FRIA model: Guide and use cases

FRIA methodology for AI design and development

DPD en xarxa

INDEX

Coordinators:
Alessandro Mantelero and Joana Marí

Authors:
Alessandro Mantelero (Polytechnic University of Turin)
Cristina Guzmán (Polytechnic University of Catalonia – BarcelonaTech (UPC))
Esther García (CaixaBank, S.A)
Ruben Ortiz (University of Barcelona)
M. Ascensión Moro (Sant Feliu de Llobregat City Council)

Participants:
Albert Portugal (Consortium of University Services of Catalonia)
Albert Serra (Catalan Data Protection Authority)
Alessandro Mantelero (Polytechnic University of Turin)
Cristina Guzmán (Polytechnic University of Catalonia – BarcelonaTech (UPC))
Esther García (CaixaBank, S.A)
Joana Judas (Department of Research and Universities. Generalitat de Catalunya)
Joana Marí (Catalan Data Protection Authority)
Jordi Escolar (University Quality Assurance Agency)
M. Ascensión Moro (Sant Feliu de Llobregat City Council)
Marc Vives (Pompeu Fabra University)
Maria José Campo (TIC Health and Social Foundation)
Mariona Perramon (Consortium of University Services of Catalonia)
Olga Rierola (Catalan Data Protection Authority)
Patricia Lozano (Open University of Catalonia)
Ruben Ortiz (University of Barcelona)
Sara Hernández (TIC Health and Social Foundation)

Barcelona, 2025

Presentation

In its Strategic Plan 2023-2028, the Catalan Data Protection Authority (APDCAT) has set as a one of its main pillars the objective of promoting the development of training aimed at institutions. In this view, one of the main lines of action was to *"promote the creation and strengthening of the Data Protection Officers' Learning Community".*

On 15 July 2023, the first Data Protection Officers (DPO) Network was launched (available at https://www.dpdenxarxa.cat/), a pioneering initiative in Catalonia and in Spain. The platform was created with the aim of contributing to the development of the culture of privacy in Catalonia, driven by Data Protection Officers (DPOs) who ensure compliance with data protection regulations, promote cooperation and collaboration among themselves, and share knowledge and expertise.

This platform brings together the DPOs of the more than 1,700 entities that are part of the APDCAT's scope of action, which includes public administrations such as the Generalitat de Catalunya, city councils, public and private universities, and professional associations, among others. It is also open to DPOs of public and private entities that provide services to the Catalan public sector as data processors, as well as to all DPOs of entities based in Catalonia.

The Network, which currently includes a large number of DPOs in Catalonia, pursues the following main objectives:

- To be an institution closer to DPOs and organisations.
- To be a space for the exchange of ideas, experiences and knowledge.
- To promote the figure of the DPO as a key player in compliance.
- To identify and promote best practices.
- To create and disseminate a model of relationships and cooperation compatible with useful and effective supervision.
- To create an environment of open interaction and collaboration services for resource generation, learning and knowledge management.
- To create a reference space for the DPOs in Catalonia and Europe.

As a result of this space for the exchange of ideas, experiences and knowledge, as well as the creation of an environment of interaction and open collaboration for the generation of resources, learning and knowledge management, the Agora section of the Network proposed to set up a working group entitled *"Methodological guidance. Impact assessment. Rights and freedoms"*

This group, coordinated by Dr. Alessandro Mantelero[1] and Ms. Joana Marí,[2] conducted its work from February to December 2024 and included the following members of the Network:

- Albert Portugal (Consortium of University Services of Catalonia)
- Albert Serra (Catalan Data Protection Authority)
- Cristina Guzmán (Polytechnic University of Catalonia – BarcelonaTech (UPC))
- Esther Garcia (Caixabank, S.A.)
- Joana Judas (Department of Research and Universities. Generalitat de Catalunya)

[1] Associate Professor of Private Law and Law and Technology at the Polytechnic University of Turin, Jean Monnet Chair of Mediterranean Digital Societies and Law.
[2] Data Protection Officer and Head of Strategic Projects of the Catalan Data Protection Authority.

- Jordi Escolar (University Quality Assurance Agency)
- M. Ascensión Moro (Sant Feliu de Llobregat City Council)
- Marc Vives (Pompeu Fabra University)
- Maria José Campo (TIC Health and Social Foundation)
- Mariona Perramon (Consortium of University Services of Catalonia)
- Olga Rierola (Catalan Data Protection Authority)
- Patricia Lozano (Open University of Catalonia)
- Ruben Ortiz (University of Barcelona)
- Sara Hernández (TIC Health and Social Foundation)

The objective of the Group was to develop a new methodology for Fundamental Rights Impact Assessments (FRIA) in the use of Artificial Intelligence systems in line with the regulatory framework established by Regulation (EU) 2024/1689 of the European Parliament and of the Council, of 13 June 2024 (Artificial Intelligence Act) and with the aim of distinguishing this methodology from the Data Protection Impact Assessments (DPIA) required by Regulation (EU) no. 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation).

This document presents the main results of this Working Group and is divided into two parts, the first describing the FRIA methodology and the second focusing on its concrete application to specific cases. The purpose of this work is to provide a useful and practical tool for entities that design, develop or use AI systems and models and, in particular, for people in charge of carrying out the fundamental rights impact assessment.

In short, preventing of the violation of fundamental rights and freedoms is a common task in respect of which the Catalan Data Protection Authority must play a key role.

Meritxell Borràs i Solé
Director
Catalan Data Protection Authority

Part I – The FRIA and the FRIA methodology

1. Introduction

This publication presents the results of a project led by the Catalan Data Protection Authority (ADPCAT) with the aim of providing AI operators, both providers and deployers, with an effective tool to develop trustworthy and human-centred AI solutions. In this view, as demonstrated by the AI Act and other national and international initiatives, such as the Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy and the rule of Law, the design and development of AI solutions that respect fundamental/human rights[3] is at the core of truly human-centred AI.

In the light of the above, the AI Act establishes the "ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter" as a one of the main objective of this regulation (Art. 1, AI Act) and, in line with the risk-based approach adopted by the EU legislator, includes the assessment of the impact of the AI on fundamental rights[4] in all the risk management procedures established by the Act. From conformity assessment to the specific fundamental rights impact assessment under Article 27 of the AI Act, including a specific provision for general-purpose AI models with systemic risk, **the impact on fundamental rights must be taken into account in the design, development and deployment of AI systems and models**.

Against this background, **the provisions on how to conduct this assessment in the AI Act, but also in the Council of Europe Framework Convention, give only a limited guidance to those who have to carry out this assessment**. On the other hand, the proposed models and the initial debate in the literature show several shortcomings from a methodological point of view [MANTELERO, 2024]. In line with an empirical approach to law, **it is therefore necessary to move from the abstract elaboration, in law and in the legal debate, to concrete applications** in order to test the feasibility and effectiveness of the models for carrying out the impact assessment on fundamental rights in the context of AI.

The Catalan project is the first initiative based on the concrete implementation of a FRIA methodology applied to real cases and based on an active interaction with public and private entities that apply AI solutions in their business and activities. **The results of this empirical approach are crucial for the effective implementation of the AI Act**, as they show that it is possible to streamline the requirements of the Act and translate them effectively into a risk analysis and risk management process that is consistent with both general risk theory and the fundamental rights framework.

In addition, **the empirical evidence provided by this publication can contribute to the EU and international debate** on the model template for fundamental/human rights impact assessment by providing evidence on crucial issues such as (i) the relevant variables to be considered; (ii) the methodology to assess them and create risk indices; (iii) the role that standard questionnaires can play in this exercise and their limitations; (iv) the role of expert-based evaluation in this assessment.

The FRIA model applied in our use cases [MANTELERO, 2024; MANTELERO-ESPOSITO 2021], as well as the use cases themselves, are made publicly available in order to contribute to the EU and international debate on the protection of fundamental rights in the context of AI, and **to serve as a source of inspiration for the many entities around Europe and in non-EU countries that wish to adopt a**

[3] See also European Union Agency for Fundamental Rights, https://fra.europa.eu/en/content/what-are-fundamental-rights ("the term fundamental rights is used in a constitutional context whereas the term 'human rights' is used in international law. The two terms refer largely to the same substance as can be seen, for instance, by the many similarities between the Charter of Fundamental Rights of the EU and the European Convention on Human Rights.").
[4] Here and in the following pages, references to fundamental rights (or simply rights) include both fundamental rights and freedoms as enshrined in the Charter of Fundamental Rights of the European Union.

fundamental rights-based approach to AI, but do not have a tested reference model and concrete cases against which to compare their experiences.

With this objective in mind, the following sections will briefly discuss the role of the FRIA in AI regulation, its interaction with data protection regulation, the model template applied in the use cases, and the case selection criteria and areas covered.

2. The role of FRIA in the AI regulation

The last spring of AI, over the last few years, has led to machine learning applications being used in a variety of operational scenarios in both the public and private sectors. One of the main uses of AI and the source of the greatest challenges relates to the role these applications play in decision-making, which is even more critical in the public sector given the imbalance of power that in many cases characterises the relationship between natural persons and public powers.

This shift from human-based decisions to AI-based decisions, either fully automated or with the AI supporting human decision makers, raises various concerns about the accuracy of the logic of such AI systems, the way they represent society, and the human-machine interaction. While recent developments in AI are bringing significant improvements in many sectors, it is important to be aware of and manage the potential side effects of this technology.[5]

Risks associated with AI technologies can have a negative impact on society due to **security issues and fundamental rights issues**. While security issues, although complicated and related to inherent limitations of current AI models (e.g. the so-called hallucinations), can be easily addressed from a regulatory perspective using established practices and tools (e.g., standardisation, certification procedures, auditing, etc.), it is much more complicated to address fundamental rights issues.

Intrinsic, extrinsic, and contextual components of an AI system can vary with respect to the socio-technical environment in which it is used. **The way the AI has been designed** (intrinsic elements), in terms of training datasets, fine-tuning, etc., can affect the way it represents the societal aspects relevant to decision making, with potential risks, for example, of discrimination against underrepresented groups in AI-based assessment of the eligibility for welfare benefits. The **way in which AI systems interact with other technologies** (extrinsic elements), depending on the nature and performance of the latter, may limit the effectiveness of the AI systems and therefore negatively affect the related rights, such as in the case of an AI-powered medical imaging tool that poorly detects cancer due to the low quality of the medical device used to generate the images, thus compromising the right to health.

Finally, properly functioning AI systems, well-integrated with other devices, may have different impacts on fundamental rights depending on the **context of use**. The same AI-based video surveillance system can be considered a proportionate measure – despite its impact on individual and collective privacy and data protection, as well as on freedom of assembly in the case of demonstrations and protests – in the presence of high crime rates in some areas, and inappropriate in the absence of this overriding public interest.

Based on this brief analysis of the interaction between the development and deployment of AI and fundamental rights, the *raison d'être* of the FRIA in AI regulation becomes clear, as well as the core elements that need to be considered in order to conduct a proper assessment of the potential impact of AI on rights. The intrinsic and extrinsic dimensions of the AI systems, the context of use, the rights potentially affected, the need for a balancing test, the individuals and groups affected, the possible

[5] For a categorisation of AI-related risks, see also UNITED NATIONS, AI ADVISORY BOARD 2024; SLATTERY ET AL. 2024.

prevention/mitigation measures, are the different areas of the FRIA to be developed in order to deal with the aforementioned issues.

In this line, Article 27(1) of the AI Act on FRIA considers (i) the context of use and the categories of actors exposed to the risk (Art. 27 (1)(a), (b) and (c)), (ii) the potential prejudice to fundamental rights (art. 27(1)(d)), and (iii) the prevention/mitigation measures to be adopted (art. 27(i)(e) and (f)). This division is reflected in the **three phases of the FRIA methodology** adopted in this study, namely (i) planning, scope definition and risk identification, (ii) risk analysis, (iii) risk mitigation and management.

The first phase (planning, scope definition and risk identification) comprises the description of the AI system and its context of use, in relation to the potential intrinsic (related to the system itself) and extrinsic risks (related to the interaction between the system and the socio-technical environment in which it is implemented). This involves a description of the process in which the high-risk AI system will be used (art. 27(1) (a)), as well as the period of time within which each system is to be used and the associated frequency (art. 27(1) (b)). Once these elements have been defined, it is possible to make an initial identification of the areas of impact of the AI system in terms of the categories of individuals and groups concerned (art. 27(1) (c)) and the related rights that may be at stake.

This phase also serves as a preliminary assessment in order to exclude from the FRIA those cases where it is clear that there is no risk of prejudice to the persons concerned. On the other hand, if a potential harm is identified, it should be examined in the **second phase (risk analysis)**, which goes beyond a general identification of potential areas of impact and estimates the level of impact for each right or freedom.

There are several reasons why individualised estimation of the level of impact is essential. First, it is a feature of all impact assessments, from environmental to cybersecurity assessments: there can be no proper risk assessment without risk estimation. Second, estimation is the basis for the **third phase, which is the definition of risk prevention/mitigation measures (risk mitigation and management)**: if the impact has not been estimated, it is not possible to identify the appropriate and effective measures to eliminate/reduce the initial impact. Third, the estimation is, therefore, functional to the implementation of the key principle of accountability: only by defining the level of risk before and after the adoption of the prevention/mitigation measures is it possible to demonstrate that the risk has been specifically and effectively addressed. For all these reasons, Article 27(1) (d) and (f) play a central role and form the basis for the development of the risk assessment methodology.

3. The interplay with the data protection regulation

The interplay between the AI Act and the GDPR in relation to the FRIA is twofold. On the one hand, **Article 35 (1) and (7) of the GDPR** already includes an assessment of the risks to the rights of data subjects [ARTICLE 29 DATA PROTECTION WORKING PARTY 2017, 6], but this part has often been poorly implemented in the practice of DPIAs. On the other hand, the AI Act emphasises the assessment of the impact on individual rights, regardless of the use of personal data in the development and deployment of AI, but there is a lack of guidance in the implementation of the FRIA.

The close link between the DPIA [APDCAT 2024] and the FRIA is also evident in **Article 27 (4) of the AI Act**, which states that “if any of the obligations laid down in this Article is already met through the data protection impact assessment conducted pursuant to Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680, the fundamental rights impact assessment referred to in paragraph 1 of this Article shall complement that data protection impact assessment”. Along the same lines, under **Article 26(9) of the AI Act**, AI deployers must use the information provided under Article 13 of the AI Act (transparency and provision of information to deployers) to comply with their obligation to carry out a DPIA under Article 35 of the GDPR.

