CECU

Towards a meaningful implementation of Article 27 under the AI Act: A call for active and impactful participation in Fundamental Rights Impact Assessments (FRIAs) & effective transparency mechanisms for their oversight and enforcement

September 2025

About this report:

This report has been prepared by **Mayra Russo Botero and Kristen M. Scott** (authors), with coordination and editing by Federación de Consumidores y Usuarios CECU (Anabel K. Arias).

This report has been developed in collaboration with IA Ciudadana[1] and aims to contribute to the coalition's ongoing work.

The research was completed in June 2025.

Supported By:

European
Artificial Intelligence
& Society Fund

[1] IA Ciudadana is a coalition of 17 organizations working to defend human rights in the context of digital technologies. Our goal is to expand spaces for societal participation in the regulation and governance of artificial intelligence and algorithms. More info: iaciudadana.org.

Executive Summary

The EU AI Act (Regulation (EU) 2024/1689), effective August 1, 2024, introduces phased obligations for AI system stakeholders through 2027. While most provisions target high-risk AI providers, Article 27—effective August 2, 2026—requires certain deployers of high-risk AI systems to conduct a Fundamental Rights Impact Assessment (FRIA). Deployers covered include public bodies and private entities delivering public services, as well as those using AI for creditworthiness, credit scoring, and insurance risk assessment (excluding fraud detection).

In the current regulation, the FRIA is a self-assessment exercise by deployers with little procedural requirements aside from submission of a filled-out FRIA template – the template will be provided by the AI Office. Oversight of FRIAs is managed nationally. EU Member States must establish National Competent Authorities, including Market Surveillance Authorities (MSAs), to whom the filled-out FRIA templates will be submitted.

Civil Rights Organizations and other experts have expressed concern regarding the limited application of FRIAs to only some high-risk systems, lack of transparency considerations regarding public access to FRIA documentation, and lack of prescriptiveness regarding inclusion of expert and public input into FRIAs. These limitations generate concerns regarding meaningful implementation of FRIAs, that is, FRIAs that prioritize the protection of fundamental rights instead of just compliance through a simplified process to further feed into "tick box culture;" hampering their effectiveness and failing to enforce AI deployers' accountability.

A meaningful implementation of FRIAs ensures protection of fundamental rights in the face of AI system development and deployment. **A meaningful FRIA must entail a process of deliberation and discussion among relevant stakeholders, including affected groups, "rights holders," domain experts, and fundamental rights experts.** This is vital for ensuring serious reflection and structured thinking about possible adverse impacts on fundamental rights, and given the risk-based nature of the AI Act, enable the definition of robust and comprehensive risk mitigation plans for them.

Moreover, thorough, **transparent documentation of the FRIA process**, including discussions and decisions taken, is required so the FRIA process can be dutifully reviewed and used as an effective mechanism for accountability of AI deployers. Transparency about FRIA processes also supports learning regarding best practices and the impacts of decisions over time, thus creating a culture of careful assessment around the development and deployment of AI systems.

Recommendations

Meaningful FRIA processes are those that include input from relevant stakeholders and that are documented and transparent in a way that allows scrutiny and accountability. Implementation policies for the AI Act and Article 27 at the EU and Member State levels must be ones that foster a meaningful FRIA process; ensure oversight, enforcement and accountability; prepare for resource requirements; and incentivize best practices by deployers in both public and private settings.

Fostering a meaningful FRIA process. To avoid FRIAs becoming only an internal self-assessment, it is necessary to include elements of third-party oversight in the form of diverse internal and external stakeholders that can support an effective and impactful process. The assessment should have a qualitative element, where the specifics of the deployment context shape the discussion, rather than only a pre-set, close-ended questionnaire. For such participation to be meaningful, there must be opportunities for the participating members of these groups, and all other participating stakeholders, to genuinely contribute to the FRIA process and outcome. This means that raised concerns are addressed, justification for decisions is communicated, and the possibility of halting or refusing deployment is always available in the case of crossing red lines or failing to justify proportionality in terms of fundamental rights infringements.

Oversight, enforcement and accountability after the FRIA process. The most rigorous implementation of a FRIA as an accountability mechanism should also enable the means to challenge the results of the FRIA and hold AI deployers accountable when their AI systems fail to comply with the results and recommendations of the FRIA upon deployment. Authorities empowered for preemptive (rather than reactive) oversight and established transparency practices will also contribute to a leaner inquiry process to ensure protection of fundamental rights as per the AI Act.

Resource allocation for meaningful FRIAS. Ensuring public participation in FRIAs will require additional funding and other resources to become available to ensure a meaningful, credible, and fair process. Preemptive review and oversight of documentation on FRIAs can prevent harm, scandal and associated costs.

Incentivising best practices. The AI Act makes legal provisions for the completion of FRIAs for certain cases specified under Article 27. Beyond a mere legal compliance exercise, FRIAS are also positioned as a key accountability mechanism for AI deployers and developers alike. Meaningfully and proactively engaging in identifying, categorizing, understanding, and mitigating for potential harms derived from the use of the AI system taking a fundamental rights approach in tight collaboration with affected people, should be seen as the baseline for these processes and primed in an ecosystem that seeks to be perceived as the gold standard in AI regulation and governance.

Content

Acronyms

AI - Artificial Intelligence

AIA - Algorithmic Impact Assessment

CSO - Civil society organisation

DPA - Data Protection Authorities

DPIA - Data protection impact assessment

EDPB - European Data Protection Board

FOI: Freedom of Information

FRIA - Fundamental rights impact assessment

FR - Fundamental right

FRAIA - Fundamental rights algorithmic impact assessment

IA - Impact assessment

Introduction and scope

Alleged benefits attributed to AI systems have contributed to the push for their development and deployment across diverse areas of the public sphere, e.g., public administration, business, health, education, law, employment. There is significant pressure from industry and governments to prioritise work on AI systems that support humans in a wide range of tasks, often with the aim to displace humans through full automation.

KEY CONCEPTS

AI SYSTEM

While a contentious definition, the AI Act[2] defines AI systems as machine-based systems that are designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infer, from the input they receive, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments

ALGORITHMIC SYSTEM FOR AUTOMATED DECISION-MAKING

Any technology that either assists or replaces the judgment of human decision-makers. These systems draw from fields like statistics, linguistics and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets.

AFFECTED GROUPS

[2] AI Act, Article 3: Definitions

Affected groups are those that will bear most of the impact and consequences derived by the deployment of a particular AI system, e.g., workers, patients, or bank clients, either directly or indirectly.

Encompassed by this definition are also vulnerable and disadvantaged individuals and groups of people, that are identified as such given their race, ethnicity, gender identity, age, disability, sexual orientation, religion, citizenship status, socioeconomic background, parental status, limited language proficiency, rural origin, renters, homeless, impoverished people, or other identities and lived experiences; including the intersection of these.

While there is a lack of legal definition for the concept, specifically under the AI Act, an approximation as per legal text interpretation understand that this concept could refer to "natural persons" to whom high-risk AI systems' decisions are "related;" and in broader terms, any consumer, citizen, end-user, migrant, etc., that might be affected by an AI system.[3]

The use of AI systems in critical and consequential decision-making scenarios such as healthcare treatment allocation, credit assessment, job suitability, incarceration, border surveillance and human mobility management, among others, has become more ubiquitous. This has occurred in the wake of efficiency, efficacy, and neutrality claims regarding AI systems and is now being reinforced in light of the popularity of generative artificial intelligence (GenAI). The enthusiasm for the deployment of these systems is widely supported by political agendas and commercial interests under the premise of digital innovation and transformation and economic growth strategies.[4]

[3] See AI Act, Article 26(11) and Article 50 & Kaminsiki M. & Malgieri G. The Right to Explanation in the AI Act. U of Colorado Law Legal Studies Research Paper No. 25-9. 8 March. 2025. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5194301

[4] OECD AI Observatory. How governments are driving AI adoption for economic growth. 20 May 2025. Available at: https://oecd.ai/en/wonk/how-governments-are-driving-ai-adoption-for-economic-growth

The growing ubiquity of AI systems has also helped to make prevalent how their use can have a harmful impact on society, i.e., algorithmic harm, and infringe on people's rights, with evidence for this supported by extensively researched instances of algorithmic discrimination, documentation of AI-related incidents,[5] and accompanying media coverage. These risks give weight to calls for guardrails for AI development and deployment in order to protect human and fundamental rights, including preemptively identifying and centring individuals and groups that will be directly affected by the use of these systems,[6] as well as prioritising regulation that above innovation

To provide a fuller picture of algorithmic harm, two 'snapshots,' or abbreviated real world cases are described below, one for an algorithmic system called RisCanvi deployed in Spain and another one called SyRI deployed in the Netherlands.

SNAPSHOT OF ALGORITHMIC HARM 1 - RisCanvi: An algorithm deployed by the Catalonian Department of Justice to predict the likelihood of crime recidivism

Despite the potential implications for fundamental rights and freedoms, ethical considerations, and questionable efficacy,[7] especially when it comes to risk scoring and recidivism predictive tasks, political and commercial interests continue to encourage the use of these systems, with European Union agency Europol going as far as to claim that AI "has the ability to significantly transform policing; from advanced criminal analytics that reveal trends in vast amounts of data, to biometrics that allow the prompt and unique identification of criminals."[8]

[5] See for example: Partnership on AI's https://incidentdatabase.ai/

[6] AI development and deployment should support values like dignity, fairness, and transparency.

[7] See for example: Garcia, Ter, et al. La Policía Nacional deja de usar Veripol, su IA estrella para detectar denuncias falsas. Civio. 19 Marzo 2025. Available at: https://civio.es/transparencia/2025/03/19/la-policia-nacional-deja-de-usar-veripol-su-ia-estrella-para-detectar-denuncias-falsas/#veripol

[8] Europol. AI and policing:The benefits and challenges of artificial intelligence for law enforcement, Europol Innovation Lab observatory report. 2024. Available at: https://www.europol.europa.eu/cms/sites/default/files/documents/AI-and-policing.pdf

In the context of this work, a case of an algorithm used for law enforcement purposes is presented, as deployers in the criminal justice system will be subjected to legal obligations as per the AI Act, due to the high-risk nature of certain areas of application, including predictive policing tools.[9]

What is it: RisCanvi is, itself, a protocol put in place by the Catalan prison system to assess incarcerated individuals in determining their risk of becoming a recidivist.

Journalistic and academic investigate efforts[10,11] have contributed for years to shedding light on some of the specifics of the protocol and the algorithmic system.

The RisCanvi protocol, co-developed by the authorities and researchers at the University of Barcelona, has been in use since 2009. The protocol is carried out by combining primary evidence on the incarcerated individuals collected by a team of multidisciplinary professionals of the justice system, e.g., jurists, educators, social workers, psychologist, and the output of an algorithmic system, generated by encoding information obtained by these human experts, in order to analyse five different sets of risk scenarios[12] based on a list of risk factors to guide decision-making regarding, temporary release (prison furlough), prison transfers, and provisional release (parole).

There are two versions of the system, RisCanvi Screening, used to screen upon detainment, and RisCanvi Complete, used mostly to re-assess and to monitor incarcerated individuals that have been flagged as high risk. RisCanvi-S assesses for less risk factors (*~10)* and yields two different levels of risk, high or low; comparatively, the assessment performed by RisCanvi-C asses for some *~40* risk

[9] See more: AI Act Annex III: High-Risk AI Systems Referred to in Article 6(2) and Recital 59

[10] Martínez Garay, Lucía, et al., Three predictive policing appraoches in Spain: Viogén, RisCanvi and Veripol. Assessment from a human rights perspective. University of Valencia. November 2022. Available at: https://regulation.blogs.uv.es/files/2024/05/Three-predictive-policing-perspectives-web-17.06.24.pdf

[11] Bellio López-Molina, Naiara. In Catalonia, the RisCanvi algorithm helps decide whether inmates are paroled. Algorithmwatch. 21 May 2021. Available at: https://algorithmwatch.org/en/RisCanvi

[12] The five Risk Scenarios include self-directed violence, intra-institutional violence, violent recidivism, risk of breaking prison permissions or leaves, and general recidivism.

factors, and yields three levels of risk, high, medium, low.

There is evidence that the system has been internally reviewed and updated at least two times since 2009. In 2023, a third party was commissioned an audit by the Department of Justice.[13]

Effectiveness: As with other predictive policing tools employed for law enforcement, RisCanvi is used and deployed with the objective to bring objectivity and neutrality to decision-making processes. External assessments based on limited sample data,[14] as well as on adversary audits,[15] have been able to obtain quantitative and qualitative evidence on the performance of RisCanvi. The results have demonstrated that RisCanvi is not a well-calibrated system. This raises doubts as to its predictive accuracy, particularly when it comes to violent crimes, given its systematic tendency to over-confidently assign higher risks of recidivism across different subgroups of the sampled populations, and an associated high number of false positives, altogether.

Social implications:

- Evidence from third-party technical audits on samples of data have unveiled that the RisCanvi system fails to perform fair assessments, reinforcing historical biases towards already marginalized groups, such as incarcerated individuals with mental health and addiction problems, as well as individuals in dire socio-economic conditions, imparting harsher outcomes in these cases.
- Another criticised aspect of RisCanvi is the lack of disclosure by the authorities of its use to those affected by the system, the incarcerated individuals. Probing

[13] Read more: Informe Tiresias: Auditoria de l’algorisme RisCanvi. January 2024. Available at: https://repositori.justicia.gencat.cat/bitstream/handle/20.500.14226/1321/auditoria-algorisme-riscanvi-informe-final.pdf?sequence=1&isAllowed=y

[14] Read more: Martínez Garay, Lucía, et al., Three predictive policing approaches in Spain: Viogén, RisCanvi and Veripol. Assessment from a human rights perspective. University of Valencia. November 2022. Available at: https://regulation.blogs.uv.es/files/2024/05/Three-predictive-policing-perspectives-web-17.06.24.pdf

[15] Eticas Foundations. Automating (In) Justice? An Adversarial Audit of RisCanvi. June 2024. Available at: https://eticasfoundation.org/automating-injustice-an-adversarial-audit-of-riscanvi/

by journalists[16] has been able to uncover details on this system that for years has had a consequential impact on thousands of lives.

Following its original deployment, the authorities maintained the use of the system "under the radar;" incarcerated individuals did not receive information on the assessment process, nor on the consequences of the resulting outcomes. These risk assessments are evasive in nature, with an impact on an individual's private life, intimacy, free development of their personality and social environment.

- In general, there is the probability that these types of systems can contribute towards automation bias and to undermine human discretion of those who wield it. Audits and external assessment of this tool have uncovered concerns and lack of trust by the professionals that use the system, as well as by criminal lawyers and advocates that have evidence on the relevance and power attributed to the outcome of the algorithmic system in judiciary decisions.[17]

SNAPSHOT OF ALGORITHMIC HARM 2 - SyRI:

An algorithm deployed by the Dutch Ministry of Social Affairs and Employment to detect possible social welfare fraud

The Dutch Government commissioned and deployed an algorithmic tool to detect welfare fraud from 2017 to 2021. The rationale behind the use of this algorithm was the promise to recover millions of euros lost over to benefits fraud,[18] employing

[16] Jiménez Arandia, Pablo et al., Un algoritmo define el futuro de los presos en Cataluña: ahora sabemos cómo funciona. El Confidencial. 24 April 2024. Available at: https://www.elconfidencial.com/tecnologia/2024-04-24/riscanvi-algoritmo-cataluna-prisiones-presos-inteligencia-artificial_3871170/

[17] Id.

[18] Geiger, G. et al. Suspicion Machines: Unprecedented experiment on welfare surveillance algorithm reveals discrimination. Lighthouse Reports. 2 March 2023. Available at: https://www.lighthousereports.com/investigation/suspicion-machines/

machine learning techniques and cross-referencing personal data.

While quantitative evidence proved the limited real life efficacy of this system, its indiscriminate use on thousands of individuals, also implied severe human rights implications, making this case a prime example of algorithmic harm.[19]

What is it: The risk scoring algorithmic system, SyRi, automatically assessed and assigned a risk score to individuals based on more than 300 input factors, including variables such as age, gender, language skills, neighbourhood, marital status, along with encoded information resulting from qualitative data collected by social workers on a case by case basis. Some of these variables were used as proxies in place of "prohibited" ones, as was the case of 'ethnicity.' To this end, GDPR restrictions on personal data processing activities were undermined, given the existence of Dutch laws that facilitate personal data processing activities under the premise of fighting welfare fraud. Flagged individuals would be subjected to further investigation and suspension of social security benefits.

Social implications:

- After a lengthy process, investigative journalists[20,21] gained access to the source code for SyRI as facilitated by the public administration of the City of Rotterdam. Following a reconstruction of the machine learning model, the journalists uncovered that:
    - SyRI discriminated based on ethnicity, age, gender, and parenthood; with elevated risk scores for individuals at the intersections of these

[19] Id.

[20] Id.

[21] Geiger, G. Inside the Suspicion Machine. Wired Magazine. 6 March 2023. Available at: https://www.wired.com/story/welfare-state-algorithms/#intcid=_wired-verso-hp-trending_145b7ab1-a36c-4aca-8059-42e2e8304e38_popular4-1

identities, for example young, single mothers.[22]

    - SyRI also discriminated based on ethnicity. Since this variable was not included in the modelling, proxy variables were used, in this case encodings for language, e.g., level of fluency in Dutch or having a maternal tongue other than Dutch. The Netherlands Institute of Human Rights stated that profiling individuals based on language constituted the basis for indirect discrimination. Implied here is that the algorithm had a disproportionate negative impact on ethnic minorities, and specifically on working class migrants.

- SyRI was deployed and used in secrecy with the intention to make high-stake decisions on people lives; being singled out by the algorithm could lead in some cases to further intrusive interrogations, coupled with lengthy administrative investigations that could upend people's lives, leading to a potential violation of their right to social security under wrongful pretences.[23]

Impact: A Dutch court ruled that the SyRI system did not comply with the right to privacy under the European Convention of Human Rights, and labelled it a "mass profiling system." The hearings also led to a highly publicized investigation that uncovered the use of similar fraud detection systems across the Netherlands, targeting low-income and immigrant families. The Dutch Childcare Benefits Scandal

[22] Klaassen, S and van Dijk, Romy. Computer zegt vrouw: hoe een Rotterdams algoritme jonge, alleenstaande moeders discrimineerde. Ver beton. 6 March 2023. Available at: https://www.versbeton.nl/2023/03/computer-zegt-vrouw-hoe-een-rotterdams-algoritme-jonge-alleenstaande-moeders-discrimineerde/

[23] Read more: Brief by the United Nations Special Rapporteur on extreme poverty and human rights as Amicus Curiae in the case of NJCM c.s./De Staat der Nderlanden (SyRI, before the District Court of The Hague). Available at: https://www.ohchr.org/sites/default/files/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf

led to the dissolution of the central government.[24,25]

These cases have also supported the efforts to increase algorithmic transparency in the Netherlands, i.e., the creation of an algorithm register. However, similar algorithms continue to be widely employed, for instance, Spain’s AI Doctor.[26] In this case, the quest for transparency continues. More on this in later sections of this report.

Fundamental Rights Impact Assessments

The inclusion of Fundamental Rights Impact Assessments (FRIAs) in the AI Act has been widely advocated for by experts, including academics and civil society organisations and rights groups, among others. For this reason, the inclusion of this framework under Article 27[27] has been welcomed, if not without accompanying critique and warnings. Concerns focus on the limited application of FRIAs, lack of transparency considerations, and lack of prescriptiveness. These limitations generate concerns about the likelihood of a meaningful implementation of FRIAs, that is, one that prioritises the protection of fundamental rights instead of just compliance through a simplified process that further feeds into “tick box culture”. These limitations risk hampering the effectiveness of FRIAs and leading to a failure to enforce AI deployers' accountability.

[24] Davidson, D. et al., The Algorithm Addiction. 20 December 2022. Available at: https://www.lighthousereports.com/investigation/the-algorithm-addiction/

[25] Geiger, G. How a Discriminatory Algorithm Wrongly Accused Thousands of Families of Fraud. Vice. 1 March 2021. Available at: https://www.vice.com/en/article/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud/

[26] Jiménez Arandia, P. et al. Spain’s AI Doctor. Lighthouse Reports. Available at: https://www.lighthousereports.com/investigation/spains-ai-doctor/

[27] AI Act, Article 27 Fundamental rights impact assessment for high-risk AI systems

KEY CONCEPTS

HIGH-RISK AI SYSTEMS

AI systems that significantly are likely to incur “high-risk” to health, safety, and fundamental rights.

Article 6,[28] denotes that an AI system is classified as high-risk if it is: (1) a product which requires third-party conformity assessment under at least one of the Union harmonisation legislations listed in Annex I;[29] (2) used as a safety component of a product mentioned in the preceding point; or (3) used in the use-cases described in Annex III.[30]

DEPLOYER

Article 3.4 defines deployer as a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity.[31,32]

PROVIDER

Article 3.3 defines provider as natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge.[33]

FUNDAMENTAL RIGHTS IMPACT ASSESSMENT (FRIA)

[28] AI Act, Article 6: Classification Rules for High-Risk AI Systems

[29] AI Act, Annex I: List of Union Harmonisation Legislation

[30] AI Act, Annex III: High-Risk AI Systems Referred to in Article 6(2)

[31] AI Act, Article 3.4

[32] AI Act, Recital 13

[33] AI Act, Article 3.3

An assessment of the potential impact of an AI system on the rights of any individual that might be affected by the operation of that system. A FRIA is a risk assessment, so it does not focus on risk elimination, but rather on *risk management*. This entails: identification of risk, assessment of likelihood and severity of impact, definition of mitigation plan.

FRIAs are a requirement under Article 27 of the AI Act,[34] to be conducted in certain circumstances for high-risk AI systems (they exclude high-risk AI systems intended to be used in the area listed in point 2 of Annex III).

AI deployers that are bodies governed by public law, or are **private entities providing public services, and deployers of high-risk AI systems** referred to in points 5 (b), that is, **AI systems intended to be used to evaluate the creditworthiness of natural persons** or **establish their credit score**, with the exception of AI systems used for the purpose of detecting financial fraud; and (c) **AI systems intended to be used for risk assessment and pricing** in relation to natural persons in the case of **life and health insurance**, as per of Annex III, are all under this obligation.

A meaningful implementation of FRIAs ensures protection of fundamental rights in the face of AI system development and deployment. **A meaningful FRIA must entail a process of deliberation and discussion among relevant stakeholders, including affected groups, "rights holders," domain experts, and fundamental rights experts.** This is vital for ensuring serious reflection and structured thinking about possible adverse impacts on fundamental rights and how to mitigate them.

Moreover, thorough, transparent documentation of the FRIA process, including discussions had and decisions taken, is required so the FRIA process can be dutifully reviewed and considered an effective mechanism for accountability of AI deployers.

[34] AI Act, Article 27

This also supports the learning of best practices, understanding the impacts of decisions and actions over time and creating a culture of careful assessment around AI systems, which can impact fundamental rights.

Achieving the goal of meaningful FRIAs requires implementation strategies at the EU and the national levels, which:

- Foster a participative FRIA process
- Ensure effective oversight, enforcement and accountability
- Provide sufficient support and incentives for conducting meaningful FRIAs

Existing reports and advocacy on FRIAs under the AI Act focus on proposing methodologies for their implementation that go beyond mere tick boxing exercises for AI deployers.[35] In this report, those efforts serve as a strong and guiding foundation, however, **the emphasis and scope here is centred on highlighting how diverse public participation in these processes and effective transparency mechanisms that ensure oversight and enforcement are essential components of a meaningful FRIA.**

Structurally, this report starts by explicating the regulatory context at a European Union and Spanish level. It looks primarily at the AI Act, as well as relevant regulation such as the General Data Protection Regulation, Consumer Protection, as well as complementary regulation and draft bills at a member state level (Section 1). Then, it presents a bird-eye view of the current state of impact assessments applied to AI systems and other areas, such as development work, and environmental protection in order to extrapolate best practices, failure points and areas for improvement for FRIA processes under the AI Act for the public administration and the private sector

[35] Some examples: European Center for Non-for-Profit Law & Society Inside. Framework for Meaningful Engagement: Human rights impact assessments of AI. 2023. Available at: https://ecnl.org/publications/framework-meaningful-engagement-human-rights-impact-assessments-ai and European Network of National Human Rights Institutions. ENNHRI calls on the European Commission to ensure effective Fundamental Rights Impact Assessments (FRIAs) under the EU AI Act. 2025. Available at: https://ennhri.org/wp-content/uploads/2025/04/ENNHRI-statement-on-ensuring-effective-Fundamental-Rights-Impact-Assessments-FRIAs-under-the-EU-AI-Act.pdf

(specifically banks and financial institutions, Section 2). Section 3 presents an overview of existing participation approaches for affected groups, and captures insights from the practices of participatory design, design justice and transdisciplinarity in order to inform the realization of a meaningful FRIA process. Last, transparency mechanisms are discussed, both through a regulatory perspective, and through the perspective of algorithmic accountability reporting. The aim is to illustrate via real world examples the existing barriers to transparency, and the potential implications this can have on oversight and enforcement of Article 27 when it goes into effect (Section 4).

Methodology

For the completion of this study, a mixed-methods approach to data collection and information gathering was carried out.

