Human rights in AI impact assessment: insights from pilot implementations, normative frameworks, and a proposal

Vanja Skoric, Giovanni Sileno, and Sennay Ghebreab
Socially Intelligent Artificial Systems (SIAS), Informatics Institute,
University of Amsterdam, Amsterdam, the Netherlands
v.skoric@uva.nls

Abstract. When assessing the impact of AI, it is imperative to consider ethical and human rights implications. Often, AI concerns involve making difficult choices. The crux of AI debate centres on understanding the nature of these trade-offs, as well as how they are perceived and addressed. Resolving these concerns involves interdisciplinary approaches, including political, legal, and technical perspectives. It is essential to inform these perspectives with robust concepts and empirical insights. This paper examines whether (potential) impacts of AI can be meaningfully addressed by leveraging an assessment based on a commonly accepted core of human rights frameworks. We address what governance and practical structure(s) would make such assessment viable by triangulating field research and qualitative desk research. Via field research, we examine lessons and challenges raised in ongoing pilots implementing AI risk assessment and management processes in the Dutch public sector. Here we have identified mismatches between discussions and expectations regarding process and implementation, and acknowledged questions critical to ensure that impact assessment becomes useful and effective, yet manageable. By means of desk research, we analyse emerging regulatory proposals as well as methodologies for AI impact assessment to understand if mandating safeguards for human rights is compatible with, if not sustained by, the current normative and methodical frame. Building upon these insights, we examine what meaningful, rights-based, context-sensitive and inclusive impact assessment for AI could mean in practice, sketching a proposal for governance processes supporting it.

Keywords: AI governance · Impact Assessment · Responsible AI · Public Sector AI · Regulatory Framework

1. Introduction

The continuous development and the increasingly widespread use of AI in all domains of human activity and experience introduce novel risks on individuals and society. The cognitive, societal, and geopolitical stakes of AI need not to be belittled, for the capacity of introducing strong path dependencies, and for its priming effects. This situation demands an adaptation, if not a reinvention, of the normative and societal settings in place, to adequately keep up with (and to some extent shape) AI deployment and impact. For these reasons, one of the key elements of such adaptation will be to embed forward-looking efforts to identify potential AI benefits and risks in advance, as well as respond with agility to emerging or approaching failure conditions. Moreover, one of the significant issues in the field of AI is the growing lack of trust in *techno-solutionism*, as it often fails to account for the complexities of human decision-making, contexts and social implications. The problem is not new: how do we adapt our systems to take into consideration societal and technological changes?

AI ethics discussions and principles have occupied a most prominent position within discourses in the public and private sectors[1] as a mean for addressing some of the mentioned challenges. However, the focus seems to be shifting to the development of appropriate strategies to operationalize values and principles into inclusive practises at the very heart of the AI design. In a rare empirical study, Stahl et al. observed that the aspect not yet well understood by AI developers is how the discourse on AI ethics translates into the practice of organisations that use these technologies.[2]

In recent years, a variety of stakeholders have increasingly voiced the need for more robust frameworks for responsible AI.[3] There is a growing recognition[4] that the international human rights, including its normative and case law guidance, could be the appropriate mechanism for its consistent framing and operationalization. For instance, main tenets of human rights comprise equitable treatment and respect for every individual's worth, safeguarding their civil, political, and social freedoms, recognizing their inherent humanity and fostering participation in the life of the community. Given their broad cross-cultural scope and acceptance, human rights may facilitate harmonization in addressing responsible AI challenges.[5] The application of core human

rights treaties creates global obligations for states to respect, protect them, and they can be seen as extra-functional requirements for any technology.[6] Yet, how these requirements should unfold in practical terms is not straightforward. Moreover, creating standardized practice can easily ignore contextual factors that may vary across regions or cultures.

First, at computational level, there is a critical translation gap between the societal and value interpretations of, among others, fairness, dignity, proportionality, unbiased, transparency, and the contemporary technological standards aiming to map these concepts into operations. Second, at governance and process level, it is not clear how AI should be governed, by whom, and by which instruments. Theoretically, internalising human rights values within AI might be achieved by identifying connections between the technical features of AI systems and their implications in terms of human rights impact. A socio-technical governance-oriented approach to this challenge would allow a steady reidentification and reassessment of organisational and technological challenges in capturing the impact and responding to it; but how to ensure a practical yet inclusive implementation? Third, at cultural level, there is no straight line for incorporating human rights frameworks into AI design or indeed a guarantee that it will be feasible or effective. Context is often messy, which calls for incorporating multiple perspectives from diverse communities affected by AI technologies and a broader framework that engages beyond normative level to ensure a more comprehensive coverage of AI impact.

A prototypical instrument for consideration against this backdrop is the human-rights based impact assessment (HRIA)[7], used in various sectors to anticipate, prevent, and mitigate harms, to improve the quality of products and services. In its general definition, HRIA is 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 rights holders.[8] As an institutional construct, HRIAs are grounded in the United Nations Guiding Principles for Business and Human Rights (UNGPs)[9], a non-binding framework unanimously endorsed by the United Nations Human Rights Council. It enjoys wide support from states, businesses, and other stakeholders, and is increasingly being reflected in relevant policies and guidelines, as well as in national, regional and international initiatives. The UNGP's core concept of human rights due

diligence refers to an ongoing process to identify, prevent, and mitigate the risk of adverse human rights impacts or actual harms that an entity may cause or contribute to through its own activities, or which may be directly linked to its operations, products, or services. As advised by the UN Business and Human Rights Guiding Principles [6], such assessment is to be conducted at appropriate intervals, aligned with critical stages of the technology lifecycle and as part of ongoing risk management processes and due diligence. Although there have been a number of emerging proposals on using impact assessment in AI design and development[10], there is no consensus on how such instrument should be operationalized or what exactly their content should include. In addition, they have not been tested or piloted in practice to an extent that would allow for coherent insights. Perhaps more importantly, there is still not a proper reflection on what are the criteria to mandate and govern impact assessments. Moreover, there is a general tendency in emerging impact assessment efforts to identify shared agreed methods and benchmarks: e.g. the correct fairness algorithm (or framework for applying various algorithms depending on context), the acceptable performance or errors depending on the tasks. In contrast, this paper addresses the necessity to identify the discussion target not as a shared method, nor a shared artefact, but as a shared, continuous socio-technical process and governance effort.

Why "human rights" in AI impact assessment?

