



MICROB·AI·OME

Federated artificial intelligence for privacy-preserving international stratification of colorectal cancer patients

Ethical and Human Rights Impact Assessment Framework

September 30th, 2024

Microb-AI-ome - Public deliverable report **D6.1**

©Microb-AI-ome consortium

Authors: *Philipp Poindl, Walter Hötendorfer, Madeleine Müller, Georg Fröwis*

This report is published on the Microb-AI-ome website www.microbaiome.net

Action number: 101079777
Picture: ADOBE: 423952248



Table of content

1	Microb-AI-ome as a whole	6
2	D6.1 Ethical and Human Rights Impact Assessment Framework	6
2.1	Rationale	6
2.2	Introduction	7
2.2.1	The Combined Impact Assessment Process (Challenge)	7
2.2.2	Acknowledgements/Background of the Methodology	9
2.2.3	Microb-AI-ome partners involved	10
2.2.4	Other Relevant Impact Assessments	10
2.3	Applicability of the Regimes and Threshold Analysis	10
2.3.1	Applicability of the AI Act and the GDPR	10
2.3.2	Threshold Analysis	16
2.3.3	Decision on Conducting Assessment(s) and Joint Performance	21
2.4	Description of Facts	22
2.4.1	Architecture	22
2.4.2	Processes in Scope	22
2.4.3	Data Processing Operations in scope	23
2.5	Identification of Stakeholders and Role Distribution	23
2.5.1	Role Distribution pursuant to GDPR	24
2.5.2	Role Distribution pursuant to AI Act	24
2.5.3	Views of Data Subjects or their Representatives (Article 35(9) GDPR)	26
2.5.4	Involvement of the Data Protection Officers	26
2.5.5	Involvement of Other Relevant Stakeholders	26
2.5.6	Role Distribution pursuant to Ethical Considerations	27
2.6	Legal Admissibility	28
2.6.1	GDPR Compliance	28
2.6.2	Measures to Ensure Compliance with the AI Act	29
2.6.3	Automated Decisions (Article 22 GDPR and Article 86 AI Act)	33
2.6.4	General Fundamental Rights Considerations	38
2.7	Fundamental Rights Risk Assessment	40
2.7.1	Methodology	40
2.7.2	Preliminary Fundamental Rights Assessment	46
2.7.3	Affected Categories of Natural Persons and Groups	47
2.7.4	Risk Table	48
2.7.5	Risk Analysis	51
2.7.6	Risk summary	51
2.8	Ethical Risk and Benefit Assessment	52
2.8.1	Introduction	52
2.8.2	Affected Persons and Groups	54

2.8.3	Responsibility.....	56
2.8.4	Theoretical Framework of Ethical Principles.....	58
2.8.5	Ethical Impact Analysis	59
2.8.6	Mitigation of Ethical Risks and Weighing with Benefits.....	64
2.8.7	Conclusion concerning Ethical Impacts	65
2.9	Residual Risk Level and Proportionality.....	65
2.9.1	Discussion and Overall Assessment of Residual Risks	68
2.9.2	Proportionality Considerations in the Case at Hand	68
2.10	General Conclusion of the Impact Assessment	68
2.10.1	Summary of the Findings.....	68
2.10.2	Following Decisions	69
2.10.3	Outlook	70
3	Conclusion, next steps	71
Annex A: Ethical Principles		72
1	Transparency.....	72
2	Fairness	73
3	Non-maleficence, Safety, Robustness	74
4	Accountability	76
5	Privacy, Data Protection.....	77
6	Beneficence.....	78
7	Autonomy, Freedom, Human Agency.....	78
8	Sustainability	80
9	Human Dignity	80
10	Solidarity, Inclusion, Accessibility	81
11	Participation.....	81
12	Democracy	82
13	Efficiency.....	83
14	Trust	83
Annex B: Ethical Recommendations		84

History of Changes

Change			Description of Change	Approval of the Report	Stage
No.	Date	Version			
1	29.09.2024	1.0	Methodological Framework	Walter Hötendorfer	Impact Assessment Template

Table of acronyms and definitions

AI	Artificial intelligence
API	Application programming interface
CFR	Charter of Fundamental Rights
CNIL	Commission nationale de l'informatique et des libertés (French data protection authority)
CRC	Colorectal cancer
DoA	Description of Action (of the Microb-AI-ome project)
DP	Differential privacy
DPIA	Data Protection Impact Assessment
DPO	Data protection officer
EC	European Commission
ECJ	European Court of Justice
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
EIA	Ethical Impact Assessment
EP	European Parliament
FIT	Faecal immunochemical test
FL	Federated learning
FRIA	Fundamental Rights Impact Assessment
GANs	Generative adversarial networks
GDPR	General Data Protection Regulation
GUI	Graphical user interface
HE	Homomorphic encryption
ICDPPC	International Conference of Data Protection and Privacy Commissioners
ICO	Information Commissioner's Office (UK data protection authority)
IEEE	Institute of Electrical and Electronics Engineers

IVDR	In Vitro Diagnostic Medical Device Regulation (Regulation (EU) 2017/746)
KPI	Key performance indicator
MDR	Medical Device Regulation (Regulation (EU) 2017/745)
ML	Machine learning
MS	Milestone
OECD	Organisation for Economic Co-operation and Development
PAML	Privacy-aware-machine-learning
PII	Personally identifiable information
PPDM	Privacy-preserving data mining
PPDP	Privacy-preserving data publishing
RI	Research Institute AG & Co KG
SMPC	Secure multi-party computation
TFEU	Treaty on the Functioning of the European Union
TOMs	Technical and organisational measures
UHAM	Universitaet Hamburg
WP	Work package

1 Microb-AI-ome as a whole

In the EU, 1 in 35 women and 1 in 23 men will be diagnosed with colorectal cancer (CRC) in their life span (ca. 340,000 cases and 156,000 deaths in 2020) causing an annual economic burden of ca. 20 billion EUR. Identifying CRC early enables better treatment options. Screening usually entails a quantitative faecal immunological test (FIT) to predict the need of colonoscopy for the detection of colorectal lesions, an expensive and invasive procedure.

We aim to predict this need with specificity increased by >20 percentage points by using metagenomic microbiomes. We hypothesise that computational microbiome profiles extracted using artificial intelligence (AI) technology will allow for optimised personal therapy stratification. However, clinicians do not have access to broad microbiome data.