Diverse secondary data sources, including academic papers, legislation, guidance, policy documents, works of investigative journalism, and shorter news pieces, were reviewed and analysed to inform our findings. The overarching topics that guided the search to compile a list of resources included: AI regulation, primarily the AI Act and the Spanish draft law, AI governance and algorithmic accountability mechanisms and tools (this includes impact assessments in their different modalities), participatory design and democracy theories, and transparency mechanisms and tools.

Additionally, eight semi-structured interviews with experts (ten experts in total), including academics and researchers, officers in the public administration, representatives from civil society organisations and an investigative journalist, were performed. The areas of expertise of the interviewees included AI regulation, policy, and governance, fundamental rights impact assessments, participatory democratic processes, and algorithmic accountability reporting. A similar approach to resource compilation was followed in order to identify experts, with more than thirty invitations being sent out. All interview material was informally coded and closely analysed in order to draw insights and inform the reported findings. Quotes have been used sparingly throughout this report, and all claims and comments derived from these interviews have been incorporated and supported through existing sources.

1. Regulatory context

1.1. Article 27 in the AI Act

The AI Act (Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence) is a European (EU) regulation on artificial intelligence (AI) which came into force on August 1, 2024, with a phased implementation between February 2025 and August 2027.[36] The Act takes a risk-based approach to regulation, with most of the requirements applying to high-risk AI systems.[37] A small set of AI uses are identified as unacceptable risk with their use prohibited (Article 5),[38] for limited risk systems there are transparency obligations, while minimal risk systems are left unregulated.[39] The majority of the obligations of the act apply to providers (developers) of high-risk AI systems, while Article 27, which requires a fundamental rights impact assessment (FRIA) for High-Risk AI Systems, applies to deployers.

Article 27 is expected to enter into force on August 2, 2026. It applies to "deployers that are bodies governed by public law, or are private entities providing public services,"[40] and deployers of "AI systems used to evaluate the creditworthiness of natural persons or for credit scoring (except for AI systems used for the detection of financial fraud), and for risk assessment and pricing in life and health insurance."[41] There are exceptions for AI systems used for critical infrastructure, as defined in point 2 of Annex III. According to Article 27, deployers of included systems must perform an ex ante FRIA, notify the market surveillance authority (MSA) of its results and submit a filled out

[36] Directorate-General for Communications Networks, Content and Technology. AI Act. 18 February 2025. Available at: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
[37] High-risk systems are defined in Article 6 of the AI Act: Classification Rules for High-Risk AI Systems
[38] Article 5, AI Act
[39] EDRi. *EU's AI Act fails to set gold standard for human rights*. 2024. Avaialbe at: https://edri.org/our-work/eu-ai-act-fails-to-set-gold-standard-for-human-rights/; BEUC The European Consumer Organization. *EU rules on AI lack punch to sufficiently protect consumers*. 12 September 2023. Available at: https://www.beuc.eu/press-releases/eu-rules-ai-lack-punch-sufficiently-protect-consumers
[40] Article 27.1, AI Act
[41] Annex III, 5(b) and (c), AI Act

template which will be developed by the AI Office. There may be exemptions of the notification requirement in cases pertaining to Article 46(1) which allow the MSA to temporarily forgo some requirements in cases they determine to be of exceptional reasons of public security or the protection of life and health of persons, environmental protection or the protection of key industrial and infrastructural assets. The FRIA needs to be kept up to date during the use of the deployed system.

1.2. FRIA procedure

In the current regulation, the FRIA is a self-assessment exercise by deployers with little procedural requirements aside from the completion of a questionnaire template to be created by the AI Office. Mantelero states that the final version of Article 27 has less detail of implementation than the original proposal, lacking “a mitigation plan, consideration of vulnerability, and a clear description of the components of this assessment.”[42] While this was intended to reduce the load on deployers, he argues that there is now a lack of specifications for deployers to follow. Article 27(1) states that the assessment must consist of descriptions of the intended purpose and use of the system, along with the specific risks of harms, including identifying groups of people particularly affected by these systems. Rights organizations have criticised the lack of explicit obligation to assess the acceptability of these risks or to prevent them–the obligation is only to assess measures to be taken when the risks materialize.[43]

[42] Mantelero, A. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. *Computer Law & Security Review*, *54*, 106020. 2024. Available at: https://doi.org/10.1016/j.clsr.2024.106020

[43] EDRi and AI coalition partners. EU’s AI Act fails to set gold standard for human rights. 3 April 2024. Available at: https://edri.org/our-work/eu-ai-act-fails-to-set-gold-standard-for-human-rights/; See also: IA Ciudadana. *Instalamos cajas transparentes frente al Ministerio de Escrivá para exigir que la legislación sobre IA tenga más luces que sombras*. 19 June 2024. Available at: https://iaciudadana.org/2024/06/19/instalamos-cajas-transparentes-frente-al-ministerio-de-escriva-para-exigir-que-la-legislacion-sobre-ia-tenga-mas-luces-que-sombras/; CECU. *Breve análisis del texto final del Reglamento de Inteligencia Artificial*. 2022. Available at: https://cecu.es/wp-content/uploads/2024/02/Breve-analisis-del-texto-final-del-Reglamento-de-Inteligencia-Artificial_CECU.pdf

1.3. Relationship of deployers and providers

Article 27 (2) allows deployers to rely on a FRIA previously completed by the provider if it reflects a "similar case." Under the AI Act, providers of high-risk systems are required to conduct a Conformity Assessment,[44] which requires providers to assess impacts on fundamental rights.[45] Mantelero argues that the relationship and information flow between AI providers and deployers regarding fundamental rights assessment is not properly framed in the AI Act; he asserts that it is not clear whether an impact assessment by the provider needs to be fully disclosed to the deployer. Article 27 specifically mentions Article 13 regarding what information must be provided to deployers by the providers, however, he argues, a specific disclosure obligation would have been more effective.[46]

Mantelero also highlights several reasons why a fundamental rights impact assessment conducted as part of the provider's Conformity Assessment (under Article 43 of the AI Act) may not fully meet the requirements of the deployer's FRIA (under Article 27). A major difference is the standards-based approach adopted by the EU legislator for the Conformity Assessment. Standardisation bodies may lack expertise in fundamental rights. Furthermore, the Conformity Assessments under Article 9 tends to focus on the AI product itself, overlooking the crucial contextual dimension where risks to individuals often manifest. AI systems are sociotechnical systems, and assessing their impact requires considering their interaction with users and with society, not just their internal design. The deployer's FRIA, addressed in Article 27 is intended to cover this relevant contextual component of risk management. Therefore, despite similarities and linkages between the two assessments, the standards-based and product-centric approach in

[44] Article 43 of the AI Act requires providers to assess conformity of systems to certain requirements of the AI Act. High-risk systems defined in Annex III points 2 to 8 require an internal assessment only. Biometric systems and those which are covered by Union harmonization legislation do require the involvement of a notified body.

[45] Article 9 of the AI Act defines the risk management system required for high-risk AI systems

[46] Mantelero, A., Guzmán, C., García, E., Ortiz, R., & Moro, M. A. FRIA model: Guide and use cases. *The Catalan Data Protection Authority*, 93. 28 January 2025. Available at: https://www.dpdenxarxa.cat/pluginfile.php/2468/mod_folder/content/0/FRIA_es_2.pdf

the Conformity Assessment means it may not sufficiently address the contextual and variable nature of fundamental rights impacts that the deployer encounters in real-world use.[47] This concern is one of the motivating factors behind calls for the inclusion of a FRIA by deployers.[48]

1.4. Enforcement regarding article 27 and fundamental rights impacts

National bodies must establish or designate and empower National Competent Authorities, consisting of one or more Notifying Authorities and Market Surveillance Authorities (MSA). These authorities are to supervise and enforce the rules for high-risk AI. Once the FRIA has been conducted, the deployer must notify the relevant MSA of the results, including submitting to them the filled-out questionnaire that the AI Office will develop. As of this writing the template is yet to be created, so the extent of documentation of the FRIA process required is still unknown. Additionally, national bodies must designate fundamental rights protection authorities under Article 77.

The deadline for establishing or designating National Competent Authorities is August 2, 2025, and for publishing a list of authorities protecting fundamental rights is November 2, 2024. As of February 18, 2025, 11 of the 27 member states had appointed at least one National Competent Authority and 15 member states had published a list of authorities protecting fundamental rights.[49]

[47] Id., see also Veale, M., & Borgesius, F. Z. Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112. 2021. Available at: https://doi.org/10.9785/cri-2021-220402; Christine Galvagna. *Inclusive AI governance*. 2023. Available at: https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/; Open Letter to the Spanish Presidency of the Council of the European Union:, Algorace, Algorights, CECU, Lafede.cat, Civio, Observatorio TAS, Institut de Drets Humans de Catalunya, & Éticas. Open Letter to the Spanish Presidency of the Council of the European Union: Ensuring the protection of fundamental rights on the AI Act. 17 May 2023. Available at: https://nextcloud.pangea.org/index.php/s/XGcgGRNssrwfD7j?dir=undefined&openfile=7442899

[48] Law 15/2022, of July 12, 2002, on equal treatment and non-discrimination. (n.d.). Retrieved 30 May 2025. Available at: https://www.boe.es/buscar/act.php?id=BOE-A-2022-11589

[49] Future of Life Institute. Overview of all AI Act National Implementation Plans. 2025. Available at: https://artificialintelligenceact.eu/national-implementation-plans/

1.4.1. Market surveillance authorities (MSA)

The authorised MSA are responsible for supervising high-risk AI systems once they are placed on the market or deployed into use. They have the power to impose sanctions for non-compliance, this includes the powers of investigation and correction outlined in of Regulation 2019/1020 on market surveillance and compliance of products[50] as well as powers to impose administrative fines defined in the AI Act, which include fines for non-compliances with prohibited AI practices or main obligations, or providing incorrect, incomplete or misleading information to notified bodies national competent authorities.[51]

According to Article 85 of the AI Act, an MSA "can act on their own initiative or upon receiving a complaint". Any person can make a complaint if they have grounds to believe that there has been an infringement related to the AI Act — it is not required that the complainant be personally affected. While Article 85 gives the right for anyone to make a complaint to an MSA, there is no specified requirement for the MSA to engage with the substance of a complaint. Rather, the requirement is that it is taken into account and "handled in line with the dedicated procedures established" by the MSA.

According to Article 79, If an MSA has "sufficient reason" to believe an AI system is risky,[52] it must evaluate the system's compliance with the AI Act, paying special attention to impacts on vulnerable groups. If the system violates fundamental rights, the MSA must also inform and cooperate with relevant national bodies (including the National Public Authorities Protecting Fundamental Rights discussed in Section 1.4.2 of this report). If non-compliance is confirmed, the MSA must require the operator to take

[50]Council of the European Parliament. Regulation—2019/1020—EN - EUR-Lex. 20 June 2019. Available at: https://eur-lex.europa.eu/eli/reg/2019/1020/oj/eng

[51] A & O Shearman. Zooming in on AI - #14: Enforcement of the AI Act. *A&O Shearman*. 2025. Available at: https://www.aoshearman.com/en/insights/ao-shearman-on-tech/zooming-in-on-ai-14-enforcement-of-the-ai-act

[52] Deployers do have a requirement to immediately notify the MSA and provider if they determine "they have reason to consider that the use of the high-risk AI system in accordance with the instructions may result in that AI system presenting a risk within the meaning of Article 79(1)" (Article 26(5) of the AI Act). The risks described in Article 79 are based on risks to health safety or fundamental rights as defined in Regulation (EU) 2019/1020 on the market surveillance and compliance of products.

corrective actions, such as bringing the system into compliance, withdrawing it, or recalling it. Should the operator fail to act in time, the MSA must take provisional steps to restrict or remove the system from the national market or promptly notify the Commission and other Member States if it is not restricted to its territory, including detailed information about the system, its risks, and the reasons for non-compliance.

There remain concerns about suitability of the MSAs for supervision and enforcement, including questions of independence and capacity. Thus, the effectiveness of MSAs in enforcing compliance remains unclear[53] and dependent on the situation at the national level.

Independence. Article 70(1) states that the authorities must be independent and able to exercise their powers “impartially and without bias”. However, there are concerns that some of the National MSAs already appointed are “politically governed or government dependent” as expressed in an open letter signed by 34 civil society organizations (CSOs).[54] A key concern is that many current governments are prioritising industry interests in AI adoption for political and or financial reasons, such as pressure to remain competitive in a global market. The letter calls for the European Commission to issue a clarifying statement regarding the requirement for independence of an MSA appointed to enforce the AI Act.

Capacity. Capacity concerns include funding and resources, as well as the profiles and expertise of the workers of the relevant agencies. A Data Protection Agency (DPA), for example, is a clear option for an existing authority that may be selected to act as an

[53] EDRi and AI coalition partners. EU’s AI Act fails to set gold standard for human rights. 3 April 2024. Available at: https://edri.org/our-work/eu-ai-act-fails-to-set-gold-standard-for-human-rights/ ; See also Veale, M., & Borgesius, F. Z. Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112. 2021. Available at: https://doi.org/10.9785/cri-2021-220402

[54] Agustin Reyna. Need for independent national market surveillance authorities under the AI Act—Commission. *BEUC – The European Consumer Organisation.* 25 June 2024. Available at: https://www.beuc.eu/letters/need-independent-national-market-surveillance-authorities-under-ai-act-commission

MSA. However, DPAs are experienced in issues of data privacy and security, but are not necessarily versed in the impacts of AI on other fundamental rights.

1.4.2. National public authorities protecting Fundamental Rights

Article 77 states that national authorities who oversee the respect of fundamental rights in relation to high-risk AI systems have the power to access any documentation created or maintained under the AI Act when such access is necessary for fulfilling their mandates within the limits of their jurisdiction. Information and documentation received must be treated with confidentiality in accordance with Article 78. If the documentation received "is insufficient to determine if there has been a breach of fundamental rights, the authority can request a testing of the AI system" to the MSA. The testing should be organised by the MSA, together with the requesting public authority, in a timely manner.[55] These national public authorities also have the right to be notified by the MSA in the case of a serious incident (defined in Article 3) related to an AI system. Member States are instructed to make a publicly available list of National Public Authorities Protecting Fundamental Rights and to keep the list up to date.[56,]

As Article 77 specifically refers to the use of high-risk AI systems, defined in Annex III of the AI Act, the text could potentially be interpreted in a way that it limits the powers of the National Public Authorities in regard to prohibited or general purpose AI systems.[57] However, Recital 157 of the AI Act explicitly states that the AI Act is "without prejudice to the competences, tasks, powers and independence of relevant national public authorities or bodies which supervise the application of Union law protecting fundamental rights" and that "[w]here necessary for their mandate, those national public authorities or bodies should also have access to any documentation created under this

[55] Article 77(3) of the AI Act

[56] Article 77(2) of the AI Act; See also: AUTHORITIES PROTECTING FUNDAMENTAL RIGHTS | SPAIN. Available at: https://digital.gob.es/dam/es/portalmtdfp/DigitalizacionIA/AuthoritiesFundamentalRights-Spain.pdf

[57] Wannes Ooms & Thomas Gils. Policy brief: Implementing the AI Act in Belgium - Scope of Application and Authorities. *Knowledge Centre Data and Society*.   December 2024. Available at: https://data-en-maatschappij.ai/en/publications/policy-brief-implementing-the-ai-act-in-be

Regulation." This may suggest that these bodies also have a role in the supervision of prohibited and general purpose systems.[58]

Further discussion of the role of Article 77 and the role of national public authorities protecting fundamental rights in AI use transparency is found in Section 4.1 of this report.

1.5. The Spanish market surveillance authority

Spain has opted to create a national agency to act as the single point of contact MSA for EU coordination, the Spanish Artificial Intelligence Supervisory Agency (The Agencia Española de Supervisión de la Inteligencia Artificial) (AESIA), according the draft bill on the implementation of the AI Act in Spain. The Spanish Council of Ministers approved the Statute creating AESIA August 22, 2023, making Spain the first country to create such an authority in Europe.[59]

There was criticism that civil society was not included in the design and establishment of the agency despite demands for such, that included a public letter signed by over 50 civil society organizations (CSOs).[60]

1.5.1. Mapping of corresponding MSA with AI application

The AESIA will share some functions with other bodies or authorities in certain areas or sectors. Additional authorities responsible for overseeing high-risk systems will be those already supervising the affected sector by default when it comes to products

[58] Id.

[59] Agencia Estatal Boletín Oficial del Estado. Real Decreto 729/2023, de 22 de Agosto, Por El Que Se Aprueba El Estatuto de La Agencia Española de Supervisión de Inteligencia Artificial, Pub. L. No. Real Decreto 729/2023, BOE-A-2023-18911 122289. 22 August 2023. Available at: https://www.boe.es/eli/es/rd/2023/08/22/729

[60] Arandia Jiménez, P. What to expect from Europe's first AI oversight agency. *AlgorithmWatch*. 1 February 2023. Available at: https://algorithmwatch.org/en/what-to-expect-from-europes-first-ai-oversight-agency/ Lafede.cat, Algorace, Fundación Eticas, CIVIO, Observatorio Trabaho, Algoritmos y Sociedad, & Komons y Algorights. Civil society organizations claim our role to the Spanish Agency for the Supervision of Artificial Intelligence (AESIA from its Spanish acronym). 14 September 2022. Available at: https://rightsinternationalspain.org/en/civil-society-organizations-claim-our-role-to-the-spanish-agency-for-the-supervision-of-artificial-intelligence-aesia-from-its-spanish-acronym/; IA Ciudadana. (2024, May 2). *Citizen AI Charter: AESIA must choose a suitable profile for its leadership*.

subject to harmonized legislation or as they are assigned under the Spanish draft national bill for the implementation of the AI Act, as shown in Table 1.

MARKET SURVEILLANCE AUTHORITY	HIGH-RISK AI APPLICATION
AESIA	Biometrics, except for those used exclusively to confirm a person's identity and when used for purposes of law enforcement, justice, democratic processes, or border control; Critical infrastructures; Education and vocational training; Employment and management of workers; Essential services and benefits, except: creditworthiness assessment or credit rating risk assessment and pricing in life and health insurance. Systems that are not high-risk or prohibited, but that fail to comply with transparency obligations or other obligations of the AI Act.
Spanish Data Protection Agency and DPA from autonomous communities	Biometrics when used for purposes of law enforcement or border control Law enforcement Migration and asylum management systems
CGPJ (General Council of the Judiciary)	AI systems in the administration of justice
Bank of Spain and CNMV (Spanish Securities Market Commission)	Evaluation of creditworthiness or credit scoring (See AI Act, Annex III.5b) CNMV - Capital markets systems
Directorate General of Insurance	Risk assessments and pricing of life and health insurance (See AI Act, Annex III.5c)
Central Electoral Board	AI systems used in democratic processes (See AI Act, Annex III 8.b)

Table 1: Current mapping of supervisory authority to system use under the Spanish draft bill.[61] This table does not consider the authorities for the prohibited practices under Art. 5 of the AI Act**.**

[61] Ministry of Digital Transformation and Civil Service. Anteproyecto de Ley para el Buen Uso y la Gobernanza de la Inteligencia Artificial. March 2025. Available at: https://avance.digital.gob.es/_layouts/15/HttpHandlerParticipacionPublicaAnexos.ashx?k=19128

1.6. Article 27 and frameworks at the European level

The AI Act and other technology regulations are not the only protections for consumers nor the only protection against fundamental rights abuses. Here, some relevant legislation, is addressed.

1.6.1. The Framework Convention on Artificial Intelligence

The Council of Europe's Framework Convention on Artificial Intelligence, opened for signature on 5 September 2024, is the first legally binding international treaty on AI. It establishes a common baseline to ensure that AI systems throughout their lifecycle uphold human rights, democracy, and the rule of law. Applicable to both public and private sector use (excluding national security, defence, and certain R&D contexts), the Convention obliges Parties to the Convention to adopt legislative and administrative measures. Parties must adopt measures to assess, prevent and mitigate risks and maintain independent oversight mechanisms. A "Conference of the Parties" will monitor implementation, with states required to report regularly. Open to Council of Europe members, the EU, and non-member states worldwide, the Convention complements EU regulation, including the AI Act, which will serve as the implementation tool within the Union.[62] The Convention introduces minimum standards from risk assessments, which should be ex ante and iterative, with particular attention to the context of deployment and changing real-world conditions.[63]

1.6.2. Charter of Fundamental Rights of the European Union

The Charter of Fundamental Rights of the European Union was proclaimed on December 7, 2000 and became legally binding December 1, 2009. This document sets out the fundamental rights and freedoms protected within the EU in 50 Articles organized into six chapters: Dignity, Freedoms, Equality, Solidarity, Citizen's Rights,

[62] Thomas Wahl. *Council of Europe Convention on Artificial Intelligence*. Eucrim. 26 September 2024. Available at: https://eucrim.eu/news/council-of-europe-convention-on-artificial-intelligence/

[63] Mantelero, A. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. *Computer Law & Security Review*, *54*, 106020. 2024. Available at: https://doi.org/10.1016/j.clsr.2024.106020

Justice.[64] The Charter applies to EU institutions and to member states when implementing EU law.

Due to the importance of fundamental rights and the strong protection they receive, the impact assessment must not compromise these rights in any way. Essentially, this means a Fundamental Rights Impact Assessment (FRIA) cannot be treated as a simple, after-the-fact formality. Instead, potential impacts on fundamental rights must be meaningfully considered and addressed throughout the AI system's design to ensure these protections are upheld.[65]

1.6.3. The General Data Protection Regulation (GDPR)

The GDPR is an EU law designed to protect the privacy and security of individuals' personal data, which went into effect on May 25, 2018.[66] Article 27(4) of the AI Act states that if any obligations of the article are already met through the data protection impact assessment (DPIA) done under the GDPR, the FRIA can complement the existing DPIA.

The DPIA of the GDPR is also an ex ante, rights-based, and iterative assessment.[67] However, the scope is different from the FRIA of Article 27 in that in practice the DPIA has a data protection focus, with rights to privacy and security of data being of primary interest, along with questions of discrimination, with a tendency to rely on legally protected categories as the focal point. Further discussion of the GDPR and the DPIA is found in Section 2.1 of this report.

[64] *Charter of fundamental rights of the European Union*. European Union. 2012. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012P/TXT

[65] *Mantelero, A. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. Computer Law & Security Review, 54, 106020. 2024. Available at: https://doi.org/10.1016/j.clsr.2024.106020*

[66] Parliament, E., & Council, of the. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). 2016. Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32016R0679

[67] Id.

1.7. Article 27 and frameworks at Spanish level[68]

1.7.1. Charter of Digital Rights

Spain's Charter of Digital Rights, released in July 2021, outlines a framework to protect individual and collective rights in the digital age. It serves as a guiding document for future legislation and public policy, aiming to adapt fundamental rights, such as freedom, equality, participation, and labour protections, to the digital environment. The Charter emphasizes a human-centred approach to technology, particularly in the development and use of AI, highlighting the right to non-discrimination, human dignity, and algorithmic impact assessments.[69]

1.7.2. Comprehensive law for equal treatment and non-discrimination

The law on equal treatment and non-discrimination entered into force July 2022. Article 23 of these laws establishes rules on AI and automated decision-making in the framework of the National Artificial Intelligence Strategy, the Digital Charter of Rights, and European initiatives on Artificial Intelligence. It mandates that public administration will promote the implementation of mechanisms that allow for taking into account criteria for minimizing bias, transparency, and accountability; mechanisms include design and training data. Assessments for determining potential discriminatory bias will be promoted. Additionally, public administrations and companies will promote the use of ethical, trustworthy, and human-rights-respecting AI and a seal of quality for algorithms will be promoted.[70]

1.7.3. General law for the protection of consumers and users of Spain

The General Law for the Protection of Consumers and Users of Spain (Ley General para la Defensa de los Consumidores y Usuarios) establishes the legal framework to

[68] For a comprehensive overview of AI related legislation in Spain, see the full text of the Spanish Digital Charter of Rights: Government of Spain. *Carta de derechos digitales*. 2021. Available at: https://citiesfordigitalrights.org/sites/default/files/140721-Carta_Derechos_Digitales_RedEs_compressed.pdf

[69] Id.