The contemporary human rights framework is centred on values and principles such as fairness, dignity, proportionality, necessity. These are salient issues in the design of AI systems. Treating everyone fairly and with equity, without discrimination, as well as ensuring dignity of the human person constitutes the basis of all rights. Furthermore, the proportionality and necessity ensure limitations on rights must be the least intrusive measure which might achieve the desired result and be proportionate to the protected interest.[11]

One of the key aspects of any AI impact assessment concerns the metrics of value constructs that are being promoted and/or protected by those developing and ultimately, deploying the AI systems. Broadly accepted international human rights *a fortiori* provide benchmarks for these value constructs, explaining what considerations

are worthy of protection. Human rights essentially extract ethical values into international norms.[12] International human rights law is not complete nor fully adapted to the AI (yet); however, it is well established as a universally agreed framework for the protection of human values and public good in diverse cultures - key international human rights treaties are well established in every continent and a few have near-universal ratification. Moreover, it has been proven over the past several decades as capable of adapting to new and evolving challenges

An additional crucial feature of the human rights framework is its potential to balance competing principles and impacts in a flexible yet predictive manner that is context-sensitive. Starting from human rights standards, including the existing interpretative doctrine and case law on proportionality and the balancing of different rights and competing interests, we could weigh societal or individual consequences and interests of using AI systems, between certain relevant rights or among other collective and societal interests,[13] depending on the specific situation. In many of the existing rights, this balancing act is explicitly written in, by means of the "three part test": considering first if there has been an interference with the right, then if yes - whether the interference was based on a law, and whether it is both necessary and proportionate to the legitimate grounds for restriction listed in the human rights law. A plethora of national, regional courts and international human rights bodies have developed extensive and precise case law on the balancing of interests in respect of each right, creating a well-understood system capable of evolving and addressing novel challenges. This capacity of the international human rights law to develop in parallel with societal progress makes it especially attractive for addressing AI development.[14] Finally, and due to its extensive jurisprudence in different contexts of life, the human rights framework is fairly clear and resilient to cultural specificities. We can list all human rights obligations and requirements transposed into practice from the legally binding instruments that apply in almost every country in the world, with its operational meaning and scope; therefore, there is less chance for the "my ethics against yours" argument. There is also a number of guidance to the countries and businesses concerning the implementation of human rights obligations.[15] Essentially, this presents a vast potential to *translate and adapt* existing practice and balancing tests within

diverse contexts into operationalisation for AI, removing most of the vagueness of value constructs. The question remains: are we able to utilize this potential?

Overview of the paper

This paper builds on this work and examines patterns and knowledge gaps in existing practical and operational strategies. Those strategies are intended to incorporate process and governance requirements into responsible development and use of AI. To this purpose, we conduct field and desk research. Through interviews with experts in public organizations and researchers, we identify gaps and mismatches of discussions and expectations on the process and implementation level. Through desk research, we analyse most recent regional and national regulatory proposals as well as relevant methodologies on AI impact assessment. With those results, we examine potential content of balancing the competing values and interests via the "three part test" borrowed from the human rights doctrine, as well as the criteria for meaningful, rights-based, context-sensitive and inclusive impact assessment in practice, sketching a proposal for governance processes supporting it. The paper is structured as follows: section 2 describes related work, section 3 our research method; section 4 presents the findings of the research, section 5 discusses them; section 6 provides actionable insights and concludes our paper.

2. Related work

Only a few efforts up to date provided discussions on emerging criteria for AI impact assessment. Selbts et al.[16] consider several aspects of AI impact assessment: early interventions, open ended questions, accountability, collaborative governance, community involvement, procedural regulation, minimum substantive standards and oversight. Similarly, Ernst and Young survey[17] explores existing methodology attempts with five key considerations, including: a) when to conduct impact assessment (as early as possible in design stage), b) team of assessors (not to be beholden to AI developer or user), publication of results (to enable independent scrutiny), c) decisions on go/no-go clause (if the assessment results indicate that AI system is too risky), d) relevant

stakeholders (mobilisation of a diverse set of participants, to help identify risks and harms).

The potential use of HRIAs in AI development builds on a history of diverse impact assessments within a range of domains such as finance, environment, data protection, with the purpose of evaluating the benefits and impacts of a business practice, technology or policy. Although impact assessments vary across domains, they share some core characteristics and pitfalls, and for this reason, they are a valuable resource for such an investigation.

Environmental Impact Assessment

EIAs are specific tools to identify and assess the actual and potential environmental implications of a project before the project commences. Nowadays EIAs are practised in many countries, with the aim to reduce the environmental impact of a project at the earliest possible stage, during the planning phase. They are typically required by law, policies, administrative orders or other types of regulations, with key elements including public participation, development of an implementation plan for monitoring and review of the project itself. Some countries are attempting to expand the EIA scope to cover sustainable development and cumulative effects on society as well.[18] However, EIAs have also been criticized[19] due to poor public consultation, low quality reports, costly, inefficient and time-consuming practises, limited scope, lack of sufficient monitoring and inconsistent implementation. These are risks that any impact assessment process has, and they should be taken into account when developing the HRIA governance framework.

Data Protection Impact Assessments

DPIAs, required by the EU's General Data Protection Regulation (GDPR)[20] of private companies collecting personal data, include cataloguing and addressing system characteristics and the risks to people's rights and freedoms presented by the collection and processing of personal data. DPIAs are a type of risk assessment process for both building and demonstrating compliance. In other jurisdictions, Privacy Impact Assessments (PIA) are, similarly to DPIA, an activity conducted by public institutions or private sector to assess the impacts on privacy of a project, technology, product, service,

policy, programme or other initiative and, in consultation with stakeholders, to take remedial actions in order to avoid or minimize negative impacts. It is considered an important 'privacy by design' process to ensure privacy considerations are built in from conception through to implementation.[21] The rules typically require a PIA post-implementation evaluation plan to monitor and review the initiative to determine whether the project is operating as envisioned in the PIA, as well as a report to be made public. Strict compliance measures for DPIA—and, to some extent, PIA—required clear guidance, and this explains why there has been a proliferation of DPIA and PIA check lists, templates and training, helping to standardize the process.[22] This approach of collectively constructed guidance and standardization has been useful and should be promoted for HRIA as well.

3. Methodology

Our overarching research objective is to examine the role of impact assessment and risk management in AI development, when used to anticipate, prevent, and mitigate harms and harness AI benefits. In doing this, we approach the AI development process from a socio-technical perspective. This paper addresses a research question central to this endeavour: How is human rights risk management of AI systems for public use conducted in practice during the AI lifecycle? We will focus in particular on two dimensions:

a) Risk or Impact Assessment (during development);

b) Risk or Impact Monitoring (during deployment).