Against this background, if AI deployers fail to comply with the FRIA obligations under the AI Act, or if these obligations are not properly enforced by the competent authorities, **Data Protection Authorities (DPAs) may in future play an active role in enforcing the FRIA of AI systems through the provisions of Article 35 of the GDPR**, to the extent that the GDPR is applicable (this is the case in many situations where AI impacts individuals and groups, given the broad notion of personal data and data processing under the GPDR and the role of data in AI development, deployment and use).

Given all these different aspects of the interplay between the FRIA in the AI Act and Article 35 GDPR, and also given the experience of DPAs in dealing with fundamental rights issues [MANTELERO-ESPOSITO 2021], **an active role of these Authorities in providing guidance on the FRIA in the context of personal data-driven AI systems is appropriate and due**.

4. The FRIA template adopted

Before examining the methodology for carrying out the FRIA in the field of AI, it is worth noting that this is an expert-based assessment, which must be consistent with both the methodological approaches commonly used in risk assessment and management and the legal theory of fundamental rights. In this sense, **analysing the potential impact of AI on individual and group rights is a complex exercise that requires different types of expertise**, combining fundamental rights, AI design and an understanding of the contextual societal dimension. As with the DPIA, the FRIA is therefore not an exercise that can be undertaken by a layperson.

The **FRIA model must be consistent with risk assessment and management methodologies**. In this respect, from the perspective of AI providers and deployers, the FRIA is not a stand-alone task, but part of an integrated set of assessments that these entities must conduct to comply with legal obligations. For example, limiting the focus to the data-related issues, which are at the core of AI systems, data protection and data security risk assessments are required by EU and national legal instruments.

In addition, AI systems are often used as a component of other technologies (e.g., in accident prevention and detection), which must comply with technical standards that include elements of risk assessment. Against this scenario, the FRIA needs to be in line with the common risk assessment methodologies,[6] not only in terms of the scientific soundness of the methodological approach, but also to ensure consistency and full interoperability between the different components of the overall risk management strategy of AI providers and deployers in all their activities.

In the same way, **the FRIA model must be consistent with the legal theory and practice of fundamental rights**. In this respect, for example, it is not possible to define the overall impact of an AI system on fundamental rights, as in the case with cybersecurity, because fundamental rights must be considered individually and **cannot be assessed cumulatively, nor can different impacts be compensated** (for example, an AI system with a low impact on data protection and a high impact on freedom of expression does not have an overall medium impact as a result of a trade-off between these different levels of impact).

In line with these assumptions, the **three main blocks of the FRIA** template are:

[6] See, for example, ISO, Risk management. Guidelines. ISO 31000. https://www.iso.org/standard/65694.html, which identifies the following three main phases, combined with three complementary tasks (recording & reporting; monitoring & review; communication & consultation): (i) scope, context and criteria; (ii) risk assessment (risk identification, risk analysis, risk evaluation); (iii) risk treatment. The same approach can also be seen in UNDP 2024.

(i) a **planning and scoping phase**, focusing on the main characteristics of the product/service and the context in which it will be placed;

(ii) a **data collection and risk analysis phase**, identifying potential risks and assessing their potential impact on fundamental rights; and

(iii) a **risk management phase**, in which appropriate measures to prevent or mitigate these risks are adopted, tested and monitored for effectiveness.

In terms of structure, in accordance with Art. 27 of the AI Act and risk assessment methodologies, the FRIA is a **contextual assessment** focused on the specific AI solution being deployed and not a technology assessment centred on AI technologies in general and their various potential uses: it looks at a specific AI application and its context of use. In addition, the FRIA is also characterised by an ***ex ante* approach**, which makes it a tool for a fundamental rights-oriented AI design, adopting the by-design approach already known in data protection.

Finally, the FRIA has a **circular iterative structure**: like all risk assessments of situations that may evolve over time, it is not a one-off prior assessment. The main phases of risk management (planning/scoping, risk analysis, risk prevention /mitigation) are therefore repeated according to a circular iterative structure, as technological, societal and contextual changes affect some of the relevant elements of a previous assessment (Art. 27(2), AI Act).

4.1 The Planning and Scoping Phase

The planning and scoping phase starts with the needs analysis and the description of the AI solutions to be developed, and goes on to consider the contextual scenario of fundamental rights (including the controls already in place) and the potentially impacted areas. **Two main areas are examined at this stage: the inherent dimension of the AI system and the contextual dimension** (see Section 2 above). With regard to **the contextual dimension** of AI solutions, it is not limited to the identification of potentially affected rights and rights holders (without quantifying the impact, which is the objective of the following phase), but also includes a preliminary analysis of the relevant elements of the existing legal protection of these rights.

For the reasons discussed above regarding the variety of potential uses of AI, contexts of use, and potentially affected individuals and groups, it is not possible to provide a comprehensive questionnaire to be used by AI operators to address all the relevant issues for this first phase of the FRIA. Building on the DPIA experience, it is possible to provide **a non-exhaustive list of potential questions for FRIA planning and scoping**, which can be further supplemented by AI operators with specific questions based on the nature and use of the AI solutions being considered (see **Planning and Scoping Questionnaire** in Section 5 - The FRIA model template). However, as demonstrated in the use cases carried out in Catalonia (see Section 6), this questionnaire is effective in guiding AI operators through the planning and scoping phases, covering all the relevant areas with questions that can then be deepened with additional case-specific questions.

4.2 The Data Collection and Risk Analysis Phase

On the basis of the information gathered in the planning and scoping and risk identification phase, it is possible to determine whether it is necessary to carry out an in-depth analysis of the level of impact on individual rights. In the case of potentially affected rights, this analysis quantifies this impact in order to prevent or reduce it through appropriate measures.

The **analysis of the level of impact** of the AI solution on potentially affected rights is therefore the first phase of the FRIA's circular approach, which also includes the following three steps, as part of the Risk

Management Phase (see Section 4.3): (i) the identification of appropriate measures to prevent or mitigate the risk, (ii) the implementation of such measures, and (iii) the monitoring of the functioning of the AI system in order to revise the assessment and the measures adopted.

Impact assessment (circular approach) [Graph 1]

With regard to the Data Collection and Risk Analysis Phase, given the nature of the assessment, the data will relate to the different aspects of the rights potentially at stake, including information on the context of use and the individuals and groups potentially affected. Despite the variety of these elements and the specific nature of each right and freedom, the analysis phase can be based on key common parameters.

These parameters make it possible to operationalise an abstract concept such as impact on rights so that it can be assessed in a way that also makes it easier to (i) compare the level of impact on different rights in order to prioritise the risk prevention/mitigation, and (ii) understand how the impact on an individual right may change if some of the system or contextual elements vary.

4.2.1 Key variables for impact assessment

In line with risk theory and the fundamental rights legal framework, the impact on rights consists of two key dimensions: the **likelihood** of an adverse impact and its **severity**.[7] The combination of variables relating to these two dimensions provides a **risk index** that is assessed for each of the potentially affected rights.

To construct these indices, it is possible to represent the relevant variables on a scale from a minimum to a maximum (assuming, in line with general risk theory, that there is no zero risk) and by using ordinal variables (e.g., low, medium, high, very high). The use of scaling and associated variables makes it possible to compare different situations using the same variables, for example, the different levels of impact on non-discrimination produced by a credit scoring system using a particular algorithm when

[7] See Article 3(2) of the AI Act, which states that " 'risk' means the combination of the probability of an occurrence of harm and the severity of that harm", and Article 25(1) of the GDPR, which refers to "the risks of varying likelihood and severity for rights and freedoms of natural person […]".

changes are made to it. These ordinal variables can therefore be used to 'measure' the impact on a range-based quantification of risk (low, medium, high, very high).

However, the fundamental rights theory does not allow for the creation of a composite index, as is common in risk assessment, where all potential impacts are combined to create an overall impact index. This approach conflicts with the legal approach to fundamental rights where each right must be considered independently, in terms of its potential prejudice, and the fact that one right is less affected than another cannot lead to any form of compensation.

As the use cases discussed in Section 5 show, it is possible to assess the impact on the different rights involved, but not to say that a given AI system has an overall impact on rights that is considered as low, medium or high. The only possible interaction between different interests is through the **balancing test** in the presence of conflicting rights, but this test follows the assessment of the level of impact on each right. The balancing test does not relate to the level of risk to the affected rights, but to the overriding importance of one interest over another. It should therefore be considered as an external factor, to be taken into account only after the impact on individual rights has been assessed, and which may influence the results of the impact assessment by making an impact on certain rights acceptable because of a prevailing competing interest.[8]

Based on these considerations, a FRIA model will define a risk index for each potentially impacted right using the dimensions of likelihood and severity. The **likelihood** is understood as a combination of (i) the probability of adverse outcomes and (ii) the exposure. The first variable relates to the probability that adverse consequences of a given risk will occur and the second variable relates to the extent to which people potentially at risk could be affected. As far as exposure is concerned, it should be noted that the focus is on those potentially exposed to the use of the AI system (the identified population) and not on the population as a whole.

The **severity** of the expected consequences is based on two variables: (i) the gravity of the prejudice in the exercise of rights and freedoms (gravity),[9] based on their attributes, including taking into account group-specific impact, vulnerability and dependency situations; and (ii) the effort to overcome it and to reverse the adverse effects (effort).

Both likelihood and severity need to be assessed on a contextual basis, and the involvement of relevant stakeholders can be of help. As is common in risk assessment, the estimation of likelihood is based both on previous cases, looking at comparable situations, and the use of analytical and simulation techniques, based on possible scenarios of use. The same approaches are also used to estimate the level of severity, but in this case with greater emphasis on legal analysis regarding the gravity of prejudice, which should be assessed with reference to the case law on fundamental rights and the relevant legal framework.

On the basis of the likelihood and severity values derived from the above variables, a risk index is determined, which indicates the overall impact for each of the rights and freedoms considered.[10] It is worth noting that, these results must be combined with any elements that justify a limitation of some rights from a legal perspective, such as the mandatory nature of certain impacting characteristics: in this case, the potential risk must be considered acceptable to the extent that the AI system complies with the given legal requirements.

---

[8] See e.g., Part II, Use Case 1, below in the case of the development and use of an advanced learning analytics platform, where some impact on privacy and data protection rights is considered acceptable in view of the benefits for the right to education.
[9] The **gravity/seriousness of prejudice to a fundamental/human right is usually assessed according to the following three elements: (i) its intensity, (ii) the consequences of the violation, and (iii) its duration**, where the intensity of the violation is related to the importance of the protected legal interest violated. See also EUROPEAN COURT OF HUMAN RIGHTS 2022.
[10] See the following section for the methodology used to combine the different variables and create the indices.

4.2.2 Variables and construction of the impact index

In many risk-based impact assessment models and standards, risk indices are constructed using matrices because they are relatively easy to use and explain.[11] As a risk matrix is a graph that combines two dimensions using colours to reflect different levels of risk, they are useful for assessing indices generated by different variables. For this reason, they can be used in the FRIA to define the level of impact on each right concerned.

The methodology proposed here uses a risk index for each potentially impacted right, based on a matrix combining two dimensions (likelihood and severity). Each of these dimensions results from the combination of two pairs of variables, also constructed using matrices: the probability of adverse consequences, and exposure, for likelihood; the gravity of prejudice, and the effort to overcome it and to reverse adverse effects, for severity. **There is no single risk matrix model** to be used in risk assessment; practice in this field shows a variety of models. The most common are 3x3, 4x4, 5x5, 5x4 and 6x4 matrices, where the pairs of numbers indicate the number of ranges of the two scales defining the dimension under consideration. As the matrix refers to two independent variables, they can be evaluated according to scales that may differ in number of ranges, for example a 6x4 scale where six different ranges are provided for one variable and only four for the other.

The **4x4 matrix** may be the most appropriate in the context of FRIA, as it reduces the risk of average positioning, gives more attention to the high and very high levels in a way that is consistent with the focus on high risk in the current regulatory approach to AI, and does not excessively fragment the lower part of the scale, which is less relevant due to the aforementioned focus.

In matrices, descriptive labels are used for the different combinations of levels in the colour scale, as follows in this example of a severity matrix:

		Gravity			
		Low (L)	**Medium (M)**	**High (H)**	**Very high (VH)**
Effort	**Low (L)**	L	L/M	L/H	L/VH
	Medium (M)	M/L	M	M/H	M/VH
	High (H)	H/L	H/M	H	H/VH
	Very high	VH/L	VH/M	VH/H	VH

Severity			
Low	Medium	High	Very high

[11] See also APDCAT 2024, 33 and 53; CNIL 2018, 23.

4.3 The Risk Management Phase

Following the risk analysis, which has defined the level of impact of the AI solution on potentially affected rights, it is necessary to manage the identified risks by adopting appropriate measures.[12] The third phase of the FRIA is therefore articulated in three steps, as follows:

(i) the **identification of appropriate measures** to prevent or mitigate the risk, taking into account their impact on the risk level according to a context-specific scenario analysis;

(ii) the **implementation** of such measures;

(iii) the **monitoring** of the functioning of the AI system in order to revise the assessment and the adopted measures should technological, societal and contextual changes affect the level of risk or the effectiveness of the adopted measures.

As the FRIA is not a final check of an AI solution, but **a design tool to guide the development and deployment of AI** towards a fundamental rights-oriented approach, monitoring the functioning of the AI system can also be part of the pre-market phase in which different design solutions are tested and implemented in order to select the most appropriate one. In line with the **circular approach to risk assessment** and AI design, it is therefore possible that several series of risk assessments, implementation of mitigation measures and re-assessments may take place until the final version of the AI product/service results in a level of **residual risk** that is satisfactory in terms of acceptability and can be placed on the market or put into service.

In addition, **changes in the technological and societal scenario or in the specific context of use may occur** after the AI tool has been placed on the market or put into service. These may have an impact on the level of risk previously assessed with respect to the rights concerned, as well as raise new concerns with respect to other rights. In such cases, the AI solutions adopted will be re-assessed and appropriate measures taken.[13]

In line with these observations, the FRIA model template (see Section 5) includes a matrix showing the impact of the risk prevention/mitigation measures adopted on the risk levels identified in the risk analysis phase and the resulting residual risk.

