Governing AI Through Assessment: A Critical Analysis of FRIA from a Technology Assessment Perspective

Pierluigi Masai, Lorenzo Carta, and Simone Arnaldi

Abstract This chapter develops a critical analysis of Fundamental Rights Impact Assessment (FRIA). FRIA is a risk-based and ex-ante mandatory procedure for evaluating potential impacts of artificial intelligence (AI) systems that makes an appearance in Article 27 of the AI Act but has not yet entered into force. Nonetheless, it is possible to evaluate it as a distinct type of governance instrument and to understand how it is intended to contribute to AI oversight. FRIA promises to integrate fundamental rights concerns in the cycle of development of AI systems and foster accountability and transparency. The issue is approached by situating FRIA within the broader historical and conceptual landscape of the governance of emerging technologies, with a specific focus on the practice of technology assessment (TA). TA has become a central tool in contemporary governance to evaluate the social, legal and economic consequences of emerging technologies. Drawing on past and current experience in TA, we will examine the methodological and implementation challenges to FRIA, as a tool to produce enhanced knowledge about AI within society and as an effective instrument for steering technological development. Overall, our contribution offers a preliminary evaluation of FRIA, its potential use, and foreseeable challenges. These insights are meant to identify new possible directions for research in legal and social sciences and to improve accountability, transparency and social responsiveness in AI governance.

All authors developed the concept for this work, conducted the literature review, and provided editorial revisions. Arnaldi drafted sections 1 and 6; Carta drafted sections 3 and 4; Masai drafted sections 2 and 5.

P. Masai (✉) · L. Carta
Applied Data Science and Artificial Intelligence, University of Trieste, Trieste, Italy
e-mail: pierluigi.masai@phd.units.it

L. Carta
e-mail: lorenzo.carta2@phd.units.it

S. Arnaldi
Department of Political Science, University of Trieste, Trieste, Italy
e-mail: sarnaldi@units.it

C. Crea and M. Infantino (eds.), *Remedies to Digital Vulnerability in European Private Law*, https://doi.org/10.1007/978-3-032-19720-7_12

Keywords Fundamental rights · High-risk artificial intelligence systems · Technology assessment · Impact assessment · Accountability

1 Introduction

The recent spread of artificial intelligence (AI) systems is accompanied by growing and legitimate concerns about their use. The last decade has witnessed an increase in the development and adoption of AI systems whose functionalities promise many benefits in a variety of fields across society. The pervasiveness of AI systems calls for appropriate regulation of them in order to balance innovation with rights protection. Governing technologies requires a proper evaluation of the outcomes of their adoption which often leads to the use of impact assessments as governance tools.

The European Union developed the AI Act which came into force in 2024. The AI Act envisages many actions to regulate AI systems among which article 27 introduces the notion of Fundamental Rights Impact Assessment (FRIA) that is the focus of the present work. FRIA is intended as a tool that should foster accountability and oversight on high-risk AI systems. FRIA has not yet been enforced but there is a lively discussion on its methodology and possible implementations in AI governance.

It is premature to evaluate the actual capacities of the AI Act or the efficacy of FRIA. Nonetheless, since there is a long tradition of studies about and practice of technology evaluation, it is possible to draw on it to think critically about the very processes of AI governance. In particular, we believe a lot may be learnt from the experience of Technology Assessment (TA). TA is "an evidence-based, interactive process to bring to light the societal, economic, environmental and legal aspects and consequences of new and emerging science technologies"[1] to influence the development trajectories of the studied technological systems. TA and FRIA represent distinct approaches to governing emerging technologies, but they share similar aims. We hold that TA's well-established theoretical foundations and practical implementation history offer valuable insights for refining FRIA's methodological framework and assessment procedures. To provide critical comments about the FRIA methodology we propose to analyse its features in line with the three conceptual dimensions identified in the literature to describe TA, namely anticipation, inclusion and complexity.

Section 2 will briefly describe TA's history and features in order to outline the elements necessary for the following analysis. Sections 3 and 4 will respectively present FRIA's characteristics as conceived in the AI Act and with regards to its methodology. Section 5 will present the analysis of FRIA through the lens of the conceptual dimensions of TA. Section 6 will present the conclusions highlighting the major takeaways of the analysis and reflect upon future research work that may be undertaken.

[1] Robinson et al. (2023, 5).

2 Technology Assessment: History and Aspects

2.1 The Aims of TA

TA aims to assess a new technology which is yet to be developed or that is currently either under development or waiting to be deployed, considering their societal implications. TA moves from the observation that technologies cannot be analyzed and discussed as limited to themselves, but they have to be recognized as embedded in society. This is in line with the approach of Science and Technology Studies (STS) that recognize technologies as the product of the interaction of many actors, both human and non-human.[2] Consequently, focusing exclusively on the technological aspects would represent a significant limitation from a critical standpoint, as it overlooks the complex interactions between people, technological artifacts and the social context. In fact, STS argue that the development of a technology can be fully comprehended only by taking into consideration the socio-materiality of the process, which is to say that actual choices and actions may be the consequence of economic, political or even personal reasons but also of structural and material constraints.[3] This is relevant not only when discussing the development of a technology but even when reasoning about its regulation and control.

The information gathered or produced thanks to the process of TA can have multiple uses. For instance, it might be used to assist decision makers in regulating a new technology or to nudge scientists throughout the design process of a new technology. Therefore, TA can have a role both in policy making and in technology development. TA is a process with a multifaceted nature. TA might be described either as a practice or as a form of analysis. The nature of TA is strictly operative: the practical dimension prevails over the theoretical one. There is no mandatory action to do though: strictly speaking there is no rigid sequence of steps to follow in order to perform TA. There are many forms of TA which reflect the variety of attitudes people may have towards it. Nonetheless, there are some general elements that characterize TA which can be outlined to illustrate its premise.