With Microb-AI-ome, we will develop a novel kind of computational stratification technology to enable microbiome-enhanced precision medicine of CRC. Metagenomic microbiome data to date is distributed over many national registries, and privacy regulations are hindering its effective integration. With Microb-AI-ome, we will overcome this barrier by establishing the first privacy-preserving federated big data network in CRC research. We will integrate isolated, national databases into one international federated database network - rather than a cloud - covering metagenomes for over 5,000 individuals screened for CRC, and an expected total of 100,000 by 2026.

Microb-AI-ome ensures that no sensitive patient data will leave the safe harbours of the local databases while still allowing for the classification of clinical CRC phenotypes, which we will demonstrate in clinical practice allowing regulatory bodies to adopt evidence-based guidelines. Our consortium combines expertise in CRC and its treatment, microbiomics, artificial intelligence, software development, and privacy protection to close the gap between privacy and big data in international medical research.

After performing colonoscopies on the FIT positive subjects, 37.4% had a normal colonoscopy or lesions without neoplasia (such as haemorrhoids), resulting in a specificity of 62.6%, which we aim to improve with Microb-AI-ome by at least 20 percentage points to >83%.

2 D6.1 Ethical and Human Rights Impact Assessment Framework

2.1 Rationale

The purpose of WP6 is the legal and ethical assessment of the activities in the Microb-AI-ome project and to provide legal and ethical guidance for these activities. This is achieved by a process of ethical and human rights impact assessment (with a special focus on data protection), which is carried out throughout the whole project and which is the centrepiece of WP6.

The present deliverable D6.1 contains the results of Task 6.1 of the project and aims to achieve its objective 6.1, the development of a project-specific ethical and human rights impact assessment framework. This framework is a methodological framework. It is the methodological basis for carrying out the ongoing (impact) assessment of the activities in the Microb-AI-ome project and its results in the years ahead until month 60. In the course of this process, the present framework document will be filled with content, i.e. assessment results and risk assessment results in particular, which will make up the final report of WP6, the Ethical and Data Protection Impact Assessment report (D6.4). In the course of this process, the framework will also be adjusted and refined where necessary.

2.2 Introduction

In the present methodology, aspects of different impact assessment methodologies have been merged together and are supposed to complement each other. The initial basis for that was a methodology and template for a Data Protection Impact Assessment (DPIA) pursuant to Article 35 GDPR, which has been created, used and further developed by Research Institute and the Authors of this document in a large number of projects. This basis has been enriched, complemented and amended by elements of the Fundamental Rights Impact Assessment (FRIA), specifically pursuant to Article 27 AI Act, where a particularly wide range of fundamental and human rights has been taken into consideration. This was a major aspect in tailoring the methodology for the use in the Microb-AI-ome project, which at its core applies AI in the medical domain. Therefore, the AI Act could be applicable to potential results of the project (see section 2.3.1.1 on the applicability of the AI Act), and irrespective of its applicability throughout project runtime it was considered important that the present methodology fulfils the criteria of a FRIA pursuant to Article 27 AI Act. As a third methodological pillar, in order to ensure a broader view on the impact of applying AI in the medical domain, the methodology is complemented by elements of Ethical Impact Assessments (EIA).

While the methodology is tailored specifically for the application of AI in sensitive areas such as medicine, in particular by taking into account the new rules of the AI Act and a strong ethical component, it is by intention published as a generic methodology document which can be applied in different projects of that kind. For this purpose, it contains **instructions and recommendations marked in turquoise** which are to be replaced with the respective explanations and assessments for the particular case at hand. In Microb-AI-ome, this will be done in the course of the project, leading to the final report of WP6, the Ethical and Data Protection Impact Assessment report (D6.4).

2.2.1 The Combined Impact Assessment Process (Challenge)

Article 27(4) AI Act expressly references to respective DPIAs. As far as it is deemed necessary or useful, the respective assessments should consequently be conducted in a complementary manner (see in detail section 2.3.3). Concerning the FRIA pursuant to Article 27 AI Act it is to be noted that the AI Office is yet to publish a template for a questionnaire (including through an automated tool) for the facilitation of the process which it shall develop pursuant to Article 27(5) AI Act. That is then presumably to be considered particularly with regard to the methodology for respective FRIA. However, the methodology at hand shall provide guidance on the performance of respective impact assessments already before that (and presumably also beyond that¹).

While having some differences and particularities (as is also apparent in the methodology at hand), both the DPIA and the FRIA² aim at the understanding of pertinent impacts and the specific mitigation of consequential risks of the object of the assessment, particularly targeted at such stemming from technologies and processes associated with a high potential of risk. In general, furthermore impact assessments can be seen as an alternative to precautionary measures concerning technology-related risks, like restrictions or bans.³

Elements of Ethical Impact Assessment are part of the methodology at hand in particular because, while all three types of impact assessments mentioned have some overlaps, the EIA addresses aspects that are not

¹ Cf. in that regard *Tünde Fülöp and Philipp Poindl*, Article 27. Fundamental rights impact assessment for high-risk AI systems, in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) sections 3.1 and 3.3.4.4.

² When, in the course of this document, it is referred to DPIA and FRIA, this particularly refers to impact assessments carried out according to the respective provisions in Art. 35 GDPR and Art. 27 AI Act.

³ Cf. in total *Mantelero*, *Beyond Data – Human Rights, Ethical and Social Impact Assessment*, Asser Press/Springer, Information Technology and Law Series, IT&LAW 36, 2022, 48 et seq.

yet covered through a fundamental rights (or a specific data protection) point of view. This, e.g., concerns aspects of societal well-being, human relationships or impacts on the environment.⁴ Also, in addition to the analysis of ethical risks, the ethical approach allows a broader view on the benefits of the object of the assessment.⁵ One such example is the principle of beneficence, which refers to the positive impact of a particular technology on societal well-being. This also illustrates that, while the DPIA and FRIA are often deemed to primarily focus on the individual level, the EIA additionally incorporates the collective dimension of the utilisation of a specific technology.⁶ E.g., when it comes to the impact of the use of an AI, questions about future generations and their well-being can be addressed. With this, the EIA takes on a broader perspective and allows for a reflection beyond existing legal frameworks. Such EIA are essentially conducted on a voluntary basis, in opposition to DPIA and FRIA, which under certain circumstances are to be conducted mandatorily (see section 2.3.2).