To answer this question, we will use both primary and secondary data.

Our field research design follows the guidelines of Yin[23]. The research follows an inductive approach, i.e., data gathering to explore the phenomenon, and an attempt at structuring this data into a conceptual framework (Saunders et al.).[24] We applied thematic analysis for the interview transcripts of the 11 use cases and two case studies. Our approach consisted of several steps: a) collecting and understanding data: by preparing, conducting, and transcribing the interviews, b) coding and labelling the data, in particular sentences or paragraphs with a three-type code method (process coding, descriptive coding and narrative coding), c) searching for categories, themes and

relationships by identifying the relevant patterns and how they relate within categories, d) interpreting the categories and themes, and e) refining and summarising the interpretations. The findings were categorised and interpreted according to the relevance for the key research question (we will present only findings from the field research relevant to the present scope.) More in detail, our field research included:
– 4 in-depth semi-structured interviews with experts from institutions selected in the two case studies, providing insights on AI development in the public sector on a national level (two governmental institutions) and on a municipal level (a large city), using different impact
assessment methods in different contexts.
– 11 semi-structured interviews with key informants directly working on AI design, development or assessment cases in the Netherlands, within both national and local level institutions.

Additionally to the field research, we conducted an extensive desk research to review key regulatory trends on human rights and AI frameworks focusing on impact assessment in AI development. We selected regulatory proposals currently being drafted or negotiated based on their maturity, relevance for impact or risk assessment framework, and anticipated regional or global impact. In addition, we selected several diverse impact assessment methodologies that correspond to or specifically invoke some of the regulatory initiatives, as being most relevant to the ongoing discussions. We conducted a comparison and a context analysis of proposed solutions, referring to what extent these proposals include an implementable framework for impact assessments.

4. Findings

4.1 Field research: AI in Dutch public institutions

We interviewed 15 participants (distributed in 11 use cases and 4 case studies) with different roles in Dutch public institutions, at local or national level, who are involved in developing or assessing AI systems. Our semi-structured questionnaire included a total of 30 questions divided into three sections: 1. General: questions about the topic of the

use-case, actors involved, and the content of the (teams) role. 2. Development process: the type of input, resources, tasks, and roles needed throughout the development process to make informed decisions, phases, risk and impact assessment issues, stakeholder engagement. 3. Considerations: perceived challenges for different phases, potential improvements or failures of the system, and communication gaps, assessing errors, bias, as well as the potential negative impact of the AI system. Interviews typically lasted one hour. We identified several key challenges that participants are facing in the AI development process.

Case Study	**Role of Interviewee**
1. National institution	Data Scientist
2. National institution	Data Scientist
3. Large City	Product Owner / Member of Innovation Team
4. Large City	Legal Officer
Use Cases	**Role of Interviewee**
1. Umbrella institution	Project Manager of AI Development
2. Municipality	Data Researcher
3. Local level institution	Team Leader - Developer
4. National level institution	Methodologist / Internal Consultant
5. Local level institution	Data Scientists
6. National level institution	Project Manager / Statistician
7. National level institution	AI Researcher / Consultant
8. Local level institution	AI Researcher
9. National level institution	Innovation Manager
10. Municipality	Product Owner / AI Team
11. National level institution	Data Scientists

Challenge 1: Assessing impact.

In most use cases, there was no official or clear governance nor process for assessing the impact of AI systems. Almost all participants were aware of the general ethic or

responsible AI guidance and principles, however, the majority stressed lack of operationalization of general guidance, or too strict scoring system for assessment. in practical terms. Six participants emphasized broader expertise as key to address the challenges regarding a broad range of interdependent human rights. Almost all participants expressed increasing understanding for the need to assess AI impacts but stressed that no clear process was available to draw on the broad expertise that is needed. Five participants revealed it is difficult for policymakers to understand how to assess AI models, as there are diverse potential interpretations of the results and too many guidelines, without clarity on which one to use and how. In general, more communication, examples, and feedback were flagged as useful and needed, not only internally but also making connections with the real world and practice.

Through the in-depth case studies, we gathered observations and lessons during the piloting of the Dutch IAMA assessment framework[25] in national institutions. The teams involved in the pilots consisted of staff from different departments (developers, data analysts, business analysts, project managers, ...). However, unexpectedly, legal and privacy officers were involved only at a later stage of the process. Participants expressed the need to involve more diverse team members (with e.g. legal background) from the beginning of the process to help internal discussions about terminology and function of certain aspects of the framework. Before the piloting, team members received an initial training on assessment but, admittedly, more skill and knowledge building was needed to fully apply this in practice. This deficit led to process delays, as the time to fill in the assessment questionnaire took much longer than scheduled. Participants expressed the additional need for a skilled facilitator role to keep the discussion moving forward. They also acknowledged that the institutional setup was wary of involving external stakeholders, e.g. there was no process or governance framework to include stakeholders outside the institution. IAMA was seen as a useful tool to initiate discussions on ethical and legal aspects of AI, however, data scientists found it too vague; for example, they considered bias to be loosely defined without specific metrics. They expressed the need for more exact instructions about translating concepts and principles into operationalisation. This was most visible in the sections that involve assessment of impact on human rights and requirements to address proportionality and necessity (IAMA Sections 4.6. and 4.7.),

where participants have difficulties to understand balancing of competing rights and principles. For example, there were no straightforward answers to the questions whether deploying the AI system is the best option, given its effects on human rights. Similarly, teams struggled with questions about trade-offs, e.g. how to balance an alternative that appears to be slightly less effective, but is less intrusive for human rights. All these issues were considered crucial by developers and policymakers for making key decisions on taking forward the development or use of AI systems, as they reflect both design and policy choices. However, it was not clear from the assessment process itself who should make those estimates within the team, how the team could reach final decisions, and what assistance they needed. Finally, teams noted IAMA could be more interactive, to follow the lifecycle of AI development through the stages of the process, as advancement in design and risks become more visible.

Challenge 2: Risk management, and roles.