To better understand the many possible applications of TA, it is appropriate to quickly look at an overview of its historical development and evolution.

2.2 The Historical Development of TA

The notion of TA was introduced within the theoretical frame of risk evaluation.[4] The need to evaluate the consequences of the development and use of a new technology both to prevent potential harms and to maximize the benefits, was one of the main

[2] Latour (1987).

[3] Orlikowski and Scott (2008).

[4] Van Est et al. (2012).

reasons. These preoccupations intensified after World War II precisely because of science and technology's rising societal role and the recognition of its unintended effects. TA was originally developed in the USA to assist politicians given the impact technological development has on economics, health and lifestyles.[5]

In 1972, the Congress approved the Technology Assessment Act, and the Office of Technology Assessment (OTA) was created.[6] One of the main figures involved was U.S. congressional representative Emilio Daddario, a member of the Democratic Party who coined the term *technology assessment* and advocated the use of TA for policy advice. OTA was meant to evaluate existing or probable impacts of technology or technological programs and identify possible actions and alternatives. Eventually, OTA presented the results of its analyses in reports made available to Congress members. OTA was closed down in 1995 as a consequence of political tensions, yet the Technology Assessment Act is still in force today, even if without assignment of budget. Nonetheless, the closure of OTA caused a form of ostracism of TA in the USA. The European Parliamentary Technology Assessment Network (EPTA), founded in 1990, shows that nowadays TA for policy-advising is adopted by many countries.[7] As of 2025, EPTA has 14 full members in Europe and 11 associate members.

Over the decades TA has been reinterpreted and implemented in different contexts. Beside the use of parliamentary TA for policy-advising, other approaches emerged such as participatory TA (PTA) and constructive TA (CTA).[8] PTA originated to meet the interest of citizens and stakeholders of being involved in the process of technology development. CTA was introduced after the SCOT program (Social Construction of Technology) that interpreted technologies as an outcome of the interaction of many actors which implies that to shape the design and development of a technology the focus cannot be solely on the technology itself.

Thus, TA can help in decision-making processes about technology. Besides TA as policy advice, Grunwald[9] identifies other two main areas of TA in action—yet, in the future more areas may arise—: TA in public dialogue and TA in the making of technology. In these two perspectives on TA, participative processes can foster reflection about values and ideas across society, raising awareness and laying the foundation for public discussions, whereas the knowledge gained through TA can inform technological development and guide decisions in accordance with expected outcomes.

[5] Grunwald (2018).

[6] Sadowski (2015).

[7] Nentwich (2016).

[8] Tran and Daim (2008).

[9] Grunwald (2018).

2.3 The Elements of TA

In his analysis of TA, Armin Grunwald[10] identifies three conceptual dimensions that characterize TA: anticipation, inclusion and complexity. These dimensions are not contingent on specific implementations of TA which depend on their social, political and institutional context. Each dimension promotes a critical attitude towards the very procedures and outcomes of TA enhancing reflexivity.

Anticipation equates to generating prospective knowledge which can guide decisions enhancing reflexivity over the temporal unfolding of socio-technical processes. The attitude toward anticipation has changed over time. Sergio Urueña[11] identifies four different approaches to anticipation which can be used to illustrate this evolving attitude of TA towards the future: *predictivist*, *strategic*, *exploratory* and *critical-hermeneutic*. Earlier versions of TA, originating from Risk Assessment procedures, adopted a *predictivist* approach which does not address elements of complexity nor take into consideration possible future changes in the social context. If traditionally TA assumed to be able to anticipate the consequences of future developments of a technology, eventually scholars recognised that uncertainty is instead a pervasive element of socio-technical trajectories,[12] especially in the case of emerging technologies. Indeed, emerging technologies are characterized by an element of ignorance about the outcomes of their implementation that will eventually be addressed only after the adoption and diffusion of the technology itself. To a certain extent prospective knowledge may help to shape the development of a technology but only if adopted from the very start of the process; this is in line with a *strategic* approach to anticipation. Yet, in the end prospective knowledge has proven to be not an end in itself: it is rather instrumental to define an orientation either comparing different possible future scenarios—in line with an *exploratory* approach—or critically understanding the many attitudes towards future scenarios existing at the present moment—in line with a hermeneutic approach to anticipation. The integration of participative approaches into TA has shown how anticipation can contribute to building public opinion. As a matter of fact, the impact of PTA on the public sphere is extremely relevant since it can prevent a rejection of the technology and public protests. It is imperative that participatory activities are performed right from the start (upstream) and not at the end (downstream) of the development process, otherwise controversies may even be fostered.

Inclusion stresses the fact that TA is not focused on a technology in itself but on a technology as embedded in society. The point is that when performing TA one should involve all interested parties, taking into consideration both their needs and values, in order to prevent biases, discrimination, monopolies and consequent forms of exclusion. This enhances reflexivity across perspectives. Inclusion also aims to overcome a neo-positivist attitude or the idea that *science knows best*. It is tempting to base decisions regarding emerging technologies exclusively on the analysis of experts, yet that

[10] Grunwald (2018).

[11] Urueña (2021).