The obligations to conduct DPIA and FRIA refer to a point in time before the respective application of the object(s) of the assessments: In other words, the impact assessments have to be conducted **beforehand** (cf. in detail Article 35(1) GDPR and Article 27(1) AI Act).⁷ Responsible entities should in a first step therefore analyse whether those obligations pursuant to the GDPR and the AI Act (as well as these regimes per se) apply to them, which after all would also entail further legal requirements (than just the initial performance of the impact assessment) that need to be complied with, particularly in case of such an initial obligation to conduct the assessment (see in detail section 2.3, particularly 2.3.2). However, it is also to be emphasised that the Combined Impact Assessment at hand is to be understood as a continuous, iterative process and the respective report as a *living document*, meaning that it is to be reviewed, updated and edited after the initial completion (see particularly also 2.10.2).⁸

Subsequently, a systematic and thorough description of facts – where applicable pursuant to relevant legal requirements – is to be conducted, also including a description of purposes and surroundings of the object to be assessed (see section 2.4). Based on that, the respective responsibilities (e.g., from a data protection perspective) are to be addressed, and other relevant stakeholders to be identified, and where appropriate, to be involved (see section 2.5).

Moreover, the principal permissibility, necessity and proportionality of the processes in scope are to be assessed, and specific measures to safeguard these aspects and relevant legal requirements to be considered accordingly (see particularly section 2.6).

Furthermore, affected persons and groups must be analysed, and – as the core of the process – a risk assessment, particularly from a human rights and data protection perspective, extended by ethical aspects, to be conducted subsequently (see particularly sections 2.7 and 2.8).

⁴ Cf. *Mantelero*, *Beyond Data – Human Rights, Ethical and Social Impact Assessment*, Asser Press/Springer, Information Technology and Law Series, IT&LAW 36, 2022, 94.

⁵ Cf. e.g. *Tünde Fülöp and Philipp Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.1.1 referring to *Mantelero*, *Beyond Data – Human Rights, Ethical and Social Impact Assessment*, Asser Press/Springer, Information Technology and Law Series, IT&LAW 36, 2022, 26; concerning the benefits of complementing different types of Impact Assessments (also ethical and human rights) see furthermore e.g.: *Mantelero*, *AI and Big Data: A blueprint for a human rights, social and ethical impact assessment*, *Computer Law & Security Review* 34 (2018), 754 <https://doi.org/10.1016/j.clsr.2018.05.017> (accessed 26. 9. 2024).

⁶ Cf. *Mantelero*, *Computer Law & Security Review* 34 (2018) 754 (765) <https://doi.org/10.1016/j.clsr.2018.05.017>.

⁷ Art. 35(1) GDPR: '[...] the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data'; Art. 27(1) AI Act: 'Prior to deploying a high-risk AI system [...] deployers [...] shall perform an assessment of the impact on fundamental rights [...]'; cf furthermore *Tünde Fülöp and Philipp Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.3.5.

⁸ Cf. concerning the FRIA and the DPIA e.g.: *Tünde Fülöp and Philipp Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.3.5; *Kastelitz/Hötzendorfer/Riedl*, *Ausgewählte Fragen der Durchführung einer Datenschutz-Folgenabschätzung gemäß Art 35 DSGVO*, in *Jahnel* (ed.), *Jahrbuch Datenschutz* 2017, 113 (130).

Based on the respective analyses, suiting measures are to be determined – where applicable, as required by the law – particularly to address the identified risks. The risk assessment should analyse and evaluate those identified risks and determine measures methodically and systematically in respect of probability of occurrence and severity of harm (see particularly section 2.7). Subsequently, ethical considerations are to be taken into account that have not yet been covered in the preceding steps. This includes – depending on the particularities of the case in question – an analysis of the ethical benefits, the applicability of and compliance with relevant ethical principles, a more extensive list of affected persons, questions of accountability, the handling of value conflicts and recommendations for ethical risk mitigation (see section 2.8).

Finally, an assessment of the residual risks and overall risk level is to be conducted and, also in comparison with potential benefits, proportionality considerations concerning the object of the assessment shall be made (see section 2.9).

In conclusion, the findings of the assessments shall be summarised (see section 2.10.1). Based on those, also certain decisions have to be reached, such as concerning further steps as well as regarding consequences resulting from residual risks. In particular, notification, registration and consultation duties are to be fulfilled, where applicable (see section 2.10.2).

All of the above-mentioned aspects of the respective assessments have to be conducted pursuant to pertinent legal provisions (particularly Article 35 GDPR and Article 27 AI Act), where applicable. At least in some (essential) respects, it is furthermore – due to legal requirements – simply necessary to document the results of the impact assessments and associated decisions somehow.⁹ Therefore, it is highly recommended to document the whole process, also in light of future iterations (cf. particularly 2.10.2). This is achieved by the report at hand.

2.2.2 Acknowledgements/Background of the Methodology

In respect to the research project *Microb-AI-ome* it is to be noted that the methodology at hand is based on previous work and long-time experience by Research Institute in various (research) projects. This includes work from the following authors: Georg Fröwis, Jan Hospes, Walter Hötendorfer, Markus Kastelitz, Elisabeth Mayer, Madeleine Müller, Philipp Poindl, Renate Riedl, Robert Rothmann, Moritz Rothmund-Burgwall, Heidi Scheichenbauer, David M. Schneeberger and Christof Tschohl. It is a particular quality of this methodology that it is based on years of experience in a lot of different projects, as described in the Description of Action of *Microb-AI-ome*. This means that parts of the methodology have been developed and applied earlier and have been further developed over several years, in particular in the research projects *FeatureCloud*, grant agreement No. 826078 and *Screen4Care*, grant agreement No. 101034427. While the document at hand is to a large extent a newly written work, some general texts in the document have been already used in other projects, in particular text describing the legal framework in data protection law and also text concerning fundamental (human) rights per se. Such descriptions and explanations of fundamentals are required over and over in the same manner as long as there is no change in the legal situation. They are continuously improved as they are re-used and this process of continuous improvement would be thwarted and it would be at the same time a waste of resources if such texts would be newly written for each project. Instead, the resources employed for this deliverable have been used for further developing the existing knowledge and

⁹ With regard to the FRIA pursuant to Art. 27 AI Act this particularly concerns notification duties to the market surveillance authority pursuant to Art. 27(3) AI Act, as well as registration obligations pursuant to Arts 26(8), 49 and Annex VIII Section C, point 4 AI Act (cf. *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) sections 3.3.2 and 3.3.4); with regard to the documentation of DPIA pursuant to Art. 35 GDPR, it is to be noted that the regulation holds data controllers generally accountable to demonstrate compliance with its requirements pursuant to Arts 5(2) and 24, and Arts 35(7) and 36 indicate that the DPIA is to be documented as well (cf. in detail *Kastelitz/Hötendorfer/Riedl* in *Jahnel* (ed.), *Jahrbuch Datenschutz 2017*, 113 (129-130)).

adding a range of new aspects to the existing DPIA methodology, in particular the FRIA and EIA and their integration into one methodology.

