



A comparative review of 10 Fundamental Rights Impact Assessments (FRIA) for AI-systems

September 2024

A comparative review of 10 Fundamental Rights Impact Assessment (FRIA) for AI-systems

| Dimensions examined | The Alan Turing Institute | ALGORITHM WATCH | ALIGNER | BSR | THE DANISH INSTITUTE FOR HUMAN RIGHTS | DEMOS HELSINKI | FORHUMANITY | Canada | unesco | AI4 | TOTAL |
|---------------------|---|-----------------|--------------------------|------|---------------------------------------|----------------|-----------------------|-----------------------|--------|-------|-------|
| Number of pages | 335 | 48 | Roughly 10 (Excel sheet) | 6 | 134 | 17 | Roughly 18 (web page) | Roughly 11 (web tool) | 51 | 95 | |
| LEGAL | Fundamental rights | ● | ● | ● | ● | ● | ● | ● | ● | ● | 9/10 |
| | Data protection | ● | ● | ● | ● | ● | ● | ● | ● | ● | 8/10 |
| | National administrative law | ● | ● | ● | ● | ● | ● | ● | ● | ● | 3/10 |
| | Proportionality test | ● | ● | ● | ● | ● | ● | ● | ● | ● | 4/10 |
| ORGANIZATIONAL | Intended purpose of algorithm | ● | ● | ● | ● | ● | ● | ● | ● | ● | 9/10 |
| | Organizational measures | ● | ● | ● | ● | ● | ● | ● | ● | ● | 9/10 |
| TECHNICAL | Documentation instructions | ● | ● | ● | ● | ● | ● | ● | ● | ● | 8/10 |
| | Data quality | ● | ● | ● | ● | ● | ● | ● | ● | ● | 8/10 |
| | Accuracy specifications | ● | ● | ● | ● | ● | ● | ● | ● | ● | 6/10 |
| | Methodological control measure | ● | ● | ● | ● | ● | ● | ● | ● | ● | 0/10 |
| SOCIAL | Stakeholder panel to identify normative questions | ● | ● | ● | ● | ● | ● | ● | ● | ● | 9/10 |
| | Stakeholder panel to resolve normative questions | ● | ● | ● | ● | ● | ● | ● | ● | ● | 3/10 |
| TOTAL | 11/12 | 9/12 | 7/12 | 2/12 | 8/12 | 9/12 | 5/12 | 6/12 | 9/12 | 10/12 | |

Key take-aways

- > Fundamental Rights Impact Assessments (FRIAs) review AI-systems mainly by posing questions that are meant to stimulate self-reflection. FRIAs do not provide answers or concrete guidelines how to realize ethical algorithms.
- > Given the inherent tension between fundamental rights and between the ethical risks and benefits of AI-systems, it is essential to provide a clear rationale for why a reasonable balance on these issues has been struck when deploying an AI-system. Many FRIAs fall short in adequately addressing the normative considerations underlying this balancing process.
- > Existing FRIAs often include common statistical measures to evaluate AI systems, such as false positive and false negative rates. However, they often fall short in connecting these metrics to assessing the conceptual soundness of the statistical methodology of the AI-system, such as (hyper)parameter sensitivity testing for machine learning and deep learning methods, as well as statistical hypothesis testing for AI-driven risk assessment of individuals.
- > The technocratic approach taken by most FRIAs does not empower citizens to meaningfully participate in shaping the technologies that govern them.
- > Digital tools and platforms should be developed to fill in lengthy FRIAs, preventing red taping and time-consuming paperwork.

Executive summary

Under the EU AI Act, Fundamental Rights Impact Assessments (FRIAs) will play an important role in safeguarding fundamental rights. We have conducted a comparative review of 10 existing FRIAs frameworks, evaluating them against 12 requirements across legal, organizational, technical and social dimensions. Our assessment shows a sharp divide regarding the length and completeness of FRIAs. We believe that future FRIAs should balance conciseness with comprehensiveness.

We recommend that future FRIA frameworks incorporate legal instruments that address the core of normative decision-making, such as the objective justification test, and that place greater emphasis on statistical methodology, including sensitivity testing of model parameters. Furthermore, we advocate for actively involving diverse stakeholder groups – not only in identifying fundamental rights tensions while conducting a FRIA, but also in resolving identified tensions. To avoid red taping, digital tools should simplify the process of conducting a FRIA, even if the judgements required by a FRIA should be made by humans and cannot be automated.

