurate with the size and nature of the human rights and environmental risks associated to their operations. A small, low-risk company would take less burdensome steps than a larger company or one with activities that are more likely to lead to human rights abuses and environmental destruction." See "Q & A on the Proposal for an EU Legislation on Mandatory Human Rights and Environmental Due Diligence", November 9 2020, < <https://www.hrw.org/news/2020/11/09/q-proposal-eu-legislation-mandatory-human-rights-and-environmental-due-diligence#One>>.

Liability and remedy are key components of ensuring human rights-based AI governance. They also have a significant impact on the overall risk level of an AI system. For example, if harm caused by an AI system is easily identifiable and reversible, the risk level will decrease. When AI systems are deployed in the context of effective and accessible grievance mechanisms – thus ensuring that people can get remedy for the harm caused by the AI system – the risk to human rights is also lower. When assessing the potential for remedy, it's important to analyze local laws and policies: those that establish a framework for person liability and put the burden of proof on the AI developer or end-user (as opposed to the person who is the target of the AI system) enable a more rights-protective space for deploying the AI system.

## ***V. Stakeholder engagement and transparency/accountability are to be evaluated for each of the criteria.***

**1. The risk assessment should evaluate stakeholder engagement.**<sup>[15]</sup> When assessing each of the criteria outlined above (see also Annex 1), it's important to consider whether 1) any individuals or groups from marginalized communities are disproportionately impacted and if so, how?<sup>[16]</sup>(UNGP 1) and 2) did these individuals or groups participate in the decision-making or assessment process?<sup>[14]</sup>(CAHAI para. 115)

**2. The risk assessment should evaluate the overall transparency of an AI system as a means towards enabling algorithmic accountability.**<sup>[17]</sup><sup>10</sup> For each of the criteria, the assessor should consider 1) how this information is documented and reported on internally (often referred to as documentation practices); and 2) how this information is shared externally. Meaningful transparency requires that all information be communicated to affected stakeholders in a way that is accessible and continuous, so that stakeholders can adequately evaluate any mitigation measure taken.<sup>[18]</sup>(UNGP 21)

## **ANNEX 1**

### ***Examples of criteria or factors influencing the level of risk to human rights of an AI system***

#### **A) Product design**

- Motive(s) for designing or using an AI system;
- Intended end-user(s) (e.g. consumer, public agency, corporation, etc.);
- Data collection processes (including issues around informed consent and opt-in vs opt-out default modes);<sup>[19]</sup>(CAHAI para. 102)

---

<sup>10</sup> “One simple and accessible approach to increasing transparency in ML lifecycles is through an improvement in both internal and external documentation norms. For an increasingly concerned public or auditing organization, externally distributed and thorough documentation on ML system components is essential to earning and maintaining trust, and minimizing the misuse of these systems. External documentation and reporting requirements can also provide teams with an argument for more resourcing to implement transparency processes. Internal documentation is also vital, serving to improve communication between collaborating teams. Internal documents also help build employee trust by outlining the nature of an individual or team’s contribution to an overall system, giving opportunity for ethical objections and a more meaningful understanding of the extent of their personal participation in the creation of an end product. Beyond the artifact itself, however, the process of documentation itself is inherently valuable towards the goal of transparency, prompting critical thinking about the ethical implications at every step in the ML lifecycle and encouraging the steps required to understand and report a complete picture on system capabilities, limitations, and risks.”

- System capabilities and functionalities (e.g. image and object recognition; text and speech analysis; risk assessment; content generation; process optimization and workflow automation)[19]( CAHAI para. 102)
- Match thresholds determination (false positives and false negatives);
- Level of human control and oversight (including the possibility to ‘disable the AI system or change its functionality’[20] and having a ‘fallback plan’[21] (CAHAI para. 104);
- Level of automation (from AI systems that assist the decision-maker to those that make discretion-based judgments[22]);
- Efficacy and accuracy of intended use;
- End-user ability to customize the product in a way that increases risk of misuse or harm;[23]
- Protection against vulnerabilities that can be exploited for adversarial use;<sup>i</sup>
- Reliability and reproducibility;[21]
- Data quality[24](CAHAI para. 106) (including presence of personal or sensitive data, representativeness of data sets, quality and integrity, accuracy of data sets, control/ownership of data, gender-disaggregated analysis, structured vs unstructured data, data labeling etc.).
- Explainability, including for those that are subject to the system;
- Reporting on the details and functions of the AI system (functioning, optimization functioning, underlying logic, type of data used).