2.2.3 Microb-AI-ome partners involved

This deliverable was created by partner Research Institute (RI) and reviewed by partner GNOME Design (GND).

2.2.4 Other Relevant Impact Assessments

Particularly considering Article 27(2) AI Act, it is indicated to also draw upon other (already existing) impact assessments concerning the object of the assessment, because that provision stipulates that “[t]he *deployer* may, in similar cases, rely on previously conducted fundamental rights impact assessments or existing impact assessments carried out by [the] *provider*”. In doing so, however, the respective particularities of such impact assessment (types and models) are to be considered accordingly.¹⁰

With regard to the Microb-AI-ome project, also existing and ongoing risk assessments shall be taken into account, in particular the risks identified already in the proposal and the continuous risk management process by the project management office.

2.3 Applicability of the Regimes and Threshold Analysis

2.3.1 Applicability of the AI Act and the GDPR

This section contains a short analysis of the scope and applicability of the AI Act and the GDPR.

2.3.1.1 AI Act

The Scope of application of the AI Act particularly combines geographical, personal and material/technical elements. The first prerequisite for its application essentially¹¹ is an “AI System”, defined in Article 3(1) AI Act as “a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments”.

[Elaboration on aspects in the case at hand, where applicable]

In consequence, the regulation applies to certain actors associated with such AI systems.¹² In the context at hand, primarily *deployers*, but to some extent also *providers* of AI systems are of relevance. The *deployer* is defined in Article 3(4) AI Act as a “natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity”, whereas the *provider* is “a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge” (Article 3(3) AI Act).

The Regulation, however, does not apply to all deployers or providers. At first, in relation to deployers, it refers to such “that have their place of establishment or are located within the Union” (Article 2(1)(b) AI Act),

¹⁰ The wording of the provision, particularly in comparison to other text parts indicates that impact assessments could be addressed generally here: Cf. in total *Fülöp/Poindl in Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.1.1.

¹¹ Certain provisions, however, e.g. also apply to *general-purpose AI Models*.

¹² Cf. particularly Art. 2 AI Act; see in detail also below.

in relation to providers on the other hand, to such *“placing on the market or putting into service AI systems or placing on the market general-purpose AI models in the Union, irrespective of whether those providers are established or located within the Union or in a third country”* (Article 2(1)(a) AI Act). It furthermore applies to both deployers and providers *“that have their place of establishment or are located in a third country, where the output produced by the AI system is used in the Union”* (Article 2(1)(c) AI Act). Concerning the particular role distribution in the case at hand see section 2.5 below (particularly 2.5.2).

Moreover, regarding *affected persons*, it applies to such *“located in the Union”* (Article 2(1)(g) AI Act; in that concern see particularly 2.7.3).

Moreover, the AI Act contains certain further specifications and exemptions. This e.g. concerns (exclusive) military/defence use cases, in the sense that the regulation on the one hand *“does not apply to AI systems where and in so far they are placed on the market, put into service, or used with or without modification exclusively for military, defence or national security purposes, regardless of the type of entity carrying out those activities”*. On the other hand, it also *“does not apply to AI systems which are not placed on the market or put into service in the Union, where the output is used in the Union exclusively for military, defence or national security purposes, regardless of the type of entity carrying out those activities”* (Article 2(3) AI Act). In that regard, also Recital 24 of the AI Act contains some explanations, particularly in light of the interpretation of the term *“exclusively”* used in Article 2(3):

- *“[I]f an AI system developed, placed on the market, put into service or used for military, defence or national security purposes is used outside those temporarily or permanently for other purposes, for example, civilian or humanitarian purposes, law enforcement or public security purposes, such a system would fall within the scope of this Regulation.”*
- *“AI systems placed on the market or put into service for an excluded purpose, namely military, defence or national security, and one or more non-excluded purposes, such as civilian purposes or law enforcement, fall within the scope of this Regulation [...]”*
- *“In those cases, the fact that an AI system may fall within the scope of this Regulation should not affect the possibility of entities carrying out national security, defence and military activities, regardless of the type of entity carrying out those activities, to use AI systems for national security, military and defence purposes, the use of which is excluded from the scope of this Regulation.”*
- *“An AI system placed on the market for civilian or law enforcement purposes which is used with or without modification for military, defence or national security purposes should not fall within the scope of this Regulation, regardless of the type of entity carrying out those activities.”*

[Where applicable, explanations on such activities in the case at hand, otherwise contrary statement]

The AI Act furthermore *“does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development”* (Article 2(6) AI Act). In that regard also Recital 25 of the AI Act contains some explanations:

- *“This Regulation should support innovation, should respect freedom of science, and should not undermine research and development activity. It is therefore necessary to exclude from its scope AI systems and models specifically developed and put into service for the sole purpose of scientific research and development.”*
- *“Moreover, it is necessary to ensure that this Regulation does not otherwise affect scientific research and development activity on AI systems or models prior to being placed on the market or put into service.”*
- *“As regards product-oriented research, testing and development activity regarding AI systems or models, the provisions of this Regulation should also not apply prior to those systems and models being put into service or placed on the market.”*

- *“That exclusion is without prejudice to the obligation to comply with this Regulation where an AI system falling into the scope of this Regulation is placed on the market or put into service as a result of such research and development activity and to the application of provisions on AI regulatory sandboxes and testing in real world conditions.”*
- *“Furthermore, without prejudice to the exclusion of AI systems specifically developed and put into service for the sole purpose of scientific research and development, any other AI system that may be used for the conduct of any research and development activity should remain subject to the provisions of this Regulation.”*

In this regard, it is therefore to be emphasised that the research and development phase of respective technologies must on the one hand be differentiated to some degree from a potential exploitation phase. On the other hand, pertinent requirements should nevertheless be considered already at the early stages of research projects (also following a privacy-by-design-approach; cf. Article 25 GDPR) in order to ensure that compliance can be achieved immediately when pertinent requirements apply without requiring complex changes to such technologies.