Motivation

The European AI Act's regulatory approach – safeguarding fundamental rights by way of product safety regulation – is groundbreaking. It obliges deployers and producers to self-assess and address risks to fundamental rights in AI-systems they develop or maintain. Historically, impact assessments have played an important role in assessing the effect of products and services on European societies, its environment and single market. Fundamental Rights Impact Assessments (FRIAs) will therefore play an important role in this decentralized self-assessment approach. This development merits an evaluation of the FRIA as a regulatory instrument in practice. How do existing legal frameworks, such as European non-discrimination law and the European Convention of Human Rights, interact with a risk-based regulatory approach? Do current FRIA frameworks contain sufficient references to statistical methodology to meaningfully evaluate the impact of AI-systems?

Our comparative analysis offers an overview of the strengths and gaps in existing FRIAs, from which recommendations for future improvements can be drawn.

- > **Existing FRIAs:** What FRIAs are currently available? How do these impact assessments prove to be valuable in AI practice?
- > **Future improvements:** What are shortcomings shared by existing FRIAs that should be addressed in future FRIA frameworks?

Our analysis of current FRIAs could be beneficial to regulators developing a standardized FRIA framework for the AI Act, as well as for public and private organizations that further develop existing and new FRIA frameworks.

Methodology

The methodology of the comparative analysis is described below.

1. A selection of 10 commonly used FRIAs has been made.¹ FRIAs are selected based on Algorithm Audit's AI policy experience and input from the international AI auditing community.
2. We compare FRIAs across 4 dimensions: i) legal, ii) organizational, iii) technical and iv) social. Every dimension includes multiple requirements. This results in a total of 12 requirements used to evaluate the selected FRIAs. Requirements map to obligations under the AI Act for high-risk AI-systems for which a FRIA needs to be conducted. See also [Table 1](#).
3. We assign a binary rubric for each of the 12 requirements. We determined whether a requirement is included (yes/no) in a specific FRIA. The selection of requirements and rubric classification is based on practical experience in AI auditing, particularly in evaluating bias in AI-systems. The rubric classification provides direct guidance on interpreting the readiness of existing FRIAs to be applied in the context of the AI Act.
4. The comparative review offers insights into the strengths and shortcomings of existing FRIA frameworks. It identifies additional aspects that should be incorporated into future versions to enhance their readiness for usage in the context of the AI Act.
5. It is important to note that not all selected FRIAs have a European origin (e.g. Government of Canada). This impact assessment includes a review of at least one similar approach from another jurisdiction that shares similar values with the European Union. Moreover, not all FRIAs are specifically designed to assess AI-systems (BSR). This FRIA is included to demonstrate that AI-systems require a specialized type of assessment, and that FRIAs cannot be easily generalized across different applications.

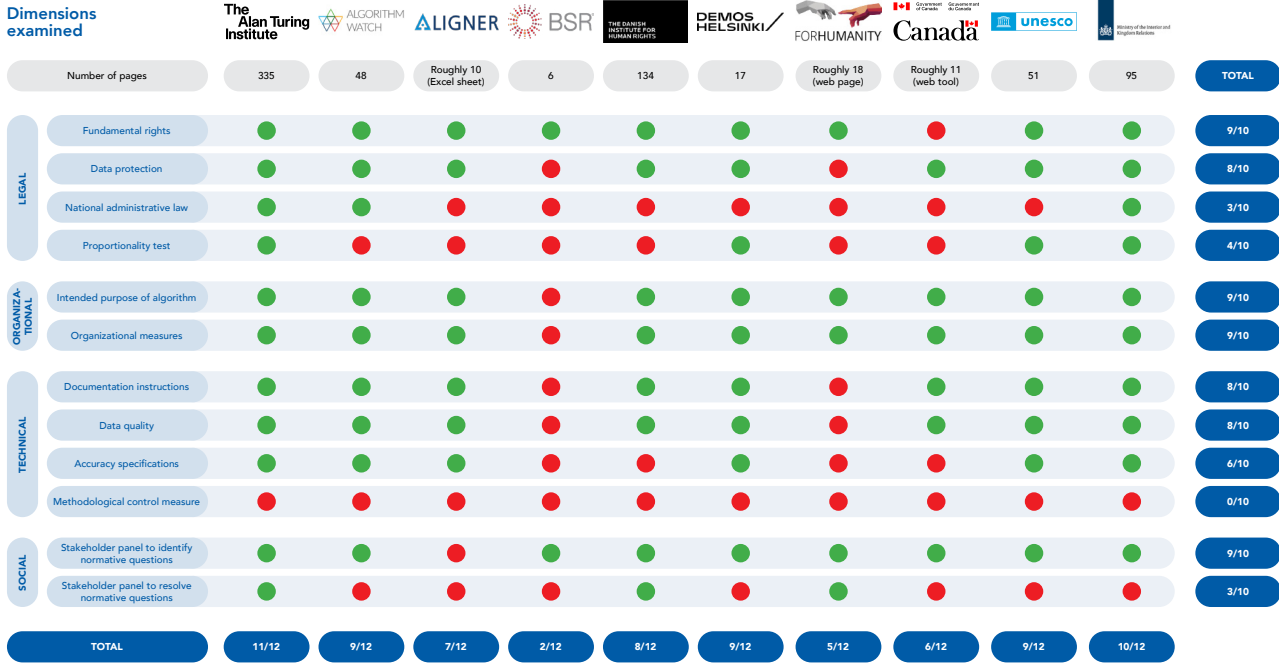
¹ Weblinks to all selected FRIAs can be found in [Appendix A](#).