5. The FRIA model template

The model template applied in the use cases is based on the three phases of the FRIA discussed in the previous section. It consists of several elements. The first one is a questionnaire covering the four main areas of the planning and scoping phase, namely the description and analysis of the AI system, the legal context, the controls already in place, and stakeholder engagement (**Planning and Scoping Questionnaire**).

The second element of the template is a set of matrices (**Tables 3, 6 and 7**) and associated variable quantification criteria that are used to assess the likelihood and the severity of the potential prejudice to each right and freedom (**Tables 1, 2, 4 and 5, and Table 1A**), and to estimate the associated overall impact.

The third element of the template consists of two tables, one indicating the level of impact on each right and freedom and the prevention/mitigation measures identified to address the risk (**Table 2A**), and another estimating the residual risk resulting from the adoption of these measures (**Table 3A).**

[12] See also Article 27(1)(f), AI Act.
[13] See also Article 27(2), AI Act.

Planning and Scoping Questionnaire

Section A Description and analysis of the AI system, including related data flows	What are the main objectives of the AI system? What are the main features of the system? In which countries will it be offered? What types of data are processed (personal, non-personal, special categories)? Identification of potential rights holders: who are the individuals or groups likely to be affected by the AI system, including vulnerable individuals or groups? Identification of duty bearers: who is involved in the design, development and deployment of the AI system? What is their role?
Section B Fundamental rights context	What fundamental rights are potentially affected by the use of the AI system? What international/regional legal instruments for the protection of human/fundamental rights have been implemented at the operational level? What are the most relevant fundamental rights courts or bodies in the context of use? What are the most relevant human/fundamental rights decisions and provisions?
Section C Controls in place	What policies and procedures are in place to assess the potential impact on fundamental rights, including stakeholder participation? Has an impact assessment been conducted, developed and implemented in relation to specific issues (e.g. data protection) or certain features of the system (e.g. use of biometrics)?
Section D Stakeholder engagement and due diligence	Who are the main groups or communities potentially affected by the AI system, including its development? Which stakeholders should be involved in addition to the individuals or groups potentially affected by the AI system (e.g. civil society and international organisations, experts, industry associations, journalists)?

	Are there other duty bearers that should be involved in addition to AI providers and deployers (e.g. national authorities, government agencies)? Have business partners, including service providers (e.g. subcontractors for AI systems and datasets), been involved in the assessment process? Has the AI provider carried out an assessment of its supply chain to determine whether the activities of suppliers/contractors involved in product/service development may affect fundamental rights? Has the provider promoted fundamental rights standards or audits to ensure respect for fundamental rights among suppliers? Have the AI provider and AI deployer publicly communicated the potential impact of the AI system on fundamental rights? Have the AI provider and AI deployer provided training on fundamental rights standards to management and procurement staff dealing with the AI system?

Risk matrices

Tab. 1 Probability

Low	The risk of prejudice is improbable or highly improbable
Medium	The risk may occur
High	There is a high probability that the risk occurs
Very high	The risk is highly likely to occur

Tab. 2 Exposure

Low	Few or very few of the identified population of rights holders are potentially affected
Medium	Some of the identified population are potentially affected
High	The majority of the identified population is potentially affected
Very high	Almost the entire identified population is potentially affected

Tab. 3 Likelihood

		Probability			
		Low	**Medium**	**High**	**Very high**
Exposure	**Low**	L	L/M	L/H	L/VH
	Medium	M/L	M	M/H	M/VH
	High	H/L	H/M	H	H/VH
	Very high	VH/L	VH/M	VH/H	VH

Likelihood			
Low	Medium	High	Very high

Tab. 4 Gravity of the prejudice

Low	Affected individuals and groups may encounter only minor prejudices in the exercise of their rights and freedoms.
Medium	Affected individuals and groups may encounter significant prejudices.
High	Affected individuals and groups may encounter serious prejudices.
Very high	Affected individuals and groups may encounter serious or even irreversible prejudices.

Tab. 5 Effort to overcome the prejudice and to reverse adverse effects

Low	Suffered prejudice can be overcome without any problem (e.g. time spent amending information, annoyances, irritations, etc.).
Medium	Suffered prejudice can be overcome despite a few difficulties (e.g. extra costs, fear, lack of understanding, stress, minor physical ailments, etc.).
High	Suffered prejudice can be overcome albeit with serious difficulties (e.g. economic loss, property damage, worsening of health, etc.).
Very high	Suffered prejudice may not be overcome (e.g. long-term psychological or physical ailments, death, etc.).

Tab. 6 Severity

		Gravity			
		Low	**Medium**	**High**	**Very high**
Effort	**Low**	L	L/M	L/H	L/VH
	Medium	M/L	M	M/H	M/VH
	High	H/L	H/M	H	H/VH
	Very high	VH/L	VH/M	VH/H	VH

Severity			
Low	Medium	High	Very high

Tab. 1A Data collection and risk analysis

Rights/ freedoms potentially affected	**Description of the impact**	**Likelihood**			**Severity**		
		Probability	**Exposure**	**Likelihood**	**Gravity**	**Effort**	**Severity**

Tab. 7 Overall risk impact

		Severity			
		Low	**Medium**	**High**	**Very high**
Likelihood	**Low**				
	Medium				
	High				
	Very high				

Overall risk impact			
Low	Medium	High	Very high

Tab. 2A Risk management (I)

Rights/freedoms affected	Likelihood	Severity	Overall impact	Impact prevention/ mitigation measures

Tab. 3A Risk management (II)

Rights/freedoms affected	Likelihood (residual)	Severity (residual)	Residual overall impact

6. From model to practice: the use cases

The ongoing methodological debate on the FRIA and its implementation has been characterised as a mainly policy and theoretical discussion, with little attention paid to empirical analysis and the full implementation in the real world of the different models proposed. In this context, the project led by the Catalan Data Protection Authority makes a difference by bringing the FRIA debate to the concrete experience of providers and deployers directly involved in the use of AI.

This case-based empirical approach is crucial to test the effectiveness of the proposed model in achieving the policy and design objectives of the FRIA as elaborated by the EU legislator in the AI Act. More specifically, the use cases in this project have shown that it is possible **to streamline the FRIA procedure by avoiding the adoption of a long checklist and focusing on the core elements of the impact on fundamental rights.**

The use cases have also demonstrated that, at least for the first round of assessment, mitigation, and reassessment, people with the appropriate background can complete the FRIA within two or three short meetings (3 hours per meeting). This confirms that the **FRIA, if properly framed, does not impose an excessive additional burden on private and public entities** in the EU in order to comply with the AI Act. Finally, each use case was based on four different interactions: (i) an initial internal analysis of the case by the experts (usually DPOs) of the entities carrying out the FRIA, with the aim of outlining the key elements of the FRIA paying, particular attention to the planning and scoping phase; (ii) a discussion with a FRIA expert to review the initial assessment; (iii) a group discussion involving the experts from all the entities involved in the project; and (iv) a final review by the experts of the entity performing the FRIA. This four-step process illustrates two key aspects of the FRIA: the importance of an expert assessment and the importance of a team-based assessment involving different expertise to improve the level of analysis.

With regard to the selection of the cases, it is worth noting that this is an ongoing project and the cases presented in this report are the first to have been discussed and in which the FRIA template has been applied. Other cases are under evaluation and will be published in the future on the website "DPD en xarxa" (https://www.dpdenxarxa.cat/) and on the official website of the Catalan Data Protection Authority (https://www.apdcat.cat). Moreover, some cases in which the FRIA template has been applied, with a relevant impact on the design of AI solutions, have not been included in this report for reasons of confidentiality, but have been useful for all participants to better elaborate the practice of the FRIA template.

In terms of the areas covered, the use cases relate to four of the key areas listed in Annex III of the AI Act, namely education (assessment of learning outcomes and prediction of student dropout), workers' management (decision support for human resource management), access to healthcare (cancer treatment based on medical imaging), and welfare services (voice assistant for elderly people), which also represent the areas where AI solutions are increasingly being used, with the greatest impact on individuals and groups. In this sense, the nature of the use cases discussed will also make them useful to many other public and private entities in other countries interested in designing AI systems/models that are compliant with fundamental rights in these core areas.

In line with the aim of this project, the use cases are presented as they were developed by the participants, rather than as best practice or standardised cases. The project was designed to test the effectiveness of the model template and associated methodology. In line with DPIA experience, we have given participants the freedom to develop the different parts of the template according to their approach, so that some analyses are more extensive and others more concise. However, the core elements (the questions, the matrices, the assessment methodology) remain the same.

The main idea is that, in this report, it is important to reflect on the exercise carried out in order to show the results obtained, and not to present the FRIAs carried out as fictional, perfect cases. The FRIA has been and will be implemented by a variety of actors, in some cases in more detail, in others with some limitations, but to the extent that it contributes to effective analysis and prevention/mitigation of the impact on fundamental rights, it will have achieved its main objective.

References

All the websites cited in this document were accessed between September and December 2024.

APDCAT. 2024. Avaluació d'impacte relativa a la protección de dades, https://apdcat.gencat.cat/web/.content/03-documentacio/Reglament_general_de_proteccio_de_dades/documents/Guia-Practica-avaluacio-impacte-proteccio-de-dades-2019.pdf.

APDCAT. 2020. Artificial Intelligence. Automated Decisions-making in Catalonia, Informe-IA-Angles-Report-Final.pdf

Article 29 Data Protection Working Party. 2017. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, WP 248 rev.01, https://ec.europa.eu/newsroom/article29/items/611236.

CNIL. 2018. Privacy Impact Assessment (PIA). Templates, https://www.cnil.fr/en/privacy-impact-assessment-pia.

European Court of Human Rights.2022. Guide on Article 8 of the European Convention on Human Rights. Right to respect for private and family life, home and correspondence, https://ks.echr.coe.int/web/echr-ks/article-8.

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

Mantelero, A. and Esposito, M.S. 2021. An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. 41 Computer Law & Security Review 105561, https://doi.org/10.1016/j.clsr.2021.105561 (open access).

Slattery, P. et al. 2024. A systematic evidence review and common frame of reference for the risks from artificial intelligence. http://doi.org/10.13140/RG.2.2.28850.00968 and https://airisk.mit.edu/.

United Nations, AI Advisory Board. 2024. Governing AI for Humanity. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.

UNDP. 2024. Visual guide. Business Process for Risk Management, https://popp.undp.org/sites/g/files/zskgke421/files/2024-05/Risk_Management_Full_Visual_Guide_13.pdf.

Part II – Use cases

Use case 1: An advanced learning analytics platform

1. The context

The following use case relates to one of the major problems of the higher education system, namely the early drop-out of the 18-24 year old population from education and training. This situation has given rise to concern at European level and is reflected in the European statistical data available in the Eurostat database.[14] This is why priority 1 of the strategic framework for European cooperation in education and training for the European Education Area (EEA) and beyond (2021-2030),[15] is "*improving quality, equity, inclusion and success for all in education and training*". Although early school leaving has been reduced over the last decade, it remains a challenge. In order to avoid limiting young people's access to future socio-economic opportunities, particular attention needs to be paid to groups at risk of low educational attainment and early school leaving.

Higher education institutions need to promote educational strategies that support successful completion of education and training pathways, reduce early drop-out rates, and address the causes of underachievement. It is therefore important to be discerning about the data available, to structure it, to extract the information it provides and to use it for the specific purpose we want to achieve.

In order to identify where each student is at a given point in their studies and to be able to predict what will happen next, the following data may be relevant:

- Historical data on student careers collected in previous years
- Data provided by the previous school
- Data provided by the student himself/herself at the time of enrolment
- Data collected during the course of his/her studies.

The information provided by these datasets can be used to identify patterns in student performance. In addition to having a global view of the situation of the entire student population in the same programme and its future development, it is also possible to have an individualised view of each student's situation and to anticipate his/her future development. This information helps to promote strategies based on the student's current situation, offering personalised treatment and taking into account the needs of each student.

The analysis of this data and its interpretation for improvement and progress in the field of education falls under the umbrella of what is known as "*learning analytics*". At the 1st International Conference on Learning Analytics and

[14] Access to Eurostat: https://ec.europa.eu/eurostat

[15] Council Resolution on a strategic framework for European cooperation in education and training for the European Education Area and beyond (2021-2030): https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021G0226(01)

Knowledge (LAK) in 2011, learning analytics was defined[16] as the measurement, collection, analysis and reporting of data about learners and their contexts for the purposes of understanding and optimising learning and the environments in which it takes place.

Although learning analytics has long been researched and used to predict students' academic success and risk of dropping out of school, the emergence of new technologies that provide alternative analytical techniques has highlighted the need to address the legal requirements arising from recent regulations in order to make proper use of them.

Given the current situation in which we find ourselves, in which artificial intelligence systems (hereafter referred to as AI systems) are becoming a regular component of our daily tasks, it is essential that institutions identify and understand their risks in order to prevent, minimise and manage them, and make the best use of them as allies in the modernisation and digitisation of the university system.

Some of the difficulties identified, apart from the expertise required to interpret the information obtained from the data, relate to the difficulty for institutions and their staff to adapt easily to new technologies in their teaching approaches and methodologies.

With the resources available in the field of education, it is a fact that higher education institutions will need to introduce AI systems into those processes where human intervention may be limited. We need to be aware of the changes that are occurring and that may occur, such as the advent of generative AI[17] in the student learning model. Thus, if we decide to use the available AI systems, appropriate tools could be put in place in advance to protect the rights and freedoms of students from an early stage and to avoid harming their integration into the university system and, therefore, the student population. However, there are still challenges to be addressed, such as data bias and ethical dilemmas that may arise, as well as issues related to AI development, barriers and resistance to the integration of AI systems in some areas that may hinder or slow down the growth in these areas. In our case, this could lead to the obsolescence of the university system and the opposite of the desired effect, which could lead to a setback in the teaching and training of young people.

When the use of AI systems is being considered, it is necessary to analyse and reflect on the different approaches and cases that we may encounter form the very begging, in order to take all the necessary precautions and to establish the technical and organisational measures to build solid solutions that guarantee the protection of young people, respecting human rights and social values.

The rapid increase in the use of AI systems in education is transforming the way we teach and learn, with a direct impact on institutions, their staff and the students themselves. This change goes hand in hand with the need to equip staff with new skills[18] to deal with this new and evolving technological landscape, and therefore to be empowered to use the information provided by AI systems.