[Where applicable, explanations on such activities in the case at hand, otherwise contrary statement]

Beyond that, the AI Act e.g. *“does not apply to obligations of deployers who are natural persons using AI systems in the course of a purely personal non-professional activity”* (Article 3(10) AI Act; see also already Article 3(4) AI Act defining *“deployer”*: *“[...] except where the AI system is used in the course of a personal non-professional activity”*). In that regard, Recital 13 of the AI Act states: *“The notion of ‘deployer’ referred to in this Regulation should be interpreted as any natural or legal person, including a public authority, agency or other body, using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity. Depending on the type of AI system, the use of the system may affect persons other than the deployer.”*

[Where applicable, explanations on such activities in the case at hand, otherwise contrary statement]

[Where applicable, explanations on other relevant specifications and exemptions particularly pursuant to Article 2 AI Act, e.g. concerning AI systems released under free and open-source licences (paragraph 12)]

Subsequent to the general applicability of the regulation and following a risk-based approach, the provisions of the AI Act are then divided into several risk-stages: prohibited AI practises (which are deemed to be unacceptable, see particularly section 2.6.2.3), high-risk AI Systems and AI Systems for which transparency provisions apply¹³ (besides certain rules for general-purpose AI models¹⁴). Practically, this classification also results in a fourth (risk-)class of AI Systems, which are not subject to any specific provisions of the AI Act.¹⁵ Thereby, a major part of the provisions, such as the obligation to conduct a FRIA (see principally section 2.3.2.2), only applies to high-risk AI systems (cf. in detail particularly Chapter III AI Act).

The classification of AI Systems as high-risk is addressed in particular in Article 6 AI Act. This provision essentially provides two alternatives for the classification of high-risk AI systems. The first refers to AI systems which are *“intended to be used as a safety component of a product”* (listed in Annex I) or which themselves are *“a product, covered by the Union harmonisation legislation listed in Annex I”* **and** such product (/the AI

¹³ Cf. particularly Recital 26 AI Act; cf. in this regard, however, also the transparency obligations of deployers of emotion recognition systems according to Art. 50(3) AI Act are to be highlighted, because these systems principally also constitute high-risk AI Systems pursuant to Art. 6(2) in connection with Annex III point 1(c) AI Act; see also Recital 132 AI Act: *“Certain AI systems intended to interact with natural persons or to generate content may pose specific risks of impersonation or deception irrespective of whether they qualify as high-risk or not.”*

¹⁴ See in detail particularly Chapter V AI Act.

¹⁵ Nevertheless, the general definition of AI Systems pursuant to the AI Act may still apply to those systems, which might have consequences in other respects.

system as such a product) *“is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation listed in Annex I”*. This is furthermore *“[i]rrespective of whether an AI system is placed on the market or put into service independently of”* such products (Article 6(1) AI Act). With regard to medical applications, the Union harmonisation legislation listed in Annex I in particular comprises the Medical Devices Regulation¹⁶, as well as the In Vitro Diagnostic Medical Devices Regulation¹⁷.

The second, for Article 27 AI Act more relevant alternative concerns *“AI systems referred to in Annex III’ AI Act, which shall also “be considered to be high-risk”* (Article 6(2) AI Act). However, also a derogation procedure from this general rule is stipulated in the provision, in the course of which *“an AI system referred to in Annex III shall not be considered to be high-risk where it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making”* (Article 6(3) AI Act). Subsequently, also some (alternative) criteria for the application of that provision/the respective assessment are stipulated (such as the intention of the AI system to only *“perform a narrow procedural task”*; see in total paragraph 3). Furthermore, the EC shall also provide guidelines concerning the specification of the practical implementation of Article 6 AI Act (paragraph 5).

[General considerations and remarks (in light of more specific explanations in sections 2.3.2.2 and 2.6.2.3) concerning the classification of AI Systems, particularly considering Annex III, concerning the case at hand]

2.3.1.2 GDPR

The scope of application of the GDPR is very broad, and is not subject to general material exemptions similar to the research exemption of the AI Act explained above (other than the exemption for purely personal or household activities, see below, which is not relevant in the present context). The GDPR, according to its Article 1(1), lays down rules relating to the protection of natural persons with regard to the processing of personal data. In order to examine the scope of application, the (very broad) definitions of the terms processing and personal data will be analysed in detail below, followed by the provisions regarding material and territorial scope of the GDPR.

2.3.1.2.1 Personal Data

The concept of personal data is the starting point for analysing whether the GDPR is applicable or not.

Article 4(1) GDPR defines the term *“personal data”* as

“any information relating to an identified or identifiable natural person (“data subject”); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.”

Thus, personal data is a very broad term and can be information such as name, address or date of birth, but also information about the private and family life (such as marital status, leisure activities, consumer behaviour or dietary habits) as well as professional or economic activities (such as employment or property relations). Other examples often include external characteristics of a person (such as height, weight, skin

¹⁶ Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, OJ L 5. 5. 2017, 117, 1.

¹⁷ Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU, OJ L 5.5.2017, 117, 176.

colour, papillary lines, iris or vein structure) as well as inner attitudes (such as motives, desires and ideological convictions), but also characteristics of things which in turn can be attributed to a particular natural person (such as information about the car or real estate a person owns). It is irrelevant whether the information is true or not, or whether it is only true with a certain statistical probability (Recitals 59 and 65 GDPR).

The GDPR further defines so called special categories of personal data, also known as sensitive data, in Article 9(1). These data are subject to a stricter data processing regime than ordinary (non-sensitive) personal data. In particular, health data, as well as genetic and biometric data fall within this special category.

As the definition in Article 4(1) GDPR implies, it is necessary that the respective information can be linked to a specific person in order to be able to speak of personal data. Information can be considered to relate to a person when it is about the respective individual, or also, when it is about an object which itself belongs to the individual or relates to it in another way (Article 29 Data Protection Working Party 2007, p 6 et seq.).