Table 1. The dimensions and requirements used to compare FRIAs, linked with the relevant articles AI Act articles

| Dimension | # | Requirement | Description | AI Act section |
|----------------|----|---|--|--|
| Legal | 1 | Fundamental rights | Charter of Fundamental Rights of the EU and national constitutions. Specifically in relation to non-discrimination law: the Race Equality Directive 2000/43/EC; the Framework Equality Directive; and the gender equality Directives 2004/113/EC and 2006/54/EC, as well as Articles 20, 21 and 23 of the Charter. | Art. 1 (Subject matter), Art. 27 (Fundamental rights impact assessment for high-risk AI-systems) |
| | 2 | Data protection | Interaction with General Data Protection Regulation (GDPR). Referred to at least 20 times in the AIA. For instance, in the context of profiling in Article 3 (52). | Art. 3 (Definitions) and Art. 10 (Data and data governance) |
| | 3 | National administrative law | Principles of sound governance, among others transparency and motivation obligations, e.g., principle of fair play, duty of care and duty to give reason. | Recital 48 + 60 (Good administration) |
| | 4 | Proportionality test | Plays a key role in balancing fundamental rights tensions and ethical quandaries, such as providing an objective justification for the differential treatment of groups by an AI-system. | Art. 7 (Amendments Annex III), Art. 14 (Human oversight) |
| Organizational | 5 | Intended purpose of algorithm | The legitimate purpose for which the AI-system is used. | Art. 9 (Risk management system) |
| | 6 | Organizational measures | Assigned roles and responsibilities, allocated resources, risk management, workplace culture <i>inter alia</i> . | Art. 26 (Obligations of deployers of high-risk AI-systems) |
| Technical | 7 | Documentation | Focus on documentation of normative data modelling decisions. | Art. 12 (Record-keeping) |
| | 8 | Data quality | Functional requirements, completeness, reliability, evaluation mechanisms. | Art. 10 (Data and data governance) |
| | 9 | Accuracy specifications | Such as confusion matrix-based evaluation metrics (FNs/FNs), but also mean average precision (MAP) or similar metrics for recommender systems, and fairness metrics. | Art. 15 (Accuracy, robustness and cybersecurity) |
| | 10 | Methodological control measure | Such as, hypothesis testing for risk assessment of natural persons and sensitivity testing of (hyper)parameters. | Art. 15 (Accuracy, robustness and cybersecurity) |
| Social | 11 | Stakeholder panel to identify normative questions | Inclusion of a diverse group of stakeholders in conducting a FRIA for a specific AI-system is promoted. | Art. 95 (Codes of conduct for voluntary application of specific requirements) |
| | 12 | Stakeholder panel to resolve normative questions | Inclusion of a diverse group of stakeholders in resolving fundamental normative tensions – as identified by conducting a FRIA – is promoted. | |

Findings

A comparative review of 10 Fundamental Rights Impact Assessment (FRIA) for AI-systems



We present the final classification of each FRIA in the figure above. Satisfying a requirement yields 1 point. In total 12 points can be obtained. A justification for every classification is [available](#). Our results demonstrate that existing FRIAs satisfy the 12 requirements to various degrees. Some FRIAs score in total 5 points or less (BSR, ForHumanity). The most complete FRIAs score 10 points or more (Alan Turing Institute, Utrecht University-Dutch Ministry of the Interior).

