DATA&
SOCIETY

European Center for
Not-for-Profit Law

Mandating Human Rights Impacts Assessments in the AI Act

Introduction

As organizations that have published research on accountability and impact assessments for artificial intelligence (AI) systems, we welcome the European Union's (EU) efforts to develop a legally binding framework on AI based on the EU's standards on fundamental rights.

However, the European Commission has notably left out the obligation of providers and users to conduct human rights impact assessments (HRIAs) from its proposed regulation, an important instrument for measuring and mitigating algorithmic harm, and ensuring accountability. Moving forward, it is essential for the EU to mandate the use of HRIAs as a mechanism for evaluating the impacts of AI systems by making HRIAs part of mandatory human rights due diligence for providers, and by mandating a separate HRIA for public sector users.

This is an essential step for achieving stated EU goals for the development and deployment of trustworthy AI. It is also central for understanding and determining the levels of risk of AI systems—without understanding the impact of the AI system on human rights, there is little evidence and knowledge for detecting the risk level. Without a commitment to regulatory approaches that center human rights, algorithmic accountability, transparency, and the protection and uplift of marginalized and vulnerable groups, AI will continue to benefit the few while threatening economic opportunity and societal well-being of many. Moving forward, it is essential for the EU to mandate the use of HRIAs as a mechanism for evaluating the impacts of AI systems by mandating a separate HRIA for users, with elevated responsibility for public sector users.

Our recommendations for mandating the use of HRIAs build on a history of impact assessments used in a range of domains such as finance, environment, data protection, and health to consider the benefits and impacts of a business practice, technology or policy. In addition, they are in line with the upcoming mandatory EU human rights and environmental due diligence framework, that includes HRIA and is broadly supported by investors.[1] Moreover, a YouGov poll reveals over 80% of EU citizens in the countries polled support

EU laws to hold companies accountable for harms to people and the environment.[2] Similarly, in a survey conducted by the Council of Europe Ad Hoc Committee on Artificial Intelligence (CAHAI), 81% of respondents identified HRIAs as an effective regulatory mechanism to protect human rights.[3]

- Impact assessments broadly serve the following objectives:
- Providing an ex ante or ex post assessment of the potential or actual impacts of a technology, policy, or business practice.
- Providing a reflexive exercise for developers of a policy or technology to question what intended outcomes they hope to achieve, and what mitigative measures they may need to put in place to address potentially harmful outcomes.
- Documenting impacts so that they can be shared with a range of stakeholders.
- Providing a mechanism for developers and policymakers to engage with affected communities.

HRIAs are a more widely used form of impact assessment that can be described as "a process for identifying, understanding, assessing and addressing the adverse effects of a [project, product, services, or activities] on the human rights enjoyment of impacted rightsholders."[4] HRIAs are grounded in the United Nations Guiding Principles for Business and Human Rights (UNGPs), a non-binding framework that was unanimously endorsed by the United Nations Human Rights Council in 2011.[5] A HRIA process can create an accountability relationship by asking actors to produce an account of how their systems may impact human rights and by empowering appropriate authorities and the public at large to act as a forum in evaluating and mitigating those impacts.[6]

Although HRIAs are an increasingly popular accountability mechanism, there's a risk that HRIAs could become performative or ineffective. To mitigate this risk, HRIAs must be designed and implemented in a way that is meaningful, with a clear governance framework that elevates civil society and affected communities' concerns instead of instrumentalizing them.[7]

Governance and methodology recommendations for establishing an HRIA process

In developing an HRIA framework, the EU must ensure it addresses **ten constitutive elements for ensuring HRIAs are meaningfully accountable to those they seek to serve.** This framework is drawn from an analysis of existing forms of impact assessment across many domains.[8] These components are crucial for developing a framework that addresses the adverse human rights impacts of AI systems.

Impact assessment processes rely on a **source of legitimacy**, which may be regulatory or normative, that establishes the validity and of the process. Accountability also hinges on who the **actors** within the process are, and what the **forum is** to which they are accountable. Accountability is also inflected by the **catalyzing event** that triggers the start of an assessment process, the circumstances under which a project might be exempted from requiring an impact assessment, and the **time frame** in which the assessment is conducted. Accountability is further shaped by **public consultation** and the degree to which

public access is given for the final assessment. In any impact assessment practices, the **methods** through which impacts are assessed, and those who are empowered as the **assessors** of impacts emerge through the development of expert practices for discerning and documenting the **impacts** of the assessed systems. Ultimately, these practices are aimed at rendering **harms** measurable as impacts, and providing the basis to provide r**edress** for those harms, either through design changes that avoid or minimize harms, or through mitigation techniques that repair unavoidable harms.

Building on these components, the EU should take the following important steps to clarify and establish the importance of measuring and evaluating the human rights impacts of AI systems.

1. **Require developers and deployers to conduct HRIAs.** The AI Act should mandate HRIAs. This is particularly important when AI systems are operating in the public sector, but should be generally applicable. HRIAs should be conducted at all stages of the AI lifecycle with different scope, starting with the ideation stage and running through post-deployment, and can include a process for reviewing the impacts in an iterative and ongoing way. Appropriate resources and capacity must be allocated for this purpose to ensure adequate classification and assessment.

2. **Determine the criteria for assessing impacts to human rights.** The AI Act should clearly identify which events or circumstances would require that an HRIA be undertaken. HRIAs should prioritize harm reduction and the adverse human rights impacts on marginalized and vulnerable groups, taking a holistic approach and assessing the impacts of AI systems on a wide range of human rights, including collective rights, economic, social and cultural rights, and environmental rights.[9] The areas of scrutiny should be assessed on a case-by-case basis, mindful of specific contexts at play, including geographical location, language, demographic group, socio-political factors, and temporal considerations.