Article 4(1) further clarifies that only natural persons count as such data subjects. In addition, Recitals 14 and 17 GDPR specify that data subjects must be living human beings, which means that data protection does not extend to deceased persons (Recital 27 GDPR) or to legal persons (Recital 14 GDPR). However, data on deceased persons, especially in the medical domain, might also contain some information relating to living persons. One example is information regarding the relatives of the deceased person, as in the case of hereditary genetic dispositions for certain diseases. Such data fall within the scope of protection of the GDPR, not because it is relating to the deceased person but because it is relating to another person that is still alive.¹⁸

The data subject to whom the respective information relates must be identified or at least be identifiable. The identification of a person can be derived directly from the given information or indirectly; that is, if the existing information is not sufficient to unambiguously identify an individual but it can still be identified by linking the existing information with additional information or by using additional criteria or means of identification such as those listed in Article 4(1) GDPR.¹⁹

The most common identifier is a person's name. However, a widely used name may not be sufficient to uniquely identify a person. In such cases, a second piece of information such as the location (address), other specific factors or a specific context are required. Article 4(1) GDPR also refers to “*identification numbers*” and “*online identifiers*”; beside IP addresses, these numbers and identifiers can also be codes such as MAC addresses, the “*International Mobile Station Equipment Identity*” (IMEI) or company codes such as Apple's “*Unique Device ID*” (UDID).

If the additional information (e.g. the identification number, the online identifier or another specific identifying factor) is stored separately to ensure that no personal reference can be made, the respective data is called pseudonymous data (see subsequent section).²⁰ According to Article 4(5) GDPR the term “*pseudonymisation*” means that “*the personal data can no longer be attributed to a specific data subject without the use of additional information*” and such additional information “*is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*”.

¹⁸ Cf. Rothmann/Kastelitz/Rothmund-Burgwall, Archive als ‚öffentliches Gedächtnis‘ personenbezogener Patientendaten? in Jahnelt (ed.), Datenschutzrecht. Jahrbuch 2021, NWV 2022.

¹⁹ Cf. Ennöckl, Der Schutz der Privatsphäre in der elektronischen Datenverarbeitung, Forschungen aus Staat und Recht – Volume 174, Verlag Österreich 2014, 107 et seq.

²⁰ Karg, Anonymität, Pseudonymität und Personenbezug revisited, Datenschutz und Datensicherheit – DuD 2015, 520–526, <https://doi.org/10.1007/s11623-015-0463-z>, 520 et seq.

Recital 26 GDPR also clarifies that personal data that have undergone pseudonymisation “*should be considered to be information on an identifiable natural person*” if these data “*could be attributed to a natural person by the use of additional information*”. This means that pseudonymised data is still considered personal data if the additional information makes it possible to attribute the data to a natural person.

Recital 26 GDPR further states that

“to ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments”.

In this regard, the European Court of Justice (ECJ) has stated that the reference to a natural person can be assumed even if an allocation is not directly possible but the relevant body has legal means to obtain the additional information required to identify the person.²¹ However, the principles of data protection should not apply to anonymous information, i.e. information that does not relate to an identified or identifiable natural person (Recital 26 GDPR). To determine whether a natural person is identifiable, not every theoretical possibility to identify the person must be taken into account but only means reasonably likely to be used to do so. To ascertain whether means are reasonably likely to be used, all objective factors should be taken into account, such as the costs of and the amount of time required for identification, the available technology at the time of the processing, and technological developments.²²

From this it can be concluded that in order to determine whether data is personal data under GDPR a practical, not a theoretical standpoint must be taken. Means reasonably likely to be used are means that not only exist theoretically but that would be used practically.

This means that in order to determine whether personal data is processed or the data is anonymous a practical assessment must be carried out: From the attack vectors on the anonymity of data only those are legally relevant that are reasonably likely to be used by an actual attacker in practice. This must be assessed on the basis of objective factors such as the costs and the amount of time required, the required skills, the potential gain and the available technology but also possible technological developments in the future. In order to assess whether an attack vector is relevant from a legal perspective, attacks that are reasonably unlikely can be ignored. An attack can be considered reasonably unlikely if it cannot be imagined that it will happen in practice in the given context because the attacker will shy away from the required effort.

For determining whether personal data is processed, in the context of pseudonymisation and anonymisation of the data, it can be important to distinguish between exactly reproducible and not exactly reproducible data. This is relevant for trying to re-identify individuals in the data by reproducing the same data from a known individual. Raw data that cannot be exactly reproduced by repeating e.g. medical examination at a later point, in particular information which is measured where there is some measurement inaccuracy and every measurement leads to a slightly different result and information which changes over time, cannot be reproduced in a way to exactly match the original raw data, which could render such an attack impossible or at least reasonably unlikely. The case is different if there is data, such as binary data or genetic data in particular, which is (theoretically) exactly reproducible when a new examination or other act of data collection is carried out.

²¹ CJEU 19. 10. 2016, C-582/14, *Breyer*, ECLI:EU:C:2016:779, para. 49.

²² *Esayas*, The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the 'all or nothing' approach, *European Journal of Law and Technology*, Vol 6, No 2, 2015.

2.3.1.2.2 Processing of Personal Data

The next concept which must be defined in order to determine the applicability of the GDPR is “*processing*”. The GDPR, according to its Articles 2 and 3, applies to the processing of personal data. Processing is defined in Article 4(2) GDPR in the broadest possible manner as “*any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction*”. This means that merely having personal data or doing anything with personal data is “*processing*” and falls within the scope of the GDPR if the following conditions are fulfilled.

2.3.1.2.3 Material Scope of the GDPR

The material scope of the GDPR is defined in Article 2 GDPR. Article 2(1) GDPR stipulates that the GDPR applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system. Article 2(2) GDPR contains several exemptions of this general rule, primarily regarding activities on the context criminal offences, public and national security on the one hand, and purely personal or household activities on the other hand.

[Explanation based on this, why the GDPR applies in the case at hand (or does not apply).]

2.3.1.2.4 Territorial Scope of the GDPR

The first – and most important – case regarding the territorial scope of the GDPR relates to the so-called “*principle of establishment*” (Article 3(1) GDPR): The GDPR applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the EU, regardless of whether the processing takes place in the EU or not.

Secondly, Article 3(2) GDPR pertains to the processing of personal data of data subjects who are in the European Union by a controller or processor not established in the European Union. In this scenario, however, processing alone is not sufficient, as it must be related to one of two possibilities. The first possibility is the offering of goods or services (to affected data subjects). The second possibility is that the processing activities are related to the monitoring of the behaviour of the data subjects as far as their behaviour takes place within the European Union. According to Recital 24 of the GDPR this applies primarily, but not exclusively, to Internet monitoring. The relevant guidelines²³ (EDPB 2019, 20) also list “[m]onitoring or regular reporting on an individual’s health status” as an example for this category.