Along the similar lines, in the majority of researched use cases, it was not always clear who takes the executive decisions on whether development or use of AI systems can proceed, or what is the (legal, procedural) basis for such decisions. Participants stressed that a more extensive system of checks and balances (more diverse, more fine-grained) would be useful, but it would inevitably make the AI development process slower. All participants expressed a clear need for an increase of communication during AI system development for enabling input on design choices from policy makers. Specifically, key decisions on AI model features that were in essence political needed detailed documentation of each step, to be able to justify the decision against the different options. Detailed input on measurements and AI model features were required from various aspects—including security, privacy, legal and technical. For decisions involving more risk from AI systems, four participants stressed that the decision was taken higher up the political chain. In such cases, developers needed to provide explanations to policymakers in various parts of the organizational structure concerning features of AI systems and their impact, to facilitate key decisions on going forward with their use. Often these decisions were deemed controversial. Finally, participants identified the engagement elements as important but missing, such as building trust with stakeholders to gain honest input, creating a meaningful process of engagement,

discussing who should be included as a stakeholder or expert, how to reach them and what methods to use.

For our case study, we interviewed experts with different roles within the Large City. The City is piloting internally developed processes and methods for AI assessment and development. Such organizational framework clearly sets out the roles, tasks and responsibilities of the involved administrative and policy actors at various levels, dividing administrative, system, and official responsibility. A process management for understanding risk before making a decision to develop or procure AI is described within the internal guidance document. According to the guidance, the necessary technical and legal expertise should be involved in this process, along with documenting a decision on whether to use AI. The decision on the development or procurement of AI systems should be made taking into account (a) established frameworks and policies on AI, (b) principles for ethical and responsible use of data and AI, and (c) frameworks for data protection, information disclosure and information security and archiving. However, there is no explicit mention of assessing impact on human rights and freedoms at large. For the impact assessment, the City uses an internally developed tool Risk Analysis, which consists of a list of questions to assess whether additional measures are necessary for managing AI systems deemed "risky". However, this only applies to AI systems that have an impact on official decision-making or services, determined via a closed list by the City. Using such predetermined (closed) list of high-risk AI systems does not allow for a comprehensive or broader human rights impact assessment that might capture other AI systems not listed. Although a variety of in-house expertise is to be consulted during the development process, in practice, there was an expressed need for more involvement and creating an environment for external stakeholders to participate directly in the AI development. Finally, there was an increased need to present risky or politically sensitive AI systems to the official city portfolio holders, who would then submit the proposal on the use of the AI to the decision makers for (political) decision. However, these processes did not always occur in practice.

4.2 Desk research: ongoing regulatory efforts

Regulatory proposals	**Key Features for Assessing Impact**
EU AI Act	Product conformity self-assessment mandated only for high-risk AI systems, lacking methodology and human rights basis
CoE Possible elements of a legal framework on AI	Two step human rights, democracy and rule of law impact assessment mandated for all AI systems
US Blueprint for an AI Bill of Rights	Non binding principles with focus on protecting rights and democratic values, based on impact
Brazil – AI Bill	Specific AI impact assessment duty solely for AI systems identified as high risk (same risk levels as the EU AI Act)
The Netherlands / IAMA	Broad assessment to supports AI development primarily for public sector, includes human rights considerations

European Union AI Act Proposal

The proposed EU Artificial Intelligence Act[26] is pioneering the safeguards for trustworthy AI by introducing a set of rules applicable to the design, development and use of certain high-risk AI systems, as well as restrictions on certain uses. High-risk AI systems, qualified as such in advance via a closed list, would need to comply with specific requirements, such as setting up a sound risk management system, the use of high-quality data sets, appropriate documentation and sharing adequate information, appropriate human oversight measures, and the highest standards of robustness, safety, cyber-security and accuracy. However, an adequate standpoint on impact assessment is missing; the 2021 EU Commission proposal requires AI providers to undergo a conformity self-assessment, solely for high-risk AI systems. Such an approach of ex-ante qualifying AI systems to different risk categories does not consider that the level of risk depends on the context of use and broader impact, which cannot be fully determined within a fixed taxonomy. Moreover, the conformity self-assessment would be conducted without clear external oversight, no rights-based methodology or engagement with external stakeholders. Finally, the proposal provides very few opportunities for public engagement and contesting this classification.

Council of Europe AI Convention negotiations

The Council of Europe's 2022 document on "Possible elements of a legal framework on artificial intelligence"[27] establishes an outline for the methodology of risk classification

of AI systems with a focus on human rights, democracy, and the rule of law. The classification would be based on an initial review to determine if a full Human Rights, Democracy and Rule of Law Impact Assessment (HUDERIA)[28] is required. In order to safeguard the proportionality of a risk-based approach, a full HUDERIA would be recommended if there are clear and objective indications of relevant risks emanating from the application of an AI system, to be determined after an initial lean HUDERIA review. However, the draft of the proposed methodology does not address external oversight, nor clear governance framework, nor meaningful stakeholder engagement as integral parts of the process. Moreover, it is unclear what legal status HUDERIA will have within the future AI Convention (mandated, guidance only, joint standard, etc.), rendering it potentially "toothless" for Member States and broader AI implementation.[29]

US Blueprint for an AI Bill of Rights

The US Government published in 2022 a Blueprint for an AI Bill of Rights[30] that includes principles to guide the design, use, and deployment of automated systems. Its focus is on protecting rights and democratic values, based on impact, as opposed to the underlying technical choices made AI systems. Specifically, the Blueprint should be applied with respect to all AI systems that have the potential to meaningfully impact individuals' or communities' rights, opportunities, or access. However, the principles are non-regulatory and non-binding, with no enforceable Bill of Rights to back it. In addition, its limited scope—applicable only to automated systems that have the potential to meaningfully impact the public's rights, opportunities, or access to critical resources or services—generally excludes many industrial and private sector applications of AI. The Blueprint is accompanied by a "technical companion"[31] that offers specific steps that industry, communities, and governments can take to put these principles into action. This includes addressing the potential AI impact, including disparity testing results and mitigation information. However, such assessment is limited in scope, referring mainly to the protection from discrimination, thereby excluding the broader human rights framework. There are no envisaged methods on how to actually assess this impact, nor how to include external (impacted) stakeholders in such discussion.