#### **B) Severity of impact on human rights, rule of law and democracy**

- Target(s) of the impact (e.g. people, environment, rule of law, democracy);
- Number of potentially affected people;
- Type of demographic groups affected (women and non-binary persons, racial or ethnic minorities, LGBTQ people, disabled persons, children, migrants, indigenous communities, lower socio-economic groups, etc.)
- Intensity of the impact (even if the number of potentially affected people is low);
- Accessibility of any positive or beneficial impacts to under-represented groups;
- Presence of elevated risk due to area/industry of high public scrutiny or litigation;[25][26]
- Presence of elevated risk due to geographic/geopolitical context;[27](UNGP 12)
- Sector where AI system is deployed (e.g. elevated risk if related to health, economic rights, social welfare, access and security);[28]
- Stakes of a decision (e.g. risk assessment tools for criminal justice are particularly high-stakes decisions, because they can lead to detention and thus impact people’s right to liberty.);

- Outcome of a necessity and proportionality test (e.g. if an AI system impacts the right to privacy for a public benefit cause, such as health concerns, to what extent is such privacy impact proportionate and necessary in relation to public health benefits?)

### C) Causal link of harm and attribution

- Causal link (causing, contributing to, or being linked to adverse human rights impacts);
- Internal mechanisms determine risk of harm at different stages of an AI's life-cycle (e.g. stages include design, development, promotion, deployment, sales, licensing and use);
- Risk of disparate impact (i.e. the risk that an AI system would reinforce existing discrimination);
- Evolution of the causality relationship between the deployment of an AI system and harm, in the short, medium and long-term. (As the AI system evolves over time (due to retraining, new data collection, engineering tweaks, debugging, etc.), how do its impacts evolve accordingly? This is especially important for reinforcement learning contexts and dynamic learning systems.[29] (CAHAI para. 158);
- Likelihood of misuse,[25] due to 1) user interests, motivations and incentives to use or misuse tech; 2) user's technological know-how and capability (including access to computing power); and 3) local policy and laws that would make the use case more likely);
- Internal mechanisms to assign individual accountability for various steps in the design, development, maintenance, and improvement of the system, as well as designating a "principal actor"[30, p. 3];

### D) Compliance mechanisms

- Existence of a clear legal or policy basis for the use of the AI system;
- Internal governance structures[31] (and risk management strategies[32]) of developers to evaluate the human rights impacts of an AI system, as part of broader human rights due diligence;[33][24, Guiding Principles 17; 18(b)]
- Internal governance structures[31] (and risk management strategies[32]) of developers to evaluate supply chains[34](UNGP 13)<sup>11</sup> (e.g. responsible sourcing and data enrichment services, "ghost work",[35] content moderation,[36] etc.);
- Paper trail and documentation of datasets[37] and models[38][39];

---

11 UNGP 13: " The responsibility to respect human rights requires that business enterprises: (a) Avoid causing or contributing to adverse human rights impacts through their own activities, and address such impacts when they occur; (b) Seek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts." Commentary to UNGP 13: " or the purpose of these Guiding Principles a business enterprise's "activities" are understood to include both actions and omissions; and its "business relationships" are understood to include relationships with business partners, entities in its value chain, and any other non-State or State entity directly linked to its business operations, products or services."