Even for the highest scoring FRIAs, there is room for improvement – particularly in the technical and social dimension. For instance, no reviewed FRIA elaborates on statistical methodological control measures for AI-systems, such as (hyper)parameter sensitivity testing for machine learning and deep learning methods, and statistical hypothesis testing for AI-driven risk assessment of natural persons. More guidance could also be provided for the

selection of specific fairness and accuracy metrics for specific types of AI applications, such as mean average precision (MAP) or comparable metrics for recommender systems. Besides, greater emphasis should be placed on actively engaging diverse stakeholder groups, not only to identify normative questions when conducting a FRIA but, more importantly, to help resolve them. This approach would make FRIAs less technocratic and would empower citizens to shape the technologies that govern them.

Legal

As expected, all FRIAs address fundamental rights, except for the Canadian government's Algorithmic Impact Assessment, which is not specifically focused on fundamental rights. In terms of data protection, only BSR and ForHumanity do not satisfy the minimum requirements for privacy and data governance in their impact assessments.

There are wider discrepancies among FRIAs when it comes to references to national administrative law provisions on sound administration, where transparency and justification are mandatory for public sector algorithm use, even before the AI Act came into effect. As a core principle of EU non-discrimination law, proportionality testing should have been more prominently featured in FRIAs. However, only 4 out of 10 organisations meet this requirement (Alan Turing, Demos Helsinki, UNESCO, Utrecht University-Dutch Ministry of the Interior).

Organizational

Many FRIAs elaborate on the intended purpose of AI-systems, recognizing that the first step toward responsible AI is to reflect on the supposed benefits of its use, in order to balance it against the risks it might pose. Equally important are organizational measures for responsible deployment of AI-systems, such as defining roles and responsibilities, implementing risk management frameworks, and fostering an open culture that encourages critical reflection on performance of the AI-system. BSR is the only organization that does not meet both of these requirements in its Human Rights Assessment.

Technical

Given the inherent tension between fundamental rights, it is essential to document why a reasonable balance between these rights has been achieved when producing or deploying an AI-system. For instance, why a specific profiling criterion is necessary to realize the legitimate aim pursued. Most FRIAs therefore contain sections relating to documentation of AI-systems, and the quality and integrity of the data (except BSR, ForHumanity). Existing FRIAs often include common statistical measures to evaluate AI-systems, such as false positive and false negative rates. However, they often fall short to connect these metrics to assessing the conceptual soundness of the statistical methodology of the AI-system, such as sensitivity testing for machine learning and deep learning methods, as well as statistical hypothesis testing for AI-driven risk assessment of individuals.

Social

Stakeholder engagement is included in many FRIAs, though the extent and focus of this engagement varies. Most FRIAs raise questions that are meant to identify fundamental rights tensions by diverse groups of stakeholders that are conducting the assessment (except BSR). However, only some FRIAs promote that these diverse group of stakeholders also resolve the identified

Box 1

Example: BSR Google Celebrity Recognition API Human Rights Assessment

BSR has published a Human Rights Assessment of Google's Celebrity Recognition API. For this type of facial recognition software, the evaluation of false positives (FPs) and false negatives (FNs) is crucial for understanding the impact of the AI-system on fundamental rights. For instance, it is important to reflect on how this type of classification systems can incorrectly label individuals – such as misidentifying non-celebrities as celebrities (false positives) or failing to recognize those who are in fact celebrities (false negatives). These errors are key to evaluate, especially across various demographic groups, as it is known from scientific research that facial recognition software has varying performance across different skin tones, and could therefore violate equal treatment provisions under EU equality directives. Besides, the assessment does not address how Google collects the celebrity image database or explain why it meets the data quality standards as mandated by Article 10 of the AI Act.

tensions (Alan Turing Institute, Danish Human Rights Institute, ForHumanity). It should be noted that some FRIAs put the emphasis on obtaining stakeholder feedback through structured processes and in multiple layers of the assessment such as panel testing as a part of conducting the FRIA (Danish Institute for Human Rights).