[12] Arnaldi (2012).

has been considered as a mere optimization of quantitative and measurable results in line with a technocratic way of reasoning. Such an attitude might not properly address the fact that different social groups, in different places or time, can define issues, decisions and their stakes in ways that are not necessarily captured by expert knowledge and drawing upon knowledge that is not leveraged by experts. Fostering inclusion has two important impacts: it helps to obtain better and more diverse knowledge about possible unintended outcomes, and it aims at building legitimacy of decisions either by selecting commonly preferred options or raising awareness. Both elements are fundamental to avoid adverse reactions. It has to be noted that inclusion is not strictly about participation. Certainly, participatory practices enhance the elements of inclusion of TA. Anyway, TA processes have to find mechanisms to represent the various epistemic and social perspectives of the many stakeholders.

The third dimension of complexity examines the tension between inclusion and exclusion that implies the necessity to judge and assess the relevance of boundaries, relationships and interactions. One must be able to draw lines and make distinctions. Grunwald clearly notices that modelling socio-technical dynamics requires systemic lenses through which we order and relate the different aspects that our perspective deem relevant for the issue we are assessing. This call for system thinking is a defining element of TA and embodies the aforementioned conceptual dimension of complexity, reflecting the difficulty to identify which aspects of life are or might be affected by a technology, in particular when ‘soft’ impacts, such as social relations, beliefs and values are considered.[13] Reflexivity is enhanced over relevance. The proper definition of what is relevant to the discussion and what is not requires a deep understanding of the dynamics that characterize technology development and implementation.

3 FRIA in the Context of the AI Act

3.1 The AI Act as Risk Regulation

Recent developments in AI governance have increasingly focused on a specific regulatory approach: risk regulation. The EU AI Act[14] exemplifies this shift, representing a significant move toward addressing the risks posed by AI systems through a horizontal approach. Risk regulation is anticipatory insofar as it is usually characterized as an ex-ante approach, aiming to mitigate potential harms before they materialize. According to Margot Kaminski,[15] in general this approach adopts a ‘techno-correctionist’ frame on regulation since the aim is mainly to ‘fix’ the technology and

[13] Urueña (2021).

[14] Regulation 2024/1689 (2024). Regulation (EU) 2024/1689 of the European Parliament and of The Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

[15] Kaminski (2023).

mitigate the potential harms. The basic assumption is that the technology will be used and there should be tools and procedures to manage the risks posed by technical artifacts. While they share these broad common features, different risk regulation frameworks can vary significantly according to the types of risks targeted for mitigation, the governance tools utilized, and the specific regulatory style adopted.

In the context of the AI Act, the risks to be addressed are indicated as “significant risk of harm to the health, safety or fundamental rights of natural persons”.[16] The incorporation of “risks to fundamental rights” in the AI Act marks a shift toward broadening the concept of ‘risk objects’ to encompass harms to normative values. This new notion was first introduced in EU digital law in the context of the GDPR and now appears also in the Digital Service Act (DSA).[17] Such expansion goes beyond traditional concerns of physical, economic, and environmental harms that have established assessment procedures.[18] The final regulation also defines different categories of risks posed by AI systems: unacceptable risks that are prohibited, high-risk systems, systems that pose limited risk, General-Purpose AI models (GPAI) that are subject to specific provisions, systems that are supposed to pose no significant risks.[19]

Specifically, for the focus of this chapter, the compliance requirements for high-risk systems are particularly relevant, as they heavily depend on risk and impact assessment procedures. These assessment tools are the most typical and widely used mechanisms in risk regulation. Under the AI Act, high-risk systems are subject to two mandatory types of risk assessment: conformity assessment (CA) for AI providers and FRIA for certain deployers.[20] Both processes are conceived as iterative, meaning they are not one-time evaluations conducted only before deployment but must be continuously carried out and updated throughout the entire life-cycle of the AI system.[21] This ensures that the risk management remains effective and up to date. If any of the evaluated elements change—such as the system’s design, its intended purpose, or the groups affected—the assessment must be revised accordingly.[22]

These tools reflect the distinctive governance approach of the AI Act, which, as suggested,[23] can be characterized as ‘meta-regulation’ or ‘enforced self-regulation’. This regulatory framework involves “imposing legally binding obligations on regulated organizations, requiring them to develop and maintain internal control systems that achieve broadly defined, outcome-oriented, and legally mandated objectives”.[24] This aligns with what is also called a ‘collaborative governance’ model, where formal

[16] Regulation 2024/1689 (2024), Art. 6(3).
[17] Malgieri and Santos (2025).
[18] Yeung and Ranchordás (2024).
[19] Smuha and Yeung (2025).
[20] Mantelero (2024).
[21] Kaminski (2023).
[22] Regulation 2024/1689 (2024), Art. 27(2) and Rec. 96.
[23] Smuha and Yeung (2025).
[24] Smuha and Yeung (2025, 24).

or informal responsibility for self-governance, such as risk assessment, is delegated to the entities subject to the regulation.[25]

3.2 FRIA in Article 27 of the AI Act

Article 27 of the AI Act establishes the requirements for conducting a FRIA for high-risk AI systems. The aim of this article is to mandate that certain deployers of high-risk AI systems "perform an assessment of the impact on fundamental rights that the use of such system may produce".[26] FRIA is required to be carried out only by specific deployers of high-risk AI systems,[27] rather than being extended to all deployers, as originally suggested by the European Parliament.