- Internal audits[26] [40] (CAHAI para. 169(3)) and a broad and detailed audit trail (including AI-generated decisions; changes/updates to the system, decision-maker; instances where the system was overridden; person(s) having access to the audit trail,[28] etc.);
- Testing and evaluation processes (e.g. sand box[41]);
- Certification and quality labelling;[42](CAHAI para. 169(2))
- Consultations between internal stakeholders (business development, legal, program management, engineering, policy and procurement teams, among others) and external stakeholders (civil society, academia, other companies, etc.);
- Existence and terms of a sunset clause;
- Systematic and continuous review of compliance mechanisms, including tracking and measuring progress,[31, UNGP 20] and periodic human rights impact assessments.[43]<sup>12</sup>

#### **E) Potential for remedy**

- Likelihood that the adverse impacts resulting from an AI-driven decision or prediction are reversible;[28][26]
- Presence of an effective and accessible internal operational grievance mechanism;[26]
- Capacity to identify algorithmic-related or caused harm;[26]
- Access to algorithmic processes and decisions to provide evidence of harm;[44](CAHAI para. 44; 108)<sup>13</sup>
- Local laws and policies that clearly regulate personal liability, put the burden of proof on the AI developer, and promote access to remedy.

---

#### **BIBLIOGRAPHY**

[1] European Commission, “White Paper On Artificial Intelligence – A European approach to excellence and trust,” COM(2020) 65 final, 02 2020. [Online]. Available: [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf).

[2] “The EU should regulate AI on the basis of rights, not risks,” *Access Now*, Feb. 17, 2021. <https://www.accessnow.org/eu-regulation-ai-risk-based-approach/> (accessed Mar. 17, 2021).

---

<sup>12</sup> Commentary to UNGP 18: Because human rights situations are dynamic, assessments of human rights impacts should be undertaken at regular intervals: prior to a new activity or relationship; prior to major decisions or changes in the operation (e.g. market entry, product launch, policy change, or wider changes to the business); in response to or anticipation of changes in the operating environment (e.g. rising social tensions); and periodically throughout the life of an activity or relationship.

<sup>13</sup> CAHAI para 44. See also “types of decisions or situations subject to automated processing, criteria relevant to a decision, information on the data used, a description of the method of the data collection. A description of the system’s potential legal or other effects should be accessible for review/audit by independent bodies with necessary competences.” (CAHAI para 108).

- 
- [3] European Union Agency for Fundamental Rights, “Getting the future right – Artificial intelligence and fundamental rights | European Union Agency for Fundamental Rights,” Report, 12 2020. Accessed: Mar. 08, 2021. [Online]. Available: <https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights>.
- [4] OECD, “XXXX (Fran’s screenshot).”
- [5] European Center for Not-for-Profit Law, “ECNL Submission for the Consultation on EC White Paper on Ai,” 06 2020. <https://ecnl.org/news/ecnl-urges-european-commission-prioritise-civic-freedoms-ai-white-paper>.
- [6] Council of Europe, Ad Hoc Committee on Artificial Intelligence, “The feasibility study on AI legal framework adopted by the CAHAI,” *Artificial Intelligence*, 12 2020. [https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset\\_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai](https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai) (accessed Mar. 08, 2021).
- [7] N. Petit and J. D. Cooman, “Models of Law and Regulation for AI,” p. 31.
- [8] United Nations Human Rights Council, Office of the High Commissioner for Human Rights, “United Nations Guiding Principles on Business and Human Rights,” international law, 2011. [Online]. Available: [https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf).
- [9] Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, “OHCHR | 2018 thematic report to the Human Rights Council on content regulation.” Accessed: Mar. 09, 2021. [Online]. Available: <https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx>.
- [10] European Parliamentary Research Service, Ionel Zamfir, “Towards a mandatory EU system of due diligence for supply chains,” Oct. 2020.
- [11] United Nations Human Rights Council, Office of the High Commissioner for Human Rights, “United Nations Guiding Principles on Business and Human Rights,” international law, 2011. [Online]. Available: [https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf).
- [12] Council of Europe, Ad Hoc Committee on Artificial Intelligence, “The feasibility study on AI legal framework adopted by the CAHAI,” *Artificial Intelligence*, 12 2020. [https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset\\_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai](https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai) (accessed Mar. 08, 2021).
- [13] United Nations Human Rights Council, Office of the High Commissioner for Human Rights, “United Nations Guiding Principles on Business and Human Rights,” international law, 2011. [Online]. Available: [https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf).
- [14] Council of Europe, Ad Hoc Committee on Artificial Intelligence, “The feasibility study on AI legal framework adopted by the CAHAI,” *Artificial Intelligence*, 12 2020. [https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset\\_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai](https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai) (accessed Mar. 08, 2021).
- [15] “Human rights impact assessment of digital activities | The Danish Institute for Human Rights.” <https://www.humanrights.dk/publications/human-rights-impact-assessment-digital-activities> (accessed Mar. 09, 2021).
- [16] United Nations Human Rights Council, Office of the High Commissioner for Human Rights, “United Nations Guiding Principles on Business and Human Rights,” international law, 2011. [Online]. Available: [https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf).
- [17] Partnership on AI, “Annotation and Benchmarking on Understanding and Transparency of Machine learning Lifecycles (ABOUT ML), v0 draft,” 2020. [Online]. Available: <https://www.partnershiponai.org/wp-content/uploads/2019/07/ABOUT-ML-v0-Draft-Final.pdf>.