The use of AI in education raises fundamental questions, which is why the European Artificial Intelligence Act (AI Act) itself has identified as high-risk AI systems "*AI systems intended to be used to evaluate learning outcomes, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions at all levels*" (Annex III, section 3(b)).

[16] https://www.solaresearch.org/about/what-is-learning-analytics/
[17] See the definition of Generative AI and how it works in UNESCO's "*Guidance for generative AI in education and research*", 2024: https://unesdoc.unesco.org/ark:/48223/pf0000386693

[18] See UNESCO. 2024. AI competency framework for teachers, https://unesdoc.unesco.org/ark:/48223/pf0000391104.

Although there are sectors that are reluctant to use AI in their processes because of what it might imply, it is undeniable that a good use of AI in education seems to be enriching in terms of bringing learning closer to the new generations of students.

In the absence of integration with the educational institution's own application, there are already learning analytics platforms on the market that compile information obtained from student self-reported data in the institution's system "*Student Information System (SIS*)" and from data generated by Learning Management Systems (LMS) to identify and/or predict students at risk of dropping out. For example:

- [19] Assessment and Learning in Knowledge Spaces[20] (ALEKS)
- DreamBox[21]
- Carnegie Learning [22]
- Smart Sparrow[23]
- The IntelliBoard

Learning analytics, by anticipating the different scenarios that may arise, reduces the effort needed to lower the drop-out rate and facilitates the adoption of more effective policies and measures that focus on the root of the problem. The provision of appropriate and personalised tools for education, tailored to the needs of individual students, can enable young people to continue their personal growth in education, which can then be transferred to their professional lives. For this reason, considering the potential of AI systems to transform the current notion of education, the use case that has been carried out has focused on the framework of learning analytics using a high-risk AI system based on a predictive algorithm.

As you can see in the following sections, the use case that has been presented and analysed has left out the use of automated decision algorithms (ADA). To learn more about the use cases of ADAs, see the report "*Artificial Intelligence. Automated Decisions in Catalonia*[24]" prepared by the Catalan Data Protection Authority (APDCAT).

2. The project

The project aims to design and development of a new 'Learning Analytics' ecosystem for the higher education system, with the creation of an advanced learning analytics platform using an AI system to assess learning outcomes and predict the risk of students dropping out. in particular, the platform will be used to:

- Manage data related to teaching and learning processes;
- Develop dashboards that monitor, analyse and visually display both the results of all students as a whole and the individual results of each student according to pre-defined indicators;
- Design and modulate the methodology used;
- Analyse students' academic performance;
- identify students at risk of dropping out.

The data used within the framework of the platform aims to detect early, quickly and effectively the risk of students dropping out of their studies. Similarly, the ultimate purpose of using the data is to improve the learning process through psycho-pedagogical counselling of the student population.

[19] Real application cases of The IntelliBoard platform: https://intelliboard.net/customers/
[20] https://www.aleks.com
[21] https://www.dreambox.com
[22] https://www.carnegielearning.com
[23] https://www.smartsparrow.com
[24] https://apdcat.gencat.cat/web/.content/03-documentacio/intelligencia_artificial/documents/Informe-IA-es.pdf

It should not be forgotten that, in relation to the academic training they receive, students have the right to tutoring and counselling by the staff of the institution in order to have appropriate guidance for their learning process.

In order to provide students with personalised and timely information about their learning, the following information is needed:

1. Aggregated historical data generated in learning management systems (*LMS*), as well as data collected from the institution's internal databases on students from previous years who have passed or dropped out of higher education.
2. Data coming directly from the educational institutions attended by the student before entering the higher education system.
3. Data requested by the higher education institution during the enrolment process and provided by the student (*SIS*).
4. Data obtained from the student's performance during their careers (academic record information).
5. Data relating to personal circumstances during the student's career (e.g. if the student combines study with work, if the student has received scholarships, if the student moves to another institution, etc.).

In order to provide a better context for the predictive algorithm, and thus obtain more reliable results, the student population has to be characterised according to the data available at any given time. Before the start of their university studies, students must be grouped according to the information provided by the secondary education institutions before the higher education entrance examination and by the students themselves at the time of enrolment. This initial profiling of the student population will provide a first insight into the potential performance of the student. Depending on the data obtained, the classification of students may vary according to the metrics and indicators defined from the historical data. To ensure compliance with the data requirements of the RIA, the datasets used to train the AI system for student classification and profiling are anonymised and untraceable.

Although the following section deals with the concrete assessment of the use case, it is worth mentioning some of the different aspects that were considered, in order to better understand the context of the use case. First, the use of certain special categories of personal data for this AI application, such as health-related data providing information on students with special educational needs and disabilities (SEND), was considered in the evaluation of the project, but excluded in this case.

Second, changes were made to the dashboards and the original idea of using different dashboards depending on the data selected and its relevance to those who could access it. A three-layered representation was proposed, based on different colours depending on the student's situation at any given moment: green colour, indicating that the student has a probability of not dropping out of more than 80%; yellow colour, indicating a probability between 20% and 60%; and red colour, indicating that the student has a probability of dropping out of higher education of more than 50%. During the discussion of the use case, the following changes were made:

- It was decided that access to the full range of information from the dashboards should be restricted to tutors, so that they could be aware of the predicted risk of students dropping out and act on it in accordance with their competences and the regulations of the higher education system.
- The information that appeared on the teachers' dashboard was restricted, limiting their access to aggregated data on the performance of the whole student population within their classes. The lack of access to individualised information about each student avoids the side effect of unconsciously stigmatising the small group of underperforming students from the outset.
- The visualisation of the dashboards by the students was also removed in order to protect the mental and emotional health of the

students and not to provoke situations of stress and/or anxiety, among others, due to the fact that they see a certain colour indicating their academic performance.

The assessment also showed that there is information about events that affect students' performance (e.g. death of a close relative) that is not taken into account by the system. It is therefore important to treat the potential risk of dropping out on a case-by-case basis and to collect the related relevant information that needs to be taken into account, not only at an individual level, but also to consider its inclusion in the variables that feed into the predictive AI system.

A Data Protection Impact Assessment (DPIA) was conducted to complement the FRIA. Although the former assessment is not included in this document, it led to changes in relation to data protection, applying the principles set out in the General Data Protection Regulation (GDPR). These changes have enriched the final version of the FRIA.

3. The FRIA

Planning and scoping

Section A Description and analysis of the AI system, including related data flows	**What are the main objectives of the AI system?**	a) To provide indicators of academic performance; b) To predict the likelihood of dropping out of higher education; c) Contribute to the improvement of the learning process. With the information provided by the AI system, higher education institutions can establish policies to reduce the rate of early drop-out, as well as help tutors with the task of counselling students
	What are the main features of the system?	AI-based learning analytics to predict situations in which action can be taken to improve the learning process.
	In which countries will it be offered?	Spain. It may be extended to foreign higher education institutions (inside or outside the European Union) with which joint degree/exchange agreements have been formalised.

	What types of data are processed (personal, non-personal, special categories)?	▪ Aggregated historical data on the student population from previous years. ▪ Personal and non-personal data obtained directly from the educational institutions where the student has been before entering the higher education system. ▪ Data requested by the higher education institution during the enrolment process and provided by the student. ▪ Data obtained from the student's behaviour during his/her studies (information related to the academic record). ▪ Data relating to personal circumstances during the student's career.
	Identification of potential rights holders: who are the individuals or groups likely to be affected by the AI system, including vulnerable individuals or groups?	Students.

	Identification of duty bearers: who is involved in the design, development and deployment of the AI system? What is their role?	Higher education institutions wishing to implement the AI system will be responsible for its design, delivery and development. The management of the data will involve the staff of the institution. Furthermore, the processing of the information provided by the data in the tracking dashboards will be visualised differently according to the user profile: (i) students, (ii) teaching staff, (iii) tutors and (iv) managers.
Section B Fundamental rights context	**What fundamental rights are potentially affected by the use of the AI system?**	Having analysed all the individual, civil, political, economic and social rights set out in the Charter of Fundamental Rights of the European Union, it has been concluded that the following rights are potentially affected: ☒ **Human dignity** *(Article 1)* Justification: Lack of complete vision; loss of personal autonomy if AI makes decisions and offers solutions for individuals without human intervention; excessive pressure on students and impact on self-perception. ☒ **Respect for private and family life** *(Article 7)* Justification: Invasion of privacy due toconstant monitoring of academic performance; impact on family privacy and impact on "*decisional privacy*". ☒ **Protection of personal data** *(Article 8)* Justification: Profiling and assessment of individuals, large-scale data and use of new technologies.

		☒ **Non-discrimination** *(Article 21)* Justification: Use of algorithms based on historical data and patterns that may be influenced by past discriminatory bias and perpetuate/exacerbate prejudices; failure to take into account certain factors or variables that may be relevant; assessment based on predictive analysis. While the **right to education** *(Article 14)* was considered, it was concluded that it is not affected as the AI system does not restrict the right to access to education.
	What international/regional legal instruments for the protection of human/fundamental rights have been implemented at the operational level?	The regulations concerning personal data protection and the university system, as well as those concerning the groups potentially affected (e.g. the Statute of University Students).
	What are the most relevant fundamental rights courts or bodies in the context of use?	The data protection supervisory authorities of the country/region where the AI system is developed and used, as well as the competent courts of that country/region.
	What are the most relevant human/fundamental rights decisions and provisions?	Not applicable (N/A).

Section C Controls in place	**What policies and procedures are in place to assess the potential impact on fundamental rights, including stakeholder participation?**	Not applicable (N/A).
	Has an impact assessment been conducted, developed and implemented in relation to specific issues (e.g. data protection) or certain features of the system (e.g. use of biometrics)?	A personal data protection impact assessment (DPIA) has been carried out.
Section D Stakeholder engagement and due diligence	**Who are the main groups or communities potentially affected by the AI system, including its development?**	Students.
	Which stakeholders should be involved in addition to individuals or groups potentially affected by the AI system (e.g. civil society and international organisations, experts, industry associations, journalists)?	Families and teachers.
	Are there other duty bearers that should be involved in addition to AI providers and deployers (e.g. national authorities, government agencies)?	Data protection supervisory authority, Department of Education/Universities; Spanish AI Supervisory Agency; AI Commission of the Generalitat de Catalunya. Competent unit/body of the institution.

	Have business partners, including service providers (e.g. subcontractors for AI systems and datasets), been involved in the assessment process?	No
	Has the AI provider carried out an assessment of its supply chain to determine whether the activities of suppliers/contractors involved in **product/service development may affect fundamental rights?** **Has the provider promoted fundamental rights standards or audits to ensure respect for fundamental rights among suppliers?**	Not applicable (N/A).
	Have the AI provider and AI deployer publicly communicated the potential impact of the AI system on fundamental rights?	Publication is not mandatory, but it is advisable to include a summary of the analysis conducted (both of the DPIA and the FRIA), in view of the principle of transparency and trust in the IA system.
	Have the AI provider and AI deployer provided training on fundamental rights standards to management and procurement staff dealing with the AI system?	Not applicable (N/A).

Risk matrices

Tab. 1 Probability

Low	The risk of prejudice is improbable or highly improbable
Medium	The risk may occur
High	There is a high probability that the risk occurs
Very high	The risk is highly likely to occur

Tab. 2 Exposure

Low	Few or very few of the identified population of rights holders are potentially affected
Medium	Some of the identified population are potentially affected
High	The majority of the identified population is potentially affected
Very high	Almost the entire identified population is potentially affected

Tab. 3 Likelihood

		Probability			
		Low	**Medium**	**High**	**Very high**
Exposure	**Low**	L	L/M	L/H	L/ VH
	Medium	M/L	M	M/H	M/ VH
	High	H/L	H/M	H	H/ VH
	Very high	VH/L	VH/M	VH/H	VH

Likelihood			
Low	Medium	High	Very high

Tab. 4 Gravity of the prejudice

Low	Affected individuals and groups may encounter only minor prejudices in the exercise of their rights and freedoms
Medium	Affected individuals and groups may encounter significant prejudices
High	Affected individuals and groups may encounter serious prejudices
Very high	Affected individuals and groups may encounter serious or even irreversible prejudices

Tab. 5 Effort to overcome the prejudice and to reverse adverse effects

Low	Suffered prejudice can be overcome without any problem (e.g. time spent amending information, annoyances, irritations, etc.)
Medium	Suffered prejudice can be overcome despite a few difficulties (e.g. extra costs, fear, lack of understanding, stress, minor physical ailments, etc.)
High	Suffered prejudice can be overcome albeit with serious difficulties (e.g. economic loss, property damage, worsening of health, etc.)
Very high	Suffered prejudice may not be overcome (e.g. long-term psychological or physical ailments, death, etc.)