Brazil's Artificial Intelligence Bill

Brazil established a Commission of Jurists[32] in 2022 tasked with drafting a new regulation on AI which will be based on the 2021 Artificial Intelligence Bill. In particular, the Senate noted that the rapporteur of this Commission will lead a study into existing experiences (such as within the EU) as a source of inspiration for key concepts. A new draft was published in December 2022[33], based on three central pillars: (i) ensuring the rights of the people affected by the AI system, (ii) risk level grading, and (iii) certain governance measures aimed at companies that provide or operate the AI system. In particular, the regulation establishes a specific AI impact assessment duty, however, limited solely to those AI systems classified as high risk (based on a taxonomy similar to that of the EU AI Act). The assessment is to be carried out by a professional or a team of professionals with technical knowledge, scientific and legal resources necessary to develop the report, including at least the following steps: preparation; risk recognition; mitigation of the risks found; monitoring. The impact assessment would consist of a continuous iterative process, running throughout the life cycle of high-risk AI, with periodic updates required. Conclusions of the impact assessment would be public, containing at least the basic information on description of the intended purpose, context of use and territorial and temporal scope of AI; risk mitigation measures, as well as residual risk level and description of the participation of different affected segments (where applicable). There are additional detailed requirements to keep the records of the assessment. However, it would be up to the competent authority to regulate the frequency of updating impact assessments and inclusion of public participation in a simplified way, without clear indication what that means in practice. Similar to the EU AI Act proposal, a closed list of pre-identified high-risk AI does not allow much contestation or stakeholder engagement.

4*.3 Desk research: impact assessment methodology proposals*

Proposed Impact Assessment Methods	**Key Features**
AI Now Institute: algorithmic impact assessment / Canadian assessment tool	Self-assessment exercise without clear or mandated governance or process structure

Human Rights Impact Assessment by Mantelero and Esposito	Complex and detailed human rights based assessment and management tool, including risk assessment matrix
Dutch Impact Assessment of Human Rights and Algorithms (IAMA)	Broad assessment to supports AI development primarily for public sector, includes human rights considerations

Developing comprehensive, coherent yet applicable methodologies for effective HRIA implementation is a core challenge for impact assessment processes, particularly when they require utilising expertise and metrics across different AI domains. The starting point for existing emerging examples is that AI designers and developers need to construct or structure their systems in order to ensure ex ante compliance with applicable human rights. Several distinct methodologies have been proposed and yet to be fully tested in the real setting, however, their analysis reveals important key features. We have reviewed five of those, based on their recent publishing, relevance of human rights features within the model and connection with relevant ongoing regulatory or policy proposals or practices:

AI Now Institute: Algorithmic Impact Assessment

In the early works, the 2018 report by Reisman at al, Jason Schultz, Kate Crawford and Meredith Whittaker from the AI Now Institute proposed a process for an algorithmic impact assessment to ensure public agency accountability.[34] The report presents both the process from pre-acquisition review to regular review during use and the content of the assessment. The Canadian government based its first Algorithmic Impact Assessment tool based on that report; however, this tool was a self-assessment tick-the-box type exercise without clear or mandated governance or process structure.[35]

Human Rights Impact Assessment by Mantelero and Esposito

The methodology proposed in 2021 by Mantelero and Esposito[36] for a Human Rights Impact Assessment (HRIA) provides both human rights assessment and management tools. It is intended for the early stage of AI development and can follow it throughout its lifecycle, providing specific, measurable and comparable evidence on potential impacts, their probability, extension, and severity. It can also facilitate comparison between alternative design options, based on risk assessment and mitigation. The process consists of two main phases: (1) planning and scoping phase, involving

identification of type of AI system, data flows and data processing purposes, overall human rights context and identification of relevant stakeholders; and (2) data collection and analysis phase, where evidence is gathered to assess the concrete AI system's impact on human rights and freedoms. This assessment includes analysing three key factors: (a) risk identification for adverse human rights impact, (b) likelihood of such adverse impact (consisting of probability of adverse consequences and the exposure to it), and (c) severity of impact (nature of potential prejudice in the exercise of rights and freedoms and their consequences). Based on this analysis, a risk assessment matrix is developed to qualify risks, and a mitigation management tool is devised to address identified risks. Although sound and complete, such HRIA methodology is resource-intensive, involving extensive research and field work, consultations with local stakeholders and experts. This process would be useful in complex multi factor scenarios (e.g., complex AI products or services especially for public sector use), however, they are likely too burdensome for smaller scale AI. For smaller AI developers, the requirements could be applied proportionally, in a less detailed process, to safeguard the key HRIA principles. However, the methodology has not yet been tested in practice, therefore, its feasibility is unclear.

Dutch Impact Assessment of Human Rights and Algorithms (IAMA)

IAMA is an AI impact assessment developed by researchers, that supports making decisions, primarily for the public sector, about the development and deployment of AI systems, including human rights consideration. IAMA provides steps and questions that should be addressed before the AI is developed and implemented in key phases. In the preparation phase, it determines why an AI system will be used for and concrete goals (if AI genuinely is the best solution for achieving the goal). In the input-and-throughput phase, technical issues of development of an AI system are addressed: how the AI should look like, its operations, transparency, and data quality. In output phase, implementation and supervision of the AI system is assessed, its use, monitoring, including the opportunity to overrule decisions. A cross-cutting phase checks whether AI affects human rights, and to what extent, and then determines how adverse impacts can be prevented or mitigated. The methodology contains a separate, elaborate questionnaire for identifying risks of infringing human rights, however, lacks corresponding guidance on engaging stakeholders and facilitating assessment by multi-

disciplinary teams. IAMA is being piloted in the Netherlands—in 2023, approximately 20 institutions (from national government, provinces, municipalities) will apply IAMA to AI development with risks to human rights. However, there is no comprehensive governance framework or clear legal basis for its implementation that would facilitate pilots and engage external stakeholders.

5. Discussion

As expected in a multi-perspective research, we gained insights that are at the same time in alignment and conflicting. There seems to be a fair amount of optimism within the proposed regulatory efforts and impact assessment methods about the possibility of achieving effective impact assessment for responsible AI; on the other hand, there is an overall impression of confusion in the practice of AI development as well as the lack of clear governance and implementation tools. The findings also indicate that, at least on a normative proposals level, we seem to be gaining an enhanced understanding of the different ways AI impact can be assessed. However, there is a lack of clarity, let alone consensus, on crucial questions guiding the practical governance and process considerations, e.g.: who should be conducting and be involved in impact assessments, at what stage, what are the scope and benchmarks for the assessment (human rights or other frameworks), what are process and methodology possibilities, how are assessment findings being communicated and published, what are different roles and actors in the process, how to involve external stakeholders, etc. Moreover, the use cases we examined show a considerable mismatch between the proposed theoretical assessment methods and practical implementations, with a number of gaps discussed above that need to be addressed. Everything suggest we are still at the beginning of this process.