- 
- [18] United Nations Human Rights Council, Office of the High Commissioner for Human Rights, “United Nations Guiding Principles on Business and Human Rights,” international law, 2011. [Online]. Available: [https://www.ohchr.org/documents/publications/guidingprinciplesbusinessshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinessshr_en.pdf).
- [19] Council of Europe, Ad Hoc Committee on Artificial Intelligence, “The feasibility study on AI legal framework adopted by the CAHAI,” *Artificial Intelligence*, 12 2020. [https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset\\_publisher/csARLoSVrbAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai](https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrbAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai) (accessed Mar. 08, 2021).
- [20] Council of Europe, Ad Hoc Committee on Artificial Intelligence, “The feasibility study on AI legal framework adopted by the CAHAI,” *Artificial Intelligence*, 12 2020. [https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset\\_publisher/csARLoSVrbAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai](https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrbAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai) (accessed Mar. 08, 2021).
- [21] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, “From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices,” *arXiv:1905.06876 [cs]*, Sep. 2019, Accessed: Mar. 09, 2021. [Online]. Available: <http://arxiv.org/abs/1905.06876>.
- [22] T. B. of C. Secretariat, “Algorithmic Impact Assessment (AIA),” *aem*, Jul. 28, 2020. <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html> (accessed Mar. 05, 2021).
- [23] OHCHR B-Tech, “Taking Action to Address Human Rights Risks Related to End-Use,” 2020. [Online]. Available: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/taking-action-address-human-rights-risks.pdf>.
- [24] Council of Europe, Ad Hoc Committee on Artificial Intelligence, “The feasibility study on AI legal framework adopted by the CAHAI,” *Artificial Intelligence*, 12 2020. [https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset\\_publisher/csARLoSVrbAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai](https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrbAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai) (accessed Mar. 08, 2021).
- [25] OHCHR B-Tech, “Identifying and Assessing Human Rights Risks Related to End-Use,” 2020. [Online]. Available: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/identifying-human-rights-risks.pdf>.
- [26] “Artificial Intelligence: A Rights-Based Blueprint for Business | Reports | BSR.” <https://www.bsr.org/en/our-insights/report-view/artificial-intelligence-a-rights-based-blueprint-for-business> (accessed Mar. 09, 2021).
- [27] United Nations Human Rights Council, Office of the High Commissioner for Human Rights, “United Nations Guiding Principles on Business and Human Rights,” international law, 2011. [Online]. Available: [https://www.ohchr.org/documents/publications/guidingprinciplesbusinessshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinessshr_en.pdf).
- [28] T. B. of C. Secretariat, “Algorithmic Impact Assessment (AIA),” *aem*, Jul. 28, 2020. <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html> (accessed Mar. 09, 2021).
- [29] Council of Europe, Ad Hoc Committee on Artificial Intelligence, “The feasibility study on AI legal framework adopted by the CAHAI,” *Artificial Intelligence*, 12 2020. [https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset\\_publisher/csARLoSVrbAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai](https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrbAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai) (accessed Mar. 08, 2021).
- [30] Business for Social Responsibility, “Artificial Intelligence: A Rights-Based Blueprint for Business Paper 3: Implementing Human Rights Due Diligence,” Aug. 2018. [Online]. Available: <https://www.bsr.org/reports/BSR-Artificial-Intelligence-A-Rights-Based-Blueprint-for-Business-Paper-03.pdf>.