Tab. 6 Severity

		Gravity			
		Low	**Medium**	**High**	**Very high**
Effort	**Low**	L	L/M	L/H	L/ VH
	Medium	M/L	M	M/H	M VH
	High	H/L	H/M	H	H/ VH
	Very high	VH/L	VH /M	VH /H	VH

Severity			
Low	Medium	High	Very high

Tab. 1A Data collection and risk analysis

Rights/freedoms potentially affected	**Description of the impact**	Likelihood			Severity		
		Probability of adverse outcomes	Exposure	**Likelihood**	Gravity	Effort	**Severity**
Human dignity	The algorithm takes into account some parameters generated by historical data collected within a given socio-economic context, but not all those that could have a direct influence on current academic performance (e.g., requested teaching improvements/adaptations; people who do not identify with a particular gender, access to new technologies, etc.).	[High] The likelihood of the risk occurring is high because the information derived from the available data does not include all information on all potentially affected groups.	[Low] The exposure is low as it concerns a limited number of cases of missing information.	**[Medium]**	[Medium] Although the omission of certain parameters may only concern small groups, the students affected may be significantly biased. The AI algorithm may predict performance for this small group that does not reflect their situation.	[Medium] The prejudices suffered can be overcome despite some difficulties. In the case of students requesting teaching improvements, teaching staff and tutors will be informed in advance. As for the other omitted parameters and the associated negative impact on minority groups, they can be taken into account when	**[Medium]**

						improving the AI system.	
Respect for private and family life	Constant monitoring of academic performance; impact on family privacy; impact on 'decisional privacy'.	[High] There is a high probability of the risk occurring. Although the different procedures in place in the relevant institution already process the different information separately, the fact that certain information is now collected together for the purposes of this project have an impact on the control of this information.	[Very high] The exposure is very high, as it would affect all students.	**[Very high]**	[Low] The students concerned may encounter minor prejudices in the exercise of their rights and freedoms, since the information is used within the framework of the institution's educational functions.	[Low] Higher education institution staff have duties and obligations to safeguard the rights of students within the framework of their functions. The institution must also train its staff in this area so that they are aware of the applicable regulations and can act in the different situations they may face.	**[Low]**
Protection of personal data	The AI system collects large-scale data and uses new technologies. It also profiles students to assess	[Medium] There is a risk of inaccurate profiling and prediction. A data protection impact	[Very high] The exposure is very high, as it would	**[High]**	[Medium] Inaccurate profiling negatively impacts on the accurate	[Medium] Applying the GDPR, appropriate organisational measures must	**[Medium]**

	and predict their risk of dropping out.	assessment is required.	affect all students.		representation of student performance and expected outcomes. Consideration has been given to whether this is a case where students also have the right not to be profiled.	be in place and protect students' rights in relation to data processing, but the way profiles are generated and used may require some changes in the AI systems design (e.g. fine-tuning) and use.	
Non-discrimination	Given that the AI system compares historical data, obtains data from other institutions and collects enrolment data, there may be historical discriminatory biases that may be perpetuated and amplified; failure to take into account certain factors or variables that may be relevant; predictive nature of the evaluation.	[Medium] The likelihood is medium given the limited weight of the variables in the consideration of the predictive model of the risk of dropout.	[Very high] The exposure is very high, as the impact would potentially affect all students to whom the algorithm would be applied.	**[High]**	[Medium] The classification of students may be biased and provide misleading information on early indicators of drop-out risk resulting in unjustified unequal treatment.	[Medium] The classification of students is not static, so the initial data will not place them in a particular cluster, but may change as they progress through their studies.	**[Medium]**

Tab. 7 Overall risk impact

		Severity			
		Low	**Medium**	**High**	**Very high**
Likelihood	**Low**				
	Medium				
	High				
	Very high				

Overall risk impact			
Low	Medium	High	Very high

Tab. 2A Risk management (I)

Rights/freedoms affected	Likelihood	Severity	Overall impact	Impact prevention/mitigation measures
Human dignity	[Medium]	[Medium]	**[Medium]**	▪ Use of predictive modelling as a decision support tool and rather than an automated decision making tool; limited use of the results provided by the AI system. ▪ Not providing students with drop-out risk rates. ▪ Provide the institution's staff with guidelines for the use of the AI system (usage policy).
Respect for private and family life	[Very high]	[Low]	**[Medium]**	▪ Design the predictive model in a way that ensures control of the data at all times. ▪ Limit access to individual profiles. Students should not be able to view other students' profiles. ▪ The predictive tool must not take into account the interactions and communications that students have with the teaching staff or with each other. ▪ The tool should be used as a support tool for the adoption of educational measures and not as an automated decision making tool.
Protection of personal data	[High]	[Medium]	**[Medium]**	▪ Restrict access to data: full access to tutors and only aggregated data to teachers.
Non-discrimination	[High]	[Medium]	**[Medium]**	▪ Periodically check that the data entered into the databases does not generate discriminatory profiles. ▪ Periodically revise the initial profiling criteria as new data is added to the database, so that new data can mitigate potential biases. ▪ Periodically check that the prediction model is not discriminatory and that the AI design is sensitive to discrimination and potential bias.

Tab. 3A Risk management (II)

Rights/freedoms affected	Likelihood (residual)	Severity (residual)	Residual overall impact
Human dignity	[Medium]	[Low]	**[Low]**
Respect for private and family life	[High]	[Low]	**[Medium]**
Protection of personal data	[Medium]	[Medium]	**[Medium]**
Non-discrimination	[Medium]	[Medium]	**[Medium]**

4. Comments

Being part of this working group, set up by the Catalan Data Protection Authority (APDCAT), has made it possible to work with other people from other sectors. This has been a great improvement in the analysis of the use case presented, as it has brought different perspectives and sensitivities to the evaluation of each of the fundamental rights.

Conducting the FRIA highlighted the importance of adopting a comprehensive approach to risk identification, covering all fundamental rights and including mitigation measures. One of the most challenging aspects has been the assessment of residual risks, as determining the resulting risk after the envisaged mitigation measures makes scenario analysis in the area of fundamental rights not easy.

The approach to the analysis from the perspective of a Data Protection Officer (DPO) highlights the priority of safeguarding a right (that of personal data protection) that is regulated in detail compared to other fundamental rights, as well as the need to address it together with other closely related rights in order to fulfil the obligations deriving from the entire legal framework.

The use case also identified the following needs:

- The need for fundamental rights training for all actors involved.
- The need to avoid checklists for fundamental rights compliance, as they cannot go into depth on the different aspects of fundamental rights.
- The need to raise awareness of the impact of AI systems in education.
- The need to understand the definition of ‘AI system’ in the AI Act (Article 3.1) and the implications of AI systems that are considered high risk in Annex III.
- The need to adopt an evaluation by default and by design.
- The need to raise awareness among those responsible for deploying high-risk AI systems of the importance of conducting and making available the FRIA.
- The need to review and adapt the assessment as the relevant context changes.
- The need to coordinate the FRIA with the DPIA in performing them.

Use case 2: A tool for managing human resources

1. The context

This use case is framed within the people selection process that the entity (a private company) has implemented and, therefore, it is guided by the principles that the entity has equipped itself with, and it is carried out by the its Human Resources department directly or through suppliers.

In particular, this entity has defined different levers in its people management master plan, in order to: i) Promote an exciting team culture, committed to the new project, collaborative and agile, while promoting close, motivating, non-hierarchical leadership, with transformative capabilities; ii) Promote new ways of working, with respect for diversity, equal opportunities, inclusion and non-discrimination, and incorporating sustainability in Human Resources processes; iii) Transform the management of the people development model: more proactive in the training of teams, with a focus on critical skills; iv) Develop a unique and differential value proposition for the employee; and v) Evolve towards a data-driven culture of the people function, through the optimisation of the data structure and the application of artificial intelligence and new technologies to facilitate the analysis of information and make data-based decisions in relation to people.

Within the framework of this last objective, the use case presented below was proposed, analysed and, finally, implemented. Before going into detail, it is worth noting that the entity is a mature organisation in terms of data protection and information security compliance schemes, and advanced in relation to artificial intelligence and its governance, to the extent that it had adopted and implemented the following measures, among others:

1. The creation and implementation of internal methodologies for the development and implementation of artificial intelligence systems, that include the 144 controls established by the Spanish Data protection Agency (AEPD) in its guide of “Audits of data processing activities that include artificial intelligence”, which allows AI systems developed internally to comply, by default and by design, with a wide range of controls such as their inventory, their relation to the data processing they serve, the evaluation of their need and proportionality, the assessment of the quality of the data (including, but not limited to, the analysis and mitigation of possible biases), their explainability, transparency and robustness under both the Artificial Intelligence Act and the General Data Protection Regulation, as well as measures in the field of validation and verification of the quality of the system.
2. The analysis of these use cases prior to their implementation within the framework of the Data Protection Impact Assessment, with the extension of its purpose, by the legal teams of Innovation and Privacy Law and Labour Law, the IT/systems’ team (CDO – Chief data officer - where the responsible AI team is included from a technical point of view) and the information security team (CISO – Chief Information Security Officer), which allows a second check on the quality of these systems.
3. The evaluation and sanctioning of these initiatives, where appropriate, by the relevant corporate committees.

2. The project

Development and application of an artificial intelligence system (4 machine learning models) based on the entity's previous experience in the area of personnel selection to fill certain vacancies. In particular, the system performs a very specific and limited task: the prediction of an additional information for each employee, consisting of the probability of his/her

suitability for a vacancy, based on data on the employment relationship as well as information on the characteristics of the centre of destination.

Thus, the result generated by the system is integrated as an additional piece of information in the personnel selection process, which is in any case managed and led by specialised human resources staff, who can use this information, together with the rest of the available information and in accordance with the company's internal processes, to perform their selection functions.

In this sense, and for the avoidance of doubt, in the event of a specific vacancy, the system will allow the aforementioned human resources staff to visualise the employees of the company ordered according to the probability of suitability for the vacancy for which they may or may not have applied. In any case, the decision to use this additional information will be made by the specialised human resources staff. The purpose of the system is therefore to support and improve the efficiency of the personnel selection process by providing the human resources department staff with systematised information that would otherwise have to be collected and structured manually.

Under no circumstances will the system make a decision.

3. The FRIA

Planning and scope

Section A Description and analysis of the AI system, including related data flows	**What are the main objectives of the AI system?**	Improve the selection process for certain vacancies by providing human resources department with additional information about the suitability of a candidate for a given vacancy based on objective criteria. In particular: I. Efficiency and time saving for recruiters in checking certain objective requirements relevant to filling a vacancy; II. Ensure the objectivity of the process; III. Promote the proactivity of the organisation in offering the vacancy to candidates who have not applied for it.
	What are the main features of the system?	It is configured as a support tool that performs a complementary and limited task consisting of generating an additional information for each employee in the context of the selection process for certain vacancies. It allows human resources department to visualised the relevant company's employees ranked according to their suitability for a given vacancy.
	In which countries will it be offered?	Spain.
	What types of data are processed (personal, non-personal, special categories)?	▪ Data relating to the work activity (level and group, assigned functions, productivity, quality and compliance data); ▪ Characteristics of the destination centre (size and type of centre).

	Identification of potential right sholders: who are the individuals or groups likely to be affected by the AI system, including vulnerable individuals or groups?	Staff of the organisation.
	Identification of duty bearers: who is involved in the design, development and deployment of the AI system? What is their role?	The human resources department, the legal department in general (including the labour law area), the DPO, the systems department (CDO), the information security department (CISO). The human resources department has developed and used the tool. The rest are evaluation teams that have accompanied the development and implementation of the system and, where appropriate, have established and implemented the necessary controls not limited to AI development.
Section B Fundamental rights context	**What fundamental rights are potentially affected by the use of the AI system?**	☒ Protection of personal data ☒ Non-discrimination ☒ Gender equality ☒ Right to information and consultation of workers in the company
	What international/regional legal instruments for the protection of human/fundamental rights have been implemented at the operational level?	The regulations on the protection of personal data, the regulations on labour relations (e.g. the Workers' Statute and the "Practical guide and tool on the corporations' obligation to provide information on the use of algorithms in the workplace" issued by the Ministry of Labour) and the AI Act.
	What are the most relevant fundamental rights courts or bodies in the context of use?	Data Protection Authorities Ministry of Labour

	What are the most relevant human/fundamental rights decisions and provisions?	Acquis communautaire both in the area of the fundamental right to data protection and in the area of the right to equality and non-discrimination.
Section C Controls in place	**What policies and procedures are in place to assess the potential impact on fundamental rights, including stakeholder participation?** **Has an impact assessment been conducted, developed and implemented in relation to specific issues (e.g. data protection) or certain features of the system (e.g. use of biometrics)?**	The AI system has been developed on the basis of an internal development methodology that incorporates, by default and by design, the controls established by the AEPD in its Guide "Requirements for audits of data processing activities that include AI" (such as inventory and registration of the system, control of data and bias, human intervention, explainability, validation and verification of the system, etc.). In addition, a Data Protection Impact Assessment (DPIA) has been carried out and it includes a statement on the impact on other fundamental rights, based on the organisation's own methodology. In particular, the DPIA covers: I. The analysis of data processing compliance with data protection regulations (including legal, security and AI system obligations) II. The analysis of the potential material or immaterial damage that could be caused and, where appropriate, the controls identified and the mitigating measures established. III. The impact on fundamental rights, the list of which includes both those enshrined in the Charter of Fundamental Rights of the EU and in the Spanish Constitution. The above-mentioned methodologies indicated and the DPIA itself were evaluated by the company's Corporate Committee at the proposal of an evaluation team composed of members of the Legal department, the DPO, the CISO and the CDO.

		In addition, the relevant information was shared with the workers' representatives.
Section D **Stakeholder engagement and due diligence**	**Who are the main groups or communities potentially affected by the AI system, including its development?**	The company's staff.
	Which stakeholders should be involved in addition to the individuals or groups potentially affected by the AI system (e.g. civil society and international organisations, experts, industry associations, journalists)?	Evaluation teams established by the company (DPO, CDO and CISO), as well as the Legal department, including the labour law area.
	Are there other duty bearers that should be involved in addition to AI providers and deployers (e.g. national authorities, government agencies)?	No
	Have business partners, including service providers (e.g. subcontractors for AI systems and datasets), been involved in the assessment process?	No
	Has the AI provider carried out an assessment of its supply chain to determine whether the activities of suppliers/contractors involved in product/service development may affect fundamental rights?	N/A

	Has the provider promoted fundamental rights standards or audits to ensure respect for fundamental rights among suppliers?	N/A
	Have the AI provider and AI deployer publicly communicated the potential impact of the AI system on fundamental rights?	The organisation has communicated the use of the AI system, as well as its purpose, logic and consequences, to its personnel and workers' representatives in accordance with the provisions of both data protection and labour law regulations.
	Have the AI provider and AI deployer provided training on fundamental rights standards to management and procurement staff dealing with the AI system?	N/A

Risk matrices

Tab. 1 Probability

Low	The risk of harm is unlikely or highly unlikely
Medium	Risk can occur
High	There is a high probability that the risk will occur
Very high	The risk is very likely to occur

Tab. 2 Exposure

Low	Few or very few of the identified population of rights holders are potentially affected
Medium	Part of the identified population is potentially affected
High	The majority of the identified population is potentially affected
Very high	Almost the entire population identified is potentially affected

Tab. 3 Likelihood

		Probability			
		Low	**Medium**	**High**	**Very high**
Exposure	**Low**	L	L/M	L/H	L/VH
	Medium	M/L	M	M/H	M/VH
	High	H/L	H/M	H	H/VH
	Very high	VH/L	VH/M	VH/H	VH

Likelihood			
Low	Medium	High	Very high

Tab. 4 Gravity of the prejudice

Low	Affected individuals and groups may encounter only minor damages or inconveniences in the exercise of their rights and freedoms.
Medium	Affected individuals and groups may encounter significant damages or inconveniences.
High	Affected individuals and groups may face serious damages or inconveniences.
Very high	Affected individuals and groups may encounter serious or even irreversible damages or inconveniences.