3. **Ensure public access.** The AI Act should require that the output of HRIA processes is made available to the public by depositing it into public registers, providing public notice through press releases, social media posts, and other available and accessible sources, and depositing physical copies at libraries and other publicly accessible archives. This is an important enabling step for public engagement, consultation, and appeal to redress harmful deployments of AI. The regulation must also develop, where necessary, legal mechanisms that protect private companies' trade secrets and intellectual property, while still providing access to assessors.

4. **Establish an oversight mechanism.** The AI Act should mandate external, iterative, and ongoing review and oversight of privately-conducted HRIAs, and determine which public authority is responsible for oversight. All information related to the oversight body and their assessments should be made publicly available and accessible. These mechanisms will ensure that developers of AI systems are not left to evaluate their own impacts.

5. **Develop methods for participatory inclusion, public consultation, and appeal.** These steps are essential for the meaningful incorporation of external stakeholders, particularly affected communities, in HRIA processes. **We therefore encourage the EU to work with a wide range of civic groups directly to develop methods for**

effective HRIA engagement. Racialized persons, women and gender non-binary persons, LGBTQ+, disabled persons, persons of lower socio-economic status, and representatives from affected and marginalized communities must be included in formulating priorities, definitions, and outcomes of the HRIA, and ultimately, the decision whether and in which ways to deploy AI.

6. Establish an HRIA research agenda. Without additional investment in a research agenda to study the effectiveness of HRIA methodologies, particularly as they are applied to AI systems, we risk fragmentation and confusion about what constitutes a high-quality impact assessment. Ongoing empirical, sociotechnical research is essential to develop standards and methods for using impact assessment methodologies on complex AI systems. This can be complemented by a regulatory sandbox and pilot programs that encourage continuous evaluation of HRIA methodologies.

7. Integrate HRIAs with other accountability mechanisms. This includes other forms of impact assessments (e.g. data protection impact assessments, human rights and environmental due diligence and conformity assessments), algorithmic auditing, transparency registers, citizen review boards, and procurement requirements.[10] In taking this holistic approach, the HRIA framework in the AI Act would be centering potential and actual harm to individuals, communities, society and the environment in the HRIA analysis.

Conclusion

We are at a turning point for the future of AI accountability. Numerous jurisdictions have proposed legislation that would implement algorithmic impact assessments as a tool for bringing accountability to the algorithmic systems increasingly used in everyday life. The European Parliament has also initiated the process for an EU mandatory human rights due diligence framework.[11]

Moving forward, it is essential for the EU to mandate the use of HRIAs as a mechanism for evaluating the impacts of AI systems.

Through the upcoming regulatory process, our organizations hope to support the EU in developing HRIA requirements, as well as deepening engagement across sectors on impact assessments as a mechanism for algorithmic governance and accountability. Establishing a human rights-based approach to impact assessments and algorithmic accountability would be a significant step forward in securing public accountability for the impacts AI technology has on society.

About the authors

Data & Society is an independent research institute focusing on the social implications of data-centric technologies & automation. We study the social implications of data and automation, producing original research to ground informed, evidence-based public debate about emerging technology.

The **European Center for Not-for-Profit Law (ECNL)** is a non-governmental organisation working to empower civil society. We aim to create legal and policy environments that enable individuals, movements and organisations to exercise and protect their civic freedoms and to put into action transformational ideas that address national and global challenges.

Endnotes

1. See https://www.europarl.europa.eu/doceo/document/TA-9-2021-0073_EN.html and https://investorsforhumanrights.org/investor-statement-support-mandated-human-rights-and-environmental-due-diligence-european-union

2. https://www.business-humanrights.org/en/latest-news/yougov-poll-reveals-over-80-of-eu-citizens-support-eu-laws-to-hold-companies-accountable-for-harms-to-people-environment/

3. https://rm.coe.int/cahai-2021-07-analysis-msc-23-06-21-2749-8656-4611-v-1/1680a2f228

4. Danish Institute for Human Rights. 2020. Guidance on Human Rights Impact Assessment of Digital Activities. Accessible at https://www.humanrights.dk/sites/humanrights.dk/files/media/document/A%20HRIA%20of%20Digital%20Activities%20-%20Introduction_ENG_accessible.pdf

5. Latest statements from UN bodies call for a UN-level legally binding instrument on mandatory and meaningful due diligence, which should be complementary to the EU instrument currently under development. https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=27672&LangID=E

6. Bovens, Mark. (2010). Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism. West European Politics - WEST EUR POLIT. 33. 946-967. 10.1080/01402382.2010.486119.

7. Sloane, Mona, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. (2020). "Participation Is Not a Design Fix for Machine Learning." In *Proceedings of the 37th International Conference on Machine Learning*, 7. Vienna, Austria. https://arxiv.org/ftp/arxiv/papers/2007/2007.02423.pdf.

8. Moss, et al. 2021

9. For example, The Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 108). https://rm.coe.int/1680078b37.

10. For a full discussion of other public sector algorithmic accountability methods, see Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). Algorithmic Accountability for the Public Sector. Available at: https://www.opengovpartnership.org/documents/ algorithmic-accountability-public-sector/

11. https://www.europarl.europa.eu/news/en/press-room/20210304IPR99216/meps-companies-must-no-longer-cause-harm-to-people-and-planet-with-impunity