Once we have acknowledged the identified weaknesses, we can sketch a preliminary map towards a solution. As we stated earlier, no static solution is plausible in this context, rather, we need to identify the right process of governance that allows the solution to be properly adapted. In general, a sound architecture of governance should: (i) start from a general but clear normative framework, (ii) translate (contextualise) it in a (functional and functioning) arrangement holistically, by engaging with

organisations from private to public sectors, and the public, especially communities that might be more knowledgeable of the actual situation and (may) have more adequate information on the potential impact.[37] In order to make it suitable to the AI context, this requires a multi-layered form of governance, with a smart interplay of international norms, national regulation, co-created technical standardisation, framework for guidance, as well as spaces for civic oversight, monitoring and engagement. An additional advantage of setting straight an ecologically oriented framework is that part of these intervention points (at the lowest layers) may be attributed to artificial devices for e.g. automated assessment, regulation, and adaptation or repair (cf. the concept of normware[38]). To address the concerns stemming from our field research and analysis of regulatory proposals, we consider adapting the human rights framework for AI impact assessment, as a prerequisite for making informative choices in the AI design and development. One of the key aspects of such assessment concerns the metrics of value constructs that are being promoted and/or protected by those developing and ultimately, deploying the AI systems. As we discussed in the introduction section, broadly accepted international human rights framework can provide potential answers for the need to balance competing interests and impacts in a flexible yet predictive manner. This was precisely one of the key challenges identified in our field research by those involved in AI development. In addition, the analysis of regulatory proposals as well as assessment methodologies revealed gaps in providing specific framework and guidance for practical usage.

We can begin to sketch the potential pathway by examining how implementing the "three-part test" (legality, legitimacy, and proportionality and necessity) of international human rights case law to the AI design context would assist with HRIA operationalisation, asking the following key questions:
– Is the intended purpose of the AI system allowed and legal? (legality test)
– Is a particular AI system effective in achieving its intended purpose? (internal legitimacy test)
– Is a particular AI system proportionate and necessary, considering its human rights impact? (external legitimacy test)

To unpack this further, we provide illustrative examples of their content and scope.

Legality test of the intended purpose could explore what is the actual driver or reason to use an AI system. The main aim would be to decide why it would be needed or useful to deploy an AI for a particular purpose, comparing that to other tools that may be available to address it. Potential sub-questions include:

Legality	**Issues to Address - Is the intended purpose of an AI allowed and legal**
Underlying cause	For which problem is the AI system to provide a solution? Is it well defined?
Purpose	What is the main purpose or goal to be achieved with the use of an AI system? It should be very specific, e.g. not (only) for 'addressing crime', or 'protection of national security', but concrete, e.g. 'using an automated process to chart and analyse indicators for criminal activity in a specific space / time'.
Reason	Why is the use of an AI system needed (as opposed to other solutions)?
Legal basis	What is the legal basis for using the AI system for the stated purpose? There should be an explicit framework that allows for the use of an AI system, in particular if it is expected to affect people's lives, freedoms or rights.
Legal process	In case AI will impact people's lives, there must be a clear legal process about the AI decisions as well as sufficient opportunities for legal protection.

Internal Legitimacy test: realistic estimate needs to be made about actual results of using an AI system against achieving the stated goal or purpose. For example, will an AI system result in cutting costs or in gaining efficiency, if the entire cost of developing and maintaining it (with all included process, legal and technical requirements) are considered?

Internal Legitimacy	**Issues to Address - Is the AI effective in achieving its intended purpose?**
Suitability	Is the AI system a suitable means to realise the set purpose? What is relevant evidence for that?
Reliability and Accuracy	Will AI be reliable and accurate from a technical perspective for its use? E.g. how many false-positives or false-negatives are expected and is that number problematic in the use context?
Performance	In what way will the system perform the relevant task more effectively compared to the current state of play, thereby improving the performance? How realistic is the benefit of using an AI relative to the present situation?
Risk of fraud	Is there a risk of the system being 'gamed' or otherwise inappropriately used? Are there unintentional uses possible in different (adversary) contexts?

External Legitimacy test: there is a need to assess the potential weight of the impact on diverse human rights and freedoms in different cultures, compared to the potential weight of the beneficial objectives and interests that could be achieved with the use of AI systems. For example, AI could be a suitable and even necessary tool to enhance the efficiency of decision-making in the public sector, however, it risks reinforcing historical discriminatory patterns and bias. The question here is whether deploying the AI is the best option, given its effects on human rights in a specific context. In such a case, there should be an assessment if it is reasonable and beneficial to still deploy the AI system, considering available mitigating measures to alleviate potential harm. Generally, the graver the expected infringement on rights and freedoms is, the heavier the benefit of use needs to be, along with measures to mitigate infringement, to outweigh harm in comparison. Ultimately, some infringements on rights and freedoms are deemed unacceptable by law, therefore, when detected, the AI system should not be considered for development at all.

External Legitimacy	**Issues to Address - Is AI proportionate and necessary, considering its human rights impact?**
Initial Scanning	Which (aspect of) a human right or freedom will be potentially affected? E.g. list all possible potential impacts on a diverse spectrum of rights and freedoms.
Scaling of Impact	For each potentially affected right or freedom, what is the scope, severity and likelihood of expected or potential harm? E.g. use scales and metrics.
Legislative Assessment	Does specific legislation apply to these potentially affected rights and freedoms, e.g. privacy regulation, constitution, bill of rights? If yes, what are the benchmarks for infringement?
Proportionality Test	Does the use of the AI system result in acceptable balance between the pursued objectives and the rights or freedoms that will be potentially infringed? Are the ultimate goals enough to justify the infringement of rights and freedoms?
Mitigation (Trade Off) Test	What are possible trade-offs in case there is an alternative means (or AI system) slightly less effective, but might affect the right and freedoms in a less harmful manner? Are there realistic and effective mitigating options / safeguards to lower harm?
Necessity Test	Is the use of AI systems really necessary to achieve the stated purpose and objective, are there no other means or mitigating measures available?

For all the elements listed above, in order to adequately address the complexities surrounding societal interactions, questions and considerations must be contextualized

within various settings. Therefore, the inclusion of socially and culturally relevant dimensions is crucial for generating precise and applicable assessments.

As we elaborated in previous sections, the human rights framework can provide a baseline for implementing normative constraints on AI system development, however, they are not sufficient for overall AI governance architecture. To implement the "*three part test*" and address challenges from the desk and field research discussed above at architectural level, key governance and process features need to be considered throughout the AI lifecycle. Therefore, we trace the emerging criteria for a human rights based impact assessment, that would allow meaningful and effective implementation.