The assessment must be carried out before the high-risk AI system is used for the first time. Deployers are also required to update it iteratively if any significant changes occur in the elements evaluated during the assessment, such as the system's design, purpose, or context of use. The components of the assessment mentioned include a description of the deployer's processes, the intended use and frequency of the AI system, the categories of individuals likely to be affected, specific risks of harm, human oversight measures, and mitigation strategies.[28]

Following Mantelero and Wernick, the different features of FRIA can be summarized as follows.[29] FRIA has:

- an ex-ante approach that should be integrated in line with a 'by design' approach to AI development;
- a rights-based focus on risk assessment with risk acceptability, "means that the risk may remain high, albeit mitigated, if justified by the potential benefits"[30];
- a circular iterative structure that follows the product/service throughout its life-cycle, with the obligation to update if something changes in the components of the assessment;
- an expert-based nature that does not require stakeholder participation;
- a sharing of responsibility and burden between providers and deployers of AI systems on risk assessment and management.[31]

Beside the difference in scope, Article 27 shows significant difference in comparison to the Parliament's proposal. There is a complete omission of stakeholder consultation, which was included in the amendment stating that deployers "shall, to best

[25] Kaminski (2023) and Wernick (2024).

[26] Regulation 2024/1689 (2024), Art. 27(1).

[27] In particular should be performed by deployers that are bodies governed by public law, private entities providing public services, systems used to evaluate the creditworthiness of natural persons, and for risk assessment and pricing in life and health insurance. Ibid.

[28] Regulation 2024/1689 (2024), Art. 27(1) (a-f).

[29] Mantelero (2024) and Wernick (2024).

[30] Mantelero (2024, 4).

[31] Regulation 2024/1689 (2024), Rec. 93.

extent possible, involve representatives of the persons or groups of persons that are likely to be affected by the high-risk AI system, as identified in paragraph 1, including but not limited to: equality bodies, consumer protection agencies, social partners and data protection agencies, with a view to receiving input into the impact assessment".[32] This omission has been criticized in particular by civil society organizations because they "will not have a direct, legally binding avenue to contribute to impact assessments".[33] Another point of contention is the planned introduction of an "automated tool", such as a questionnaire, by the AI Office to simplify compliance. Several scholars argue that this approach risks reducing the assessment to a mere "box-ticking exercise" deprivazioneing it of substantive value and failing to account for its context-dependent nature.[34]

4 Methodological Aspects of the FRIA

After having provided a brief overview of FRIA's role within the AI Act, this section will delve into some of the methodological aspects that define FRIA as an assessment process. Regarding FRIA or similar rights-based approaches to AI evaluation, there are still no established common procedures or shared practices.[35] Moreover, the AI Office has not yet provided an official template, and as far as we are aware, only two FRIA models aligned with Article 27 have been developed at the moment in which this page is being written: one by the Catalan Data Protection Authority (APDCAT)[36] and the other by researchers at the Intesa San Paolo Data Science & Artificial Intelligence unit.[37]

Yet, despite being in its early stages, the FRIA does not arise from a vacuum but is grounded in extensive discussions on impact assessments for AI systems. Numerous academic proposals, standardization efforts, and industry initiatives have been developed in recent years.[38] From a legal standpoint, the FRIA is influenced by two established frameworks: the Data Protection Impact Assessment (DPIA) under the GDPR and the Human Rights Impact Assessment (HRIA), which is part of the UN Guiding Principles on Business and Human Rights.[39]

In the following parts of this section, we will outline the most significant methodological aspects that define FRIA. Specifically, we will focus on two key issues central to the debate: the question of which risks are intended to be assessed, and the features and phases that compose the assessment procedure.

[32] European Parliament (2023), Amendment 413.

[33] ECNL (2024).

[34] Mantelero (2024) and Smuha and Yeung (2025).

[35] Skoric et al. (2024).

[36] APDCAT (2025).

[37] Cosentini et al. (2025).

[38] Stahl et al. (2023) and Skoric et al. (2024).

[39] Mantelero (2024).

4.1 Risk and Fundamental Rights

As mentioned earlier, the risks to fundamental rights are framed as 'risks of harm' to those rights. However, as Malgieri and Santos point out, this concept—and similar formulations—is "inherently ambiguous",[40] as it is open to varied interpretations and presents several problems. This ambiguity stems from the fact that the concepts of risk and rights have very different origins and approaches, which are often seen as conflicting frameworks.[41] Thus, two distinct approaches to assessing impacts on fundamental rights can be distinguished[42]: a rights-based approach and a risk-based approach.

The rights-based approach follows a binary logic as it focuses on the dichotomy between the respect and violation of rights and obligations. In other words, a right is either respected or violated, operating on a 'yes or no' assessment regarding violations. As highlighted by Yeung and Bygrave,[43] fundamental rights are conceived as "moral boundary markers", that draw ethical demarcation lines that should not be crossed. This means that they define protected spheres that aim to safeguard intangible moral values, associated with the inherent dignity of individuals. This requires a "qualitative assessment method, which should be informed by the interpretations by the relevant regulatory authorities and case-law".[44] Conversely, the risk-based approach has an anticipatory and probabilistic nature. Its focus is on the anticipation of undesirable events or 'harms' and on measuring the probability and severity of their potential occurrence. It is therefore about managing potential future risks. Article 3(2) of the AI Act similarly conceptualizes risks through this conventional definition adopted in risk assessment. However, focusing solely on 'harms' or 'damages' can be considered a reductionist and limited approach when it comes to fundamental rights, which should have an intrinsic value and are hard to quantify since they are intangible values related to human dignity.[45]

In the literature, various methods have been proposed to address the tensions between these two approaches.[46] Specifically, the debate has largely treated the notion of likelihood as unproblematic, given its reliance on established statistical methods, and has instead concentrated on defining severity. Multiple variables and approaches have been suggested to capture both the tangible and intangible impacts on fundamental rights. The debate remains open, and its full exploration goes beyond the aim of this chapter. What is crucial to emphasize are the methodological assumptions underlying these interpretations and some of the main tensions that emerge in operationalizing FRIA.