Tab. 5 Effort to overcome harm and reverse adverse effects

Low	The damages suffered can be overcome without problems (e.g. time spent modifying information, discomfort, irritation, etc.)
Medium	The damages suffered can be overcome despite some difficulties (e.g. additional costs, fear, misunderstanding, stress, small physical ailments, etc.)
High	The damages suffered can be overcome although with serious difficulties (e.g. economic loss, material damage, deterioration of health, etc.)
Very high	The damages suffered may not be overcome (e.g. long-term psychological or physical ailments, death, etc.).

Tab. 6 Severity

		Gravity			
		Low	**Medium**	**High**	**Very high**
Effort	**Low**	L	LM	L/A	L/VH
	Medium	M/L	M	M/H	M/VH
	High	H/L	H/M	H	H/VH
	Very high	VH/L	VH/M	VH/H	VH

Severity			
Low	Medium	High	Very high

Tab. 1A Data collection and risk analysis

Rights/freedoms potentially affected	**Description of the impact**	Likelihood			Severity		
		Probability of adverse outcomes	Exposure	**Likelihood**	Gravity	Effort	**Severity**
Data protection	The algorithm requires the use of data relating to the employment relationship of the company's employees. Therefore, any use that infringes data protection regulations could affect this right.	[Low] Both the creation of the system and its use have been subject to a data protection impact assessment.	[Very high] The impact potentially affects everyone to whom the algorithm is applied.	**[Medium]**	[Low] Although data processing is carried out in the context of personnel selection, the system performs a very limited task, using objective data limited to the work context.	[Medium] In case of non-compliance, measures can be taken to comply with the transparency and information obligations, to interrupt the processing where appropriate, and even to delete the data generated by the system.	**[Low]**
Non-discrimination and gender equality	The algorithm is trained on historical data, so if biases exist and are perpetuated, there may be situations of discrimination that	[Low] The system and its use have been verified to be free of bias, including controls to correct for historical biases	[Medium] The potential impact would affect a part of the company's staff.	**[Low]**	[Medium] The system provides additional data that is available to the human resources department, who	[High] Although the potential damage of an incorrect suitability result could be corrected in the vacancy	**[Medium]**

	include and affect gender equality.	contained in the training data			are the subject matter experts and, in any case, take the lead in the selection of personnel. The system does not make decisions.	management process itself or subsequently by human resources staff, it is rated as high due to the possibility of it being detected after the vacancy has been filled.	
Workers' right to information and consultation	The system is included in the framework of the management of the Entity's personnel selection, so a breach of labour regulations, in particular, in relation to the obligations to inform workers or their representatives, could affect this right.	[Low] The Legal Department in general, including the labor law area, has been involved in both the creation and use of the system. The procedures in place ensure that information obligations towards employees and their representatives are met.	[Very high] The impact potentially affects everyone to whom the algorithm is applied.	[**Medium]**	[Low] The potential damage consists in the lack of mandatory information to workers' representatives.	[Low] Mitigating this potential damage can be easily achieved by addressing the lack of information.	**[Low]**

Tab. 7 Overall Risk Impact

		Severity			
		Low	**Medium**	**High**	**Very high**
Likelihood	**Low**				
	Medium				
	High				
	Very high				

Overall risk impact			
Low	Medium	High	Very high

Tab. 2A Risk management (I)

Rights/freedoms affected	Likelihood	Severity	Overall impact	Impact prevention/mitigation measures
Data protection	[Medium]	[Low]	**[Low]**	N/A. Measures already taken in the development of the AI algorithm and before its use (e.g. carrying out the DPIA, providing information, etc.).
Non-discrimination and gender equality	[Low]	[Medium]	**[Low]**	Train human resource staff to avoid over-reliance on the output of the system.
Workers' right to information and consultation	[Medium]	[Low]	**[Low]**	N/A. Measures already taken, both in the development of the algorithm and before its use. Mandatory information has been provided to both the company's staff and the workers' representatives in accordance with the models established by the Ministry of Labour.

4. Comments

As explained above, the company is equipped with structures, procedures and controls, by default and by design, that focus on several of the issues covered by the FRIA. This has enabled the FRIA of the use case to result in low levels of impact risks. These structures, procedures and controls allow the system to be developed in a controlled framework, which leads the data scientist to include, by default, certain measures in the development and implementation of the AI system that mitigate the risks identified by the company from the outset.

In addition, the involvement of the various evaluation teams and their support in the development of the system will allow risks to be identified and mitigated at the time of development/implementation, where no risks or measures have been identified initially.

Use case 3: An AI-powered medical imaging tool for cancer detection

1. The context

In Europe, significant progress is being made in the development of AI tools using cancer images. For example, in 2012 a research team at Universiteit Maastricht proposed the concept of ‘radiomics’, which refers to the method of extracting a large number of features from medical images using data characterisation algorithms. The increasing development of AI systems aimed at using medical images to treat cancer can be illustrated by looking at the number of publications on ‘AI radiomics’ on the PubMed portal (42 results in 2019, 99 results in 2020, 165 results in 2021, 235 results in 2022, 309 results in 2023 and 338 results in 2024). Therefore, there is a type of artificial intelligence system that will become increasingly common not only in the academic world, but also in the world of healthcare.

In addition, there are common types of cancer in the world where patients receive a high degree of overtreatment and a high degree of considerable avoidable effects.

Consequently, the use of AI systems to analyse patients' medical images would provide healthcare professionals with a support tool for predicting the response to therapy and, consequently, adjusting the therapy to be as efficient as possible, i.e. to achieve the target goal with the minimum treatment of the patient. Moreover, these AI systems would provide both healthcare professionals and patients with a prediction of the patient's evolution over the coming years.

2. The project

The project is divided into two phases. The first phase consists of the development of an AI system based on medical images, which is trained with data from 5,000 patients from ten countries in Europe. The training dataset is therefore a multi-centre dataset.

In addition, a second phase of the project will involve the validation of the AI system in eight healthcare centres around the world outside Europe. The aim of this second phase is to test the AI system in one health centre in Asia, one in Africa and one in South America.

3. The FRIA

Planning and scope

Section A Description and analysis of the AI system, including related data flows	**What are the main objectives of the AI system?**	Improving the treatment of patients with cancer X [anonymised] by predicting: a) Patient response to treatment b) Side effects (toxicity and sensitivity) c) Projections for the next five years
	What are the main features of the system?	AI-based image recognition: using medical images to predict a patient's response to a given treatment and help healthcare professionals determine its application in specific cases, as well as the level of use once it is applied.
	In which countries will it be offered?	Global distribution
	What types of data are processed (personal, non-personal, special categories)?	・ Demographic data (gender, age and country) ・ Cancer characteristics (type of cancer and site affected) ・ Cancer stage and molecular subtype ・ Information regarding previous treatment ・ Treatment regimen (schedule and duration) ・ Pathology report (post-treatment)

	Identification of potential rights holders: who are the individuals or groups likely to be affected by the AI system, including vulnerable individuals or groups?	People between 18 and 85 years of age. As all the people involved are affected by cancer, they should be considered vulnerable because of their health conditions and the relationship between these conditions and the purpose of the AI system.
	Identification of duty bearers: who is involved in the design, development and deployment of the AI system? What is their role?	Hospitals and research centres are involved in the design, the latter only in the design of the AI system and the former also in the related health treatment.
Section B Fundamental rights context	**What fundamental rights are potentially affected by the use of the AI system?**	☒ Right to data protection ☒ Freedom from discrimination ☒ Right to an adequate standard of living (including the right to physical and mental health)
	What international/regional legal instruments for the protection of human/fundamental rights have been implemented at the operational level?	Universal Declaration of Human Rights, EU Charter of Fundamental Rights, applicable data protection regulations.
	What are the most relevant fundamental rights courts or bodies in the context of use?	Data protection supervisory authorities in the country/region where AI systems are developed and used, as well as courts, the Court of Justice of the European Union and the European Court of Human Rights.
	What are the most relevant human/fundamental rights decisions and provisions?	N/Arevisa

Section C Controls in place	**What policies and procedures are in place to assess the potential impact on fundamental rights, including stakeholder participation?**	A specific ethics committee will be established for the project.
	Has an impact assessment been conducted, developed and implemented in relation to specific issues (e.g. data protection) or certain features of the system (e.g. use of biometrics)?	A data protection impact assessment must be carried out.
	Who are the main groups or communities potentially affected by the AI system, including its development?	Patients with cancer X [anonymised].
	Which stakeholders should be involved in addition to the individuals or groups potentially affected by the AI system (e.g. civil society and international organisations, experts, industry associations, journalists)?	Cancer patient associations.
	Are there other duty bearers that should be involved in addition to AI providers and deployers (e.g. national authorities, government agencies)?	Data protection supervisory authority, local health department, scientific research ethics committee, AI supervisory authority.
	Have business partners, including service providers (e.g. subcontractors for AI systems and datasets), been involved in the assessment process?	No

	Has the AI provider carried out an assessment of its supply chain to determine whether the activities of suppliers/contractors involved in product/service development may affect fundamental rights? **Has the provider promoted fundamental rights standards or audits to ensure respect for fundamental rights among suppliers?**	N/A
	Have the AI provider and AI deployer publicly communicated the potential impact of the AI system on fundamental rights?	No
	Have the AI provider and AI deployer provided training on fundamental rights standards to management and procurement staff dealing with the AI system?	N/A
Section D Stakeholder engagement and due diligence	**Who are the main groups or communities potentially affected by the AI system, including its development?**	Patients with cancer X [anonymised].
	Which stakeholders should be involved in addition to the individuals or groups potentially affected by the AI system (e.g. civil society and international organisations, experts, industry associations, journalists)?	Cancer patient associations.

	Are there other duty bearers that should be involved in addition to AI providers and deployers (e.g. national authorities, government agencies)?	Data protection supervisory authority, local health department, scientific research ethics committee, AI supervisory authority.
	Have business partners, including service providers (e.g. subcontractors for AI systems and datasets), been involved in the assessment process?	No
	Has the provider promoted fundamental rights standards or audits to ensure respect for fundamental rights among suppliers?	N/A

Risk matrices

Tab. 1 Probability

Low	The risk of prejudice is improbable or highly improbable
Medium	The risk may occur
High	There is a high probability that the risk occurs
Very high	The risk is highly likely to occur

Tab. 2 Exposure

Low	Few or very few of the identified population of rights holders are potentially affected
Medium	Some of the identified population are potentially affected
High	The majority of the identified population is potentially affected
Very high	Almost the entire identified population is potentially affected

Tab. 3 Likelihood

		Probability			
		Low	**Medium**	**High**	**Very high**
Exposure	**Low**	L	L/M	L/H	L/VH
	Medium	M/L	M	M/H	M/VH
	High	H/L	H/M	H	H/VH
	Very high	VH/L	VH/M	VH/H	VH

Likelihood			
Low	Medium	High	Very high

Tab. 4 Gravity of the prejudice

Low	Affected individuals and groups may encounter only minor prejudices in the exercise of their rights and freedoms.
Medium	Affected individuals and groups may encounter significant prejudices.
High	Affected individuals and groups may encounter serious prejudices.
Very high	Affected individuals and groups may encounter serious or even irreversible prejudices.

Tab. 5 Effort to overcome the prejudice and to reverse adverse effects

Low	Suffered prejudice can be overcome without any problem (e.g. time spent amending information, annoyances, irritations, etc.)
Medium	Suffered prejudice can be overcome despite a few difficulties (e.g. extra costs, fear, lack of understanding, stress, minor physical ailments, etc.).
High	Suffered prejudice can be overcome albeit with serious difficulties (e.g. economic loss, property damage, worsening of health, etc.).
Very high	Suffered prejudice may not be overcome (e.g. long-term psychological or physical ailments, death, etc.).

Tab. 6 Severity

		Gravity			
		Low	**Medium**	**High**	**Very high**
Effort	**Low**	L	L/M	L/H	L/VH
	Medium	M/L	M	M/H	M/VH
	High	H/L	H/M	H	H/VH
	Very high	VH/L	VH/M	VH/H	VH

Severity			
Low	Medium	High	Very high

Tab. 1A Data collection and risk analysis

Rights/freedoms potentially affected	**Description of the impact**	Likelihood			Severity		
		Probability of adverse outcomes	Exposure	**Likelihood**	Gravity	Effort	**Severity**
Data protection	The development of the AI system is based on the use of special categories of personal data and other personal information of patients. Any processing operation that does not comply	[Low] The project is subject to specific ethical and data protection impact assessments.	[Very high] The impact potentially affects all individuals to whom the algorithm is applied.	**[Medium]**	[Medium] Illegal processing of cancer-related health data and the unlawful use of this information can be invasive and affect the	[Medium] Illegal data collection and processing can be detected and stopped, with unlawfully collected	**[Medium]**

	with the applicable personal data protection regulations may affect this right.				privacy of individuals.	information deleted.	
Non-discrimination	The algorithm was trained on data from European health centres, so it is possible that discrimination may occur when it is used in the three non-EU health centres.	[High] Ethnicity may cause some differences in medical imaging, that may affect diagnostic accuracy.	[Very high] All persons in the relevant group (ethnic group) to whom the algorithm applies.	**[Very high]**	[Very high] Negative impact on equal access to healthcare and on the quality of cancer treatment received.	[High] It would be necessary to adapt or even retrain the algorithm with data that avoids discrimination.	**[Very high]**
Right to physical and mental health	Incorrect functioning of the algorithm may result in ineffective and harmful medical treatment for the patient, resulting in a prejudice to the right to health.	[Medium] when used in European patients. [High] when used in non-European patients.	[Very high] The impact potentially affects all individuals to whom the algorithm is applied.	**[High] when used in European patients.** **[Very high] when used in non-European patients.**	[Very high] Incorrect functioning of the algorithm may result in ineffective and harmful health treatment for the patient.	[Medium] Pathologies where subsequent follow-up can correct the system error. [High] Pathologies where subsequent follow-up cannot correct the system error.	**[High] when subsequent cancer screening can correct the system error** **[Very high] when subsequent cancer control cannot correct the system error.**

Tab. 7 Overall risk impact

		Severity			
		Low	**Medium**	**High**	**Very high**
Likelihood	**Low**				
	Medium				
	High				
	Very high				

Overall risk impact			
Low	Medium	High	Very high

Tab. 2A Risk management (I)

Rights/freedoms affected	Likelihood	Severity	Overall impact	Impact prevention/mitigation measures
Data protection/privacy	[Medium]	[Medium]	**[Medium]**	·Publish information on the procedures used to obtain and process the original data used to train the AI system.
Non-discrimination	[Very high]	[Very high]	**[Very high]**	·Expand the training dataset, avoiding under-representation of relevant groups.
Right to physical and mental health	[High] when used in European patients. [Very high] when used in non-European patients.	[High] when subsequent cancer screening can correct the system error [Very high] when subsequent cancer screening cannot correct the system error	**[High]** only when it is used in European patients and when the pathology is such that subsequent follow-up can correct the system error. **[Very high]** for the other three scenarios.	·Inform health professionals of the limitations of the tool. For example, indicating the type of errors. ·Differentiate between pathologies with and without rapid progression. ·Inform health professionals that the error rate of the imaging equipment used should be taken into account.