[70] Full text of the law is found here: https://www.boe.es/buscar/act.php?id=BOE-A-2022-11589

safeguard the rights of consumers in Spain.[71] The law establishes six fundamental rights for consumers and users in Spain. These include the right to protection against risks to their health or safety, as well as the protection of their legitimate economic and social interests. Consumers are also entitled to compensation and redress for any damage they suffer, and to receive accurate information about goods and services. Additionally, they have the right to participate in the development of legislation that directly affects them, and to have their rights protected through effective legal and administrative measures.[72]

1.7.4. The Spanish AI Strategy

Spain introduced the National Strategy for Artificial Intelligence (ENIA) on December 2, 2020 and released an updated strategy in 2024. The strategy promotes AI innovation while aiming to ensure responsible and ethical use. It is structured around three main pillars: Strengthening levers for AI development; Facilitating the expansion of AI in the public and private sectors by promoting innovation and cybersecurity; Developing transparent, responsible and humanistic AI. The strategy places notable emphasis on accelerating AI deployment and competitiveness. Specific initiatives include encouraging use of AI in the General State Administration for decision-making and administrative processes and allocating funds for supporting SMEs in the uptake of AI.[73]

[71] Government of Spain. *Royal legislative decree 1/2007, of November 16, approving the consolidated text of the general law for the defense of consumers and users and other complementary laws*. 2007. Available at: https://www.boe.es/buscar/doc.php?id=BOE-A-2007-20555

[72] Read more: Government of Spain. *Consumer rights, including product safety—Commercial practices and consumer rights—Starting, running and closing a business—Business—Your rights and obligations in the EU - Tu espacio europeo—Punto de Acceso General*. 4 April 2025. Available at: https://administracion.gob.es/pag_Home/en/Tu-espacio-europeo/derechos-obligaciones/empresas/inicio-gestion-cierre/practicas-comerciales/derechos-consumidores.html

[73] Ministry for Digital Transformation. 2024 artificial intelligence strategy. 2024. Available at: https://digital.gob.es/dam/en/portalmtdfp/DigitalizacionIA/1_DOSSIER_AI_ENGLISH_15_JULIO.pdf; *Key features of the Spanish AI Strategy for 2024: Reinforcement of the factors for the development of AI, promotion in public and private sectors and strengthening supervision for sustainable and ethical AI*. 21 May 2024. Available at: https://www.garrigues.com/en_GB/garrigues-digital/key-features-spanish-ai-strategy-2024-reinforcement-factors-development-ai

1.7.5. Draft bill on the implementation of the AI Act in Spain

The Draft Bill constitutes the provisions for the Implementation of the AI Act in Spain, and designates the Secretary of State for Digitalization and AI as the notifying authority, while the National Accreditation Body (ENAC) will handle monitoring and assessment of the notifying bodies.[74]

The draft bill sets narrow conditions for the lawful use of "real-time" remote biometric identification (RBI) systems in publicly accessible areas for law enforcement purposes. The use of RBI is prohibited under Article 5(1) of the AI Act, except in three scenarios: locating missing persons and victims, preventing terrorist threats, and identifying suspects of certain crimes – and it must be explicitly authorised in implementing national legislation. The Spanish draft bill authorises RBI use only for identifying individuals suspected of specified serious criminal offences, if judicially approved, an allowance which not all member states are expected to give. Any RBI use outside this sole exception is classified as a "very severe" infringement under the draft bill, which introduces a three-tier system for violations: minor, severe, and very severe.[75] As public authorities that use RBI need to conduct FRIAs, the Federación de Consumidores y Usuarios CECU, called for the FRIA documentation to be required as part of the request to the judicial authority.

The Spanish AI draft bill allows for product withdrawal, disconnection, or bans of AI systems in cases of very serious infringements or incidents causing significant harm, such as death—measures that can be triggered by a complaint (per Article 85 of the AI

[74] Consejo de Ministros. *El Gobierno da luz verde al anteproyecto de ley para un uso ético, inclusivo y beneficioso de la Inteligencia Artificial*. Ministerio para La Transformación Digital y de la Función Pública. 11 March 2025. Available at: https://www.administracionpublicadigital.es/normativas/2025/03/el-gobierno-da-luz-verde-al-anteproyecto-de-ley-para-un-uso-etico-de-la-ia; The draft bill: Ministry of Digital Transformation and Civil Service. Anteproyecto de Ley para el Buen Uso y la Gobernanza de la Inteligencia Artificial. March 2025. Available at: https://avance.digital.gob.es/_layouts/15/HttpHandlerParticipacionPublicaAnexos.ashx?k=19128https://avance.digital.gob.es/_layouts/15/HttpHandlerParticipacionPublicaAnexos.ashx?k=19128

[75] Cabrera, L. L., Duprat-Macabies, A., & Maier, M. CDT Europe's AI Bulletin: March 2025. *Center for Democracy and Technology*. 26 March 2025. Available at: https://cdt.org/insights/cdt-europes-ai-bulletin-march-2025/

Act). It treats infringements regarding deepfakes, including ultra-spoofing and disinformation, as serious infringements, applicable to both providers and deployers.[76]

Article 20 of the draft bill recognizes failure of the deployers (pubic or private) to comply with the obligation under Article 27 of the AI Act as a serious infringement. In comments to the draft bill, CECU also recommended that the compliance of this obligation should be evaluated by taking into account whether the performed FRIA has included substantial evaluation of risks, a criterion for which should be the participation of affected groups , or their representatives.

2. Accountability mechanisms and AI: an overview of impact assessments in theory and in practice

This section introduces the concept of an algorithmic accountability mechanism and examples of the different types of accountability mechanisms defined in the context of responsible AI development and deployment. It then has a closer look at algorithmic impact assessments, with a special focus on how these mechanisms are present in European Union regulation related to the AI Act, such as the General Data Protection Regulation and Digital Services Act.

Following this, existing methodologies for Fundamental Rights Impact Assessments for AI are described, with a focus on three models currently positioned as potential gold standards. Impact assessments as they manifest in the private sector, particularly in finance and baking, are then examined.

[76] Pablo García Mexía, Rebeca Oriol, & Iván Pinheiro. Alert: Spain implements the European AI regulation ahead of schedule. *Herbert Smith Freehills Notes*. 2 February 2025. Available at: https://www.herbertsmithfreehills.com/th/notes/madrid/2025-posts/alert-spain-implements-the-european-ai-regulation-ahead-of-schedule-marzo-2026

KEY CONCEPT

ACCOUNTABILITY MECHANISM

Accountability mechanisms[77] are a set of mechanisms that exist with the aim of ensuring the responsible actors that are building, procuring, and using algorithms are answerable, can justify the use of the algorithmic system, and are capable of facing consequences for their use.

Types of Accountability Mechanisms

- Principles and guidelines: Policy documents providing non-binding normative guidance.
- Prohibitions and moratoria: Banning or prohibition of use of particular kinds of ‘high risk’ algorithmic systems.
- Public transparency: Transparency mechanisms that provide information about algorithmic systems to the general public.
- Impact Assessments: mechanisms intended for actors building, procuring, and using AI systems to better understand, categorize and respond to potential harms or risks of AI systems before deployment.
- Audits and Inspections: mechanisms intended to provide insights into the functioning of an algorithmic system.

[77] Reisman, Dillon, Jason Schultz, Crawford Kate, & Whittaker Meredith. “Algorithmic Impact Assessments Report: A Practical Framework for Public Agency Accountability.” AI Now Institute. 9 April 2018. Available at: https://www.opengovpartnership.org/wp-content/uploads/2021/08/executive-summary-algorithmic-accountability.pdf

2.1. Impact assessments in existing AI regulation

Regulation overseeing the development and deployment of artificial intelligence systems is still in its early stages. This means that where regulation exists, empirical proof of its efficacy in terms of compliance, as well as in terms of minimization of algorithmic harms, is scarce.

Outside AI-specific regulation in the European Union, accountability mechanisms that oversee algorithmic systems and data-processing activities exist in analogous regulation, i.e., General Data Protection Regulation and Digital Services Act. On top of this, a number of best practice guidance, government directives and charters have also been proposed with the aim to govern AI systems and to facilitate the operationalization of risk and impact assessment in this context.

2.1.1. Algorithmic impact assessments

Borrowing from environmental protections, human rights,[78] and data protection and privacy[79], algorithmic impact assessments (AIAs) are often self-assessment tools recommended as a way to prompt reflection on the intended use of an AI system and to proactively elaborate mitigation plans for possible harmful impacts associated to an AI system before it is deployed (ex-ante); followed up by their recurrent revision across time (post-ante) to make sure the initial results obtained remain current, especially if there have been any changes in the way the AI system is used.[80]

Academic research and international organisational efforts have positioned AIAs as recommended best practices for AI developers and deployers. This is of relevance,

[78] Read more: The World Bank. Human Rights Impact Assessments: a review of the literature, differences with other forms of assessments and relevance for development. 2013. Available at: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/834611524474505865/human-rights-impact-assessments-a-review-of-the-literature-differences-with-other-forms-of-assessments-and-relevance-for-development

[79] Leonardo Horn Iwaya, Ala Sarah Alaqra, Marit Hansen & Simone Fischer-Hübner. Privacy impact assessments in the wild: A scoping review, Array, Volume 23, 2024. Available at: https://doi.org/10.1016/j.array.2024.100356.

[80] Reisman, Dillon, Jason Schultz, Crawford Kate, & Whittaker Meredith. "Algorithmic Impact Assessments Report: A Practical Framework for Public Agency Accountability." AI Now Institute. 9 April 2018. Available at: https://ainowinstitute.org/publications/algorithmic-impact-assessments-report-2

given that in a legal context, AIAs and similar assessments often fall into what is deemed to be soft-law, denoting non-binding implications. This aspect changes when AIAs come to be prescribed by regulation, which in turn means that AIAs are now part of legally binding requirements; for that reason, it is important to emphasize that without some element of real legal enforceability, the efficacy of AIA's will vary.[81]

In regard to the execution of AIAs, recommendations also call for AIAs to be expert-based exercises, where different types of expertise come together, for example, business representatives, AI designers, fundamental rights lawyers, affected groups and or their representatives. Other recommendations include that the results of these assessments be dutifully documented and accessible, in order to promote transparency and facilitate external oversight.

This particular type of algorithmic accountability mechanism sets out to serve as a process where the deployer of an AI system meaningfully engages in identifying, categorizing, understanding, and mitigating for potential harms or risks derived from the use of the AI system such as bias, discrimination, and fairness. Moreover, AIAs should support the creation of appropriate governance and oversight mechanisms by AI deployers in line with the particular characteristics of the assessed AI system.

Theoretically, algorithmic impact assessments also serve as a mechanism for including the participation of affected groups. The idea is that their feedback, lived experiences, opinions, and concerns can influence and are incorporated in the results of the assessments. Not only that, but in opening these processes to members of affected groups, the most rigorous implementation of an algorithmic impact assessment should habilitate mechanisms to challenge the results of the AIA, and to challenge AI deployers when their AI systems fail to comply with the results and recommendations of the AIA once upon deployment.

[81] Read more: Reuben Binns. Data protection impact assessments: a meta-regulatory approach. International Data Privacy Law 7 (1): 22-35. 2017. Available at: https://doi.org/10.1093/idpl/ipw027

In practice, it is common for algorithmic impact assessments to be used as an internal self-assessment exercise for companies and public administration offices, with limited or zero engagement with affected groups. Moreover, a survey of 38 documents defining algorithmic impact assessments frameworks specifically concerned with AI systems, found AIAs are seldom prescriptive and often lack standardization.[82] Other types of concerns include the lack of preparedness or topical knowledge of teams engaging in AIAs processes, jeopardizing the results of these as they depend on the ability of the deployers to provide "substantive guidance and organisational support"[83] to these parties, as full discretion is given to the deployers as they assess whether their AI systems may or may not elicit harm and to what degree. The worst case scenario of this dynamic is that deployers can end up completing these impact assessments without addressing any harmful implications associated with their systems.

Furthermore, in an ecosystem where there is also very little access to the results of these assessments by external stakeholders, such as civil society organizations and rights groups, researchers, journalists, and affected people or their representatives, two problematics arise:

- a reliance on self-disclosed documentation, especially in the case of private companies, making it more difficult to gain a clear and real understanding of what type and how many AI systems are in deployment; and
- the impediment to an objective appraisal of how AI deployers are performing against their own self-assessments.[84]

[82] Bernd Carsten Stahl et al. "A Systematic Review of Artificial Intelligence Impact Assessments." Artificial Intelligence Review56(11):12799–831. 2023. Available at: https://doi.org/10.1007/s10462-023-10420-8
[83] Ashar, A., Ginena, K., Cipollone, M., Barreto, R., & Cramer, H. Algorithmic Impact Assessments at Scale: Practitioners' Challenges and Needs. *Journal of Online Trust and Safety*, *2*(4). 2024. Available at: https://doi.org/10.54501/jots.v2i4.206
[84] Lara Groves et al. Auditing Work: Exploring the New York City algorithmic bias audit regime. In The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil. 2024. Available at: https://doi.org/10.1145/3630106.3658959

In the following subsections, the most common components of algorithmic impact assessments will be discussed, specifically those designed to assess the impact of AI systems on fundamental rights.

EXAMPLE CASE 1: Six Years of the Canadian Algorithmic Impacts Assessment[85]

Canada's AI regulation strategy saw the instauration of the Canadian Directive's on Automated Decision-Making (ADMs), which went into effect on April 2019.[86] One of the characteristics of this directive is that it makes it a requirement to complete an Algorithmic Impact Assessment (AIA) prior to the production of any automated decision-making system (ex-ante) used for a government application. The final results of these AIAs are to be released in an accessible format in both official languages, i.e., English and French, on the Canadian Open Government Portal.[87] At the time of writing this report, there are 26 records.

The Canadian AIA proposes a scoring algorithm that correlates the level of risk attributed to an automated decision-making system, to the severity of the requirements for its use. The AIA algorithm assigns points to questionnaire answers, a higher number of points implies a higher level of impact; there are four levels, ranging from Level I (little impact) to Level IV (very high impact), systems with a level II or higher require their AIA undergoes a peer review process. A per the Canadian

[85] Canada's Algorithmic Impact Assessment Tool. Government of Canada. 2024. Available at: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html

[86] Directive on Automated Decision-Making. Government of Canada. 2024. Available at: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592

[87] Canada's Open Government Portal. Government of Canada. 2024. Available at: https://search.open.canada.ca/opendata/?collection=aia&page=1&sort=date_modified+desc

Government, “peer review is a quality assurance mechanism in which the project is subject to scrutiny by experts in the relevant domain.”[88]

Since the launch of the Canadian AIA, the scoring algorithm and the questionnaire have been reviewed and amended. New questions have been added, and the algorithm has been recalibrated, penalizing more the use of certain types of personal data. Further, more questions have been added to evaluate the impact of ADMs on different populations according to biological sex, gender, age, disability, race, religion, sexual orientation, among others.

Selbst [89] elaborates on some of the shortcomings of the Canadian AIA model. First, questionnaires are inherently constricting, more so if they have a fixed number of questions; he goes on to qualify the questions as generic in nature, and is critical of the inclusion of close-ended or multiple choice format questions. Opting to forego an open-ended formatted question, limits the possible nuance of the responses, and hinders any attempt to elaborate on a possible justification or explanation that could contribute towards a meaningful assessment of a particular issue.

Moreover, the Canadian AIA model contemplates the involvement of designers at an early stage of the process (ex-ante); ideally, effective AIAs should consider the whole lifecycle of AI development, where other stakeholders, e.g., engineers, can be probed and expected to rationalize over pressures, choices, and trade-offs.[90]

[88] Guide to Peer-Review: Directive on Automated Decision-Making. Government of Canada. 2024. Available at: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-peer-review-automated-decision-systems.html

[89] Selbst, A. D. An Institutional View of Algorithmic Impact Assessments. Harvard Journal of Law & Technology (Harvard JOLT), 35, 117. 2021. Available at: https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf

[90] Id.

Key takeaways derived from a report by the *World Privacy Forum* in conversation with the Canadian AIA oversight team[91]

- Risk quantification can in itself give rise to the risk of misinterpretation of the obtained results of the AIA. They can also contribute towards the oversimplification of real-world complex problems.
- AIA tools should be released with thorough documentation and guidance for use to avoid misuse.
- The assessment should be periodically reviewed for efficacy, and their content updated for real-world changes and nuance.
- The robustness of the results obtained depend on the composition, knowledge and predisposition of the teams completing the AIAs. While the perceived calibre of the AIAs has improved over the past six years, the quality of the content and results vary across all AIAs on record.
- Multidisciplinary teams contribute to improving the quality of AIAs results.

Canada's AIAs and Real-world Impact

AI systems are becoming ubiquitous in immigration procedures.[92] William Tao, an immigration and refugee lawyer and founder of Heron Law Offices in British Columbia,[93] is cognizant of this. He has seen how these systems are able to influence high-stake decisions on immigration cases that can dictate whether immigrants or

[91] Kate Kaye. AI Governance on the Ground: Canada's Algorithmic Impact Assessment Process and Algorithm has evolved. World Privacy Forum. 2024.
Available at: https://www.worldprivacyforum.org/2024/08/ai-governance-on-the-ground-series-canada/

[92] Read more: Petra Molnar & Lex Gill. Bots at the Gate: A Human Rights Analysis of Automatic Decision-Making in Canada's Immigration and Refugee System. University of Toronto's International Human Rights Program (IHRP), Citizen Lab, and Information Technology, Transparency, and Transformation Lab. 2018. Available at: https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf

[93] Heron Law Office. Will Tao has Been Cited in the World Privacy Forum on AI Governance in Immigration. 2024. Available at:
https://heronlaw.ca/will-tao-has-been-cited-in-the-world-privacy-forum-on-ai-governance-in-immigration/

refugees can legally work in Canada, or even if they should be separated from their spouses and/or children.
For years, legal watchdogs, such as Tao, have tried to gain access to meaningful information on algorithmic systems used by the Immigration, Refugees and Citizenship Canada (IRCC) to little avail.

Under the context of the revised AIA model, and requirement to make the results public, lawyers like Tao were able to get a tangible example as to how one of these AI systems works, gaining critical information as to what type of predictive variables are used to determine if an applicant should be granted or not a work permit in Canada.[94]

NOTE: For some time, Canada was at the forefront of the AI regulation conversation.[95] Following the above-mentioned Directive, the next step was the anticipated and also criticised[96] Artificial Intelligence and Data Act (AIDA). Given a turbulent political climate, the draft bill failed to become a law, leaving behind lessons learned on the process, and a future of possibilities in regulatory matters.[97]

[94] International Experience Canada Work Permit Eligibility Model - Gender-Based Analysis Plus - Automated triage and positive eligibility determinations of International Experience Canada Work Permit Applications. Government of Canada. 2024.
Available at: https://open.canada.ca/data/en/info/b4a417f7-5040-4328-9863-bb8bbb8568c3

[95] Government of Canada. Canada moves toward safe and responsible artificial intelligence. 6 March 2025. Available at: https://www.canada.ca/en/innovation-science-economic-development/news/2025/03/canada-moves-toward-safe-and-responsible-artificial-intelligence.html

[96] Read more: Teresa Scassa. Regulating Ai In Canada: A Critical Look At The Proposed Artificial Intelligence And Data Act. The Canadian Bar Review. 2023. Available at: https://cbr.cba.org/index.php/cbr/article/view/4817/4539

[97] Read more: Blair Attard-Frost. The Death of Canada's Artificial Intelligence and Data Act: What Happened, and What's Next for AI Regulation in Canada?. The Montreal AI Ethics Institute. 17 January 2025. Available at: https://montrealethics.ai/the-death-of-canadas-artificial-intelligence-and-data-act-what-happened-and-whats-next-for-ai-regulation-in-canada/

2.1.2. Accountability mechanisms in the GDPR and the DSA

2.1.2.1. Data Protection Impact Assessments under the GDPR

The General Data Protection Regulation (GDPR), adopted in 2018, is the risk-based European Union data privacy and security law, that imposes obligations on data controllers anywhere in the world that target or collect personal data related to data subjects based in the EU.[98]

KEY CONCEPTS
DATA CONTROLLERS
Article 4.730 of the GDPR defines data controllers as the 'person' who determines the purposes and means of the processing of personal data, and the scope of their obligations, as defined by Article 24, include, among others, to "implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation."
DATA SUBJECTS
Article 4.1. of the GDPR defines data subject as an "identified or identifiable natural person" from whom or about whom information is collected. A company or organization cannot be a data subject.
PERSONAL DATA
Article 4.1. of the GDPR defines personal data as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by

[98] Ben Wolford. What is GDPR, the EU's new data protection law? GDPR.eu. Available at: https://gdpr.eu/what-is-gdpr/

reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".

The GPDR establishes requirements for data controllers, and defines the obligation for them to implement appropriate measures that demonstrate the elaboration of risk management plans in relation to personal data processing operations in terms of perceived risk to the rights and freedoms of individuals.

In the context of this report, a relevant mechanism defined in the GDPR are data protection impact assessments (DPIA).[99] The DPIA Guidelines, written by the Article 29 Data Protection Working Party, predating the European Data Protection Board (EDPB),[100] define this mechanism as "a process designed to describe the processing, asses its necessity and proportionality,[101] and help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data by assessing them and determining the measures to address them."

Prior to the implementation of the AI Act, legal scholars[102] and data protection authorities[103] encouraged for DPIAs to be seen as mechanisms that elicit algorithmic accountability, specifically when an AI processes personal data, conducts profiling on

[99] Article 35 of the GDPR

[100] European Data Protection Board. Guidelines and Recommendations on Data Protection impact assessments, High risk processing. 2019. Available at: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/data-protection-impact-assessments-high-risk-processing_en

[101] Read more: European Data Protection Supervisor. Assessing the necessity of measures that limit the fundamental right to the protection of personal data. 11 April 2017. Available at: https://www.edps.europa.eu/data-protection/our-work/publications/papers/necessity-toolkit_en

[102] Margot E Kaminski & Gianclaudio Malgieri, Algorithmic impact assessments under the GDPR: producing multi-layered explanations, International Data Privacy Law, Volume 11, Issue 2, Pages 125–144. April 2021. Available at: https://doi.org/10.1093/idpl/ipaa020

[103] Agencia Española de Protección de Datos. Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. February 2020. Available at: https://www.aepd.es/guias/adecuacion-rgpd-ia.pdf

natural persons or makes decisions regarding such natural persons, as these activities are subjected to the provisions laid down by the GDPR.

In a broader sense, under the regulation, DPIAs are encouraged as general good practice for all data controllers, however are only mandatory when "a type of data processing activity is likely to result in a high risk"[104] to fundamental rights,[105] with non-compliance resulting in penalties ranging from fines to prohibition of processing operation. Making DPIAs regulatory requirements leads to a conversation as to how a "self-regulation tool becomes one of meta-regulation." As per legal theory, meta-regulation allows actors with legal obligations to manage themselves, while being subjected to legal consequences in case of non-compliance. The implication here is that enforcement and external oversight mechanisms are in place to guarantee efficacy.[106]

DPIAs are also seen as expert-based assessments that should mimic methodologies used in risk assessment and risk management theory, which bring to the table experts across different areas. Further, they also follow an ex-ante approach, with the presupposition that data controllers are continuously assessing the risks created by their processing operations in order to identify when a type of data processing can be subjected to this requirement.

On top of the initial guidelines stated by the Working Party, national data protection authorities also provide guidance and habilitate communication mechanisms to impart advice on a case by case basis pertaining to data processing operations and DPIAs. As per Article 36 of the GDPR, the supervisory authority must be consulted when a high risk is identified during the elaboration of a DPIA prior to carrying out the data processing

[104] See page 9 of Article 29 Working Group's DPIA guidelines. There, criteria to guide the identification of a data processing operation likely to result in a high risk to fundamental rights are defined. Available at: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/data-protection-impact-assessments-high-risk-processing_en

[105] Article 35 of the GDPR

[106] Read more: Reuben Binns. Data protection impact assessments: a meta-regulatory approach. International Data Privacy Law 7 (1): 22-35. 2017. Available at: https://doi.org/10.1093/idpl/ipw027

operation. Along with this, a number of templates that support the operationalization of self-assessments have also been introduced. Below the minimum four elements set out in Article 35:

1. A systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;
2. An assessment of the necessity and proportionality of the processing operations in relation to the purposes;
3. An assessment of the risks to the rights and freedoms of data subjects; and
4. The measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with - the GDPR - taking into account the rights and legitimate interests of data subjects and other persons concerned.

A CLOSER LOOK 1: DPIAS IMPLEMENTATION OVER THE PAST 7 YEARS

Shortcomings associated to the implementation of the DPIAs focus on:

- The minimalistic approach some national data protection authorities and data controllers have taken when it comes to their implementation. A consequence of this is the risk of turning these self-assessments into a check-boxing exercise. Minimalistic approaches and simplification can also contribute to the misinterpretation of the purpose of the assessment altogether. The Spanish Data Protection Office has reported in their Annuals how the quality of DPIAs they receive are evidence of how these assessments are seen as simple formalities, with data controllers mostly focusing on meeting legal requirements associated with data protection, while failing to consider the assessment of other fundamental rights.[107]

- Perceived lack of expertise or comprehension of the law by those in charge of performing risk assessments, as well as by those entrusted with overseeing compliance with the GDPR in general (the Data Protection Officers),[108] by private companies and public administrations. As per the Annual Reports published by the Spanish Agency of Data Protection, as recent as 2023, they continue to report the seemingly incorrect interpretation of the legal text by data controllers,[109] and express doubt whether they are receiving proper advice on these matters.[110]
- Inconsistent or lack of enforcement, with a focus on monetary sanctions, e.g., fines.[111] Temporary suspension of processing activity is seen as a last resort measure. Other problems associated to this are the self-reported lack of resources at a national data protection authority level.[112] The NGO *noyb* published this January a comprehensive analysis of statistics by the European Data Protection Board on fines and budgets over the 2018-2023 period.[113] Lastly, inefficient cross border

[107] Agencia Española de Protección de Datos. Memoria 2021, p.51. Available at: https://www.aepd.es/memorias/memoria-anual-2021-aepd.pdf

[108] Agencia Española de Protección de Datos. Memoria 2022, p. 12. Available at: https://www.aepd.es/memorias/memoria-aepd-2022.pdf

[109] Agencia Española de Protección de Datos. Memoria 2023, p. 42. Available at: https://www.aepd.es/memorias/memoria-aepd-2023.pdf

[110] Agencia Española de Protección de Datos. Memoria 2021, p.10. Available at: https://www.aepd.es/memorias/memoria-anual-2021-aepd.pdf

[111] See more: an interactive map of Data Protection Authority Activity across the EU (2018-2023). Available at: https://www.datawrapper.de/_/6KLw9/

[112] European Data Protection Board. Overview on resources made available by Member States to the Data Protection Supervisory Authorities. 2022. Available at: https://www.edpb.europa.eu/system/files/2022-09/edpb_overviewresourcesmade_availablebymemberstatestosas2022_en.pdf

[113] *noyd.* Data Protection Day: Only 1.3% of cases before EU DPAs result in a fine. 28 January 2025. Available at: https://noyb.eu/en/data-protection-day-only-13-cases-eu-dpas-result-fine

cooperation that can increase investigatory powers of the authorities is also seen as an impediment to enforcement.