The third option concerning the territorial scope is that the GDPR applies to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by virtue of public international law, which is particularly relevant for diplomatic missions and consular posts (Article 3(3) GDPR).

[Explanation based on this, why the GDPR applies in the case at hand (or does not apply).]

2.3.2 Threshold Analysis

In the course of the so-called *threshold analysis*, it is to be analysed whether pertinent legal preconditions for respective duties of the responsible entities apply.²⁴ The entity obliged to conduct a DPIA pursuant to the

²³ European Data Protection Board, Guidelines 3/2018 on the territorial scope of the GDPR (Article 3), https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-32018-territorial-scope-gdpr-article-3-version_en.

²⁴ Cf. e.g. with regard to the DPIA: *Kastelitz/Hötzendorfer/Riedl* in *Jahnel* (ed.), *Jahrbuch Datenschutz* 2017, 113 (130, 136-137).

GDPR is the data controller, whereas the FRIA pursuant to the AI Act is to be conducted by the deployer (of a respective high-risk AI system).²⁵ Even in case the responsible entity²⁶ already took the decision to conduct a respective impact assessment, it is nevertheless important to consider whether there is an obligation to do so for several reasons. Primarily, in case conducting a DPIA or FRIA is mandatory, these obligations entail further legal specifications that need to be complied with particularly in that case.²⁷ This e.g. concerns update/review obligations concerning impact assessments (Articles 35(11) GDPR and 27(2) AI Act), as well as duties to involve competent authorities in the process (consultation/notification, Articles 36 GDPR, 27(3) AI Act) and to register findings of the impact assessments (Article 49(3), Annex VIII AI Act).²⁸

Moreover, acknowledging that an impact assessment is not just conducted voluntarily but because it is required by law, may naturally lead to increased awareness and seriousness in the process.

Inter alia in light of the accountability principle pursuant to Article 5(2) GDPR, it is furthermore indicated to also document the threshold analysis, even in case no impact assessment is conducted.²⁹

2.3.2.1 Scope of DPIA pursuant to Article 35 GDPR

Subsequently, the criteria of Article 35 GDPR for the mandatory performance of a DPIA are to be addressed. Conducting a DPIA pursuant to Article 35 GDPR principally is obligatory “[w]here a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons” (paragraph 1). As mentioned above the entity obliged principally is the data controller. [respective statement referring to the analysis in section 2.5.1]

According to Article 35(3) GDPR, a DPIA “shall in particular be required in the case of”:

- “a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person”
- “processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; or”³⁰
- “a systematic monitoring of a publicly accessible area on a large scale.”

[respective analysis of relevant factors in the case at hand]

²⁵ Cf. Fülöp/Poindl in Pehlivan/Forgó/Valcke (eds.), Artificial Intelligence Act: A Commentary (Kluwer Law International forthcoming) section 3.3.5 with further references.

²⁶ This term is (following the aforementioned legal stipulations) used subsequently to address the entity in charge of the Combined Impact Assessment.

²⁷ Although it could even be discussed whether some of the legal stipulations concerning impact assessments also apply in case an impact assessment is conducted voluntarily (e.g., at least the German wording Art. 36(1) GDPR on the consultation of the supervisory authority might be interpreted in a way that the provision would apply to every DPIA conducted (methodologically) in accordance with Art. 35 GDPR).

²⁸ Concerning all of these aspects see particularly section 2.10 and, furthermore, section 2.6.2.2.

²⁹ Cf. Kastelitz/Hötendorfer/Riedl in Jahnelt (ed.), Jahrbuch Datenschutz 2017, 113 (121, 130); in principle, similar considerations are to be deduced concerning the FRIA pursuant to the AI Act; concerning the different approaches for the scope of the impact assessments see however: Fülöp/Poindl in Pehlivan/Forgó/Valcke (eds.), Artificial Intelligence Act: A Commentary (Kluwer Law International forthcoming) section 3.3.5.

³⁰ Special categories of data referred to in Article 9(1) are such „revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation”; with regard to „personal data relating to criminal convictions and offences referred to in Article 10“ the CJEU has ruled that “information relating to legal proceedings brought against an individual” (even if there is no criminal offence shown to be committed in the course of such), also concerning the coverage of such court proceedings, could be included by that (see CJEU 24. 9. 2019, C-136/17, ECLI:EU:C:2019:773).

Apart from that, national supervisory authorities “shall establish and make public a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment” (Article 35(4) GDPR; also called “blacklist”). Moreover, the national supervisory authorities optionally “may also establish and make public a list of the kind of processing operations for which no data protection impact assessment is required” (Article 35(5) GDPR; also called “whitelist”). These lists are therefore to be taken into account respectively as well.³¹

[respective analysis of respective white- and blacklists and relevant factors in the case at hand]

Beyond that, the former Article 29 Working Party has also published guidelines on DPIA and determining whether processing is “likely to result in a high risk”.³² Here, some criteria are given, which principally shall indicate that a DPIA must be conducted when two or more of such are fulfilled (with a higher probability of an obligation the more criteria are fulfilled). However, in certain cases the guidelines assume that already one criterium might suffice to result in the requirement to conduct a DPIA.³³ The criteria concern:

- *Evaluation or scoring*
- *Automated-decision making with legal or similar significant effect*
- *Systematic monitoring*
- *Sensitive data or data of a highly personal nature*
- *Data processed on a large scale*
- *Matching or combining datasets*
- *Data concerning vulnerable data subjects*
- *Innovative use or applying new technological or organisational solutions*
- *When the processing in itself “prevents data subjects from exercising a right or using a service or a contract”*

[respective analysis of relevant factors and criteria in the case at hand]

Lastly it is to be highlighted again, that the obligation to conduct a DPIA pursuant to Article 35 GDPR in the first place is connected with certain other implications and duties pursuant to the GDPR, such as the review obligation pursuant to Article 35(11) GDPR and the consultation procedure pursuant to Article 36 GDPR, and that the threshold analysis should be documented in any case, even when no obligation to conduct a DPIA is found in the course of it (see already section 2.3.2 above).

2.3.2.2 Scope of FRIA pursuant to Article 27 AI Act

The AI Act as well encompasses some criteria that must be fulfilled in order that a FRIA becomes mandatory. The requirements for the applicability of the AI Act per se have already been discussed above (cf. section 2.3.1.1). Beyond that, Article 27 virtually contains personal and material requirements for its applicability. Firstly, the provision only applies to certain high-risk AI Systems, namely such listed in Annex III, with the

³¹ Cf. in total detailed with regard to Austria: *Trieb* in *Knyrim*, *DatKomm Art 35 DSGVO* paras 39 et seq. (status: first September 2019, rdb.at).