[40] Malgieri and Santos (2025, 2).

[41] Yeung and Bygrave (2022).

[42] Malgieri and Santos (2025).

[43] Yeung and Bygrave (2022).

[44] Wernick (2024, 23).

[45] Malgieri and Santos (2025).

[46] Yeung and Bygrave (2022) and Malgieri and Santos (2025).

4.2 The Assessment Procedure

In accordance with risk assessment and management practices, FRIA should be structured into three distinct phases aligned with Article 27 of the AI Act[47]: planning, scoping, and risk identification; data collection and risk analysis; finally, risk management.

The stage of planning, scoping, and risk identification involves a detailed description of the AI system, its intended purposes, and the context in which it will be used. It requires an evaluation of both intrinsic risks (linked to the system's inherent design) and extrinsic risks (arising from the system's interaction with the social context and other technologies).[48] This step helps identify the potential domains that may impact rights, as well as the individuals and groups likely to be affected by the system.

During the stage of data collection and risk analysis, the potential impacts on rights identified in the previous phase are quantified. This evaluation may employ risk matrices—graphical tools that plot likelihood against severity on intersecting axes—to develop specific risk indices for each affected right. These matrices categorize distinct risk levels through the combined analysis of both dimensions.

After completing the risk analysis, the identified risks must be managed through the implementation of appropriate measures. The risk management phase is structured into three key steps: (1) identifying measures to prevent or mitigate risks, considering their impact on the overall risk level; (2) implementing these measures; and (3) monitoring the AI system's performance to ensure the assessment and measures remain up to date, revising them if "technological, societal and contextual changes affect some of the relevant elements of a previous assessment".[49]

According to Mantelero and APDCAT, this procedure is inherently expert based since it requires multidisciplinary expertise including specialists in fundamental rights, AI system design, and the contextual aspects of deployment. However, as previously noted, stakeholders are absent in the final version of the AI Act, marking a notable omission. Also other impact assessment frameworks that evaluate impacts on rights often incorporate stakeholder input to identify normative concerns, yet they rarely assign stakeholders an active role in resolving tensions between competing rights or interests.[50]

5 FRIA from a TA Perspective

The preceding analysis of TA and FRIA provides the groundwork for the following discussion. Rather than undertaking a direct comparison between FRIA and TA, this study seeks to initiate a critical dialogue about FRIA by discussing its distinctive

47 Mantelero (2024).

48 APDCAT (2025, 8).

49 APDCAT (2025, 10).

50 Algorithm Audit (2024).

features. This analysis is informed by the ongoing debate on TA, which has been rigorously investigated and well-documented over several decades. FRIA and TA are both procedures that aim at performing an evaluation of a technology in order to discuss the outcomes of the uses of such technology, though they differ in many aspects. For instance, TA is usually performed for a technology that is yet to be envisioned or that is currently under development while FRIA is an evaluation that occurs before the actual implementation and use of a technology but likely after its development.

This analysis is not a comprehensive assessment of FRIA, nor is this analysis grounded in empirical data about its uses, since FRIA is yet to be implemented. Drawing from the experience of TA, our goal is instead to stimulate a reflection on the potential and limitations of FRIA to perform two important tasks, namely, to steer technology development and to gain further knowledge about the technology in a hermeneutic perspective, which is to say to obtain knowledge about the social and epistemic configuration within which the technology is situated. To do so, our analysis will use the three theoretical dimensions of TA, previously described in Sect. 2.3 as heuristic concepts. While definitive conclusions about FRIA's governance implications remain premature, we present open questions about FRIA in order to cultivate critical engagement with its assumptions, and foster awareness of both its limitations and possibilities.

5.1 Complexity

The notion of complexity introduces the issue of relevance: TA involves sets of decisions about the inclusion or exclusion of specific actors, expertise, interests and priorities to steer the process through the multiple interrelations that characterize socio-technical trajectories related to AI. Applying this complexity dimension to FRIA means understanding what is relevant for its methodology and why. Certainly, an element of complexity characterizing FRIA is the very focus on fundamental rights that, as discussed in Sect. 4.1, draw ethical demarcation lines. Yet, fundamental rights are not always interpreted univocally and coherently: what interpretations are valid and relevant for the process? This would necessarily involve examining also how the laws are interpreted by legislators and legal professionals, an analysis that extends beyond the scope of this study.[51]

Within this context, the crucial focus instead becomes the role of experts and expert knowledge, both legal and technical, in FRIA's implementation. The primary reflexive challenge emerges from FRIA's fundamental design as an expert-based process, which constitutes its most significant element of complexity. By rigidly delimiting what constitutes legitimate knowledge and marginalizing non-expert perspectives, this approach depoliticizes risk evaluation.[52] This orientation conflicts

[51] Malgieri and Santos (2025).

[52] Van Est et al. (2012).

with TA's established understanding that socio-technical complexity necessitates questioning default assumptions about what constitutes 'relevant' or 'sufficient' expert knowledge. As we will see when discussing the inclusion dimension, this requires integrating stakeholder's perspectives to foster and open political discussions on the use of technology.