- The regulation recommends the participation of stakeholders and affected people and groups in processes carried out to complete DPIAs. However, there is little to no evidence of this being an adopted practice, and if it is, to what degree. All this highlights general concerns associated with self-regulation initiatives in general.
- While they are mandatory in cases stipulated by the law, there is no obligation to publish the results of these. The absence of this requirement in the regulation contributes to there not being a common baseline understanding of what content DPIAs should constitute and its quality. Moreover, this poses a real barrier to independent and external oversight as well as to transparency efforts sought by civil society organizations, researchers, journalists, etc., investigating potential harmful activities concerning personal data processing and AI systems.
    - As of 2023, a research group compiled manually 130 publicly available DPIAs to create a public DPIA repository.[114] Their objective is to turn this project into a community-supported and maintained effort, to contribute to the better understanding of these processes, and to provide empirical proof of their efficacy as accountability mechanisms in the absence of enforcement.[115]

[114] DPIA Repository Annotated collection of public DPIAs. 2024. Available at: https://dpiarepository.distrinet-research.be/

[115] Sion, L., Van Landuyt, D. &Joosen, W. A DPIA Repository for Interdisciplinary Data Protection Research. In: Garcia-Alfaro, J., *et al*. Computer Security. ESORICS 2024 International Workshops. ESORICS 2024. Lecture Notes in Computer Science, vol 15263. Springer, Cham. 2025. Available at: https://doi.org/10.1007/978-3-031-82349-7_13

2.1.2.2. Risk assessment under the Digital Services Act

The European Union's Digital Services Act went into effect in August 2023. It is a first of its kind regulation, as it seeks to address illegal content, transparent advertising and disinformation in EU online environments.

KEY CONCEPTS
VERY LARGE ONLINE PLATFORMS (VLOPs)
Platforms that have more than 45 million users per month in the EU.
VERY LARGE ONLINE SEARCH ENGINES (VLOSEs)
Search engines that have more than 45 million users per month in the EU.

Relevant to this report, Article 34 of the DSA stipulates that "providers of very large online platforms (VLOPs) and of very large online search engines (VLOSEs)[116] shall diligently identify, analyse, and assess any systemic risks in the [European] Union stemming from the design or functioning of their service and its related systems, including algorithmic systems, or from the use made of their services."[117] While there is an ongoing discussion by researchers, and other experts, on what constitutes the

[116] European Union. Supervision of the designated very large online platforms and search engines under DSA. 2025. Available at: https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses

[117] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R2065

concept of systemic risk, and how they should be researched,[118] the DSA specifies the following four categories of systemic risks:[119]

- The dissemination of illegal content;
- Negative effects for the exercise of fundamental rights;
- Negative effects on civic discourse and electoral processes, and public security;
- Negative effects in relation to gender-based violence, the protection of public health and minors and serious negative consequences to the person's physical and mental well-being.

Apart from this, there are no further provisions by the DSA that establish specific guidelines that set harmonized rules for completing a risk assessment (RAs).[120]

In terms of participation, the DSA recital 90, indicates that VLOPs and VLOSEs must perform the completion of risk assessments and mitigation plans based on "the best available information and scientific insights and that they test their assumptions with the groups most affected by the risks and the measures they take". Therefore, very much like the GDPR, the recommendation is to seek the "involvement of representatives of the recipients of the service, representatives of groups potentially affected by their services, independent experts and civil society organizations." [121]

[118] Oliver Marsh. Researching Systemic Risks under the Digital Services Act. 26 July 2024. Available at: https://algorithmwatch.org/en/wp-content/uploads/2024/08/AlgorithmWatch-Researching-Systemic-Risks-under-the-DSA-240726_v2.pdf

[119] Global Network Initiative. Implementing risk assessments under the Digital Services Act. 2023. Available at: https://globalnetworkinitiative.org/wp-content/uploads/2023/06/Discussion-summary-%E2%80%93-GNI-and-DTSP-workshops-on-implementing-risk-assessments-under-the-DSA-June-2023.pdf

[120] Eliška Pírková (Access Now), Marlena Wisniak & Karolina Iwańska (European Center for Not-for-Profit Law). Towards Meaningful Fundamental Rights Impact Assessments Under the DSA. Available at: https://www.accessnow.org/wp-content/uploads/2023/09/DSA-FRIA-joint-policy-paper-September-2023.pdf

[121] Official Journal of the European Union. REGULATION (EU) 2022/2065 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)
Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065#d1e4142-1-1

As opposed to the GDPR's DPIAs, Article 42(4) of the DSA[122] establishes that VLOPs and VLOSEs must publish, on an annual basis, extensive documentation of the risk assessments, along with the specific mitigation measures implemented. Furthermore, in accordance with these risk assessments, there is also the obligation to undergo compliance audits by an independent auditing organization. The audit reports and audit implementation reports must also be published.[123] At the moment of writing this report, there is no centralized database to deposit these reports. For the time being, independent initiatives have been created to facilitate their compilation, such as the Tremau T&S Research Team's DSA Database[124] and DSA: Risk Assessment & Audit Database,[125] maintained by independent researcher Alexander Hohlfeld.

A CLOSER LOOK 2: FIRST IMPRESSIONS FROM FIRST ITERATION OF RAs

The DSA Civil Society Coordination Group, an informal coalition of civil society organizations across the European Union, the Recommender Systems Taskforce and People vs Big Tech published their initial analysis on the first round of Risk Assessments in March 2025.[126] The reporting focused on the following companies: Google (Search and YouTube); Meta (Facebook and Instagram); TikTok and X. Additionally, the DSA Observatory[127] also prepares periodical analysis of the implementation of the DSA and including the first rollout of RAs.

[122] See also Section 5, Chapter III of the DSA for specific DSA obligations for VLOPs and VLOSEs.

[123] European Union. Q&A on risk assessment reports, audit reports and audit implementation reports under DSA. 2025. Available at: https://digital-strategy.ec.europa.eu/en/faqs/qa-risk-assessment-reports-audit-reports-and-audit-implementation-reports-under-dsa

[124] DSA Database. Available at: https://tremau.com/resources/dsa-database/

[125] DSA: Risk Assessment & Audit Database. Available at: https://docs.google.com/spreadsheets/d/12hJWpCFmHJMQOlz1qkd6OgGsMW82YcsWgJHXD7BHVps/edit?gid=0#gid=0

[126] Center for Democracy & Technology. Assessment Reports: An Initial Feedback Brief. Available at: https://cdt.org/insights/dsa-civil-society-coordination-group-publishes-an-initial-analysis-of-the-major-online-platforms-risks-analysis-reports/

[127] Read more: Digital Services Act (DSA) Observatory. Available at: https://dsa-observatory.eu/

The identified gaps and recommendations on the assessed RAs focus on how these reports fall short as transparency tools, and specifically allude to:

- Omission or failure to mitigate against specific systemic risks. Concerning fundamental rights in particular, "several providers have either failed to conduct broader assessments or have chosen not to specify how this assessment was conducted, and their resulting conclusions."
- Little new data on mitigation effectiveness, in particular concerning the lack of "insightful metrics" to quantify harm, and by extension, omission to include "quantifiable mitigation metrics." The absence of metrics makes it difficult to track and compare the effectiveness of a mitigation strategy systematically.
- No evidence of meaningful engagement with experts and affected groups. For example, the coalition and their partners consist of about 200 global, local, and European CSOs and academic researchers; they have self-reported that none of them were consulted in the process of conducting the RAs nor the reports. External links citing independent research were also scarce across the analysed documents.
- Impediments to scaled quantitative analysis. Given the lack of harmonization and templates, most of the reports were presented as PDFs with no encoded machine-readability that could facilitate data extraction and analysis.

2.2. Existing tools and frameworks for FRIAs for AI

There are numerous FRIA-for-AI tools, models and templates, many of which overlap in terms of format and content. The Algorithm Audit has conducted a Comparative Review of 10 Fundamental Rights Impact Assessments (FRIAs),[128] measuring the frameworks

[128] Algorithm Audit. *A comparative review of 10 Fundamental Rights Impact Assessments (FRIA) for AI-systems*. 2024. Available at: https://algorithmaudit.eu/knowledge-platform/knowledge-base/comparative_review_10_frias/

against twelve legal, technical, organisational, and social criteria. The review found that many FRIAs lack key legal instruments for normative assessment, such as the proportionality test. In terms of technical criteria, the assessments were lacking statistical analysis, such as hyperparameter sensitivity and statistical hypothesis testing. Most frameworks were found to adopt a technocratic approach that excludes meaningful public engagement, limiting citizen involvement in shaping the normative decisions behind data modelling. The report calls for more interdisciplinary, rigorous, and participatory FRIA methods to ensure AI systems respect fundamental rights.

The Algorithm Audit report also determines that the HUDEIRA methodology from the Committee on Artificial Intelligence of the Council of Europe and the Alan Turing Institute met eleven of the twelve criteria points (only losing a point in the technical category); the Dutch Fundamental Rights Algorithmic Impact Assessment (FRAIA) met ten of the twelve criteria (lacking the same technical element as well as one stakeholder inclusion related element). Here, these two models as well as a recent addition, purposely built for the FRIA process for the Catalan region, will be looked at in closer detail.

2.2.1. Case Study: Dutch FRAIA

The Fundamental Rights and Algorithms Impact Assessment (FRAIA) is a tool developed by Utrecht University for the Dutch Ministry of the Interior and Kingdom Relations to identify the impact of specific algorithms on fundamental rights.[129] It is intended to serve as a discussion and decision-making instrument for government organisations considering the development, procurement, or use of an algorithm, or for evaluating those already in use. The tool takes the form of a 99-page interactive document, providing the option of filling the questions in directly on the document.

[129] FRAIA tool

The questions are accompanied by explanations, guidance and links to resources. This includes indicating, with red print, responses that may entail a serious risk. The questions that are structured into four main parts:

1) Why: intended effects, objectives, preconditions,
2) What: data input and algorithm throughput,
3) How: implementation, use, supervision, and output
4) Assessing the impact on fundamental rights and considering if the algorithm is suitable to be used for a given purpose.

Part four focuses on defining the seriousness of any fundamental right infringement, however, it is intended that the earlier parts must be completed prior to this section as the responses in previous sections will inform the assessment in the last part. Any potential infringements are to be weighed against factors such as the objective of the algorithm, whether the algorithm is suitable to meet that objective, and what other methods to meet the objective exist.

The creators of the FRAIA have conducted a pilot study where the FRAIA process was conducted, with a facilitator, for 15 existing public sector algorithms. The process took five hours in total for each FRAIA, where time was spent discussing and filling out the FRAIA questions. They state that ideally this time would be spread over two or three meetings over several weeks. They maintain the five hours is a suitable amount of time for this process, though, they are referring specifically to the time spent in the room completing the FRAIA questions. Other activities such as citizen panels, surveys or other research would not count towards that time, rather all reports and outcomes of such an activity should be brought to these meetings. The creators maintain that with proper preparation, the filling out of the actual FRAIA document can be done in five hours or less.[130]

[130] FRAIA in Action

2.2.1.1. A multidisciplinary team

The introductory instructions explain that the first step is determining who is involved in the FRAIA meetings. Emphasis is placed on composing a multidisciplinary team, where multidisciplinary is defined as consisting of a variety of disciplines and expertise. Potential participants are listed and include a range of roles from within the institution, the developer (even in the case that the system was procured from an outside company), representatives of an interest group or citizen panel and a legal advisor.

A key role identified by the creators is the process supervisor, who guides the discussion. This person should be knowledgeable about the FRAIA process and be "be substantively independent of the case to ensure objectivity." [131] This role is deemed so important and valuable that the creators have suggested that the government take on creating a pool of available trained FRAIA supervisors.

2.2.1.2. Experiences so far

The results of the pilot study of the FRAIA are that participants found it useful to engage in the reflective process, even in cases when they were having initial doubts of its usefulness. In terms of concrete impact, there were very few cases where no expected harm was identified at all, which led to action points for things like improvements on technical validation, improved transparency and training of the users. One case involving fraud detection resulted in the algorithm being terminated.

One of the creators of the FRAIA, Mirko Tobias Schäfer, reflected that a large part of the value of the FRAIA is that the practice builds capacity in government organisations. This includes facilitating the sharing of insights and good practices gleaned from how different teams and organisations undertake the assessment. By making these experiences visible, participants gain a clearer understanding of effective methods and common pitfalls, fostering a collective "assessment culture" and a shared notion of a

[131] Id.

"proper way of looking at algorithms" within the public sector.[132] This reflection is aligned with both of the two principal goals of the algorithmic impact assessment approach, as identified by Selbst.[133] The first being getting builders (or in this case deployers) of systems to methodically reflect on potential impacts before they occur and the second, the creation of documentation of decisions and their rationales, in order to improve accountability for the decisions and to inform future policy (as the impacts of the decisions emerge).

The creators of the FRAIA, in conjunction with the Dutch Ministry of the Interior and Kingdom Relations and Rijks ICT Gilde, are actively involved in updating and refining the tool; key changes are to streamline it based on user feedback and to incorporate new elements, including environmental impact.

2.2.1.3. Relevance for FRIA Implementation

FRAIA is not only intended for high-risk AI applications, as it was designed to assess any algorithm, but the experiences so far do not fully reflect that situation. As the FRAIA pilot study was conducted only on algorithms in use in public administration, the experiences do not reflect what the process would look like if conducted with, or within, a private company. Additionally, participation in the pilot study was entirely voluntary and thus participating teams were unlikely to include those that had high levels of scepticism or aversion to an impact assessment process.

While the FRAIA is ultimately a questionnaire-based tool, the creators of the FRAIA state that the document should be used as a tool to facilitate ethical discussions from which informed choices can be made. This involves a structured discussion guided by process supervisors or moderators, a role that they deem essential. The FRAIA document is intended not just for noting down answers, but also the key considerations and choices

[132] Interview with Dr. Mirko Tobias Schäfer. 4 April 2025.

[133] Selbst, A. D. An Institutional View of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology (Harvard JOLT)*, *35*, 117. 2021. Available at: https://heinonline.org/HOL/Page?handle=hein.journals/hjlt35&id=123&div=&collection=

made during this discussion process. These best practice instructions, along with other 'suggested' activities such as engaging in relevant public consultation, risk being overlooked in the case that the sole hard requirement for a FRIA is the filling out of a questionnaire. The creators of the FRAIA do note that in their experience, it is possible for an experienced person to assess whether a questionnaire has been filled out in a minimalistic way, without adherence to best practices or completion of minimum additional activities outside of filling in the form. This speaks to the need for oversight and review of submitted FRIA questionnaires, which requires an ecosystem of trained personnel.

Another critique of the FRAIA is the lack of depth in its guidance of how to answer the normative questions, particularly in regard to measuring the likelihood or severity of impacts on fundamental rights. This issue is mentioned in a report on the FRAIA pilot study, where it is acknowledged that some participants noted seeming ambiguities in the questions. However, the report authors state that these ambiguities were often cleared up through discussion with other participants and the moderator, or by additional reading material. This points to the importance of having a team with a comprehensive set of multidisciplinary expertise involved in the FRAIA process.

Malgieri and Santos[134] see room for significant improvements in the FRAIA, and other Impact Assessments, approach to assessing severity of fundamental rights infringements, stating that there could, for example, be more guidance that reflects how to incorporate case law from the Court of Justice and state-of-the-art scholarly discussion around approaches to assessing impacts on fundamental rights.[135] This issue, common with rights impact assessments, inspired Malgieri and Santos[136] to develop a framework for assessing the severity of impacts on fundamental rights. They

[134] Malgieri, G., & Santos, C. Assessing the (severity of) impacts on fundamental rights. *Computer Law & Security Review*, *56*, 106113. 2025. Available at: https://doi.org/10.1016/j.clsr.2025.106113

[135] Interview with Dr. Gianclaudio Malgieri. 9 April 2025.

[136] Malgieri, G., & Santos, C. Assessing the (severity of) impacts on fundamental rights. *Computer Law & Security Review*, *56*, 106113. 2025. Available at: https://doi.org/10.1016/j.clsr.2025.106113

propose specific operational parameters for assessing severity of these interferences, which are organized across three categories:

- objective normative evaluation (based on legal rules),
- subjective perception (across individual, group, and societal levels)
- real-life consequences or adverse effects on individuals' lives.

2.2.2. Case Study: Catalan Data Protection Authority FRIA model

The Catalan Data Protection Authority (APDCAT) has put forward a FRIA model, which was created specifically with an eye towards meeting the FRIA requirements of the AI Act.[137] The model was developed by a working group led by Alessandro Mantelero, a professor and legal scholar who has been an advocate of the inclusion of FRIAs in the AI Act and has advised the European Commission on the topic of FRIAs in general.[138] The model has also been adopted by The Personal Data Protection Agency of Croatia as a recommended methodology.[139]

The model consists of three phases, with a template provided:

1) Planning and scoping (a questionnaire prompting reflection on the system, the deployment context and potential risks).
2) Data collection and risk analysis (filling in a series of risk matrices to determine likelihood and severity of each identified risk).

[137] Mantelero, A., Guzmán, C., García, E., Ortiz, R., & Moro, M. A. FRIA model: Guide and use cases. *The Catalan Data Protection Authority*, 93. 28 January 2025. Available at: https://www.dpdenxarxa.cat/pluginfile.php/2468/mod_folder/content/0/FRIA_es_2.pdf

[138] See for example: https://digimed.polito.it/2025/02/21/12-15-june-2024/ and Mantelero, A., & Esposito, M. S. An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. *Computer Law & Security Review*, *41*, 105561. 2021. Available at: https://doi.org/10.1016/j.clsr.2021.105561

[139] Croatia: AZOP publishes guidance on FRIA under AI Act. 17 March 2025. Available at: https://www.dataguidance.com/node/643108

3) Risk management (identifying mitigation measures for each identified risk, as well as implementing and monitoring these measures).

Key requirements highlighted in the guide are that the assessment be expert based - it needs to be completed by a multidisciplinary team of relevant experts and that it should be iterative – monitoring, mitigation and assessment should be a continuous process. Also highlighted is the fact that the FRIA is a contextual assessment rather than a technological one - the impact of a system in its context of use is the focus.[140]

The model was tested on multiple use cases, with the results of four of them being publicly shared in the released guide. Reasons of confidentiality are given for not providing details of the other use cases that were conducted. The reported use cases consist of three algorithms for public services (education, health, and welfare benefits) and one for a private company (a hiring related algorithm). Mantalero has stated that a case study was conducted with CaixaBank,[141] and one of the report authors is from CaixaBank, however it is not clear whether one of the undisclosed use cases is about CaixaBank, or similar banking and financial services. This is of relevance given the requirements for banks and financial institutions to perform FRIAS as potential deployers of AI systems applied to creditworthiness or insurance pricing tasks. This type of case study would have been beneficial to publicly make available as a teaching tool, given the lack of self-disclosed IAs from this sector and the urgent need to establish a baseline for best practices.

The guide for the model emphasizes the relative quickness with which the FRIA can be conducted. The given time estimate for completing the FRIA is very similar to the FRAIA: two to three three-hour meetings, if the right team of people is present. This

[140] Mantelero, A., Guzmán, C., García, E., Ortiz, R., & Moro, M. A. FRIA model: Guide and use cases. *The Catalan Data Protection Authority*, 93. 28 January 2025. Available at: https://www.dpdenxarxa.cat/pluginfile.php/2468/mod_folder/content/0/FRIA_es_2.pdf

[141] Catalan Data Protection Authority. Alessandro Mantelero: "A technology that goes against fundamental rights is not a good technology." *Catalan Data Protection Authority*. 10 April 2025. Available at: http://apdcat.gencat.cat/en/sala_de_premsa/notes_premsa/noticia/Entrevista-Alessandro-Mantelero-FRIA

short time is highlighted as evidence that the process "does not impose an excessive additional burden on private and public entities in the EU in order to comply with the AI Act." [142] This time does not reflect the future iterations of the FRIA nor the time spent on system changes and mitigation activities that may be identified during the FRIA process. The (suggested and non-exhaustive) list of questions for the planning and scoping phase does include prompting to identify groups and communities who will be affected by the model and who should be part of the assessment, but no specific guidance on public participation on the FRIA is given.

2.2.3. HUDERIA methodology

The risk and impact assessment of artificial intelligence (AI) systems from the point of view of human rights, democracy and the rule of law (HUDERIA Methodology)[143] has been adopted by the Committee on Artificial Intelligence of the Council of Europe.

The methodology is described in four stages. Interestingly, there are two risk assessment stages, with the first being a preliminary background research and information collection stage, with a focus on identifying potential risk factors in the application, development and deployment contexts. Importantly, an outcome of this stage is to make an initial determination about whether a system should be developed or deployed at all. There is also a distinct stakeholder development stage, with a focus on identifying at-risk stakeholders that are particularly vulnerable to potential harms and or have particularly limited influence on the systems design and deployment.

Five steps towards stakeholder engagement are explicated:

- Stakeholder Analysis
- Positionality Reflection
- Establishment of Engagement Objectives

[142] Id
[143] Committee on Artificial Intelligence of the Council of Europe. HUDERIA: New tool to assess the impact of AI systems on human rights. 2024. Available at: https://www.coe.int/en/web/portal/-/huderia-new-tool-to-assess-the-impact-of-ai-systems-on-human-rights

- Determination of Engagement Method
- Implementation

Additional resources for stakeholder engagement are still being developed.

In the second risk assessment stage, risks are assessed by the variables of scale, scope, reversibility and probability. Key aspects of the mitigation stage include assessing mitigation options in a "mitigation hierarchy" and revisiting the question of whether the system should be developed or deployed at all.

2.3. Draft "*Fundamental Rights Impact Assessment (FRIA) fit for the AI Act*" led by the European Center for Non-Profit Law[144]

Another FRIA methodology is currently being drafted by the European Center for Non-Profit Law. It is being created in a joint effort by technical, legal, and social experts and draws from the existing FRIA methods and case studies discussed above. This methodology is meant to highlight what is mandatory for a FRIA process to effectively guard against AI impact on fundamental rights. Mandatory features include deliberation by a multidisciplinary group, documentation of description and justification of all choices regarding risk levels and mitigation measures.[145]

2.4. Accountability mechanisms and the banking and finance sector

The use of AI systems in banking and finance encompass different applications in areas such as advisory services, internal process, trading, and risk mitigation. Further, this sector is highly regulated, with specific provisions related to their use of data-driven systems and processes, including data management and governance. Some examples of frameworks companies in this sector are subjected to include: Basel Framework

[144] Organizations involved in drafting and providing input include leading academics, fundamental rights experts, equality bodies and technical experts, such as Algorithm Watch, Amnesty International, Avaaz, CDT Europe, Civil Liberties Union for Europe, Danish Institute for Human Rights, European Centre for Not for Profit law, Equinet, European Network of Human Rights Institutions, etc..

[145] Personal communication from the European Center for Non-Profit Law, April 2025

Alignment standard number 239 (BCBS 239),[146] GDPR Article 22,[147] European Banking Authority Guidelines on Loan Origination and Monitoring.[148] Concerning the AI Act, banking and finance is of particular interest, as AI systems intended for the evaluation of creditworthiness, credit-scoring, and for insurance pricing, i.e., health and life, are classified as high-risk, and deployers of these systems are obliged to perform FRIAs.

A CLOSER LOOK 3: FINANCIAL SERVICES, AI AND CONSUMER PROTECTION

The use of AI systems in the banking and finance sector raises concerns, primarily in relation to consumer protection. A report, published by the Alan Turing Institute,[149] identified the following four areas where the use of AI could have an impact on consumer protection:

- financial inclusion;
- unwarranted denials of service in the context of financial crime prevention;
- unlawful discrimination and unfair differential treatment;
- mismatches between products and customer needs;
- performance of investments; and
- consumer empowerment.

The report elaborates on all six points by first identifying the relevant context where contention points may manifest, followed by a discussion of both, the advantages of using AI systems, and the potential harms that can arise from its use on consumers.

[146] Basel Committee on Banking Supervision. Principles for effective risk data aggregation and risk reporting. January 2013. Available at: https://www.bis.org/publ/bcbs239.pdf

[147] European Data Protection Board. Automated decision-making and profiling. 25 May 2018. Available at: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/automated-decision-making-and-profiling_en

[148] European Banking Authority. Guidelines on loan origination and monitoring. 2025. Available at: https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/credit-risk/guidelines-loan-origination-and-monitoring

[149] Ostmann, F., & Dorobantu C. AI in financial services. The Alan Turing Institute. 2021. Available at: https://doi.org/10.5281/zenodo.4916041

Below, a relevant example is detailed in the context of financial inclusion, drawing from additional sources.