- 
- [31] I. D. Raji *et al.*, “Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing,” *arXiv:2001.00973 [cs]*, Jan. 2020, Accessed: Mar. 09, 2021. [Online]. Available: <http://arxiv.org/abs/2001.00973>.
- [32] T. Meek, H. Barham, N. Beltaif, A. Kaadoor, and T. Akhter, “Managing the ethical and risk implications of rapid advances in artificial intelligence: A literature review,” in *2016 Portland International Conference on Management of Engineering and Technology (PICMET)*, Sep. 2016, pp. 682–693, doi: 10.1109/PICMET.2016.7806752.
- [33] J. Metcalf, E. Moss, E. A. Watkins, R. Singh, and M. C. Elish, “Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts,” Sep. 2020, Accessed: Mar. 05, 2021. [Online]. Available: <https://papers.ssrn.com/abstract=3736261>.
- [34] United Nations Human Rights Council, Office of the High Commissioner for Human Rights, “United Nations Guiding Principles on Business and Human Rights,” international law, 2011. [Online]. Available: [https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf).
- [35] Mary Gray and Siddharth Suri, “Ghost Work. How to Stop Silicon Valley from Buidling a New Global Underclass,” *Ghost Work*. <https://ghostwork.info/> (accessed Mar. 09, 2021).
- [36] Sarah T. Roberts, *Behind the Screen. Content Moderation in the Shadows of Social Media*, Yale University Press. 2019.
- [37] T. Gebru *et al.*, “Datasheets for Datasets,” *arXiv:1803.09010 [cs]*, Mar. 2020, Accessed: Mar. 09, 2021. [Online]. Available: <http://arxiv.org/abs/1803.09010>.
- [38] “Google Cloud Model Cards.” <https://modelcards.withgoogle.com/about> (accessed Mar. 09, 2021).
- [39] “IBM FactSheets.” <https://www.ibm.com/blogs/research/2020/07/aifactsheets/> (accessed Mar. 09, 2021).
- [40] Council of Europe, Ad Hoc Committee on Artificial Intelligence, “The feasibility study on AI legal framework adopted by the CAHAI,” *Artificial Intelligence*, 12 2020. [https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset\\_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai](https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai) (accessed Mar. 08, 2021).
- [41] Salesforce, “Sandboxes: Staging Environments for Customizing and Testing.” [https://help.salesforce.com/articleView?id=sf.deploy\\_sandboxes\\_parent.htm&type=5](https://help.salesforce.com/articleView?id=sf.deploy_sandboxes_parent.htm&type=5) (accessed Mar. 09, 2021).
- [42] Council of Europe, Ad Hoc Committee on Artificial Intelligence, “The feasibility study on AI legal framework adopted by the CAHAI,” *Artificial Intelligence*, 12 2020. [https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset\\_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai](https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai) (accessed Mar. 08, 2021).
- [43] United Nations Human Rights Council, Office of the High Commissioner for Human Rights, “United Nations Guiding Principles on Business and Human Rights,” international law, 2011. [Online]. Available: [https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf).
- [44] Council of Europe, Ad Hoc Committee on Artificial Intelligence, “The feasibility study on AI legal framework adopted by the CAHAI,” *Artificial Intelligence*, 12 2020. [https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset\\_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai](https://www.coe.int/en/web/artificial-intelligence/newsroom/-/asset_publisher/csARLoSVrBAH/content/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai) (accessed Mar. 08, 2021).