Tab. 3A Risk management (II)

Rights/freedoms affected	Likelihood (residual)	Severity (residual)	Residual overall impact
Data protection/privacy	[Medium]	[Low] as the effort has been reduced to Low.	**[Medium]**
Non-discrimination	[Medium] as the probability has been reduced to Low.	[Medium] as the effort has been reduced to Low.	**[Medium]**
Right to physical and mental health	[Medium] when used in European patients, as the probability has been reduced to Low. [High] when used in non-European patients, as the probability has been reduced to Medium.	[Medium] in the case of a pathology where a subsequent follow-up can correct the system error, as the effort has been reduced to Low. [High] in the case of a pathology where subsequent monitoring cannot correct the system error, as the effort has been reduced to Medium.	**[Medium]** only when it is used in European patients and when the pathology is such that subsequent follow-up can correct the system error. **[High]** for the remaining three scenarios.

4. Comments

The main difficulties encountered in the use of the FRIA methodology in this use case are described below.

This methodology requires, at a minimum, identifying the fundamental rights and freedoms that will be impacted by the artificial intelligence system. This therefore requires expert knowledge of the essential content of each of the fundamental rights and freedoms under scrutiny. It can be assumed that Data Protection Officers (DPO) have such knowledge in relation to the fundamental right to the protection of personal data and the fundamental right to personal and family privacy. However, this expert knowledge is not necessarily required of a DPO in relation to the other fundamental rights and freedoms. Consequently, the first difficulty is that a thorough knowledge of each of the fundamental rights and freedoms is required in order to identify which ones will be affected. Once the fundamental rights and freedoms affected have been identified, the DPO can seek advice from experts in these rights and freedoms.

The above difficulty becomes more complex when it is envisaged to use the artificial intelligence system outside the European Union. While it is possible to define a common framework with regard to the content of fundamental rights and freedoms within the European Union, taking into account the EU Charter of Fundamental Rights and the rulings of the CJEU, without going into the details of the differences established by national courts, it is hardly possible to speak of a common framework when examining the content of fundamental rights and freedoms worldwide. Applying this methodology in the EU and outside the EU with the same level of detail would require a comparative law analysis that the vast majority of organisations could hardly undertake given the human, time and financial resources it would require. Hence, it would be useful to have a guidance on the minimum content of each fundamental right and freedom at the global level, or by region or legal tradition, which would allow the use of the present methodology without requiring resources disproportionate to its purpose, i.e. to have an *ex ante* analysis that facilitates the design of the artificial intelligence system.

The final difficulty faced was identifying the people who should be involved in carrying out the impact assessment, both in terms of their expertise and their role in the development and implementation of this type of artificial intelligence (e.g. identifying which people with expertise in developing AI systems for healthcare purposes should be involved without risk of breaching of confidentiality).

Use case 4: ATENEA: AI at the service of the elderly

1. The context

Public administrations in general are immersed in a process of digital transformation, with the idea of reforming public services by taking advantage of the benefits provided by the exponential evolution of technologies. However, these digital transformation policies must be formulated and implemented with a positive impact in terms of social inclusion, combining the promotion of digitalisation with social policies to minimise, as far as possible, the digital divide that inevitably arises in such disruptive processes of change. It is therefore crucial to put technology at the service of people, with the aim of improving relations with citizens and social care, tackling the inequalities caused by increasing digitalisation, guaranteeing equal opportunities and, in general, improving the living conditions of citizens.

It is in this context and under these conditions and social commitments that the ATENEA project has been promoted. The project (now in its pilot phase) aims to contribute to the digital transformation of the territories, putting the most vulnerable citizens at the centre, with the specific objectives of reducing the existing digital divide and unwanted loneliness, increasing the safety of people, especially at home, promoting inclusion, well-being, health and, ultimately, the quality of life of people. The project specifically targets people over 65 who live alone and who suffer most from the vulnerabilities caused by the digital divide.

2. The project

ATENEA is a project based on the development of a generative AI system (neural networks), voice assistant, voice biometrics and robotic process automation, combined with mature technologies such as data analytics, cloud computing and smartphones. Using biometric voice recognition, it can respond to the needs of elderly users in different use cases: call and video call to a family member, emergency calls (112), call to the call to the municipality and/or automatic booking of an appointment with the social services, booking of an appointment with the Primary Care Centre, diary reminders, transport route information and, in the future, on-line shopping, banking and supply management.

The ATENEA solution does not require any digital skills or physical interaction from the user, the identification is biometric in order to guarantee exclusive individual use and security. ATENEA is an artificial intelligence in the form of a tablet, without buttons, without touch screens, it works only with an interaction as simple as voice. It provides an elderly person, probably in a situation of dependency, with agile responses to their basic needs. This artificial intelligence makes it possible to carry out everyday tasks (e.g. checking your bank statement, making a medical appointment, having a video conference with a family member, calling emergencies or social services) through a conversation.

ATENEA has been designed with the co-creation of elderly people, professional carers and family carers from the user's environment, building a bond of trust and support in case of need.

The initiative is led by a strategic alliance between technological and social entities that are responsible for the design of the solution, contact with users and the deployment of socio-digital integrators in the field, cloud services, provision of devices (tablets), speech recognition technology, robotic automation of processes, communication services, security, evaluation of results and impact, and guaranteeing users' rights.

The project is a public-private partnership. The current phase of the project is funded by the Department of Social Rights of the Government of Catalonia as part of the EU-funded Next Generation EU Recovery, Transformation and Resilience Plan. This experience is supported by various municipalities and public administrations, which are using their territory as a pilot for testing the solution and are responsible for identifying potential users.

3. The FRIA

Planning and scope

Section A Description and analysis of the AI system, including related data flows	**What are the main objectives of the AI system?**	Automated response to care requests made by the user's voice: - Call and video call to a family member - Emergencies calls (112) - Call to the municipality and/or automatic booking of an appointment with the social services - Appointment booking with the Primary Care Centre of reference - Agenda reminders - Transport route information - Medication reminders - City events calendar · In development: online grocery shopping, TV and radio on demand
	What are the main features of the system?	An artificial intelligence-based voice assistant interacts with users through voice commands, providing information, performing tasks and offering services in real time. Its design combines speech recognition, natural language processing (NLP), and speech synthesis technologies to understand, process and respond to user requests in a natural way. User voice requests are processed by a robotic process automation system. This technology uses software robots or bots to automate repetitive rule-based tasks normally performed by humans. These bots mimic and execute human actions in digital environments, such as clicking, moving files, filling out forms, copying and pasting information,

		or processing massive amounts of data. In short, ATENEA makes it possible to carry out everyday tasks with a conversation with the AI.
	In which countries will it be offered?	Spain
	What types of data are processed (personal, non-personal, special categories)?	・Demographic data (age, city, family situation, country). ・Personal information, not including data that users process via the services accessed through ATENEA ・Device configuration ・Language of dialogue (Catalan or Castilian) ・Access information for family doctor appointments ・Municipal social services/OAC appointment telephone number ・Emergency and tele-assistance telephone numbers ・Telephone number and family relationship ・Telephone number of assigned socio-technological operators ・Information about reminders and personalised alerts
	Identification of potential rights holders: who are the individuals or groups likely to be affected by the AI system, including vulnerable individuals or groups?	Over 65 users with a sufficient cognitive level to interact with the device. As all the people involved are affected by the digital divide and loneliness at home, they should be considered vulnerable, due to their socio-demographic conditions and the relationships between these conditions and the purpose of the AI system.
	Identification of duty bearers: who is involved in the design, development and deployment of the AI system? What is their role?	The initiative is led by a strategic alliance between technological and social entities responsible for the design of the solution, contact with users and the deployment of socio-digital integrators in the field, cloud services, provision of devices (tablets), speech recognition technology, robotic automation of processes, communication services, security,

		evaluation of results and impact, and guaranteeing users' rights. It is a public-private partnership. The current phase of the project is funded by the Department of Social Rights of the Government of Catalonia as part of the EU-funded Next Generation EU Recovery, Transformation and Resilience Plan. This experience has the support of various municipalities and public administrations, which have made their territory a pilot for testing the solution and are responsible for identifying potential users.
Section B Fundamental rights context	**What fundamental rights are potentially affected by the use of the AI system?**	☒ The right to data protection ☒ Freedom from discrimination ☒ Right to health care ☒ Right to social assistance ☒ Right of access to services of general economic interest ☒ Right of access to services of general interest
	What international/regional legal instruments for the protection of human/fundamental rights have been implemented at the operational level?	The project started before the publication of the AI Act, which will also apply to it. At that time, the only applicable regulations were the GDPR and the Spanish Organic Law 3/2018 of 5 December on the protection of personal data and guarantee of digital rights (LOPD-GDD) and complementary regulations.
	What are the most relevant fundamental rights courts or bodies in the context of use?	In Spain, the protection of fundamental rights is ensured by a national legislative framework and by judicial and non-judicial institutions. Spain is part of international, EU, and regional frameworks on fundamental and human rights. The following are the main courts and bodies in the field of fundamental rights in Spain:

		▪ Constitutional Court ▪ Supreme Court ▪ National High Court ▪ Ordinary courts and tribunals ▪ Ombudsman and similar institutions in the regions ▪ Autonomous regional agencies for the defence of rights ▪ General State Prosecutor's Office ▪ Data protection supervisory authorities in the Country/Region
	What are the most relevant human/fundamental rights decisions and provisions?	Constitutional Court: ▪ Judgment 135/2024: This decision dealt with the violation of the right to effective judicial protection in a case of the application of legal provisions that had been declared unconstitutional. ▪ Ruling 113/2021: In this decision, the Court recognised the violation of the right to effective judicial protection in relation to the protection of the family and minors, emphasising the need for a stronger motivation in cases affecting substantial fundamental rights. ▪ Ruling 58/2018: This ruling dealt with the protection of personal data and freedom of expression, and established criteria for the balancing of the two rights. Legislative provisions: ▪ Organic Law 3/2007, of 22 March, which establishes measures to eliminate gender discrimination. ▪ Law 15/2022, of 12 July, which strengthens the legal framework against all forms of discrimination.

		· Organic Law 1/2004, of 28 December, which deals with the prevention, protection and punishment of violence against women.
Section C Controls in place	**What policies and procedures are in place to assess the potential impact on fundamental rights, including stakeholder participation?**	ATENEA is a project based on the voluntary collaboration of the users to improve their wellbeing and quality of life, and users have been involved in the development of the project's features since its inception. In order to participate in the project, users are asked to give an initial informed consent, exercising their self-determination on the basis of prior verbal and written information. This initial informed consent is then validated by the users, who are again informed of the risks and benefits associated with using the system.
	Has an impact assessment been conducted, developed and implemented in relation to specific issues (e.g. data protection) or certain features of the system (e.g. use of biometrics)?	The project has carried out a personal data protection impact assessment (DPIA) and is compliant with current regulations on security and personal data protection. This is part of the cooperation agreements with the pilot areas.
Section D Stakeholder engagement and due diligence	**Who are the main groups or communities potentially affected by the AI system, including its development?**	Persons aged 65 and over living alone.
	Which stakeholders should be involved in addition to the individuals or groups potentially affected by the AI system (e.g. civil society and international organisations, experts, industry associations, journalists)?	Public administrations, private companies and third sector entities.

	Are there other duty bearers that should be involved in addition to the AI provider and deployers (e.g. national authorities, government agencies)?	Data protection authority, local and regional public administrations, scientific research ethics committee of a public university, AI supervisory authority.
	Have business partners, including service providers (e.g. subcontractors for AI systems and datasets), been involved in the assessment process?	The evaluation process has involved the project partners, which are a group of companies not constituted as an autonomous legal entity. From the outset, all the parties involved in the project (a public-private partnership) have been very aware of the need to ensure compliance with existing regulations and to protect the fundamental rights of citizens potentially affected by the implementation of this project. A specific accredited ethics committee has been identified for the project (the bioethics and law committee of a public university in Catalonia), which is responsible for the ethical evaluation of the project. On the other hand, the agreements signed with the participating public administrations provide for the obligation to carry out impact assessments, continuous monitoring, training, transparency and accountability of this project.
	Has the AI provider carried out an assessment of its supply chain to determine whether the activities of suppliers/contractors involved in product/service development may affect fundamental rights? **Has the provider promoted fundamental rights standards or audits to ensure respect for fundamental rights among suppliers?**	No (it is not a legal requirement)
	Have the AI provider and deployer publicly communicated the potential impact of the AI system on fundamental rights?	No (it is not a legal requirement)

	Have the AI provider and AI deployers provided training on fundamental rights standards to management and procurement staff dealing with the AI system?	Only in relation to data protection

Risk matrices

Tab. 1 Probability

Low	The risk of harm is improbable or highly improbable.
Medium	The risk may occur
High	There is a high likelihood that the risk will occur
Very High	The risk is very likely to occur

Tab. 2 Exposure

Low	Few or very few of the identified population of rights holders are potentially affected
Medium	Some of the identified population is potentially affected
High	Most of the identified population is potentially affected
Very High	Almost all of the identified population is potentially affected

Tab. 3 Likelihood

		Probability			
		Low	**Med**	**High**	**Very high**
Exposure	**Low**	L	L/M	L/H	L/VH
	Medium	M/L	M	M/H	M/VH
	High	H/L	H/M	H	H/VH
	Very High	VH/L	VH/M	VH/H	VH

Likelihood			
Low	Medium	High	Very high

Tab. 4 Gravity of the prejudice

Low	Affected individuals and groups may encounter only minor prejudices in the exercise of their rights and freedoms.
Medium	Affected individuals and groups may encounter significant prejudice.
High	Affected individuals and groups may face serious prejudice.
Very High	Affected individuals and groups may encounter serious or even irreversible prejudice.