Normative framework: Scope, Content, Benchmarks, Priorities

Regulatory proposals for assessment would need to present a clear scope and criteria for the whole-of-rights approach assessing a wide range of human rights, including collective rights, economic, social and cultural rights, and environmental rights, all serving in principle as benchmarks and criteria for the assessment. According to the UN Guiding Principles on Business and Human Rights, a credible assessment would draw on independent expertise, lived experiences and the concerns of affected stakeholders. In addition, the assessment could prioritise harm reduction and the adverse impacts on marginalised and vulnerable groups, and amplifying potential AI benefits.

Process: Phases, Procedures, Roles, Iterations

During HRIA implementation, procedural rights would need to be respected alongside substantive ones; in particular, principles of non-discrimination, participation and transparency. For doing so, assessment could be conducted iteratively throughout the AI lifecycle starting at the earliest design phase, reviewing the impacts in an agile manner. Clear step by step process, roles and actors could to be defined to guide the assessment. Appropriate resources (including time) and capacity would need to be allocated for this purpose as well.

Methodology: Teams, Indicators, Scales, Balancing (Trade-offs)

Developing a standardised yet flexible set of methods for various AI contexts will be crucial to enable translation of human rights into the operational assessment process and quantitative and qualitative metrics. Methodology could be based on an analysis of the severity and likelihood of risks relevant to human rights, however, these need to recognize the societal and cultural context of the AI system. There is also a need for the indicators to help with balancing and proportionality tests and trade-offs regarding competing interests. Moreover, the impact assessment process can become severely limited if conducted as a self-assessment exercise. To resolve this, involvement of external assessors and/or oversight can provide assurances. Assessment teams need to include a variety of socio-technical expertise with potential assistance available to make it economically sustainable (e.g. partial subsidisation, engagement of civil society, research institutes, dedicated agencies, etc.).

Context: Inclusion, Frameworks for engagement

Process of assessment could be context sensitive and embed meaningful external stakeholder inclusion as a central part of impact identification during design phase. AI system developers would need to engage with different groups during design and development to deliver practical solutions, with diverse forms of expertise, lived experience and lessons from other sectors. Such processes should not be seen as a burden, but an integral part of the AI lifecycle. Specific methods for engagement and co-creation are needed as guidance, particularly for involving vulnerable and affected communities. There are different learnings from inclusion in policy making processes that can serve as a baseline and warnings how to avoid "participation-washing".

Oversight: Documentation, Reviews, Feedback

Establishing an oversight framework for assessment to ensure they are run in a consistent and criteria-based manner remains critical. This could include external, iterative, or ongoing review of conducted assessment. All information related to the oversight body and their review would ideally be made publicly available (note that this is a distinct transparency requirement than the one on impact assessment process: we

may have a transparent impact assessment process, but an opaque oversight). Furthermore, steps could be incorporated for receiving feedback from the public on the AI impact, for making revisions when necessary. This also involves potential tracking the effectiveness of responses to impacts on individuals or groups that may be at increased risk of vulnerability.

Accountability: Publication, Monitoring, Revisions

The findings of assessment processes would need to be presented to the public, to facilitate (re)building trust in AI use. This is an important step to enable public scrutiny, engagement, and potential concerns. This might include, where necessary, mechanisms to protect trade secrets and intellectual property, while providing access to key findings. When AI is likely to affect certain communities in a disproportionate way, potential public consultation boards composed of individuals directly affected could be established and consulted on the impact. A smart mix of policy measures should be envisaged for authorities, standardisation bodies, private sector and interested public, to monitor, assess and revise AI impacts.

6. Conclusion and future work

This paper explores the comprehensive topic of human rights in AI impact assessment as a baseline for potential normative constraints on AI system development. It delves into whether AI impacts can be adequately addressed by utilizing a commonly accepted core of human rights frameworks. We examine ongoing pilots implementing AI risk assessment and management procedures in the Dutch public sector to identify mismatches between expectations, process, and implementation that need to be overcome to make impact assessments useful, effective, yet manageable. Additionally, emerging regulatory proposals and methodologies for AI impact assessment were analysed through desk research to determine whether mandating safeguarding measures for human rights is consistent with current normative frames. Based on these insights and analyses, the paper sketches a governance process supporting a meaningful, rights-based, context-sensitive and inclusive assessment approach towards AI impacts. To implement it and address challenges from the desk and field research,

following our analysis, the key architectural features need to be considered as relevant emerging criteria:

- Clear scope and criteria for the whole-of-rights approach and a wide range of human rights as benchmarks for assessment.
– Prioritising harm reduction and the adverse impacts on marginalised and discriminated groups.
– Assessment at all AI lifecycle phases iteratively, starting at design.
– Easy-to-use step by step process, roles, actors and responsibilities.
– Allocated resources and learning capacities for all involved.
– Standardised yet flexible methodology and indicators for balancing and trade-offs.
– External assessors and/or oversight involved with multidisciplinary teams.
– Mandated engagement of external stakeholders to ensure contextualisation and specific methods for inclusion.
– Robust monitoring mechanisms, documenting and reviews.
– Publication of assessment findings, feedback mechanisms and scrutiny of results.

In conclusion, leveraging a human rights-based impact assessment could be the means to address the potential societal impacts of AI. However, this requires clear governance structures and practical processes that ensure its viability and effectiveness. The question of meaningful and successful implementation of human rights impact assessment as an instrument for responsible AI hinges on key normative, technical and guidance features including appropriate scope, structure, process, timing, methodology, and administrative burden. For future work, it is urgent to settle upon these criteria to ensure it will become an effective and sustainable process. Developing those should involve sound concepts as well as empirical insights informed by ethical, political, legal, and technical perspectives.

Endnotes:

[1] E.g. OECD AI Principles, Montréal Declaration for a Responsible Development of Artificial Intelligence, UNESCO Recommendation on the Ethics of Artificial Intelligence.

[2] Stahl, B.C., Antoniou, J., Ryan, M. et al., Organisational responses to the ethical issues of artificial intelligence, AI & Soc 37, 23–37 (2022) https://doi.org/10.1007/s00146-021-01148-6

[3] Philip Pullella and Jeffrey Dastin, 'Vatican joins IBM, Microsoft to call for facial recognition regulation' Reuters (London, 28 February 2020); Kate Crawford, 'Time to regulate AI that interprets human emotions' (2021) 593 Nature 167; Business & Human Rights Resource Centre, 'Tech companies call for govt. regulation of artificial intelligence' (2021)

[4] McGregor, L., Murray, D. and Ng, V. (2019), 'International Human Rights Law as a Framework for Algorithmic Accountability', International & Comparative Law Quarterly, 68(2), April 2019, pp. 309–43, https://doi.org/10.1017/S0020589319000046; Yeung, K., Howes, A. and Pogrebna, G. (2019), 'AI Governance by Human Rights-Centred Design, Deliberation and Oversight: an end to Ethics Washing', in Dubber, M. and Pasquale, F. (eds) (2019), The Oxford Handbook of AI Ethics, Oxford: Oxford University Press; Jones, Kate: AI governance and human rights: Resetting the relationship, 2022

[5] The interpretation of key definitions and principles of human rights has been assisted by the Universal Declaration of Human Rights, corresponding international and regional human rights instruments, case law and related guidelines.