FRIA is expected to evaluate the impacts of AI systems with respect to the context of use, but the context is narrowly conceived as it relates to the alignment between the intended purposes and the "deployer's processes".[53] The evaluation does not extend on reflecting and deciding on the purpose of the technology considered, to how these purposes are embedded in technology design, and what unexpected (longer term) impacts are generated by the interaction of the same technology with a changing social context from a co-evolutionary perspective. These aspects appear to be taken for granted and artificially separates technology's development and implementation phases, a division that primarily exists in analytical frameworks rather than in actual technological practices. The separation of development and implementation phases is also codified by the AI Act which establishes FRIA for deployers and Conformity Assessment (CA) for providers. CA focuses on the results of the development of the technology while FRIA focuses on the possible consequences of the use of the technology. But there is no clear indication about the interaction providers and deployers should have, if any.[54]

By failing to adequately account for technology's social embeddedness, FRIA structurally reduces complexity through its conventional risk-based approach. It assumes that development and use are separable phases whereas TA tries to overcome such a simplification analysing the complexity of the entire processes of design and implementation of a technology.

A further dimension of complexity concerns FRIA's relationship to risk and its implications for evaluation outcomes. FRIA operates within already delineated decisions. The AI Act identifies high-risk systems according to criteria of relevance listed in Annex III. While these criteria are not explicitly part of the FRIA procedure, they play an implicit role by determining which systems require evaluation. Also, the fact that the AI regulatory framework defines a restricted category of prohibited AI uses which poses unacceptable risks, has the complementary consequence that the range of acceptable risky uses of AI is quite broad and encompasses a wide range of risk severity. As we mentioned before, in order to be put on the market, high risk systems require that Conformity Assessment (CA) and FRIA are performed respectively by providers and deployers. This leaves the risk acceptability decisions mainly in the hands of these actors that in their assessment heavily rely on experts' knowledge and technical standards.[55]

Since FRIA should be performed for high risks systems, it can conceal or exclude normative decisions of risk acceptability restricting its scope to the identification and mitigation of possible impacts on rights in deployment. Thus, as mentioned

---

[53] Regulation 2024/1689 (2024), Art. 27(1) (a).

[54] Mantelero (2024) and Cosentini et al. (2025).

[55] Laux et al. (2024).

before, FRIA does not assess the societal justification of technology's purposes and contextual uses. This aligns with the AI Act's techno-correctionist approach,[56] which prioritizes technical mitigation, 'fixing' flawed systems, over critical engagement with technology purposes and systemic effects. This highlights the fundamental tension inherent in translating rights protections into risk calculus—a tension rooted in the differences between risk-based and rights-based paradigms analyzed in Sect. 4. Such framing seems to presuppose that impacts on rights can be rendered permissible if 'managed' through mitigation measures.

These considerations open two questions for FRIA. Does FRIA permit reflections on the thresholds that distinguish 'manageable' from 'unacceptable' risks to rights and who gets to define them? Does FRIA risk considering impacts on fundamental rights as inherently 'manageable' thereby eroding the protections traditionally afforded to rights?

FRIA is meant to have a circular iterative structure: re-evaluation is mandatory if changes in systems' purpose, affected populations, or identified risks emerge.[57] However Article 27 does not provide what are the criteria for assessing these changes. TA recognizes that the background against which analyses are conducted is an element of complexity since socio-technical systems are inherently complex: risks and technologies are embedded in contexts that evolve over time. Stakeholders may change their view about values and rights. Even the access to a technology and its use may change. How can such changes be addressed? Also, focusing on risk assessment, is it actually feasible to evaluate the risk of an event that will eventually occur in the future in a context which could be significantly different from the present one? This calls for a reflexive attitude that constantly questions the dynamics of inclusion and exclusion. The question arises: is there room for reflexivity with FRIA?

5.2 Anticipation

The dynamics of socio-technical systems, which is an element of complexity, requires assessment of future evolutions in order to be critically investigated. This calls for anticipatory practices. There are two elements of anticipation in the FRIA procedure as it can be derived from the description of its methodology outlined in Sect. 4.2. The first one is related to identification and analysis. The second one is related to the guidelines or actions envisioned to mitigate the risks. They both represent a risk assessment approach which was once at the core of TA as well. TA moved on to adopt a proactive approach, though. Following Urueña's classification,[58] in order to guide technological development, one has to visualize desirable futures (*strategic* approach) and be critical about how the technology is conceived and understood at

56 See Sect. 3.1.

57 Cosentini et al. (2025).

58 Urueña (2021).

the present time (*hermeneutic* approach). In principle, developers do not explicitly incorporate reflections on values and visions of the future into the design of the technology. The question arises: how can FRIA guide the development of a technology toward the production of desirable outcomes?

In the case of FRIA, the focus is on the technology itself, not on the technology as embedded in society. FRIA fosters no discussion about desirable outcomes; it concentrates on possible damages and harms. There is another limit to stress, though. Future harms are considered in relation to expected or intended uses of the technology. It is in fact on the intended uses of the technology that risks are assessed.[59] Yet, STS have shown how technology may have many unintended uses and emergent effects that TA tries to account for thanks to its more general approach. The co-evolution of technologies and society implies that uncertainty cannot be avoided. For AI systems and other new technologies there is also another, specific source of uncertainty: divergence about the evaluation of the outcomes, intended and unintended. The Rathenau Institute has introduced the notion of *Intimate Technology* to discuss AI systems.[60] The concept expresses the fact that such technologies are nested within society, reshaping people's habits, attitudes and beliefs, and gathering more information about them. For intimate technologies, uncertainty does not depend only on the ignorance about actual outcomes but also on people's divergences about the perceived goodness of such outcomes and the subjective trade-offs between them. The evolution, adoption and diffusion of intimate technologies define a new landscape, a new interpretive framework not only for the relationship we humans have with technologies but for the way we conceive our daily lives and relate to the world as well. This is especially true for information technologies that change radically the way we experience and interpret the world, but the case can be made for many other technologies as well. Such considerations are relevant to the present discussion because intimate technologies for their very nature, as it is recognized under the lens of this definition, produce soft effects that question personal and cultural values rather than hard, tangible and quantifiable effects such as physical, economic and environmental harms. A proper analysis of intimate technologies and an assessment of their impact call for trade-offs on values upon which there are in principle many divergences within society. Ultimately, the adoption of intimate technologies is not just a matter of assessing potential harms: it cannot be just a consequence of a judgment based on a common understanding of what is harmful and what is not, which risks can be taken and which cannot. The question arises: can FRIA account for such considerations?