ON FINANCIAL INCLUSION/EXCLUSION

Relevant contexts: product eligibility, product pricing, some of the forms of denials of service related to financial crime prevention

How can it materialize: Access or lack of access to a financial product or service takes form in a context where consumers are deemed eligible or ineligible following an AI-enabled risk-profiling assessment, to name an example. The price of products and services and their general affordability is also a determinant of financial inclusion. AI-enabled systems can be used to determine suitability and personalized prices, drawing from the explosion of personal data used for these purposes, which at the same time could be at odds with GDPR compliance and data minimization principles.

It goes without saying that the performance of AI systems is highly dependent on data availability and on the quality of the data. Potential consumers that require access to financial products and services, or that would benefit from lower prices, could be at a disadvantage if their "data footprints" are not deemed large enough. Besides this practicality, in considering financial institutions as for-profit organizations, potential consumers at the margins can also fall into the category of unprofitable.

Potential advantages of AI use:

- AI could enable reductions in operational costs that firms could pass on to consumers in the form of lower prices.
- Improved risk profiling capabilities enabled by AI could translate into favourable eligibility decisions or price reductions for customers that would otherwise lack access.
- AI systems and non-traditional data-driven processes may enable more

granular or personalized forms of price differentiation.

Potential Harms of AI use:

- Risk profiling, poorly performing systems, or problems with competent use and human oversight could result in flawed risk profile assessments capable of leaving consumers out of the market, with no access to recourse mechanisms.[150]
- Consumers flagged by AI-systems as "high-risk" can also be subjected to inaccessible or excessively expensive products and services.[151] Cases such as these have already materialized in the Netherlands insurance sector, as the use of alternative and faulty data have caused consumers to be priced out of home insurance, or simply become "uninsurable." [152, 153] Similarly, the use of certain attributes by AI systems to determine the price of financial products contribute to significant price increases in car insurance premiums. Different research efforts in European Union countries, as well as Great Britain, and the United States of America have uncovered how ethnicity or proxies for it, e.g., zip cope,

[150] Jiménez Arandia, P. & Russo, M. Cuando un algoritmo dicta (erróneamente) el bloqueo de tu cuenta bancaria. EL PAÍS. 24 September 2024. Available at: https://www.proquest.com/newspapers/cuando-un-algoritmo-dicta-erróneamente-el-bloqueo/docview/3109579334/se-2 .

[151] Federico Oliveira da Silva, Kasper Drazewski. Regulating AI to protect the consumer. The European Consumer Organisation. 6 October 2021. Available at: https://www.beuc.eu/sites/default/files/publications/beuc-x-2021-088_regulating_ai_to_protect_the_consumer.pdf

[152] Joyce Donat. Premies woonhuisverzekeringen stijgen door gebruik big data. Consumenten Bond. 12 October 2018. Available at: https://www.consumentenbond.nl/nieuws/2018/premies-woonhuisverzekeringen-stijgen-door-gebruik-big-data

[153] See also: The European Consumer Organisation. The Use Of Big Data And Artificial Intelligence In Insurance. 19 May 202 Available at: https://www.beuc.eu/sites/default/files/publications/beuc-x-2020-039_beuc_position_paper_big_data_and_ai_in_insurances.pdf&sa=D&source=docs&ust=1748375672001995&usg=AOvVaw19Jjabm2O7mVK3Lf5LrXSo

country of origin, contribute to significant price hikes.[154,155,156]

- Different instances of algorithmic bias have also been reported and researched with relation to unfavourable creditworthiness assessments that penalize individuals by granting them lower credit limits, or altogether refusing them a loan or credit, resulting in discriminatory action.[157,158,159] Research has uncovered that the algorithmic systems gave greater importance to variables such as place of residence, gender, age and mother tongue, instead of income or the individuals' credit history.

EU residents have a series of rights and protections applicable to opening a bank account, transferring money, taking out loans and buying insurance products; notwithstanding, banks and financial institutions are still capable of biased and discriminatory practices, consumer protections infringements, and have been involved in corporate scandals with consequences reverberating across society world-wide for

[154] Gonzalez, A. Car insurance quotes 33% higher in most ethnically diverse areas. Motor Finance online. 26 February 2024. Available at: https://www.motorfinanceonline.com/news/car-insurance-quotes-33-higher-in-most-ethnically-diverse-areas-bbc/?cf-view&sa=D&source=docs&ust=1748375672001728&usg=AOvVaw1LdkzPRVlE_-DhRcfkSYte

[155] Angwin, J., et al. Minority Neighborhoods Pay Higher Car Insurance Premiums Than White Areas With the Same Risk. ProPublica. 5 April 2017. Available at: https://www.propublica.org/article/minority-neighborhoods-higher-car-insurance-premiums-white-areas-same-risk

[156] Gender Equality Commission And The Steering Committee On Anti-Discrimination, Diversity And Inclusion (Cdadi). Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination. Council of Europe. 2023. Available at: https://rm.coe.int/study-on-the-impact-of-artificial-intelligence-systems-their-potential/1680ac99e3&sa=D&source=docs&ust=1748375671852358&usg=AOvVaw03FDfL6tIHGJgr_P-hPVG2

[157] Id.

[158] Vincent, J. Apple's credit card is being investigated for discriminating against women. The Verge. Available at: https://www.theverge.com/2019/11/11/20958953/apple-credit-card-gender-discrimination-algorithms-black-box-investigation

[159] Matzat Lorenz. Finnish Credit Score Ruling raises Questions about Discrimination and how to avoid it. Algorithm Watch. 21 November 2018. Available at: https://algorithmwatch.org/en/finnish-credit-score-ruling-raises-questions-about-discrimination-and-how-to-avoid-it/

many years.[160] Beyond regulatory compliance, their behaviour demands constant scrutiny, as banks and financial institutions are integral parts of the global economy, as well as everyday life given the hyper-dependence of society on their services and infrastructure for basic and consequential operations such as salary payments, bills, loans, and other types of transactions.

At this crossroads, the rest of the subsection examines existing accountability mechanisms, within and outside the context of AI systems, with the objective of drawing insights as to what meaningful FRIAs could look like for banks and financial institutions.

2.4.1. Private accountability mechanisms

When discussing the banking and finance sector, it is important to delve into the concept of private actors' accountability. Micro-enterprises, cooperatives, and multinationals that carry out their activity as private actors, exist outside traditional accountability systems, and thus the need for alternative mechanisms, is created as a way to get these companies to engage in more accountable business behaviour,[161] beyond what is constituted as corporate social responsibility.[162] Some examples include, codes of conduct, disclosure directives,[163] corporate reporting, and impact assessments. Concisely, social taxonomies, such as those presented by the Platform

[160] For example, see: Transparency International. Banking scandals: On corporate culture, public interest and role of Western governments. 15 October 2021 Available at: https://www.transparency.org/en/blog/banking-scandals-corporate-culture-public-interest-western-governments

[161] Read more: TAP Network. SDG ACCOUNTABILITY HANDBOOK: Accountability of the Private Sector. Available at: https://sdgaccountability.org/wp-content/uploads/2019/05/Accountability-of-the-Private-Sector.pdf

[162] Tamvada, M. Corporate social responsibility and accountability: a new theoretical foundation for regulating CSR. *Int J Corporate Soc Responsibility* 5, 2. 2020. Available at: https://doi.org/10.1186/s40991-019-0045-8

[163] Read more: Danish Institute for Human Rights. Reviews of mandatory human rights due diligence and disclosure laws. Available at: https://humanrightseducation.dk/Methodologies%20for%20assessing%20business%20respect%20for%20human%20rights%20/index.html#/lessons/e7OtdNatyRsQz32lCnorbNCFIJxkwiOU

on Sustainable Finance,[164] identify key areas and criteria that bind an economic activity to the idea of a sustainable business model; incorporating evaluation areas such as labour rights, working conditions, social inclusion, no discrimination, consumer protection, and tracing them back to key stakeholders affected by them. Further, voluntary agreements such as the Equator Principles,[165] also bind banks to periodically evaluate and self-report on their investment projects in terms of criteria that pertain to social, environmental, and human rights risks.

EXAMPLE RESOURCES: A HUMAN RIGHTS IMPACT ASSESSMENT TOOL FOR AI-INFORMED DECISION-MAKING SYSTEMS IN BANKING[166]

In 2023 the Australian Human Rights Commission partnered with the National Australia Bank (NAB) to co-develop and produce a human rights impact assessment tool for AI-informed decision-making systems in banking. The proposed tool is not ideated as prescriptive, but it is highly encouraged.

Additional resource Even though, it is not AI-specific, in the referenced document,[167] the Office of the United Nations High Commissioner for Human Rights provides a response to a request from BankTrack,[168] an international tracking, campaigning and civil society support organization targeting private sector commercial banks and the

[164] Read more: Platform on Sustainable Finance. Available at: https://finance.ec.europa.eu/sustainable-finance/overview-sustainable-finance/platform-sustainable-finance_en

[165] Read more: Equator Principles Limited. 2025. Available at: https://equator-principles.com/signatories-epfis-reporting/

[166] Read more: Australian Human Rights Commission. HRIA Tool: AI in Banking. 28 September 2023. Available at: https://humanrights.gov.au/our-work/technology-and-human-rights/publications/hria-tool-ai-banking

[167] Read more: Office of the United Nations High Commissioner for Human Rights. Application of the UNGPs in the context of the banking sector. 12 June 2017. Available at: https://www.ohchr.org/sites/default/files/Documents/Issues/Business/InterpretationGuidingPrinciples.pdf

[168] Read more: https://www.banktrack.org/page/about_banktrack

activities they finance, on advice regarding the application of the UN Guiding Principles on Business and Human Rights in the banking sector.

Of particular relevance to this report, businesses are also expected to act in accordance with the United Nations Guiding Principles on Business and Human Rights, especially in reference to large and influential enterprises. This framework establishes, for businesses, the responsibility to respect human rights by adopting human rights policies. They are also expected to proactively identify and measure impacts on human rights, propose mitigation strategies,[169] as well as monitor, and perform audit processes to track performance. An instrument that accompanies this framework is a human rights impact-assessment (HRIA). As already discussed, these assessments are instrumental in supporting the operationalization of human rights considerations prior (ex-ante) to the adoption of new projects, agreements, programs, etc. The Guide to Human Rights Impact Assessment and Management (HRIAM) includes practical human rights scenarios for companies in the banking sectors looking for practical guidance prior to completing a HRIA.[170]

A CLOSER LOOK 4: THE FINANCIAL SYSTEM BENCHMARK AND SPANISH BANKS PERFORMANCE

The World Benchmarking Alliance (WBA), a non-profit organization, tracks and measures the social and environmental impact of the 2000 most influential companies in the world across industries. Their objective is to advance the United Nations Sustainable Development Goals, by attempting to hold companies

[169] Read more: Transparency, Accountability & Participation (TAP) Network. Campaign for a Decade of Accountability. 2021. Available at: https://www.sdgaccountability.org/wp-content/uploads/2021/06/GlobalSDGAccountabilityReport_pages_hRes-1.pdf

[170] Read more: The International Business Leaders Forum and the International Finance Corporation. Guide to Human Rights Impact Assessment and Management (HRIAM). 2010. Pages 75-81. Available at: www.ifc.org/hriam

accountable across different measurement areas, with assessments performed every two years.[171]

In the context of the finance industry, they assess 400 companies, and define the following measurement areas: governance, financing climate and nature protection and restoration, environmental footprint, inclusive finance, and responsible business conduct, with scores comprised between 0-100.[172]

KEY FINDINGS The WBA states that the 2025 results show a slight improvement in comparison to the results obtained for 2022.[173] However, they consider that the real world impact of the industry is still very limited, with all companies scoring in the lower tier of the scoring range (<50). This is discouraging, considering that alignment of the industry could "trigger a domino effect of positive change across the financial system."

The WBA identifies an industry-wide lack of strategy and processes aimed at driving real change. The WBA also identifies a gap, as there is still room for overt and meaningful collaboration with external stakeholders. Also, despite ongoing discussions and promises to adhere to global principles, companies in the industry tend to not disclose their processes used to assess human rights risks from their financial activity. Moreover, only 6% of assessed companies disclose the processes they have in place to protect workers' rights and livelihood, denoting here the shortcoming associated to 'voluntary due diligence disclosures.'

Below, an overview at industry level in Europe and Spain.

[171] The benchmarks developed by the WBA are put together based solely on the data that is disclosed by companies and is publically available. This means that third party data does not factor into the assessments. As part of WBA process, they offer companies the opportunity to provide feedback on the findings they obtain.

[172] Read more: Details on methodology. Available at: https://www.worldbenchmarkingalliance.org/publication/financial-system/methodology/

[173] See also: SpainSIF. La Inversión Sostenible y Responsable en España: ESTUDIO DE MERCADO. 2024. Available at: https://www.spainsif.es/wp-content/uploads/2024/10/Estudio_Anual_Spainsif_2024.pdf

Overall measurement score for banks and insurance companies (64) at a European level during the last reporting cycle (2022-2025)

Overall score	***Number of Companies***
Score > 40	*0*
Score 25–40	*23*
Score 10–25	*32*
Score < 10	*9*

Overview of ranking and overall measurement score for banks (3) and insurance companies (1) at a Spain level, during the last reporting cycle (2022-2025)

Bank/Insurance Co.	***In-group ranking (out of 400)***	***Overall score***
Banco Bilbao Vizcaya Argentaria (BBVA)	6	35.6
CaixaBank	14	32.8
Banco Santander	51	24.2
MAPFRE España	87	20.6

Relevant highlights:

- BBVA
  - While there are internal processes for performing human rights risk impact assessment, no examples were found of actions taken in the past years to address human rights issues identified in relation to the products, services, and capital it offers.
- Caixabank
  - Could disclose more details on the processes they have put in place to avoid negative impacts on low-income countries as a consequence of

their sustainability strategy. Similarly, there is no disclosure of processes in play to identify social risks in consequence of net-zero transition.

    - While an internal human rights policy that addresses the International Labour Organization rights[174] exists, there is a lack of evidence these are incorporated into their risk impact assessments. Also, no examples of actions taken on salient human rights issues of their activity were found.

- Banco Santander
    - Insufficient disclosure on important topics related to their sustainability strategy.
    - While disclosures on their commitment to respect the International Labour Organization rights at work have been made, this does not translate into the existence of internal policy documents.
    - There is no description of the methodology used to define living wage across all the territories where they operate. In a similar line, they do not disclose if risk assessment processes incorporate risks associated with the International Labour Organization fundamental rights at work for those affected by their products, services, and capital.
    - No examples found of the conclusions reached nor actions taken or planned in response to at least one of significant human rights issues resulting from assessment processes within one of its activities in the past three years.
- MAPRE España

[174] Read more: International Labour Organization. ILO Declaration on Fundamental Principles and Rights at Work. 2022. Available at: https://www.ilo.org/about-ilo/mission-and-impact-ilo/ilo-declaration-fundamental-principles-and-rights-work

- Lack of supportable evidence and criteria employed for impact identification processes and prioritization with regard to their material sustainability impacts.
- No examples found of actions taken to address human rights issues resulting from risk assessment processes related to their activity, product or services.

2.5. Opportunities for improvement

The ever evolving meaning as to what entails AI and the different types of technology encompassed by this term, requires a constant assessment of their potential impact on fundamental rights and further emphasizes the importance of rigorous FRIA processes. As described throughout this section, there are different approaches and methodologies to perform impact assessments, as well as different expectations when it comes to the actions of public administrations and private companies, such as Big Tech and those in the finance industry. With sights set on a meaningful implementation of Article 27, given the presented case examples, salient best practices and opportunities for improvement are identified below, highlighting when necessary differences between the public and private sector.

Best practices. Key agreements across the presented examples related to socially-focused impact assessment include:

- the essential need for a multidisciplinary expert team comprising diverse specializations and backgrounds to conduct the assessment;
- the recognition of the importance of involving stakeholders outside the AI deployers' institution, including potentially affected rights-holders and CSOs that represent them;
- the role of documentation during IAs processes and how it should serve the purpose of fostering accountability, transparency, and knowledge transfer – this

requires that the documentation about decisions made and reasoning behind them should be complete and informative; and

- the importance of making IAs public in order to promote the idea of third-party oversight and facilitate enforcement duties. Here, it is important to highlight that there are transparency provisions in the AI Act and other laws and considerations that affect public and private actors differently (more on this in Section 4 of this report); however, relying on the good faith of private actors to both, self-regulate and self-disclose reliable information is not advisable, as evidence shows. For that reason, it is important to explicitly define requirements, obligations and mechanisms that enforce and incentivize transparency in a meaningful way for all actors.[175]

Opportunities for improvement. Given the characteristics of the FRAIA and Canadian AI Directive, there is more evidence for the appraisal of public administration participants in IAs processes; this also applies to the APDCAT model case studies, as there was only one private sector participant. Inherently, another factor that contributes to making this comparative analysis is that the FRAIA and APDCAT pilot studies were all done under what could be considered optimal conditions, with intake processes, expert guidance, and moderators, facilitating the identification of best practices.

At the other end of the spectrum are the lessons derived from ‘unsupervised’ instances. For example, the evaluation of DPIAs over the years at the hands of the Spanish Data Protection Authority, as well as the self-reported RIs published under the DSA, completed mostly, if not only, by private and powerful corporations. From assessing these cases, it is possible to see how the lack of guidance and prescriptiveness results in:

[175] Pénicaud, S. Making Algorithm Registers Work for Meaningful Transparency. IA Ciudadana. 2025. Available at: https://iaciudadana.org/2025/03/13/making-algorithm-registers-work-for-meaningful-transparency/

- FRIAs being misinterpreted as more bureaucracy or red tape to deal with. This can lead to adopting a minimalistic approach to FRIAs, facilitated by close-ended questionnaires that do not elicit critical or reflective thinking, aggravating the already perceived lack of expertise or comprehension of the law in analogous processes, such as the GDPR's DPIAs, by those responsible for its implementation, e.g., data controllers, DPOs. This has the potential to translate into ineffective FRIAs given how AI systems engage a wide range of fundamental rights and there is still a general misconception regarding the negative impact of AI systems and a lack of knowledge with regard to fundamental rights and mitigation strategies.[176]

  Findings published by the EU Agency for Fundamental Rights already identify the tendency to define a limited scope for addressing impacts on FRs; and while this was observed in both, officers at public administration and staff at private companies, it was more evident in participants from private companies. In some instances, AI developers in private settings noted how the responsibility to assess for potential fundamental rights issues fell on "their clients;" [177] This brings into focus how AI provider and AI deployer relationships could play out, and the need to emphasize how anticipating the negative impact of AI systems is the preemptive responsibility of all parties involved in the planning, development, and deployment of these systems.
- IAs not meeting a desirable baseline when it comes to the quality of the IA process itself, as well as for the quality of the content that is documented and published;

[176] European Union Agency for Fundamental Rights. Getting The Future Right Artificial Intelligence And Fundamental Rights. 2021. Available at: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2021-artificial-intelligence-summary_en.pdf

[177] European Union Agency for Fundamental Rights. Getting The Future Right Artificial Intelligence And Fundamental Rights. 2021. Available at: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2021-artificial-intelligence-summary_en.pdf

- the absence of predefined measurable parameters to quantify harm or severity of impact, lack of detailed methodology on how risks were assessed, as well as lack of actionable plans for risk mitigation and redress; and
- the limited participation or altogether exclusion of affected parties from deliberation processes. This can be attributed to the varied level of explicit focus and detailed guidance on achieving meaningful public participation from affected groups across different regulation and frameworks.

While there are evident differences between the private and public sector's approach to FRIAs or analogous processes, a general takeaway that became evident in the undertaken pilot exercises, is the seemingly shared understanding that the core objective of an FRIA process is not to be a mere checklist exercise, but rather an opportunity to spark informed discussion and reflection about the assessed AI project and its impacts among the involved parties. For that reason, there is an undeniable need to make certain resources, guidelines, and infrastructure become available to support AI deployers during the FRIA process.

3. Participation of affected groups

3.1. Motivating participation in AI regulation and impact assessments

Public participation is generally being put forward as a best practice in technology design and governance, including in EU regulations. In regard to impact assessments, Recital 95 of the AI Act explicitly states:

> *"Where appropriate, to collect relevant information necessary to perform the impact assessment, deployers of high-risk AI system, in particular when AI systems are used in the public sector, could involve relevant stakeholders, including the representatives of groups of persons likely to be affected by the AI system, independent experts, and civil society organisations"*

The need for civil society engagement has also been highlighted in design and enforcement of other recent technology regulations from the EU, such as the Digital

Services Act[178] and in regulation of AI In the financial sector[179]. When specifically considering FRIAs, it is important to note that public consultation and participation is a best practice in impact assessments in other industries. Impact assessments in other industries, including their relationship to FRIAs is discussed in Section 2 of this report.

In this report, participation in FRIAs refers to the inclusion of third party participants beyond the deployer organization, including representation of relevant expertise and of interests of members of the public (or "rights holders). This can manifest as the inclusion of CSOs, Fundamental Rights experts, affected individuals and representatives of affected groups. These participants must have the opportunity to influence deployment decisions away from solely the interest of the AI deployer, towards achieving a more balanced deliberation process that accounts for the interests of all stakeholders, particularly those of the public.

Participation in design and governance of AI systems is required for what amounts to democratic participation in important social decisions. Stakeholder participation is essential in fundamental rights impact assessments for AI systems because it ensures that the voices of those most vulnerable to fundamental rights infringements are heard and accounted for in the deployment, and oversight of these technologies. Without direct input from affected groups, it is impossible to fully understand the specific ways AI systems may help or harm people,[180] especially in contexts where data-driven tools intersect with structural inequalities.

There are also tangible benefits for AI system deployers: engaging with public expertise can surface risks early, help avoid harm or scandal, and mitigate reputational or legal

[178] Jahangir, R. *EU Steps Up Civil Society Engagement On the Digital Services Act — Is It Enough? | Tech Policy Press*. Tech Policy Press. 16 April 2025. Available at: https://techpolicy.press/-eu-steps-up-civil-society-engagement-on-the-digital-services-act-is-it-enough

[179] European Commision. (2024, June 18). *Targeted consultation on artificial intelligence in the financial sector—European Commission*. https://finance.ec.europa.eu/regulation-and-supervision/consultations-0/targeted-consultation-artificial-intelligence-financial-sector_en

[180] Example initiative: Connected by Data. (2024, February 12). *People's Panel on AI*. https://connectedbydata.org/projects/2023-peoples-panel-on-ai

fallout. There are numerous cases of algorithms being implemented by public or private bodies only to be identified as discriminatory or otherwise harmful, with resultant public outcry[181] – along with the resultant real harm to affected individuals, consequences have included legal action, public disgrace, and even toppling of political powers (as was the case in the Netherlands welfare fraud scandal discussed in Section 1).

Participation in AI impact assessments is not without significant limitations. There is a risk of "participation washing," where public involvement is superficial, serving more to legitimize decisions than to influence them meaningfully. In some cases, the process may become extractive, exploiting the knowledge or experiences of participants without offering real influence or reciprocity.

Participation washing activities can be fostered in cases of regulatory capture. This refers to situations where regulators are overly influenced, or even controlled, by the industry that it is meant to regulate. In cases of capture, the regulation can actually be used to benefit the regulated industry, while not effectively regulating it in the intended way.[182] This would be the case for example, if AI systems that appeared to have included a participatory element were viewed more favourably, or more likely to be taken up, even if the actual participation did not meet minimum standards,[183] (such as the public having no actual impact on the decision-making process involving the deployment of a system). The self-regulatory nature of FRIAs in the AI Act make this a particular risk. While participation should be viewed as a form of tripartism, where stakeholders, including the public, have an active role in regulatory processes, participation requirements are also at risk of capture.[184]

[181] A few high profile examples, of many: COMPAS, Amazon, AMAS, SyRI, France welfare system (see: https://www.wired.com/story/algorithms-policed-welfare-systems-for-years-now-theyre-under-fire-for-bias/), Cambridge Analytica.

[182] Selbst, A. D. An Institutional View of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology (Harvard JOLT)*, *35*, 117. 2021. Available at: https://heinonline.org/HOL/Page?handle=hein.journals/hjlt35&id=123&div=&collection=

[183]Veale, Micheal [mikarv]. Tweet.2023. Retrieved 1 May 2025. Available at: https://x.com/mikarv/status/1734297345519468969

[184]Kaminski, M. E., & Malgieri, G. Impacted Stakeholder Participation in AI and Data Governance (SSRN Scholarly Paper No. 4836460). Social Science Research Network. 2024. Available at: https://papers.ssrn.com/abstract=4836460

3.2. Defining meaningful participation

Ensuring that public participation is not merely symbolic but actually provides a mechanism for shaping the deployment of high risk AI systems towards public interest, is essential to building accountable and rights-respecting AI governance. This requires that participation activities imbue a level of power to the relevant interested public, in relation to the system deployers (which in the case of FRIAs will be public institutions or private entities such as banks or insurance companies). Okidegbe refers to the three dimensions identified by Jocelyn Simonson and K. Sabeel Rahman for measuring the extent to which an institutional structure is able to shift power in this way: the nature of authority, the composition of authority, and the moment of authority." [185] These dimensions are useful for measuring the extent to which participation shifts power given to the participating public, and is thus meaningful.