Tab. 5 Effort to overcome the prejudice and to reverse adverse effects

Low	The harm suffered can be overcome without problems (e.g. time spent on changing information, inconvenience, irritation, etc.).
Medium	The harm suffered can be overcome despite some difficulties (e.g. additional costs, fear, misunderstanding, stress, minor physical ailments, etc.).
High	The harm suffered can be overcome although with serious difficulties (e.g. financial loss, material damage, deterioration of health, etc.).
Very High	The harm suffered may not be overcome (e.g. long-term psychological or physical ailments, death, etc.).

Tab. 6 Severity

		Gravity			
		Low	**Med**	**High**	**Very high**
Effort	**Low**	L	L/M	L/H	L/VH
	Medium	M/L	M	M/H	M/VH
	High	H/L	H/M	H	H/VH
	Very High	VH/L	VH/M	VH/H	VH

Severity			
Low	Medium	High	Very high

Tab. 1A Data collection and risk analysis

Rights/freedoms potentially affected	**Description of the impact**	Likelihood			Severity		
		Probability of adverse outcomes	Exposure	**Likelihood**	Severity	Effort	**Severity**
Right to data protection	The development of the AI system is based on the use of speech recognition. In this respect, AI is used to elaborate the user's requests (voice-to-text) to other non-AI based robotic technologies that will perform the requested task. ATENEA listens to all conversations waiting for the wake up word 'Hello ATENEA'. This listening process is only kept on the device (tablet), it is not processed and stored in the cloud. Only the command 'Hello ATENEA' activates the service and its communication with the cloud, where the	[Medium] The project is subject to specific ethical and data protection impact assessments. Users must have an adequate cognitive level to give their informed consent to the use of this system. The project has been evaluated by a recognised ethics committee of a public university in Catalonia and the system complies with the security measures laid down in the technical regulations in force	[Very high] The impact potentially affects all individuals to whom the algorithm is applied.	**[High]**	[High] Risks associated with the use of biometrics are limited due to the optional nature of biometric ID; biometric ID can be a proportionate solution for elderly people with limited mobility. There is a risk that voice interactions could be used for profiling. The main risk is the unlawful processing of data, including	[Medium] Illegal data collection and processing can be detected and stopped, with unlawfully collected information deleted.	**[Medium]**

	conversation is processed and stored to train the algorithm used. All the user's conversations related to a specific service accessed via ATENEA, e.g. emergency calls or other calls, are kept private. Voice biometrics is only used to identify the person and is an option that is activated based on the user's own decision. Identification can be based on voice biometrics or on the traditional combination of username and password. Any processing operation that does not comply with the applicable regulations on personal data protection may affect this right.	in Spain (ENS, UNE standards...) following the report prepared by the Cybersecurity Agency of Catalonia.			special categories of data related to the vulnerable situation of users, which may be invasive and affect the privacy of individuals.		

Non-discrimination	The AI mechanism uses speech recognition technology. Discrimination may occur if the voice assistant does not understand the speaker due to a communication issues (e.g. speech impediment or impairment). Discrimination may also occur when users do not speak the language of the system correctly (CAT/ES).	[High] Possible voice and speech problems of users can lead to their exclusion.	[Medium] There will be a limited number of people with speech or language problems.	**[Medium]**	[Very high] Negative impact on equal access to the support provided by the system and on quality of service.	[High] It would be necessary to adapt the voice assistant to integrate cases of speech or language problems. Users must have a certain cognitive level and it is also possible to exclude people with speech or language problems from the project.	**[Very High]**
Right to health	Inadequate functioning of the system may result in a failure to adequately guarantee this right.	[Low] The project is subject to continuous evaluation and specific monitoring to detect possible malfunctions. Only in some cases can malfunctioning significantly affect	[Very high] The impact potentially affects all individuals to whom the algorithm is applied.	**[Low]**	[High] Not being able to access health services when needed can cause serious harm to users, who are potentially more exposed to risky situations	[Medium] This system is not the only solution that people can use in case of need, as there are other channels of access to health systems (as well as to other services),	**[Medium]**

		the right to health (e.g. emergency call).				taking into account that potential users must have a certain cognitive level.	
Right to social assistance	Inadequate functioning of the system may result in a failure to adequately guarantee this right.	[Low] The project is subject to continuous evaluations and specific monitoring to detect possible malfunctions. Only in certain cases of use can improper operation affect this right.	[Very high] The impact potentially affects all individuals to whom the algorithm is applied.	**[Low]**	[High] Not being able to access social assistance services when needed can cause serious harm to users, who are potentially more exposed to risk situations.	[Low] This system is not the only solution that people can use in case of need, as there are other channels of access to social assistance services, taking into account that potential users must have a certain cognitive level. In addition, these are not emergency services and users have a contact person from social	**[Medium]**

						services who monitors their situation.	
Right of access to services of general economic interest	This right could be undermined if not all people can have access to this system due to a lack of resources of public administrations.	[Medium] The project is in a pilot phase with public funding from the Next Generation funds.	[Medium] The impact is potentially on all people who are not yet beneficiaries of this project due to lack of resources or information.	**[Medium]**	[Medium] Lack of access to this system, whether due to lack of information or lack of resources, can undermine equal opportunities and social cohesion in the area.	[Medium] It would be a matter of providing sufficient resources so that all those who are likely to have access to this system can do so.	**[Medium]**

Tab. 7 Overall risk impact table

		Severity			
		Low	**Medium**	**High**	**Very High**
Likelihood	**Low**				
	Medium				
	High				
	Very High				

Overall risk impact			
Low	Medium	High	Very high

Tab. 2A Risk management (I)

Rights/freedoms affected	**Likelihood**	**Severity**	**Overall impact**	**Impact prevention/mitigation measures**
Right to data protection	[High]	[Medium]	**[High]**	▪ Publication of information on the procedures used to obtain and process the original data used to train the AI system (algorithmic transparency and public record of these algorithms).

				▪ Voice biometric identification as an option based on prior consent. This identification will not be used for any other purpose, including identification shared with third parties. ▪ Exclusion of any profiling activity based on the use of the services accessible through ATENEA. ▪ The implementation and maintenance of a risk and security management system throughout the entire life cycle of the system. ▪ Ensure the quality of training, validation and test data through appropriate data governance and management practices. ▪ Accountability: develop and maintain up-to-date technical documentation for the system. ▪ Enable effective human supervision during use by trained personnel. ▪ Ensure the exercise of users' rights.
Non-discrimination	[Medium]	[Very high]	**[High]**	▪ Train the voice assistant to avoid some of the discrimination situations described. ▪ Restrict potential users to avoid these situations. ▪ The implementation and maintenance of a risk and security management system throughout the entire life cycle of the system.

				▪ Ensure the quality of training, validation and test data through appropriate data governance and management practices. ▪ Accountability: develop and maintain up-to-date technical documentation for the system. ▪ Enable effective human supervision during use, with trained personnel and periodic testing of system effectiveness in critical cases. ▪ Ensure the exercise of users' rights.
Health care	[Low]	[Medium]	**[Medium]**	▪ Ensure adequate technical assistance and improve the technical quality of the system. ▪ Provide information on all the channels available to users to access health care.
Social assistance	[Low]	[Medium]	**[Medium]**	▪ Ensure adequate technical assistance and improve the technical quality of the system ▪ Provide information on all the channels available to users to access social assistance.
Access to services of general economic interest	[Medium]	[Medium]	**[Medium]**	▪ Ensure sufficient resources and means so that all potential users have access to this service.

Tab. 3A Risk management (II)

Rights/freedoms affected	Likelihood (residual)	Severity (residual)	Residual impact
Data Protection/Privacy	[Low] The probability of an impact on this right (probability of adverse outcomes) has been reduced to Low because the measures adopted have limited the possibility of negative consequences.	[Medium] The severity of the impact has been reduced to Medium, as the measures adopted have limited the scope of the negative consequences that could occur. The severity level remains Medium, but is lower than the initial level.	**[Medium]**
No discrimination	[Low] The probability has been reduced to Low, because the probability of adverse effects has been reduced to Medium, due to the decrease in the number of possible malfunctions, and the exposure has been reduced to Low, due to the decrease in the number of people affected (some limitations of use related to language skills have been identified).	[Medium] The severity level has been reduced from Very High to Medium due to the implementation of complementary/alternative measures to access services (reducing the gravity of the impact from Very High to High) and the reduced effort required to react due to the measures already in place (reducing the effort from High to Medium).	**[Medium]**
Health care	[Low]	[Medium] The severity level has decreased, due to complementary/alternative measures to access services (reduction in gravity from Very High to High) and the reduced effort required to react due to the measures already in place	**[Medium]**

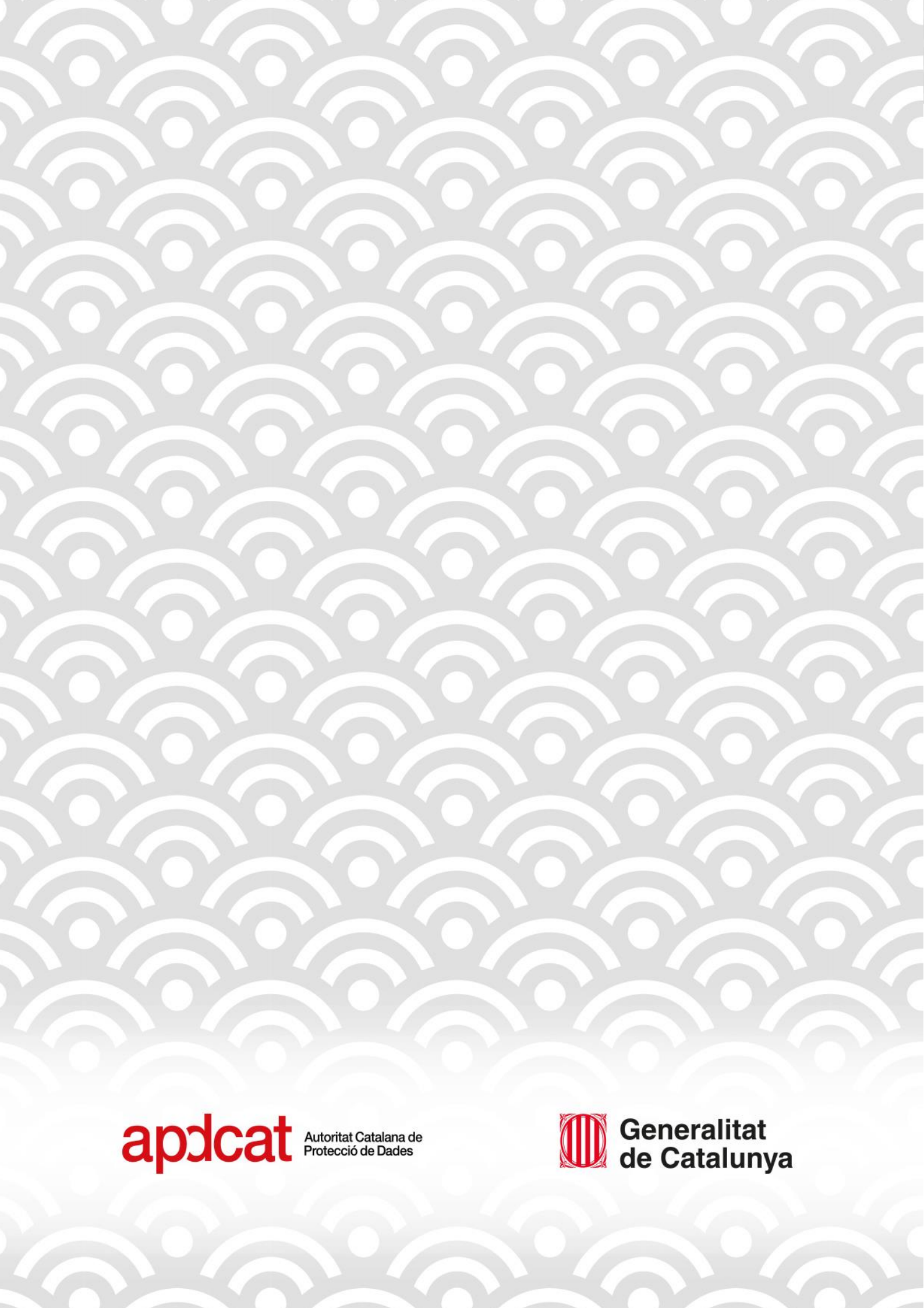
		(reduction in effort from Medium to Low). The severity remains Medium, but is lower than the initial level.	
Social assistance	[Low]	[Medium] The severity has decreased, due to complementary/alternative measures to access services (reduction in gravity of impact from Very High to High) and the reduced effort required to react due to the measures already in place (reduction in effort from Medium to Low). The severity remains Medium, but is lower than the initial level.	**[Medium]**
Access to services of general economic interest	[Medium]	[Medium] The level of severity is maintained because guaranteeing this right depends on the resources that governments allocate to it.	**[Medium]**

4. Comments

The emergence of new technologies is not only causing a revolution in the way we work and deliver public services, but also the need to rethink, from the design stage, the impact of these applications on citizens' fundamental rights. In fact, this paradigm shift was already evident with the entry into force of the General Data Protection Regulation (GDPR) and is now gaining momentum.

A methodology for carrying out a fundamental rights impact assessment (FRIA) is an essential tool for identifying the risks and the organisational and technical measures to be applied, with a transversal approach that must be integrated into the multidisciplinary work teams responsible for implementing AI in public administrations. To this end, as already observed with the application of the GDPR, it is absolutely necessary that data protection officers are involved, from the outset, in the digital transformation strategies and projects to be implemented in each organization. In this way, these risks can be analysed from the outset and by default, and the necessary organisational and technical measures taken.

On the other hand, it is also advisable to review job descriptions to include these types of skills and responsibilities, to invest in training and to develop appropriate professional profiles. Furthermore, at this time of transition to 'smart administration' (based on the use of AI), some kind of practical guide should be developed as an internal instruction, so that everyone in the organisations is aware of these risks, including guidelines on best practice in the use of AI in administrative activity.


apdcat
Autoritat Catalana de
Protecció de Dades
Generalitat
de Catalunya