[59] The AI Act explicitly pursues this objective, requiring mitigation measures for "the known and reasonably foreseeable risks of AI systems to the health, safety and fundamental rights in light of their intended purpose and reasonably foreseeable misuse". Regulation 2024/1689 (2024), Recital 65.

[60] van Est et al. (2014).

5.3 Inclusion

Over time TA has recognized not only the importance of including actors other than solely scientists, technicians, legal experts and politicians, but to do so from the very start of technological development. The reason is twofold. The inclusion of as many stakeholders as possible is meant to take into consideration experiential knowledge and preferences about values and evaluation processes. The importance of starting immediately is meant to prevent situations in which it is very difficult if not impossible to work on the technology to change it and adapt it to the needs of the stakeholders. In other words, actions should be taken from the start of the innovation process (upstream) and not at the end to avoid so-called locked-in phenomena when the technology has already stabilized.[61] The question arises: how does FRIA take into consideration the knowledge and preferences of non-experts?

Scholars and policymakers have increasingly acknowledged that vulnerable populations—including the elderly, poor people, children, migrants, and persons with disabilities—often face disproportionate direct and indirect consequences from AI systems.[62] The exclusion of these groups' perspectives in impact assessments could undermine the validity of fundamental rights evaluations. When risk evaluations are conducted solely by experts, decisions about risk acceptability can become problematic, as they disregard the values and preferences of affected communities which might disagree with the implementation of the technology itself. As Laux and his colleagues[63] argue, these questions are not appropriately addressed by the AI Act which largely adopts a 'paternalistic' and 'technocratic' approach to AI risk assessment, prioritizing expert judgment and self-governance mechanisms over the participatory inclusion of non-experts in decision-making processes.

In relation to these reflections on inclusion, in particular Constructive Technology Assessment (CTA) incorporates methodologies like value-sensitive design and participatory design, which try to integrate ethical considerations and stakeholder perspectives throughout the technological development process.[64] However, applying these principles to FRIA presents unique challenges. As a post-development evaluation conducted by deployers—who typically have limited capacity to modify system design—FRIA risks encountering technological lock-in effects, where problematic design choices become entrenched. While providers could more effectively address these concerns during the conformity assessment (CA), critical questions remain regarding FRIA's role. How can FRIA mitigate lock-in effects when evaluating already developed systems? To what extent can FRIA evaluations drive meaningful design changes post-development?

---

[61] Macnaghten et al. (2005).

[62] FRA (2022).

[63] Laux et al. (2024).

[64] Grunwald (2018) and Buijsman et al. (2025).

6 Conclusion

This chapter provides a critical examination of the FRIA, interrogating its role in AI governance and AI systems evaluation. We are aware that FRIA acts within the pre-established theoretical framework of the AI Act that defines boundaries such as the identification of the systems to be analyzed. We also acknowledge that something similar holds true for TA as well: TA may be commissioned by a governmental institution or may be limited by precise requirements upon its focus. Nonetheless, the tradition of TA was proven useful to gather insights on FRIA given they are both procedures for the evaluation of a technology. By establishing a structured dialogue between FRIA and TA methodologies, we have shed light on FRIA's assumptions, its potentialities and limitations. Our analysis employed Grunwald's[65] three constitutive dimensions of TA—anticipation, complexity, and inclusion—as analytical lenses to systematically examine FRIA. We have stressed that TA adopts a reflexive attitude capable of stimulating considerations on the criteria of relevance selected (complexity), the multiple temporal dimensions considered (anticipation) and the different perspectives included (inclusion). The full implications of FRIA's application to AI risk governance are still unfolding, yet the study has produced critical insights that could inform future FRIA revisions and implementations.

Our analysis has, however, identified key conceptual elements essential for understanding AI systems assessment procedures and has led us to raise open questions about FRIA, which we summarize below. Criteria of relevance enable and constrain reflexivity over the assessment process. To question the very criteria of relevance is important to adapt the procedure to the current situation and address complexity. Can FRIA account for such a reflexive attitude? Unintended uses of technologies cannot be foreseen beforehand. Also, AI systems require trade-offs between values and impacts' perception could change over time. To address such uncertainty, it is necessary to move from a *predictivist* approach to anticipation to a *hermeneutic* one. Can FRIA integrate these attitudes in its framework?

Finally, we have shown in Sect. 5.3 that inclusion is not only instrumental to evaluate value conflicts and divergences, but also to integrate different experiential knowledge on technological uses. Although the AI Act encourages the participation of stakeholders it does not mandate their consultation for conducting FRIA.[66] This poses some doubts on its commitment toward participation, which is also demonstrated by the FRIA models we analyzed which are expert-based procedures. Restricting participation in this way also affects on the other dimensions shaping what is relevant to evaluate and what futures are considered. How can FRIA effectively address these inclusion-related issues?

[65] Grunwald (2018).

[66] Regulation 2024/1689 (2024), Rec. 96.