The nature of authority. This refers to the level of power in decision-making; what is the level of decision-making power of the public, and to what extent can they influence the processes and plans that are in place? In the context of the deployment of a high-risk AI system, this includes whether the public has the option of refusal; is there any route for the participating public to say no to the deployment of a given system?

The composition of authority. This refers to who is being represented in the process. Are the relevant-affected communities, ethicists, and advocacy groups part of the process?

Moment of authority. This refers to the stage of influence and how early or late in the process are the public's opportunities for involvement and influence. In the context of the entire process of AI system design and development, public involvement can start as early as the decision of whether to build a system for a given purpose, or even that the public proposes the use of AI for a given purpose. Similarly, in the case of AI system

[185] Rahman & Simonson (2020), as cited in: Okidegbe, N. To Democratize Algorithms. *UCLA Law Review*, *69*. 2022. Available at: https://www.uclalawreview.org/to-democratize-algorithms/

deployment, participation can occur before procurement or in-house development, or even be the driver of what problems will be addressed through AI.

Visible impact and trustworthy process. Visible impact of the participation is one of the key factors defining meaningful engagement. It is also important for avoiding 'participation fatigue' and other negative perceptions of participatory processes. Visible impact can be present even in cases of low levels of power in making final decisions, in raising concerns, offering insights, and influencing decisions. In addition, the process must be a trustworthy one, which requires transparency about the process[186]. The institution must be transparent about the level and type of impact that participants have, and about the various other interests that have influence in the process.[187]

3.3. Participation methods

There are many forms of public participation. A wide range of formats and methods exist; they can be used at different points in a development and deployment process, and used for different purposes, and the activities can target different members of the public and different interests.

A useful way to conceptualize different participation methods is through the IAP2 Spectrum of Public Participation, which defines five levels: Inform, Consult, Involve, Collaborate, and Empower. Below, are outlined each level with examples relevant to AI policy. These levels are not mutually exclusive, nor are the example methods mentioned only and always relevant to the given category – holding a specific activity, such as a workshop, does not guarantee that the requirements of the category are met.

Inform: Building awareness and transparency. The public is kept informed, but does not influence decision-making. While informing does not give specific powers of influence

[186] The European Center for Not-for-Profit Law & Society Inside. Framework for Meaningful Engagement: Human rights impact assessments of AI. 2023. Available at: https://ecnl.org/publications/framework-meaningful-engagement-human-rights-impact-assessments-ai
[187] Id.

to participants, it is a fundamental aspect of participation and is required for all subsequent levels. Common methods include:

- Publishing algorithmic impact assessments or audit results.
- Providing informational content for the general public about a specific AI system.
- Maintaining open data portals on a website.

Consult: Gathering public feedback. This level involves soliciting public input, usually at a single point in time, while the institution conducting the activities retains full control over final decisions. Some acknowledgement of where the input impacted the final decisions is still expected. Common methods include:

- Surveys and focus groups to gather opinions on the use of AI in policing or healthcare.
- Public consultations, often online or in-person, to comment on draft AI legislation or guidelines.

Involve: Integrating public input. Here, the public is included more systematically throughout the process. Their feedback is reflected in decisions, although final authority remains with the institution. Common methods include:

- Hosting a series of participatory workshops where citizens help shape ethical AI principles.
- Including civil society representatives during the drafting of Fundamental Rights Impact Assessments (FRIAs)

Collaborate: Shared decision-making structures. This reflects a deeper, sustained partnership with citizens. Power is shared, and citizens are engaged across all stages of policymaking. Common methods include:

- Establishing community advisory boards that meet regularly to guide AI policy in local government.

- Implementing mini-publics (e.g. citizen assemblies or juries) to deliberate on complex AI issues.

Empower: Public holds final authority. The public has the final say in decision-making.

- Binding referendum on high-risk AI uses (e.g. facial recognition in public spaces).
- Citizen assemblies with decision-making power over specific AI projects or budget allocations.

Another dimension by which to categorize participation activities is by the extensiveness and complexity of the process. Activities at any of the above levels can vary greatly in this regard. An example of a lower complexity process is having a small number of representatives of an affected group in the room while filling out a questionnaire. An example of a higher complexity process is holding a mini-publics. This is a multi-year process where participants are randomly selected from the population and asked to engage in a long and structured deliberation process that includes learning from expert input. The purpose is for the cohort to make a statement or recommendation about a highly complex problem – outcomes influence, or even co-determine policy.

A CLOSER LOOK 5: PARTICIPATION IN POLICY MAKING THROUGH MINI-PUBLICS

A mini-public is a method for public engagement in policymaking. A cohort of randomly selected members of the public are brought together to provide input on a specific question about policy for a large, complex regulatory problem (*wicked problem*).

This process is usually initiated by a governing body and is led with the support of experts in the method, such as Deliberativa.[188] The process follows established methodology which involves the selected cohort in a carefully structured process of

[188] See more: https://deliberativa.org/en/

learning and deliberation taking place over months or years. The outcome is some form of recommendation from the cohort – the level to which this recommendation is, or is not, binding, is determined beforehand.

Mini-publics have been held in countries worldwide, as well as at the European level[189]. There is currently interest in using them to determine policy to address climate change, with several countries having already held mini-publics on this topic.

When is a mini-publics appropriate?

If a question of policy meets the following three criteria, a mini-publics may be an appropriate method:

1. The policy question is related to values.
2. The potential answers involve dilemmas like crossed interests.
3. The issue, and potential answers, relate to the personal experiences of the participants.

Prominent example: Ireland abortion legislation

Ireland held a mini-public in 2016– 2017 on the question of the legalization of abortion. It took the form of a citizens' assembly of 99 citizens, meant to be representative of the population. The assembly recommended a new legislative framework for legalized abortion, which informed a draft bill. A referendum on the draft legislation was held in 2018, where it passed by a majority.

Relevance to AI governance and FRIAs

The three criteria for an appropriate policy question apply to open questions related to AI and fundamental rights impacts, such as:

[189] Boswell, J., Dean, R., & Smith, G. (2023). Integrating citizen deliberation into climate governance: Lessons on robust design from six climate assemblies. Public Administration, 101(1), 182–200. https://doi.org/10.1111/padm.12883

- Determining the fundamental rights impacts of a specific AI System and/or the use of AI in a specific domain
- Informing decisions on which values should be embedded in AI systems
- Providing guidance on restrictions on use and red lines regarding AI Systems

3.4. Challenges of participation in FRIAs

There are barriers and challenges to meaningful participation in the FRIA process. Key challenges often brought up include: defining who participates and who counts as a representative of a relevant group, the complexity of incorporating differing views, funding and resource requirements and lack of interest or knowledge from the public.

Defining who participates and who counts as a representative of a relevant group. For meaningful participation in FRIAs, the relevant stakeholders need to be identified and included. This often means identifying specific groups of people, such as those who will be the most affected by the use of the deployed AI system or those that are most vulnerable to potential harms from it. Resources for identifying affected groups are available, for example, The Danish Institute of Human Rights has a guide which includes a stakeholder "power map"[190] which is a guide for assessment of which stakeholders are the most affected, while having least impact or power in the AI system deployment decisions. People falling into this category are key stakeholders and *"rights-holders."* Existing representative associations such as CSOs may be able to serve as effective representatives, however it is important to avoid including only the most privileged members of the affected groups, a phenomenon referred to by Kaminski and Malgieri as

[190] The Danish Institute for Human Rights. (2017, March 1). *A collaborative approach to Human Rights Impact Assessments*. https://www.humanrights.dk/publications/collaborative-approach-human-rights-impact-assessments

"elite capture of social marginalization." [191] In order to address the complexity and unsureness of capturing all the appropriate stakeholder representatives, it is important to continuously query who is not at the table.

Incorporating differing views. A diverse range of stakeholders is likely to have a diverse range of perspectives – the range of knowledge gained from diverse stakeholders is a primary motivator for including them, and the many differing views must ultimately inform one Impact Assessment outcome. The presence of conflicting views is not inherently a problem, the intention is that participants engage in dialogue together to jointly construct the assessment and decisions.[192]

Ensuring sufficient resources. Increasing public engagement and participation in a FRIA process requires funding and resources. Asking people to participate in these processes as representatives is a burden of time and effort. Some commonly contacted affected groups may experience participation fatigue. If CSOs become go-to representatives called in for these assessments, it becomes a drain on their resources. Compensation for participation in FRIAs as external parties needs to be considered. One proposed solution is for legislation or regulatory bodies to allocate dedicated public funds to support stakeholder engagement—such as financing stakeholder associations or participatory activities. These could be funded through mechanisms like fines imposed on large tech companies or taxes on firms whose technologies pose risks to fundamental rights. Such funding would help level the playing field, especially for small and medium-sized enterprises (SMEs), which may otherwise find participatory processes financially burdensome and view them as barriers to market entry.[193]

[191] Kaminski, M. E., & Malgieri, G. *Impacted Stakeholder Participation in AI and Data Governance* (SSRN Scholarly Paper No. 4836460). Social Science Research Network. 2024. Available at: https://papers.ssrn.com/abstract=4836460

[192] Id.

[193] Id.

Ensuring the credibility of participatory processes and avoiding conflicts of interest or undue influence requires careful consideration of who funds them.[194]

Public interest and engagement. A critique of participation is that members of the public may not be informed enough to either actively participate in a FRIA process, or to have any interest in participating in one, as they are not sure why it is relevant to them. One way to address this can be through increasing digital literacy. The AI Act refers to AI literacy, including provisions for providers and deployers of AI systems to ensure adequate digital literacy among their ranks. This definition of AI literacy in the Act also includes users and affected persons of AI systems having "skills, knowledge and understanding that allow […] awareness about the opportunities and risks of AI and possible harm it can cause." [195] Following this definition, AI literacy can be conceptualized as capacity building of affected persons– efforts to fulfil AI literacy requirements of the AI Act should by definition provide members of the public who may be affected by AI systems, with the motivation for, and tools for participation in a FRIA.

3.5. Collaborative (human rights) impact assessments and lessons of participation from environmental and social development projects

Human Rights Impact Assessments (HRIAs) are perceived as the gold standard in methodologies and processes that aim at incorporating the analysis of social issues in impact assessments, as they inherently incorporate a human rights framework. HRIAs are predated by Environmental Impact Assessments (EIAs) and social impact assessments (SIAs); the former having its roots "in [United States] environmental law, in which the National Environmental Policy Act ('NEPA') imposed the requirement to document the choices made in project development, and the rationales for them, in an environmental impact statement." [196] As already alluded to earlier in this report,

[194] The Danish Institute for Human Rights. *A collaborative approach to Human Rights Impact Assessments*. 1 March 2017. Available at: https://www.humanrights.dk/publications/collaborative-approach-human-rights-impact-assessments

[195] Article 3(56) of the AI Act

[196] Selbst, A. D. An Institutional View of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology (Harvard JOLT)*, *35*, 117. 2021. Available at: https://heinonline.org/HOL/Page?handle=hein.journals/hjlt35&id=123&div=&collection=

Algorithmic Impact Assessments, to varying degrees of similarities, are all NEPA-based models, by default this also applies to FRIAs.

Public participation is seen as an essential feature of the HRIA, however, as extensively covered throughout this work, there are different ways to meaningfully involve affected people and stakeholders. In a nutshell, all this is dependent on companies making a genuine commitment to engage beyond legal requirements. In this context, collaborative HRIAs[197] emerge as a novel theoretical approach to a joint process to be undertaken by a company and project-affected people. Furthermore, the idea is that government representatives, and other stakeholders, are also involved in different roles, primarily those concerned with providing mediation, external oversight and specialized expertise. The concept of a collaborative HRIA is presented as a way to bring reliability, and validity, to otherwise company-commissioned HRIAs, by co-designing and co-implementing a HRIA between companies and project-affected people. This type of approach, if materialized, could contribute to the realization of human rights goals in shorter time spans, while also enabling longer-term goal-setting, where companies become privy to project-affected people grievances and concerns first hand, thus possibly contributing to higher commitments from their side.

Lessons from practice. Drawing from the literature, different example cases are described below. The objective of including these examples is two-fold, first illustrate different participatory techniques in practice, and second, implicitly show how even despite achieving a certain degree of meaningful participation in HRIAs, that is not the be-all, and end-all, given how enforcement ultimately plays a defining role in the legitimizing this type of processes, especially when political and financial interests are at play.

[197] Columbia Center on Sustainable Investment, Danish Institute for Human Rights, and Sciences Po Law School Clinic. A Collaborative Approach To Human Rights Impact Assessments. March 2017. Available at: https://ccsi.columbia.edu/sites/ccsi.columbia.edu/files/content/docs/publications/A-Collaborative-Approach-to-HRIAs_Web.pdf

Two examples cases have been reproduced from the World Bank,[198] to the extent of including context, participatory techniques, impact, and limitations of SIAs. A third example case was driven by the humanitarian organization Oxfam, and relates to a land conflict. The novelty here is that it presents a comparative of two HRIAs performed for the same project, but from different perspectives. This example case could be used to drive the conversation forward in terms of collaborative HRIAs, where community members and companies co-own the HRIAs, process, and results. Lastly, the example case of the Goldcorps's Marlin Mine in Guatemala[199] is also included, as an example of a failed non-binding implementation of both, an ESIA and a HRIA.

EXAMPLE CASE 2: SIAs IN PRACTICE ACCORDING TO THE WORLD BANK

Project 1: *Argentina rural poverty alleviation project*

Context: The project focused on the assessment of needs faced by the targeted population, people living in rural areas lacking access to essential resources and means for living, and the development of efficient mechanisms to channel resources to them. Other stakeholders included, indigenous people, women's groups, and small producers.

Participatory techniques: Regional workshops with participants representative of the broad spectrum of the rural economy; followed by surveys to 1000 households across three different provinces.

The workshops were organized and facilitated by a team of local experts across fields, sociology, agricultural economics, agronomy, and community organization.

[198] Read more: Rietbergen-McCracken, J. & Narayan, D. *Participation and social assessment : tools and techniques (English).* Washington, D.C. : The World Bank. 2010. Available at: http://documents.worldbank.org/curated/en/673361468742834292

[199] Canadian Network on Corporate Accountability. Case Study: Goldcorp Inc.'s Marlin mine – Environmental contamination and human rights abuses. 2023. Available at: https://cnca-rcrce.ca/2023/02/14/case-study-goldcorp-inc-s-marlin-mine-environmental-contamination-and-human-rights-abuses/

Impact: The workshops counted with strong participation by poor rural communities, resulting in strengthened relationship between local committees and NGOs to co-decide on section and implementation of rural development subprojects.

The workshops had an allotted budget, co-financed by the central government, however there was no disclosure of how these funds were used.

Challenges: While engagement was strong, the initial planning had to be adapted and reduced to account for women participants that could not attend for the duration of the 3-day workshop schedule.[200]

Analysis and processing of workshop findings resulted in a time-consuming exercise (4 months).

Project 2: *Azerbaijan Baku water supply project*

Context: Sustainable implementation of a water supply system for 2.5 million people by the central government, with the need to prioritize the needs of disadvantaged populations, whilst minimizing negative impact on other groups.

Participatory techniques: Researchers at the Baku University and the Baku Water Agency partnered to impart two rounds of rapid user surveys, approximately 800 households participated. Additionally, other affected parties were also surveyed, i.e., industry and agriculture members. The results of the surveys were used to perform various consultative activities, e.g, informal discussions with individuals from affected communities, formal discussions with international organizations and national and local government; interviews with private sector providers of water, and focused

[200] As an observation, in performing any variation of a HRIA, it is not feasible nor acceptable to expect project-affected participants to dedicate long periods of time to participation. It is advisable that the implementation of an HRIA accounts for potential wage loss or other responsibilities, such as child-rearing. Parts of the budgets could be destined to a fair compensation for community participants.

assessments with a small subset of households.

Following this, a stakeholder workshop was imparted with 72 participants across all sectors of society, user groups, government ministries, local NGOs, academic researchers, local experts, media members, and donors.

The workshops had an allotted budget, used to cover data collection, analysis, and write up costs.

Impact: The impact assessment made a significant contribution to policy dialogue between the central government and the World Bank. Specifically, the stakeholder workshop was found to be instrumental in generating a high level of consensus on the project. Wide media coverage of the event also contributed to strengthening the support and ownership by different segments of the population.

A practical outcome resulted in the development of a project component to meter and bill for water, and additional measures to support and improve water conservation and cost recovery.

Challenges: Most of the challenges identified with this impact assessment were associated with funding available. The timeline of the funders placed constraints on the data analysis phase and impacted the overall extent as to how the results were incorporated in the final recommendations for the project.

EXAMPLE CASE 3: LAND CONFLICT IN BRAZIL & COMMUNITY - AND COMPANY-BASED HRIA COMPARATIVES[201]

Context: Usina Trapiche is a company dedicated to providing agriculture processing

[201] Adapted from: Tamir I, Zoen S. Human Rights Impact Assessments in a Brazil Land Conflict: Towards a Hybrid Approach. *Business and Human Rights Journal*. 2(2):371-377. 2017. Available at: https://doi.org/10.1017/bhj.2017.16

services, with a sugar cane mill factory that operates from the islands of the Sirinhaém estuary in Brazil;[202] these lands are also home to local fishing communities. For many years, both co-existed, however from the 1980s on, the company started on their efforts to drive these communities out of their homelands, under the premise that the land is under their concession, and supported by the local governments.

Local communities have documented forceful and violent evictions from their homes for years; with the impact of these actions also contributing to the loss of their livelihood.

With the support of a local organization, CPT, the local communities have presented a petition to designate the area as a protected reserve to the central government, a move that has been in contention for years and has seen no favourable resolution.

While the government remains on the side of the sugar mill, Oxfam's Behind the Brands campaign contributed to get Coca-Cola and PepsiCo to engage in this conflict, under the premise of their "zero tolerance to land grabs in their supply chains." Both of these companies have an influential role and stakes in the sugar supply chain, and both agreed to perform a land impact assessment in Brazil; however, each company chose a different approach, yielding different results.[203]

Company-led HRIA carried out by Coca-Cola and Pepsi

Participatory technique: Coca-Cola commissioned a baseline study, made up of 21 visits to the mill, and 120 visits to farms. During this time, 111 interviews were performed with different stakeholders, and an additional 929 with farmworkers.

PepsiCo carried out an audit on three mill sites, the audits allotted three days per site.

Findings: External evaluators of the land assessments found that Coca-Cola's efforts

[202] Brazil is the number one producer of sugar from sugarcane in the world.

[203] Read more: Oxfam. Oxfam Briefing Note: Land rights and soda giants Reviewing Coca-Cola and PepsiCo's land assessment in Brazil. October 2016. Available at: https://www-cdn.oxfam.org/s3fs-public/bn-land-rights-soda-giants-brazil-201016-en.pdf

to the assessment were comparatively stronger in terms of scope and quantity and quality of engagement with stakeholders. Coca-Cola also self-disclosed the entire report.

The actual results of the study were also deemed to be robust, with recommendations to work towards better policies around land ownership and forced labour, and for the need to put in place mechanisms for effective grievance systems.

On the other hand, the efforts by PepsiCo's were deemed limited across different dimensions, especially in terms of stakeholder engagement and the disclosure of the results; they only published a summary of the audit.

The audit heavily focused on legal compliance, with the findings reporting no evidence of human rights, land rights, health, safety, or environmental violations.

Community-led HRIA carried out by CPT and Oxfam

Participatory Technique: Local allies and community members closely collaborated with CPT, focusing on documenting positive and negative impacts of all duty-bearers involved in the case. For the HRIA, a series of interviews were conducted, 60 with community members, eight with government offices, nine with civil society organizations, and two companies. Usina Trapiche declined to participate.

Findings: The expulsion of community members from the land constitutes a violation of Brazilian law and international human rights conventions. Further, the results recommended the local government grant protected status to the land, so locals could return to their homes and traditional ways of life, with legal backing. Additional findings were in relation to evidence of the effects water supply pollution derived from the industrial activity of the factory has on the local communities and the environment.

EXAMPLE CASE 4: GOLDCORP INC.'S MARLIN MINE

Context: The Marlin Mine is a gold mine located in the San Marcos Department in Guatemala. In 2005 and prior to the takeover of the mining operation by Goldcorp, Inc., a Canadian gold production company, local residents filed a complaint against the previous owners of the mining operation. The local community alleged to the World Bank Group that the operation was taking place without adequate consultation and in violation of the rights and wishes of indigenous people.

The representative of the World Bank Group uncovered the existence of an Environmental and Social Impact Assessment (ESIA), which was deemed to be highly technical and to lack evidence of any engagement with the local Indigenous population affected by the mine.

Following the Goldcorp, Inc. takeover, the company commissioned a Human Rights Impact Assessment (HRIA) for the mining operation in 2007. Originally undisclosed, the resulting report[204] was released in 2010 given the suggestion of the Inter-American Commission on Human Rights (IACHR) to suspend all mining activity in the Marlin Mine.

Techniques: The assessment team spent 180 days in Guatemala, and 80 days spread out across the San Marcos Department. They claimed to have been involved in continuous meetings with local organizations, municipal and community authorities, and residents building capacity to foster trust and put in place a mechanism to operationalize the assessment. Following this, interviews and focus groups were conducted for a shorter stint with people from most of the stakeholder groups. The commissioned company reported to have accumulated more than 250 hours of

[204] Read more: On Common Ground Consultants, Inc. Human Rights Assessment Of Goldcorp's Marlin Mine. May 2010. Available at: https://s24.q4cdn.com/382246808/files/doc_downloads/2020/09/OCG_HRA_Marlin_Mine_June_7.pdf

interviews with local participants, including land sellers, businesses, contractors, employees, community residents, and authorities working in the mines. The report also explicitly states that local communities most opposed to the mines rejected invitations to participate in the HRIA, thus their perspective could not be incorporated in the results.

Impact: The HRIA was carried out during a time of high polarization with regard to mining operations in Guatemala. Tensions were already high and brought at odds the interests of the Canadian corporation, the Guatemalan Government and the Maya peoples. In the report itself, the commissioned company acknowledged how the HRIA had become “a proxy for the larger debate over mining in Guatemala.” They also deemed the company had been reactive and defensive in dealing with the complaints uncovered by the HRIAs.

Some of the many recommendations of the HRIA included the call for an immediate halt to all land acquisition, exploration activities, mine expansion projects, and to identify and support at-risk families that had been impacted by the mining operation and were being denied access to basic services.

Legacy: Despite the HRIA recommendations and political and social pressure, the Guatemalan Government did not adhere to IACHR guidelines and mining activity continued until 2017. During this time, the company continued to carry out their activity in opacity, demonstrating little evidence of efforts carried out to mitigate the negative impact of their activities on the local communities and on the environment.[205]

[205] Read more: Canadian Network on Corporate Accountability. Case Study: Goldcorp Inc.’s Marlin mine – Environmental contamination and human rights abuses. 14 February 2023. Available at: https://cnca-rcrce.ca/2023/02/14/case-study-goldcorp-inc-s-marlin-mine-environmental-contamination-and-human-rights-abuses/

3.6. Meaningful participation in FRIAS: Takeaways

Meaningful means that the public's input actually has an impact on decisions made. This includes having mechanisms for refusal of uses of AI and specific AI Systems. Participation can be part of all stages of AI system design, development and assessment, including the continuous monitoring of deployed systems.

There are challenges that come with participation, as well as open questions regarding best practices. However, there is also a large body of resources, guides and previous experiences to draw from. Additionally, there are many options for types and methods of participation, allowing for identification and application of the most contextually relevant processes.

4. Transparency and accessibility of FRIAs

Transparency regarding the use, function, and risks of AI systems, high-risk or otherwise, is crucial for enabling public awareness, scrutiny, and resistance—particularly in contexts where these systems are deployed by the state, given that they are subjected to higher standards when it comes to transparency rules; however sometimes corporations are also voluntarily transparent or required to be, in the existence of transparency policies, e.g., inspections, labelling obligations. As Okidegbe[206] emphasizes, transparency can significantly increase public knowledge about whether and how algorithmic tools are used, informing litigation, policy, and resistance efforts, especially by affected people and communities. Similarly, Selbst highlights that access to information about decisions and identified risks is necessary for enabling recourse and democratic input.[207]

[206] Okidegbe, N. To Democratize Algorithms. *UCLA Law Review*, *69*. 2022. Available at: https://www.uclalawreview.org/to-democratize-algorithms/

[207] Selbst, A. D. An Institutional View of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology (Harvard JOLT)*, *35*, 117. 2021. Available at: https://heinonline.org/HOL/Page?handle=hein.journals/hjlt35&id=123&div=&collection=

The AI Act defines a series of transparency requirements for a given AI system, with these varying depending on the system's risk level.[208] However, in the absence of mechanisms for accountability and power redistribution,[209] transparency alone offers no guarantee that affected communities can alter the systems shaping their lives. Thus, while transparency is a right and a necessity, it must be pursued critically and in tandem with efforts to empower affected groups, including the most vulnerable among them.[210]

The rest of this section contextualizes transparency in relation to Article 27, and how it interacts with other relevant articles in the AI Act, and other transparency mechanisms found in EU regulation. The section closes by identifying barriers to transparency through the perspective of algorithmic accountability reporting.

4.1. The availability of FRIAs according to the AI Act

Article 27 requires the relevant deployers to share a completed FRIA template with the MSA (discussed in Section 1). As the template has not yet been created, it is not clear what level of detail will be required by the template and if all documentation created under the FRIA will be included.