[6] See for example: Institute of Electrical and Electronics Engineers (IEEE) Global Initiative on Ethics of Autonomous and Intelligent Systems, Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligence Systems, 1st ed.2019, https://standards.ieee.org/content/ieeestandards/en/industry-connections/ec/autonomous-systems.html; UNESCO Recommendation on the Ethics of Artificial Intelligence, 2021, https://en.unesco.org/artificial-intelligence/ethics

[7] It is 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 rights-holders. Danish Institute for Human Rights. 2020. Guidance on Human Rights Impact Assessment of Digital Activities, https://www.humanrights.dk/sites/humanrights.dk/files/media/document/A

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

[9] United Nations Guiding Principles for Business and Human Rights. 2011, https://www.ohchr.org/documents/publications/hr.pub.12.2 en.pdf

[10] E.g. Alessandro Mantelero, Samantha Esposito (2021), An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data- Intensive Systems,

https://papers.ssrn.com/sol3/papers.cfm?abstract id=3829759. See also: Mirko Tobias SchÅNafer, Arthur Vankan, Iris Muis, Janneke Gerards. 2021. Impact Assessment Mensenrechten en Algoritmes (IAMA), https://www.uu.nl/en/news/iama-makes-careful-decision-aboutdeployment-of-algorithms-possible

[11] UN Human Rights Committee: General Comment No. 27, 1999, CCPR/C/21/Rev.1/Add.9.

[12] "There is no conflict between ethical values and human rights, but the latter represent a specific crystallisation of these values that are circumscribed and contextualised by legal provision and judicial decisions." Mantelero, A. and Esposito, S. (2021), 'An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems', Computer & Security Law Review, 2021, https://ssrn.com/abstract=3829759

[13] Yeung K., Howes A. and Pogrebna G. (2019), 'AI Governance by Human Rights-Centred Design, Deliberation and Oversight: an end to Ethics Washing', in Dubber, M. and Pasquale, F. (eds) (2019), The Oxford Handbook of AI Ethics, Oxford: Oxford University Press.

[14] The European Court of Human Rights has repeatedly stressed that the ECHR is a "living instrument which…must be interpreted in the light of present-day conditions". Tyrer v United Kingdom, ECtHR App No 5856/72, judgment of 25 April 1978, Series A No 26, para. 31.

[15] UN Human Rights Committee general comments, UN Human Rights Council reports, UN Guiding Principles on Business and Human rights, etc.

[16] Selbst, Andrew D., An Institutional View of Alghoritmic Impact Assessment, Harvard Journal of Law & Technology Volume 35, Number 1 Fall 2021, https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf

[17] Ernst & Young LLP, A survey of artificial intelligence risk assessment methodologies, 2021, https://www.trilateralresearch.com/wp-content/uploads/2022/01/Asurvey-of-AI-Risk-Assessment-Methodologies-full-report.pdf

[18] Hari Srinivas. 2015. Sustainable Development Concepts: Environmental Impact Assessment. https://www.gdrc.org/sustdev/concepts/08-eia.html

[19] Ibid.

[20] European Commission. Data Protection Obligations. https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/when-data-protection-impact-assessment-dpia-required_en

[21] Office of the Australian Information Commissioner, 2021, Guide to undertaking privacy impact assessments, https://www.oaic.gov.au/privacy/guidance-and-advice/guide-to-undertaking-privacy-impact-assessments

[22] E.g. United Kingdom Government, Information Commission Office, 2018, Sample DPIA template, https://ico.org.uk/media/2258461/dpia-templatev04-post-comms-review-20180308.pdf

[23] R. K. Yin, 'Case study research: Design and methods, applied social research', Methods series, vol. 5, 1994.

[24] M. Saunders, P. Lewis, and A. Thornhill, 'Research methods', Business Students 4th edition Pearson Education Limited, England, 2007.

[25] Impact Assessment Mensenrechten en Algoritmes (IAMA), https://www.uu.nl/en/news/iama-makes-careful-decision-aboutdeployment-of-algorithms-possible

[26] 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, COM/2021/206 final

[27] Ad hoc Committee on Artificial Intelligence (CAHAI), Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law, 2022, https://rm.coe.int/possibleelements-of-a-legal-framework-on-artificial-intelligence/1680a5ae6b

[28] HUDERIA draft, https://rm.coe.int/huderaf-coe-final-1-2752-6741-5300-v-1/1680a3f688

[29] Minutes CAI plenary 2023, potential options being discussed for HUDERIA legal status.

[30] U.S. Blueprint for an AI Bill of Rights, https://www.whitehouse.gov/ostp/ai-billof-rights/

[31] U.S. Blueprint for an AI Bill of Rights, https://www.whitehouse.gov/wpcontent/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf#page=12

[32] Brazil Senate announcement, https://www.dataguidance.com/news/brazil-senateannounces-commission-draft-regulatory

[33] Brazil draft AI Law, https://legis.senado.leg.br/sdleggetter/documento/download/777129a2-e659-4053-bf2e-e4b53edc3a04

[34] Algorithmic Impact Assessments Tool, https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html

[35] Algorithmic Impact Assessments Report: A Practical Framework for Public Agency Accountability, AI Now Institute, https://ainowinstitute.org/publication/algorithmic-impact-assessments-report-2

[36] Alessandro Mantelero, Samantha Esposito (2021), An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data- Intensive Systems, https://papers.ssrn.com/sol3/papers.cfm?abstract id=3829759

[37] Amandeep S. Gill, Stefan Germann, 2021, Conceptual and normative approaches to AI governance for a global digital ecosystem supportive of the UN Sustainable Development Goals (SDGs), https://link.springer.com/article/10.1007/s43681-021-00058-z

[38] Giovanni Sileno, 2020, Of duels, trials, and simplifying systems, European Journal of Risk Regulation 11(3)