Recommendations

We formulate five recommendations to FRIA developers.

- i. Share conducted FRIAs** – By sharing conducted FRIAs more broadly, others can learn how complex issues related to responsible AI can be addressed and how best-practices can be harmonized.
- ii. Focus on normative data modelling decisions** – Assessing fundamental rights is inherently a balancing act. FRIAs should include additional questions to identify the normative aspects of AI modelling. For example, proportionality testing (a legal requirement under the objective justification test) should be linked with quantitative metrics related to predictive value and statistical significance.
- iii. Include statistical methodology** – Without incorporating statistical terminology, FRIAs won't fully realize their potential in assessing the impact of AI-systems on fundamental rights. For instance, questions should be included in FRIAs what type of sensitivity testing² is performed for machine learning and deep learning methods, as well as statistical hypothesis testing for algorithmic-driven risk assessment of individuals.³
- iv. Inclusive decision-making** – Impactful

normative decisions about AI should not be taken behind closed doors by technical experts only. It calls for a more democratic and less technocratic approach. FRIAs can support this goal by actively engaging a diverse range of stakeholders in resolving fundamental rights tensions identified by conducting a FRIA.⁴

- v. Digital FRIA tools** – FRIAs can be lengthy and overly comprehensive, like HUDERIA exceeding 300 pages. To encourage adoption and practical use, FRIA developers should aim to keep their documents concise. Digital tools can help to make conducting FRIAs less burdensome. There are limits to how far this process can be automated, however, since the motivated judgements within a FRIA need to be consciously taken by humans.

Disclaimer

While Algorithm Audit is a European knowledge platform for AI bias testing and normative AI standards, our reading of the AI Act is not a formal legal analysis. Nonetheless, it may provide valuable insights that may help inform such analyses, particularly in areas where the law is ambiguous. We encourage creators of reviewed FRIA to provide feedback to us and respond to these scores.⁵

² An example of sensitivity testing of a Balanced Random Forest (BRF) model can be found in our [Github](#).

³ An example of statistical hypothesis testing in the context of algorithmic-driven risk assessment can be found in our [Github](#).

⁴ A diverse and inclusive AI auditing methodology can be found on our [website](#).

⁵ Reach out to info@algorithmaudit.eu

Appendix A – Weblinks to 10 FRIAs

- > [Alan Turing Institute](#)
- > [Algorithm Watch](#)
- > [Aligner](#)
- > [BSR](#)
- > [Danish Institute for Human Rights](#)
- > [Demos Helsinki](#)
- > [ForHumanity](#)
- > [Government of Canada](#)
- > [UNESCO](#)
- > [Utrecht University-Dutch Ministry of the Interior](#)

About Algorithm Audit

Algorithm Audit is a European knowledge platform for AI bias testing and normative AI standards.

The goals of the NGO are three-fold:



Normative advice commissions

Forming diverse, independent normative advice commissions that advise on ethical issues emerging in real world use cases, resulting over time in [algotprudence](#)



Technical tools

Implementing and testing technical tools for bias detection and mitigation, e.g, [bias detection tool](#), synthetic data generation



Knowledge platform

Bringing together experts and knowledge to foster the collective learning process on the responsible use of algorithms, see for instance our [AI Policy Observatory](#) and [position papers](#)

Structural partners of Algorithm Audit

SIDNfonds

SIDN Fund

The SIDN Fund stands for a strong internet for all. The Fund invests in bold projects with added societal value that contribute to a strong internet, strong internet users, or that focus on the internet's significance for public values and society.

European Artificial Intelligence & Society Fund

European AI&Society Fund

The European AI&Society Fund supports organisations from entire Europe that shape human and society centered AI policy. The Fund is a collaboration of 14 European and American philanthropic organisations.



Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

Dutch Ministry of the Interior and Kingdom Relations

The Dutch Ministry of the Interior is committed to a solid democratic constitutional state, supported by decisive public management. The ministry promotes modern and tech-savvy digital public administrations and governmental organization that citizens can trust.

Building **AI auditing** capacity
from a **not-for-profit** perspective



www.algorithmaudit.eu



www.github.com/NGO-Algorithm-Audit



info@algorithmaudit.eu



Stichting Algorithm Audit is registered as a non-profit organisation at
the Dutch Chambre of Commerce under license number 83979212