It is important to highlight that Article 71 of the EU AI Act establishes an EU-wide public database for certain AI systems, the setup and maintenance of which is the responsibility of the European Commission and Member States. This requires providers of specific high-risk AI systems (and some systems in high-risk areas deemed not high-risk) to register information such as system purpose and technical details before placing them on the market or into service. Furthermore, deployers "that are public authorities, Union institutions, bodies, offices or agencies or persons acting on their behalf" [211] must register their use of high-risk systems and are specifically required to

[208] See more: The EU AI Act. Transparency Obligations and The interplay between different transparency-related provisions. Available at: https://www.euaiact.com/key-issue/5

[209] Critique of the AI Act mentions how it is a law that lacks "in individual rights and recourse" mechanisms.

[210] Okidegbe, N. To Democratize Algorithms. *UCLA Law Review*, *69*. 2022. Available at: https://www.uclalawreview.org/to-democratize-algorithms/

[211] AI Act Article 49(3)

include a summary of their FRIA and DPIA, where applicable.[212] Limiting the requirement to publish a summary of the FRIA excludes deployers in sectors like banking and insurance[213]. Additionally, AI systems in the areas of biometrics, law enforcement, migration, asylum and border control management[214] are to be registered in a non-public section of the database.

The exceptions to registration in the public database, as well as the requirement to include only a summary of the FRIA may impede the goals of public accountability and knowledge transfer. However, according to Recital 131 deployers not mandatorily required should still be entitled to register the systems they use voluntarily.[215] A more complete and thorough repository would improve its effectiveness as an oversight and accountability tool.

Pénicaud, in a report completed for the Spanish IA Ciudadana coalition, recommends using more complete national AI registers to complement the EU AI Act database in three ways[216]:

- Explicitly record high-risk AI systems (as defined by the AI Act) and label them accordingly, while also encouraging registration of other impactful systems.
- Use the AI Act's voluntary provisions to allow and promote registration of AI systems and fundamental rights impact assessments (FRIAs) by private sector deployers, especially in finance and insurance where registration isn't mandatory.
- Include information (e.g., accuracy, limitations) received by deployers from third-party AI providers during activities such as procurement in the publicly

[212] The full list of data required from these deployers is found in Annex VII Section C of the AI Act

[213] Pénicaud, S. Making Algorithm Registers Work for Meaningful Transparency. IA Ciudadana. 2025. Available at: https://iaciudadana.org/2025/03/13/making-algorithm-registers-work-for-meaningful-transparency/

[214] These categories are defined in points 1, 6 and 7 of Annex III of the AI Act

[215] AI Act Recital 131

[216] Pénicaud, S. Making Algorithm Registers Work for Meaningful Transparency. IA Ciudadana. 2025. Available at: https://iaciudadana.org/2025/03/13/making-algorithm-registers-work-for-meaningful-transparency/

available database.``[217]

4.1.1. Fundamental Rights Oversight Bodies' access to documentation

As discussed in Section 1.4.2, Article 77 of The AI Act states that Fundamental Rights Oversight Bodies can "request and access any documentation created or maintained under this Regulation [...] when access to that documentation is necessary for effectively fulfilling their mandates." [218] It is not specified if this documentation does or does not include FRIA documents, the questionnaire, or the FRIA summaries stored in the non-public area of the AI database created in Article 71. One interviewee stated that access to the submitted FRIA documents would be a minimum of what these authorities can expect to receive. In addition to the filled out template, the AI Act refers to a summary of the FRIA, which in some cases must be made publicly available.

He also highlighted that other documents created under Article 26: "Obligations of Deployers of High-Risk AI Systems" are another source of deployer related documents that should be expected. Article 26 obliges deployers to implement technical and organisational measures to ensure proper use of AI systems and includes obligations related to human oversight, data management, continuous monitoring, logging and informing affected people of the use of AI in the workplace and for decision-making about people.

4.1.2. Transparency mechanisms in the EU Principles and GDPR

In the European Union, existing transparency mechanisms are encoded in regulation. For instance, foundational EU treaties and more specifically the EU Freedom of Information (FOI) Law entitles EU citizens to some relevant data and document access

[217] AI Act Article 13(2) requires providers to provide this and other information about high-risk systems to deployers
[218] Article 77(1) of the AI Act

by EU institutions, bodies and agencies.[219,220] This entitlement also trickles down to Member State level. In the case of Spain, for instance, the transparency law came into effect on 10 December 2014, encoding the right of access to information for its citizens.

The right to information is also baked into the GDPR, as this law was enacted with transparency as a core principle.[221] The GDPR thus gives rights of access to data subjects over their personal data[222] in addition to the affordances aforementioned with public administrations via FOI laws, as private businesses are also obliged to a series of legally-binding requirements with regard to the information they give data subjects: intelligible and easily accessible, using a clear language, free of charge, in written form or by other means, containing all relevant information requested; lastly, it needs to be provided by an appropriate measure and in an appropriate time. There are four exceptions to these requirements, however, their interpretation should be restrictive. They are: when the data subject already has the information, when it is impossible to do so, when it is mandated by law, or when there is an obligation of secrecy.[223]

4.1.3. The right to explanation

Another relevant piece of regulation in terms of transparency, is the right to explanation, which is explicitly mentioned in the AI Act's Article 86, and was implicitly alluded to in Article 22 of the GDPR, which mentions automated individual decision-making on data subjects.

In very broad strokes, the AI Act does not grant affected persons explicit mechanisms to contest decisions derived from the use of an AI system. However, explanations are needed in order for them to be able to exercise their rights. While the scope and interplay of these regulations is a current topic of legal text interpretation, it is understood that the AI Act and GDPR are complementary, and should effectively enable

[219] Read more: List of all EU bodies subjected to this law. Available at: https://eur-lex.europa.eu/summary/chapter/01.html?expand=0105,010504,010502#arrow_010502
[220] See exceptions: in Article 4 of the EU FOI law.
[221] GDPR Article 5 (1) (a)
[222] GDPR Article 15
[223] Read more: GDPR, Article 12

an affected person or data subject to obtain an explanation that is "clear and meaningful," meaning, it should provide enough information to enable an effective claim on grounds of anti-discrimination or violations of any other right.[224] Kaminski and Malgieri draw a straight line between the main elements of an AI system necessary for deployers to complete a FRIA (collective level) and the necessary elements related to an AI systems' decision-making process that can prompt invoking the rights of an affected person (individual level).[225]

Discussing meaningful explanations is outside the scope of this report, however it is important to highlight how for a long time explanations in regard to individual decision-making supported by semi- and automated means has been a hot topic, with already legal precedent in the Court of Justice.[226] Additionally, there is also the buzz worthy case of the SCHUFA credit scoring algorithm, which has been in the press and courts for years, closely binding the topic of transparency and explanations when it comes to consequential decision-making via algorithmic systems.[227,228,229,230]

Whilst an open debate regarding necessary explanation desiderata from a technical, legal, and social perspective remains,[231] explanations are a cornerstone in the pursuit of transparency and accountability from those deploying AI systems.

[224] Kaminsiki M. & Malgieri G. The Right to Explanation in the AI Act. University of Colorado Law Legal Studies Research Paper No. 25-9. 8 March. 2025. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5194301

[225] Id.

[226] Read more: Court of Justice of the European Union. Judgment of the Court in Case C-203/22 | Dun & Bradstreet Austria. Press Release N 22/25. 27 Febraury 2025. Available at: https://curia.europa.eu/jcms/upload/docs/application/pdf/2025-02/cp250022en.pdf

[227] See more: Open SCHUFA Campaing by AlgorithmWatch. Available at: https://openschufa.de/english/

[228] Matthias Spielkamp. EuGH-Urteil zum Scoring: Eine Ohrfeige nicht nur für die Schufa AlgorithmWatch. 8 December 2023. Available at:https://algorithmwatch.org/de/eugh-urteil-schufa-scoring/

[229] Matthias Spielkamp. LinkendIn Post. May 2025. Retrieved 10 May 2025. Available at: https://www.linkedin.com/posts/matthiasspielkamp_schufa-activity-7313568912099479553-FMZ6?utm_source=share&utm_medium=member_desktop&rcm=ACoAABclHfsBqyNrYmNjIcQUWhyZUS5rtJbYMC8

[230] See more: SCHUFA Score-Simulator. Available at: https://www.schufa.de/scorechecktools/scoresimulator/

[231] Bringas Colmenarero, A. et al. How should an explanation be? A mapping of technical and legal desiderata of explanations for machine learning models. *International Review of Law, Computers & Technology*, 1–32. 2025. Available at: https://doi.org/10.1080/13600869.2025.2497633

Moreover and despite the existence of laws that require transparency from governments and private actors, the absence of good will or adequate enforcement will limit or impede exercising any rights.

4.2. Transparency through algorithmic accountability reporting

Journalists have been news breakers on consequential cases of algorithmic harm, some snapshots of the impact of their reporting on algorithmic accountability have already been described earlier in this report via the case examples of the RisCanvi and SyRI systems.

KEY CONCEPT

ALGORITHMIC ACCOUNTABILITY REPORTING[232]

Algorithmic Accountability Reporting has a closer look at the increasing number of ‘opaque’ algorithms that lawmakers, government officials, private companies rely on, to find out how they work and the effects they have.

These two examples build on earlier investigative work, such as the now quintessential ‘Machine Bias’ story broken by Julia Angwin and other journalists at ProPublica in 2016. There, they shed light on the COMPAS algorithm, and on the consequences of its use on the lives of incarcerated individuals in North American prisons.[233] The impact of this story in the debate concerning AI systems and its responsible use is unquestionable. Inherently part of this debate, is that of transparency and accountability in relation to how these systems work from a strictly technical perspective; but also how they are ideated and then deployed for a wide array of high-stake tasks at scale on assuming populations. These processes, to this day, continue to be often riddled by secrecy and

[232] Diakopoulos, N. Algorithmic Accountability Reporting: On the Investigation of Black Boxes. Tow Center for Digital Journalism. February 2014.

[233] Angwin, J. et al. Machine Bias. ProPublica. 23 May 2016. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

opacity at the hands of both public and private actors, and thus perpetuating the myth of the 'black-box'.

In light of this, the experiences of journalists navigating existing transparency mechanisms are used in this report as a proxy, to illustrate the barriers to achieving meaningful transparency.

4.2.1. Barriers to transparency: Lessons from the field

As already stated, FOI laws facilitate access to some data from all levels of government in countries that have transparency laws in place; a good proxy to measure the efficacy of these laws are indexes that assess the level of freedom given to journalists to carry out their reporting duties with independence from political influence and integrity.[234,235] The preemptive expectation is that this mechanism will be of use in the event the public and their advocates necessitate access to substantial information regarding FRIAs. However, the reality might be another one, given the limitations of the existing text in prescribing stronger transparency requirements in this context.

For cases when there is no incentive to disclose data, or when they pertain to private actors, access is further constrained and investigation requires other approaches to data gathering, e.g., industry insiders, open-source intelligence tools, and personal stories of individuals affected by algorithmic decision-making. Different combinations of these approaches have and can contribute to reverse engineering algorithms to further characterise them, albeit within the existing limitations of this methodology.[236] The collaboration of the insider, for example, can be crucial to shed light on key information that is withheld from the public, even if their participation is kept off-the-record or credited anonymously; the latter though can raise concerns of credibility. This also relates to the lack of whistleblower protection at a public and

[234] See the World Press Freedom Index 2025 as an example. Available at: https://rsf.org/en/index

[235] See also: Reporters without Borders. World Press Freedom Index 2025: Spain Fact-File. Available at: https://rsf.org/en/country/spain

[236] Diakopoulos, N. Algorithmic Accountability Reporting: On the Investigation of Black Boxes. Tow Center for Digital Journalism. February 2014.

private level, and the need for safeguards that can empower industry and government employees to report and expose unethical, illegal, and anti-constitutional practices associated with AI development and deployment.

In all cases, different challenges arise as journalists set out to probe for data and information on algorithmic systems. Below, some of the most prominent ones regarding FOI requests in practice are enumerated, including real world examples.

1) *Complex legal structure and absence of a culture of transparency*
    - One of the biggest critiques of the Spanish Transparency Law has to do with how it is operationalized. Employing a decentralised approach has resulted in the existence of different laws across the Spanish territory as per its 17 autonomous regions, all with their different request systems. In addition, public administration at local levels also have their own ordinances as well. The overstated lack of a culture of transparency contributes to both responses of poor quality to the requests, and also to a high number of unanswered ones. [237,238]
2) *Trade secrets, data privacy concerns and other confidentiality protections*
    - Private actors are more likely to allude to trade secrets and privacy concerns when asked to voluntarily disclose details on their AI systems or data infrastructures. However, these can also be reasons for denial of requests under FOI laws by the public administration. They are also often used as an excuse to exclude certain data from disclosed documents, for

[237] Access Info Europe. Spain: Five Years of the Transparency Law. 10 December 2019. Available at: https://www.access-info.org/2019-12-10/spain-five-years-of-the-transparency-law/

[238] Álvarez del Vayo, M. & Adrián Maqueda, A. Más de un millar de resoluciones del Consejo de Transparencia han sido ignoradas desde 2016. Civio. 22 May 2025. Available at: https://civio.es/transparencia/2025/05/22/mas-de-un-millar-de-resoluciones-del-consejo-de-transparencia-han-sido-ignoradas-desde-2016/

example the training data used for model development or the source code.[239]

- Similarly, this problem also arises when public actors are involved with private actors, e.g., a procurement contract to develop an AI system, regulation negotiations; in these instances, governments may opt to not disclose any information alluding to trade secrets or privacy concerns. A timely and related example of this corresponds to reporting by FragDenStaat fellows investigating the influence of Big Tech companies, e.g, Google, Meta, Microsoft, on the AI Act negotiations. They report to have been in contact with European governments, including Spain's Ministry of Economic Affairs and Digital Transformation, for over a year via FOI requests. The negative replies received overwhelmingly cited reasons concerning commercial interests or privacy concerns. In the case of Spain specifically, the Ministry alluded to the nonexistence of official records in the event that meetings took place.[240]

3) *Inadequate, untimely, or no answers to FOI requests*

- Stonewalling, or the refusal to cooperate by denying requests or not answering to them, is also a very typical response from the government in topics related to AI. Journalists involved in investigating the SyRI case reported to have contacted different governments across the European Union for at least two years. From all of them, only one responded positively to the request, the Government of Rotterdam. Luckily, they released the source code, the list of variables, evaluations of the model's

[239] Geiger, G. et al. Suspicion Machines: Unprecedented experiment on welfare surveillance algorithm reveals discrimination. Lighthouse Reports. 2 March 2023. Available at: https://www.lighthousereports.com/investigation/suspicion-machines/

[240] Vidal, N., & Muftić, N. AI regulation and industry influence: The public is locked out. FragDenStaat. 14 May 2025. Available at: https://fragdenstaat.de/en/articles/exclusive/2025/05/ai-regulation-and-industry-influence-the-public-is-locked-out/

performance, and Rotterdam's handbook for data scientists. At a later day, they also received further material.[241]

- The case of the BOSCO algorithm in Spain is a good example of the lack of cooperation by the government, with journalists and researchers at odds with the Ministry of Ecological Transition since 2018.[242] The algorithm-supported process to allocate a social bond to front electricity costs leaves an elevated number of eligible applicants with denied applications; this kickstarted investigative efforts to uncover details as to why this was happening, and a public call for algorithmic transparency to the public administration.[243] The government initially refused to cooperate by altogether denying access to the system to undergo independent auditing, alluding to public safety concerns.

4) *Lawsuits*

- Elaborating further on the BOSCO case, journalists and the government have continued their exchanges over the source code, going as far as to the Supreme Court. Following appeals and counter appeals to lower courts and the Transparency Council, journalists continue their now legal efforts to make disclosable the source code used to train AI systems deployed by the public administration for high-risk decision-making;[244] on the other hand, the State Attorney General's Office continues to deny access, citing reasons concerned with intellectual property with sights set

[241] Geiger, G. et al. Suspicion Machines: Unprecedented experiment on welfare surveillance algorithm reveals discrimination. Lighthouse Reports. 2 March 2023. Available at: https://www.lighthousereports.com/investigation/suspicion-machines/

[242] Civio. La transparencia de los algoritmos públicos, en juego: Civio presenta el recurso sobre BOSCO ante el Tribunal Supremo. 30 January 2025. Available at: https://civio.es/novedades/2025/01/30/la-transparencia-de-los-algoritmos-publicos-en-juego-civio-presenta-el-recurso-sobre-bosco-ante-el-tribunal-supremo/

[243] Civio. El Gobierno se niega a explicar cómo funciona su aplicación para conceder o no el bono social. 28 November 2018. Available at: https://civio.es/transparencia/2018/11/28/el-gobierno-se-niega-a-explicar-como-funciona-su-aplicacion-para-conceder-o-no-el-bono-social/

[244] Id.

on creating a legal precedent that could have serious consequences for algorithmic accountability in Spain.[245,246]

- The BOSCO case is not the only ongoing legal dispute between journalists and the government. Similarly, an algorithm deployed and used by Spain's National Institute of Social Security (INSS) to detect social security fraud from 2018, has been under scrutiny since 2022. Given the denied requests by the Ministry of Inclusion, Social Security and Migration to disclose details such as training data or the source code, a positive answer to an appeal to the Transparency Council triggered the Ministry to challenge the Council's decision via a lawsuit through the National Audience earlier this year; the case is still undergoing litigation.[247,248,249]

In including these real world examples, the goal is to illustrate the reality of the often complex, long, and sometimes costly process that professional and trained individuals and, in some cases, teams of people have to undergo to uncover crucial details on AI systems. Ultimately, underscoring the long road ahead to attain meaningful transparency for all.

Closing remarks

The findings in this report have been presented in an attempt to elaborate on the question of what is considered a "meaningful" implementation of Article 27 under the AI Act. As already established, this constitutes active and impactful participation by those

[245] Civio. Por la transparencia de las decisiones automatizadas. 2025. Available at: https://civio.es/transparencia-decisiones-automatizadas/

[246] Civio. Mientras el Constitucional colombiano reconoce el derecho a la transparencia algorítmica, España trata de blindar su opacidad. 14 April 2025. Available at: https://civio.es/novedades/2025/04/14/transparencia-algoritmica-colombia-abogacia-estado-bosco/

[247] Jiménez Arandia, P et al. Spain's AI Doctor. Lighthouse Reports. 17 April 2023. Available at: https://www.lighthousereports.com/investigation/spains-ai-doctor/

[248] Interview with Mr. Pablo Jiménez Arandia, May 12 2025

[249] Id.

affected by AI system deployment and putting in place clear and effective mechanisms of transparency with regard to the FRIA process and to the obtained results in order to promote and create a sustainable culture of enforcement and accountability.

At a more fundamental level, the purpose of FRIAs is to prevent fundamental rights infringements at the hands of a high-risk application of AI systems. Something that requires, not comprehensively:

- implementing a robust and reflective FRIA process that goes beyond a mere tick boxing exercises and standardized close-end templates;
- prioritizing the careful identification of all affected groups;
- including affected groups and or their representatives as active participants in FRIA processes, and granting them consequential decision-making power;
- assessing critically the necessity and proportionality of using an AI system, considering non-automated alternatives, and the preemptive and clear definition of red lines that should not be trespassed;
- being prepared to decide against deployment if the system poses unacceptable risks to fundamental rights;
- clarifying oversight mechanisms during the FRIA process, and also with regard to the evaluation of the obtained results by independent parties.
- defining clear responsibilities of the enforcement authorities, e.g., MSAs, Fundamental Rights Authorities, and ensuring their full independence to promote the enforcement of the regulation;
- creating efficient and effective mechanisms for complaints by third parties, such as CSOs, rights groups, and the public, to the MSAs;
- making sufficient information on the results of the FRIAs available to the public.

Future work must examine the range of potential parameters across which to assess risks to fundamental rights to determine which are most effective for identifying and preventing fundamental rights violations by deploying AI systems. Additionally, further case studies on the FRIA process for AI systems deployed by private sector actors, particularly for systems related to creditworthiness and insurance risk, which will be subject to the requirements of Article 27, are also needed.

Bibliography

Reports and journalistic articles

1. A & O Shearman. Zooming in on AI - #14: Enforcement of the AI Act. A&O Shearman. 2025. Available at: https://www.aoshearman.com/en/insights/ao-shearman-on-tech/zooming-in-on-ai-14-enforcement-of-the-ai-act
2. Agencia Estatal Boletín Oficial del Estado. Real Decreto 729/2023, de 22 de Agosto, Por El Que Se Aprueba El Estatuto de La Agencia Española de Supervisión de Inteligencia Artificial, Pub. L. No. Real Decreto 729/2023, BOE-A-2023-18911 122289. 22 August 2023. Available at: https://www.boe.es/eli/es/rd/2023/08/22/729
3. Algorithm Audit. A comparative review of 10 Fundamental Rights Impact Assessments (FRIA) for AI-systems. 2024. Available at: https://algorithmaudit.eu/knowledge-platform/knowledge-base/comparative_review_10_frias/
4. Álvarez del Vayo, M. & Adrián Maqueda, A. Más de un millar de resoluciones del Consejo de Transparencia han sido ignoradas desde 2016. Civio. 22 May 2025. Available at: https://civio.es/transparencia/2025/05/22/mas-de-un-millar-de-resoluciones-del-consejo-de-transparencia-han-sido-ignoradas-desde-2016/
5. Angwin, J., et al. Minority Neighborhoods Pay Higher Car Insurance Premiums Than White Areas With the Same Risk. ProPublica. 5 April 2017. Available at: https://www.propublica.org/article/minority-neighborhoods-higher-car-insurance-premiums-white-areas-same-risk
6. Attard-Frost, B. The Death of Canada’s Artificial Intelligence and Data Act: What Happened, and What’s Next for AI Regulation in Canada?. The Montreal AI Ethics

Institute. 17 January 2025. Available at: https://montrealethics.ai/the-death-of-canadas-artificial-intelligence-and-data-act-what-happened-and-whats-next-for-ai-regulation-in-canada/

7. Australian Human Rights Commission. HRIA Tool: AI in Banking. 28 September 2023. Available at: https://humanrights.gov.au/our-work/technology-and-human-rights/publications/hria-tool-ai-banking
8. Basel Committee on Banking Supervision. Principles for effective risk data aggregation and risk reporting. January 2013. Available at: https://www.bis.org/publ/bcbs239.pdf
9. Bellio López-Molina, Naiara. In Catalonia, the RisCanvi algorithm helps decide whether inmates are paroled. Algorithmwatch. 21 May 2021. Available at: https://algorithmwatch.org/en/RisCanvi
10. Cabrera, L. L., Duprat-Macabies, A., & Maier, M. CDT Europe's AI Bulletin: March 2025. Center for Democracy and Technology. 26 March 2025. Available at: https://cdt.org/insights/cdt-europes-ai-bulletin-march-2025/
11. Canadian Network on Corporate Accountability. Case Study: Goldcorp Inc.'s Marlin mine – Environmental contamination and human rights abuses. 2023. Available at: https://cnca-rcrce.ca/2023/02/14/case-study-goldcorp-inc-s-marlin-mine-environmental-contamination-and-human-rights-abuses/
12. Catalan Data Protection Authority. Alessandro Mantelero: "A technology that goes against fundamental rights is not a good technology." Catalan Data Protection Authority. 10 April 2025. Available at: http://apdcat.gencat.cat/en/sala_de_premsa/notes_premsa/noticia/Entrevista-Alessandro-Mantelero-FRIA
13. Center for Democracy & Technology. Assessment Reports: An Initial Feedback Brief. Available at:

https://cdt.org/insights/dsa-civil-society-coordination-group-publishes-an-initial-analysis-of-the-major-online-platforms-risks-analysis-reports/