References

Algorithm Audit (2024) A comparative review of 10 Fundamental Rights Impact Assessments (FRIA) for AI-systems. https://algorithmaudit.eu/pdf-files/knowledge-base/20240918_Comparative%20review%2010%20FRIAs%20Algorithm%20Audit.pdf

Arnaldi S (2012) Oltre il microcosm. Innovazione e valutazione tecnologica partecipata. In: Chiocca M, Valli L (eds) L'innovazione responsabile. Retecamere scrl, Roma, pp 91–105

Buijsman S, Klenk M, van den Hoven J (2025) Ethics of AI: toward a "design for values" approach. In: Smuha NA (ed) The Cambridge handbook of the law, ethics and policy of artificial intelligence. Cambridge University Press, Cambridge, pp 59–78

Catalan Data Protection Authority (APDCAT) (2025) FRIA model: guide and use cases. https://apdcat.gencat.cat/ca/documentacio/intelligencia_artificial/

Cosentini A et al (2025) Assessing the impact of artificial intelligence systems on fundamental rights. MediaLaws. https://www.medialaws.eu/wp-content/uploads/2025/03/20250714_FRIAct_MediaLaws.pdf

European Center for Not-for-Profit Law (ECNL) (2024) Packed with loopholes: why the AI Act fails to protect civic space and the rule of law | ECNL. https://ecnl.org/news/packed-loopholes-why-ai-act-fails-protect-civic-space-and-rule-law

European Parliament (2023) Artificial intelligence act amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=EP%3AP9_TA%282024%290138

European Union Agency for Fundamental Rights (FRA) (2022) Bias in algorithms—artificial intelligence and discrimination. Publications Office of the European Union, Luxembourg

Grunwald A (2018) Technology assessment in practice and theory. Routledge, London

Kaminski ME (2023) Regulating the risks of AI. 1348 Boston Univ Law Rev 103(5):1347–1411

Latour B (1987) Science in action: how to follow scientists and engineers through society. Harvard University Press, Cambridge

Laux J, Wachter S, Mittelstadt B (2024) Trustworthy artificial intelligence and the European Union AI Act: on the conflation of trustworthiness and acceptability of risk. Regul Govern 18(1):3–32. https://doi.org/10.1111/rego.12512

Macnaghten P, Kearnes MB, Wynne B (2005) Nanotechnology, governance, and public deliberation: what role for the social sciences? Sci Commun 27(2):268–291. https://doi.org/10.1177/1075547005281531

Malgieri G, Santos C (2025) Assessing the (severity of) impacts on fundamental rights. Comput Law Sec Rev 56:106113. https://doi.org/10.1016/j.clsr.2025.106113

Mantelero A (2024) The fundamental rights impact assessment (FRIA) in the AI act: roots, legal obligations and key elements for a model template. Comput Law Sec Rev 54:106020. https://doi.org/10.1016/j.clsr.2024.106020

Nentwich M (2016) Parliamentary technology assessment institutions and practices. A systematic comparison of 15 members of the EPTA network. https://epub.oeaw.ac.at/0xc1aa5572%200x00345db6.pdf

Orlikowski WJ, Scott SV (2008) Sociomateriality: challenging the separation of technology, work and organization. Annals 2(1):433–474. https://doi.org/10.5465/19416520802211644

Robinson DKR, Winickoff D, Kreiling L (2023) Technology assessment for emerging technology: meeting new demands for strategic intelligence. OECD Publishing, Paris

Sadowski J (2015) Office of technology assessment: history, implementation, and participatory critique. Technol Soc 42(1):9–20. https://doi.org/10.1016/j.techsoc.2015.01.002

Skoric V, Sileno G, Ghebreab S (2024) Roles of standardised criteria in assessing societal impact of AI. In: 2024 IEEE conference on artificial intelligence (CAI). IEEE, New York, pp 1240–1245

Smuha NA, Yeung K (2025) The European Union's AI act: beyond motherhood and apple pie? In: Smuha NA (ed) The Cambridge handbook of the law, ethics and policy of artificial intelligence. Cambridge University Press, Cambridge, pp 228–258

Stahl BC et al (2023) A systematic review of artificial intelligence impact assessments. Artif Intell Rev 56(11):12799–12831. https://doi.org/10.1007/s10462-023-10420-8

Tran TA, Daim T (2008) A taxonomic review of methods and tools applied in technology assessment. Technol Forecast Soc Change 75(9):1396–1405. https://doi.org/10.1016/j.techfore.2008.04.004

Urueña S (2021) Responsibility through anticipation? The 'future talk' and the quest for plausibility in the governance of emerging technologies. NanoEthics 15(3):271–302. https://doi.org/10.1007/s11569-021-00408-5

van Est R (2014) Intimate technology. The battle for our body and behaviour. Rathenau Instituut, The Hague

van Est R, Walhout B, Brom F (2012) Risk and technology assessment. In: Roeser S, Hillerbrand R, Sandin P, Peterson M (eds) Handbook of risk theory: epistemology, decision theory, ethics, and social implications of risk. Springer, Dordrecht, pp 1067–1091

Wernick A (2024) Impact assessment as a legal design pattern—a "timeless way" of managing future risks? DISO 3(2):29. https://doi.org/10.1007/s44206-024-00111-4

Yeung K, Bygrave LA (2022) Demystifying the modernized European data protection regime: cross-disciplinary insights from legal and regulatory governance scholarship. Regul Govern 16(1):137–155. https://doi.org/10.1111/rego.12401

Yeung K, Ranchordás S (2024) An introduction to law and regulation: text and materials, 2nd edn. Cambridge University Press, Cambridge