14. Civio. El Gobierno se niega a explicar cómo funciona su aplicación para conceder o no el bono social. 28 November 2018. Available at: https://civio.es/transparencia/2018/11/28/el-gobierno-se-niega-a-explicar-como-funciona-su-aplicacion-para-conceder-o-no-el-bono-social/
15. Civio. La transparencia de los algoritmos públicos, en juego: Civio presenta el recurso sobre BOSCO ante el Tribunal Supremo. 30 January 2025. Available at: https://civio.es/novedades/2025/01/30/la-transparencia-de-los-algoritmos-publicos-en-juego-civio-presenta-el-recurso-sobre-bosco-ante-el-tribunal-supremo/
16. Civio. Mientras el Constitucional colombiano reconoce el derecho a la transparencia algorítmica, España trata de blindar su opacidad. 14 April 2025. Available at: https://civio.es/novedades/2025/04/14/transparencia-algoritmica-colombia-abogacia-estado-bosco/
17. Civio. Por la transparencia de las decisiones automatizadas. 2025. Available at: https://civio.es/transparencia-decisiones-automatizadas/
18. Columbia Center on Sustainable Investment, Danish Institute for Human Rights, and Sciences Po Law School Clinic. A Collaborative Approach To Human Rights Impact Assessments. March 2017. Available at: https://ccsi.columbia.edu/sites/ccsi.columbia.edu/files/content/docs/publications/A-Collaborative-Approach-to-HRIAs_Web.pdf
19. Committee on Artificial Intelligence of the Council of Europe. HUDERIA: New tool to assess the impact of AI systems on human rights. 2024. Available at: https://www.coe.int/en/web/portal/-/huderia-new-tool-to-assess-the-impact-of-ai-systems-on-human-rights
20. Court of Justice of the European Union. Judgment of the Court in Case C-203/22 | Dun & Bradstreet Austria. Press Release N 22/25. 27 Febraury 2025. Available at:

https://curia.europa.eu/jcms/upload/docs/application/pdf/2025-02/cp250022en.pdf

21. Danish Institute for Human Rights. Reviews of mandatory human rights due diligence and disclosure laws. Available at: https://humanrightseducation.dk/Methodologies%20for%20assessing%20business%20respect%20for%20human%20rights%20/index.html#/lessons/e7OtdNatyRsQz32lCnorbNCFIJxkwiOU
22. Davidson, D. et al., The Algorithm Addiction. 20 December 2022. Available at: https://www.lighthousereports.com/investigation/the-algorithm-addiction/
23. Department of Justice of the Government of Catalonia. Informe Tiresias: Auditoria de l'algorisme RisCanvi. January 2024. Available at: https://repositori.justicia.gencat.cat/bitstream/handle/20.500.14226/1321/auditoria-algorisme-riscanvi-informe-final.pdf?sequence=1&isAllowed=y
24. Diakopoulos, N. Algorithmic Accountability Reporting: On the Investigation of Black Boxes. Tow Center for Digital Journalism. February 2014.
25. Directorate-General for Communications Networks, Content and Technology. AI Act. 18 February 2025.Available at: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
26. Donat, J. Premies woonhuisverzekeringen stijgen door gebruik big data. Consumenten Bond. 12 October 2018. Available at: https://www.consumentenbond.nl/nieuws/2018/premies-woonhuisverzekeringen-stijgen-door-gebruik-big-data
27. EDRi. EU's AI Act fails to set gold standard for human rights. 2024. Avaialbe at: https://edri.org/our-work/eu-ai-act-fails-to-set-gold-standard-for-human-rights/; BEUC The European Consumer Organization. EU rules on AI lack punch to sufficiently protect consumers. 12 September 2023. Available at: https://www.beuc.eu/press-releases/eu-rules-ai-lack-punch-sufficiently-protect-consumers

28. Eticas Foundations. Automating (In) Justice? An Adversarial Audit of RisCanvi. June 2024. Available at: https://eticasfoundation.org/automating-injustice-an-adversarial-audit-of-riscanvi/
29. Eticas Foundations. Automating (In) Justice? An Adversarial Audit of RisCanvi. June 2024. Available at: https://eticasfoundation.org/automating-injustice-an-adversarial-audit-of-riscanvi/
30. European Banking Authority. Guidelines on loan origination and monitoring. 2025. Available at: https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/credit-risk/guidelines-loan-origination-and-monitoring
31. European Center for Non-for-Profit Law & Society Inside. Framework for Meaningful Engagement: Human rights impact assessments of AI. 2023. Available at: https://ecnl.org/publications/framework-meaningful-engagement-human-rights-impact-assessments-ai
32. European Data Protection Board. Automated decision-making and profiling. 25 May 2018. Available at: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/automated-decision-making-and-profiling_en
33. European Data Protection Board. Guidelines and Recommendations on Data Protection impact assessments, High risk processing. 2019. Available at: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/data-protection-impact-assessments-high-risk-processing_en
34. European Network of National Human Rights Institutions. ENNHRI calls on the European Commission to ensure effective Fundamental Rights Impact Assessments (FRIAs) under the EU AI Act. 2025. Available at: https://ennhri.org/wp-content/uploads/2025/04/ENNHRI-statement-on-ensuring-

effective-Fundamental-Rights-Impact-Assessments-FRIAs-under-the-EU-AI-Act.pdf

35. European Union Agency for Fundamental Rights. Getting The Future Right Artificial Intelligence And Fundamental Rights. 2021. Available at: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2021-artificial-intelligence-summary_en.pdf
36. European Union. Charter of fundamental rights of the European Union. 2012. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012P/TXT
37. European Union. Q&A on risk assessment reports, audit reports and audit implementation reports under DSA. 2025. Available at: https://digital-strategy.ec.europa.eu/en/faqs/qa-risk-assessment-reports-audit-reports-and-audit-implementation-reports-under-dsa
38. Europol. AI and policing: The benefits and challenges of artificial intelligence for law enforcement, Europol Innovation Lab observatory report. 2024. Available at: https://www.europol.europa.eu/cms/sites/default/files/documents/AI-and-policing.pdf
39. Future of Life Institute. Overview of all AI Act National Implementation Plans. 2025. Available at: https://artificialintelligenceact.eu/national-implementation-plans/
40. García Mexía, P., Oriol, R., & Pinheiro, I. Alert: Spain implements the European AI regulation ahead of schedule. Herbert Smith Freehills Notes. 2 February 2025. Available at: https://www.herbertsmithfreehills.com/th/notes/madrid/2025-posts/alert-spain-implements-the-european-ai-regulation-ahead-of-schedule-marzo-2026
41. Garcia, Ter, et al. La Policía Nacional deja de usar Veripol, su IA estrella para detectar denuncias falsas. Civio. 19 Marzo 2025. Available at: https://civio.es/transparencia/2025/03/19/la-policia-nacional-deja-de-usar-veripol-su-ia-estrella-para-detectar-denuncias-falsas/#veripol

42. Geiger, G. et al. Suspicion Machines: Unprecedented experiment on welfare surveillance algorithm reveals discrimination. Lighthouse Reports. 2 March 2023. Available at: https://www.lighthousereports.com/investigation/suspicion-machines/

43. Geiger, G. How a Discriminatory Algorithm Wrongly Accused Thousands of Families of Fraud. Vice. 1 March 2021. Available at: https://www.vice.com/en/article/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud/

44. Gender Equality Commission And The Steering Committee On Anti-Discrimination, Diversity And Inclusion (Cdadi). Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination. Council of Europe. 2023. Available at: https://rm.coe.int/study-on-the-impact-of-artificial-intelligence-systems-their-potential/1680ac99e3&sa=D&source=docs&ust=1748375671852358&usg=AOvVaw03FDfL6tIHGJgr_P-hPVG2

45. Global Network Initiative. Implementing risk assessments under the Digital Services Act. 2023. Available at: https://globalnetworkinitiative.org/wp-content/uploads/2023/06/Discussion-summary-%E2%80%93-GNI-and-DTSP-workshops-on-implementing-risk-assessments-under-the-DSA-June-2023.pdf

46. Gonzalez, A. Car insurance quotes 33% higher in most ethnically diverse areas. Motor Finance online. 26 February 2024. Available at: https://www.motorfinanceonline.com/news/car-insurance-quotes-33-higher-in-most-ethnically-diverse-areas-bbc/?cf-view&sa=D&source=docs&ust=1748375672001728&usg=AOvVaw1LdkzPRVlE_-DhRcfkSYte

47. Government of Canada. Guide to Peer-Review: Directive on Automated Decision-Making. 2024. Available at: https://www.canada.ca/en/government/system/digital-government/digital-gover

nment-innovations/responsible-use-ai/guide-peer-review-automated-decision-systems.html

48. Horn Iwaya, L., Alaqra, S., Hansen, M. & Fischer-Hübner, S. Privacy impact assessments in the wild: A scoping review, Array, Volume 23, 2024. Available at: https://doi.org/10.1016/j.array.2024.100356.

49. International Labour Organization. ILO Declaration on Fundamental Principles and Rights at Work. 2022. Available at: https://www.ilo.org/about-ilo/mission-and-impact-ilo/ilo-declaration-fundamental-principles-and-rights-work

50. Jahangir, R. EU Steps Up Civil Society Engagement On the Digital Services Act — Is It Enough? | Tech Policy Press. Tech Policy Press. 16 April 2025. Available at: https://techpolicy.press/-eu-steps-up-civil-society-engagement-on-the-digital-services-act-is-it-enough

51. Jiménez Arandia, P. et al. Spain's AI Doctor. Lighthouse Reports. 17 April 2023. Available at: https://www.lighthousereports.com/investigation/spains-ai-doctor/

52. Jiménez Arandia, P. et al., Un algoritmo define el futuro de los presos en Cataluña: ahora sabemos cómo funciona. El Confidencial. 24 April 2024. Available at: https://www.elconfidencial.com/tecnologia/2024-04-24/riscanvi-algoritmo-cataluna-prisiones-presos-inteligencia-artificial_3871170/

53. Jiménez Arandia, P. What to expect from Europe's first AI oversight agency. AlgorithmWatch. 1 February 2023. Available at: https://algorithmwatch.org/en/what-to-expect-from-europes-first-ai-oversight-agency/

54. Kaye, K. AI Governance on the Ground: Canada's Algorithmic Impact Assessment Process and Algorithm has evolved. World Privacy Forum. 2024.Available at: https://www.worldprivacyforum.org/2024/08/ai-governance-on-the-ground-series-canada/

55. Klaassen, S and van Dijk, Romy. Computer zegt vrouw: hoe een Rotterdams algoritme jonge, alleenstaande moeders discrimineerde. Ver beton. 6 March 2023. Available at: https://www.versbeton.nl/2023/03/computer-zegt-vrouw-hoe-een-rotterdams-algoritme-jonge-alleenstaande-moeders-discrimineerde/
56. Lorenz, M. Finnish Credit Score Ruling raises Questions about Discrimination and how to avoid it. Algorithm Watch. 21 November 2018. Available at: https://algorithmwatch.org/en/finnish-credit-score-ruling-raises-questions-about-discrimination-and-how-to-avoid-it/
57. Mantelero, A., Guzmán, C., García, E., Ortiz, R., & Moro, M. A. FRIA model: Guide and use cases. The Catalan Data Protection Authority, 93. 28 January 2025. Available at: https://www.dpdenxarxa.cat/pluginfile.php/2468/mod_folder/content/0/FRIA_es_2.pdf
58. Martínez Garay, Lucía, et al., Three predictive policing appraoches in Spain: Viogén, RisCanvi and Veripol. Assessment from a human rights perspective. University of Valencia. November 2022. Available at: https://regulation.blogs.uv.es/files/2024/05/Three-predictive-policing-perspectives-web-17.06.24.pdf
59. Marsh, O. Researching Systemic Risks under the Digital Services Act. 26 July 2024. Available at: https://algorithmwatch.org/en/wp-content/uploads/2024/08/AlgorithmWatch-Researching-Systemic-Risks-under-the-DSA-240726_v2.pdf
60. Molnar, P. & Gill, L. Bots at the Gate: A Human Rights Analysis of Automatic Decision-Making in Canada's Immigration and Refugee System. University of Toronto's International Human Rights Program (IHRP), Citizen Lab, and Information Technology, Transparency, and Transformation Lab. 2018. Available at:

https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf

61. OECD AI Observatory. How governments are driving AI adoption for economic growth. 20 May 2025. Available at: https://oecd.ai/en/wonk/how-governments-are-driving-ai-adoption-for-economic-growth
62. Office of the United Nations High Commissioner for Human Rights. Application of the UNGPs in the context of the banking sector. 12 June 2017. Available at: https://www.ohchr.org/sites/default/files/Documents/Issues/Business/InterpretationGuidingPrinciples.pdf
63. Oliveira da Silva, F., Drazewski, C. Regulating AI to protect the consumer. The European Consumer Organisation. 6 October 2021. Available at: https://www.beuc.eu/sites/default/files/publications/beuc-x-2021-088_regulating_ai_to_protect_the_consumer.pdf
64. Ostmann, F., & Dorobantu C. AI in financial services. The Alan Turing Institute. 2021. Available at:
65. Oxfam. Oxfam Briefing Note: Land rights and soda giants Reviewing Coca-Cola and PepsiCo's land assessment in Brazil. October 2016. Available at: https://www-cdn.oxfam.org/s3fs-public/bn-land-rights-soda-giants-brazil-201016-en.pdf
66. Pírková, E. (Access Now), Wisniak, M. & Iwańska, K. (European Center for Not-for-Profit Law). Towards Meaningful Fundamental Rights Impact Assessments Under the DSA. Available at: https://www.accessnow.org/wp-content/uploads/2023/09/DSA-FRIA-joint-policy-paper-September-2023.pdf
67. Platform on Sustainable Finance. Available at: https://finance.ec.europa.eu/sustainable-finance/overview-sustainable-finance/platform-sustainable-finance_en

68. Reisman,D., Schultz, J., Crawford, K., & Whittaker M., "Algorithmic Impact Assessments Report: A Practical Framework for Public Agency Accountability." AI Now Institute. 9 April 2018. Available at: https://ainowinstitute.org/publications/algorithmic-impact-assessments-report-2
69. Reyna, A. Need for independent national market surveillance authorities under the AI Act—Commission. BEUC – The European Consumer Organisation. 25 June 2024. Available at: https://www.beuc.eu/letters/need-independent-national-market-surveillance-authorities-under-ai-act-commission
70. Rietbergen-McCracken, J. & Narayan, D. Participation and social assessment : tools and techniques (English). Washington, D.C. : The World Bank. 2010. Available at: http://documents.worldbank.org/curated/en/673361468742834292
71. Scassa, T. Regulating AI In Canada: A Critical Look At The Proposed Artificial Intelligence And Data Act. The Canadian Bar Review. 2023. Available at: https://cbr.cba.org/index.php/cbr/article/view/4817/4539
72. Pénicaud, S. (2025). Making Algorithm Registers Work for Meaningful Transparency. IA Ciudadana. https://iaciudadana.org/2025/03/13/making-algorithm-registers-work-for-meaningful-transparency/
73. SpainSIF. La Inversión Sostenible y Responsable en España: ESTUDIO DE MERCADO. 2024. Available at: https://www.spainsif.es/wp-content/uploads/2024/10/Estudio_Anual_Spainsif_2024.pdf
74. Spielkamp, M. EuGH-Urteil zum Scoring: Eine Ohrfeige nicht nur für die Schufa AlgorithmWatch. 8 December 2023. Available at:https://algorithmwatch.org/de/eugh-urteil-schufa-scoring/
75. TAP Network. SDG ACCOUNTABILITY HANDBOOK: Accountability of the Private Sector. Available at:

https://sdgaccountability.org/wp-content/uploads/2019/05/Accountability-of-the-Private-Sector.pdf

76. Ter, et al. La Policía Nacional deja de usar Veripol, su IA estrella para detectar denuncias falsas. Civio. 19 Marzo 2025. Available at: https://civio.es/transparencia/2025/03/19/la-policia-nacional-deja-de-usar-veripol-su-ia-estrella-para-detectar-denuncias-falsas/#veripol
77. The Danish Institute for Human Rights. (2017, March 1). A collaborative approach to Human Rights Impact Assessments. https://www.humanrights.dk/publications/collaborative-approach-human-rights-impact-assessments
78. The European Consumer Organisation. The Use Of Big Data And Artificial Intelligence In Insurance. 19 May 202 Available at: https://www.beuc.eu/sites/default/files/publications/beuc-x-2020-039_beuc_position_paper_big_data_and_ai_in_insurances.pdf&sa=D&source=docs&ust=1748375672001995&usg=AOvVaw19Jjabm2O7mVK3Lf5LrXSo
79. The International Business Leaders Forum and the International Finance Corporation. Guide to Human Rights Impact Assessment and Management (HRIAM). 2010. Pages 75-81. Available at: www.ifc.org/hriam
80. The World Bank. Human Rights Impact Assessments: a review of the literature, differences with other forms of assessments and relevance for development. 2013. Available at: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/834611524474505865/human-rights-impact-assessments-a-review-of-the-literature-differences-with-other-forms-of-assessments-and-relevance-for-development
81. Transparency, Accountability & Participation (TAP) Network. Campaign for a Decade of Accountability. 2021. Available at: https://www.sdgaccountability.org/wp-content/uploads/2021/06/GlobalSDGAccountabilityReport_pages_hRes-1.pdf

82. UN Human Rights Office. Brief by the United Nations Special Rapporteur on extreme poverty and human rights as Amicus Curiae in the case of NJCM c.s./De Staat der Nderlanden (SyRI, before the District Court of The Hague). Available at: https://www.ohchr.org/sites/default/files/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf
83. Vidal, N., & Muftić, N. AI regulation and industry influence: The public is locked out. FragDenStaat. 14 May 2025. Available at: https://fragdenstaat.de/en/articles/exclusive/2025/05/ai-regulation-and-industry-influence-the-public-is-locked-out/
84. Vincent, J. Apple's credit card is being investigated for discriminating against women. The Verge. Available at: https://www.theverge.com/2019/11/11/20958953/apple-credit-card-gender-discrimination-algorithms-black-box-investigation

Regulation

1. Law 15/2022, of July 12, 2002, on equal treatment and non-discrimination. (n.d.). Retrieved 30 May 2025. Available at: https://www.boe.es/buscar/act.php?id=BOE-A-2022-11589
2. Council of the European Parliament. Regulation—2019/1020—EN - EUR-Lex. 20 June 2019. Available at: https://eur-lex.europa.eu/eli/reg/2019/1020/oj/eng
3. Ministry of Digital Transformation and Civil Service. Anteproyecto de Ley para el Buen Uso y la Gobernanza de la Inteligencia Artificial. March 2025. Available at: https://avance.digital.gob.es/_layouts/15/HttpHandlerParticipacionPublicaAnexos.ashx?k=19128
4. Wahl, T. Council of Europe Convention on Artificial Intelligence. Eucrim. 26 September 2024. Available at: https://eucrim.eu/news/council-of-europe-convention-on-artificial-intelligence

5. Parliament, E., & Council, of the. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). 2016. Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32016R0679
6. Gobierno de España. Carta de derechos digitales. 2021. Available at: https://citiesfordigitalrights.org/sites/default/files/140721-Carta_Derechos_Digitales_RedEs_compressed.pdf
7. Government of Spain. Royal legislative decree 1/2007, of November 16, approving the consolidated text of the general law for the defense of consumers and users and other complementary laws. 2007. Available at: https://www.boe.es/buscar/doc.php?id=BOE-A-2007-20555
8. Government of Spain. Consumer rights, including product safety—Commercial practices and consumer rights—Starting, running and closing a business—Business—Your rights and obligations in the EU - Tu espacio europeo—Punto de Acceso. 4 April 2025. Available at: General.https://administracion.gob.es/pag_Home/en/Tu-espacio-europeo/derechos-obligaciones/empresas/inicio-gestion-cierre/practicas-comerciales/derechos-consumidores.html
9. Ministry for Digital Transformation. 2024 artificial intelligence strategy. 2024. Available at: https://digital.gob.es/dam/en/portalmtdfp/DigitalizacionIA/1_DOSSIER_AI_ENGLISH_15_JULIO.pdf
10. Garrigues. Key features of the Spanish AI Strategy for 2024: Reinforcement of the factors for the development of AI, promotion in public and private sectors and strengthening supervision for sustainable and ethical AI. 21 May 2024. Available at:

https://www.garrigues.com/en_GB/garrigues-digital/key-features-spanish-ai-strategy-2024-reinforcement-factors-development-ai

11. Consejo de Ministros. El Gobierno da luz verde al anteproyecto de ley para un uso ético, inclusivo y beneficioso de la Inteligencia Artificial. Ministerio para La Transformación Digital y de la Función Pública. 11 March 2025. Available at: https://www.administracionpublicadigital.es/normativas/2025/03/el-gobierno-da-luz-verde-al-anteproyecto-de-ley-para-un-uso-etico-de-la-ia
12. Ministry of Digital Transformation and Civil Service. Anteproyecto de Ley para el Buen Uso y la Gobernanza de la Inteligencia Artificial. March 2025. Available at: https://avance.digital.gob.es/_layouts/15/HttpHandlerParticipacionPublicaAnexos.ashx?k=19128https://avance.digital.gob.es/_layouts/15/HttpHandlerParticipacionPublicaAnexos.ashx?k=19128
13. Government of Canada. Directive on Automated Decision-Making. 2024. Available at: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592
14. European Union. Supervision of the designated very large online platforms and search engines under DSA. 2025. Available at: https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses
15. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R2065
16. Official Journal of the European Union. REGULATION (EU) 2022/2065 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065#d1e4142-1-1

Academic Papers

1. Ashar, A., Ginena, K., Cipollone, M., Barreto, R., & Cramer, H. Algorithmic Impact Assessments at Scale: Practitioners' Challenges and Needs. Journal of Online Trust and Safety, 2(4), Article 4. 2024. https://doi.org/10.54501/jots.v2i4.206
2. Binns, R. Data protection impact assessments: A meta-regulatory approach. International Data Privacy Law, 7(1), 22–35. 2017. https://doi.org/10.1093/idpl/ipw027
3. Boswell, J., Dean, R., & Smith, G. Integrating citizen deliberation into climate governance: Lessons on robust design from six climate assemblies. Public Administration, 101(1), 182–200 2023. .https://doi.org/10.1111/padm.12883
4. Bringas Colmenarejo, A., State ,Laura, & and Comandé, G. (n.d.). How should an explanation be? A mapping of technical and legal desiderata of explanations for machine learning models. International Review of Law, Computers & Technology, 0(0), 1–32. 2024. https://doi.org/10.1080/13600869.2025.2497633
5. Carsten Stahl, B., et al. "A Systematic Review of Artificial Intelligence Impact Assessments." Artificial Intelligence Review 56 (11):12799–831. 2023. Available at: https://doi.org/10.1007/s10462-023-10420-8
6. Groves, L. et al. Auditing Work: Exploring the New York City algorithmic bias audit regime. In The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil. 2024. Available at: https://doi.org/10.1145/3630106.3658959
7. Kaminski, M. E., & Malgieri, G. Algorithmic impact assessments under the GDPR: Producing multi-layered explanations. International Data Privacy Law, 11(2), 125–144. 2021. https://doi.org/10.1093/idpl/ipaa020
8. Kaminski, M. E., & Malgieri, G. Impacted Stakeholder Participation in AI and Data Governance (SSRN Scholarly Paper No. 4836460). Social Science Research Network. 2024. https://papers.ssrn.com/abstract=4836460

9. Kaminski, M. E., & Malgieri, G. The Right to Explanation in the AI Act (SSRN Scholarly Paper No. 5194301). Social Science Research Network. 2025. https://papers.ssrn.com/abstract=5194301
10. Malgieri, G., & Santos, C. Assessing the (severity of) impacts on fundamental rights. Computer Law & Security Review, 56, 106113 2025. https://doi.org/10.1016/j.clsr.2025.106113
11. Mantelero, A. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. Computer Law & Security Review, 54, 106020. 2024. https://doi.org/10.1016/j.clsr.2024.106020
12. Mantelero, A., Guzmán, C., García, E., Ortiz, R., & Moro, M. A. FRIA model: Guide and use cases. The Catalan Data Protection Authority, 93. 2025.
13. Okidegbe, N. To Democratize Algorithms. UCLA Law Review, 69. 2022. https://www.uclalawreview.org/to-democratize-algorithms/
14. Ooms, W., & Gils, T. Policy brief: Implementing the AI Act in Belgium - Scope of Application and Authorities. Knowledge Centre Data and Society. 2024. https://data-en-maatschappij.ai/en/publications/policy-brief-implementing-the-ai-act-in-be
15. Selbst, A. D. An Institutional View of Algorithmic Impact Assessments. Harvard Journal of Law & Technology (Harvard JOLT), 35, 117. 2021. https://heinonline.org/HOL/Page?handle=hein.journals/hjlt35&id=123&div=&collection=
16. Sion, L., Van Landuyt, D., & Joosen, W. A DPIA repository for interdisciplinary data protection research. In J. Garcia-Alfaro, K. Barker, G. Navarro-Arribas, C. Pérez-Solà, S. Delgado-Segura, S. Katsikas, F. Cuppens, C. Lambrinoudakis, N. Cuppens-Boulahia, M. Pawlicki, & M. Choraś (Eds.), Computer security. ESORICS 2024 international workshops (pp. 181–192). Springer Nature Switzerland. 2025.
17. Tamvada, M. Corporate social responsibility and accountability: A new theoretical foundation for regulating CSR. International Journal of Corporate Social Responsibility, 5(1), 2. 2020. https://doi.org/10.1186/s40991-019-0045-8

18. Veale, M., & Borgesius, F. Z. Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International, 22(4), 97–112. 2021. https://doi.org/10.9785/cri-2021-220402

Annex I -Interview participants

1. Pablo Jiménez Arandia - Freelance journalist
2. Grace S. Thompson - AI Policy clinic director at Center for AI and Digital Policy
3. Maurice Guiaux - Independent researcher
4. Marc Woltering - GDPR specialist
5. Mirko Tobias Schaefer - Associate Professor in Media and Culture Studies at Utrecht University
6. Iris Muis - Expert in responsible AI and ethical data governance at Utrecht University
7. Gianclaudio Malgieri - Associate Professor of Law & Technology at Leiden University
8. Luis de Salvador Carrasco - Director of Innovation and Technology Division at Spanish Data Protection Authority
9. Lara Groves - Senior researcher at Ada Lovelace Institute
10. Arantxa Mendiharat - co-founder of Deliberativa

Acknowledgements

The authors would like to thank everyone that participated in the interviews conducted under the context of this research for their generosity with their time and for the shared insights.

FEDERACIÓN DE CONSUMIDORES Y USUARIOS
CECU

Federación de Consumidores y Usuarios CECU
C/ Gran Vía, 69, 1ª planta, oficina 103 (Madrid)