³² Article 29 Working Party, 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, WP248 rev.01 (17/EN); see in total detailed (also concerning the endorsement by the EDPB) also *Müller/Poindl/Scheichenbauer*, *Folgenabschätzungs-Methodologie: Datenschutz, Grundrechte, Ethik*, in *Tagungsband der ÖFG-Tagung “KI-VO: Exekutive Rechtsetzung, Standardisierung, Zertifizierung und Grundrechte-Folgenabschätzung” (forthcoming)* section 2.d.

³³ See detailed in total: Article 29 Working Party, WP248 rev.01 (17/EN), pages 9 et seq.

exception of systems “intended to be used in the area listed in point 2 of Annex III” (certain applications as safety components in critical infrastructure).³⁴

The remaining areas and applications of high-risk systems in Annex III that Article 27 AI Act refers to are the following:

“

1. *Biometrics, in so far as their use is permitted under relevant Union or national law:*

(a) *remote biometric identification systems.*

This shall not include AI systems intended to be used for biometric verification the sole purpose of which is to confirm that a specific natural person is the person he or she claims to be;

(b) *AI systems intended to be used for biometric categorisation, according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics;*

(c) *AI systems intended to be used for emotion recognition.*

[...]

3. *Education and vocational training:*

(a) *AI systems intended to be used to determine access or admission or to assign natural persons to educational and vocational training institutions at all levels;*

(b) *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;*

(c) *AI systems intended to be used for the purpose of assessing the appropriate level of education that an individual will receive or will be able to access, in the context of or within educational and vocational training institutions at all levels;*

(d) *AI systems intended to be used for monitoring and detecting prohibited behaviour of students during tests in the context of or within educational and vocational training institutions at all levels.*

4. *Employment, workers’ management and access to self-employment:*

(a) *AI systems intended to be used for the recruitment or selection of natural persons, in particular to place targeted job advertisements, to analyse and filter job applications, and to evaluate candidates;*

(b) *AI systems intended to be used to make decisions affecting terms of work-related relationships, the promotion or termination of work-related contractual relationships, to allocate tasks based on individual behaviour or personal traits or characteristics or to monitor and evaluate the performance and behaviour of persons in such relationships.*

5. *Access to and enjoyment of essential private services and essential public services and benefits:*

(a) *AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for essential public assistance benefits and services, including healthcare services, as well as to grant, reduce, revoke, or reclaim such benefits and services;*

(b) *AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems used for the purpose of detecting financial fraud;*

(c) *AI systems intended to be used for risk assessment and pricing in relation to natural persons in the case of life and health insurance;*

³⁴ Cf. Art. 27(1), 6(2) AI Act; cf. detailed in total, also concerning the differences to the scope of Art. 35 GDPR: *Fülöp/Poindl in Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International forthcoming) section 3.3.

- (d) *AI systems intended to evaluate and classify emergency calls by natural persons or to be used to dispatch, or to establish priority in the dispatching of, emergency first response services, including by police, firefighters and medical aid, as well as of emergency healthcare patient triage systems.*
6. *Law enforcement, in so far as their use is permitted under relevant Union or national law:*
- (a) *AI systems intended to be used by or on behalf of law enforcement authorities, or by Union institutions, bodies, offices or agencies in support of law enforcement authorities or on their behalf to assess the risk of a natural person becoming the victim of criminal offences;*
- (b) *AI systems intended to be used by or on behalf of law enforcement authorities or by Union institutions, bodies, offices or agencies in support of law enforcement authorities as polygraphs or similar tools;*
- (c) *AI systems intended to be used by or on behalf of law enforcement authorities, or by Union institutions, bodies, offices or agencies, in support of law enforcement authorities to evaluate the reliability of evidence in the course of the investigation or prosecution of criminal offences;*
- (d) *AI systems intended to be used by law enforcement authorities or on their behalf or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for assessing the risk of a natural person offending or re-offending not solely on the basis of the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680, or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups;*
- (e) *AI systems intended to be used by or on behalf of law enforcement authorities or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of the detection, investigation or prosecution of criminal offences.*
7. *Migration, asylum and border control management, in so far as their use is permitted under relevant Union or national law:*
- (a) *AI systems intended to be used by or on behalf of competent public authorities or by Union institutions, bodies, offices or agencies as polygraphs or similar tools;*
- (b) *AI systems intended to be used by or on behalf of competent public authorities or by Union institutions, bodies, offices or agencies to assess a risk, including a security risk, a risk of irregular migration, or a health risk, posed by a natural person who intends to enter or who has entered into the territory of a Member State;*
- (c) *AI systems intended to be used by or on behalf of competent public authorities or by Union institutions, bodies, offices or agencies to assist competent public authorities for the examination of applications for asylum, visa or residence permits and for associated complaints with regard to the eligibility of the natural persons applying for a status, including related assessments of the reliability of evidence;*
- (d) *AI systems intended to be used by or on behalf of competent public authorities, or by Union institutions, bodies, offices or agencies, in the context of migration, asylum or border control management, for the purpose of detecting, recognising or identifying natural persons, with the exception of the verification of travel documents.*
8. *Administration of justice and democratic processes:*
- (a) *AI systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution;*

(b) *AI systems intended to be used for influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda. This does not include AI systems to the output of which natural persons are not directly exposed, such as tools used to organise, optimise or structure political campaigns from an administrative or logistical point of view.*

“

[Explanations and analysis in the case at hand/concerning the given context, particularly assessing the different (pertinent) areas of use cases listed in Annex III of the AI Act]

Furthermore, Article 27 AI Act only applies to certain entities. In that regard, the provision on the one hand only addresses *deployers* pursuant to Article 3(4) AI Act (concerning which see in principle already section 2.3.1.1; see in detail furthermore 2.5.2 below). On the other hand, only such deployers are covered that are “*bodies governed by public law, or are private entities providing public services*”, as well as “*deployers of high-risk AI systems referred to in points 5 (b) and (c) of Annex III*” (following Recital 96, the latter essentially concerns “*banking or insurance entities*”).

Beyond Article 27 AI Act per se, also Article 5 AI Act, principally addressing the prohibition of certain (unacceptable) AI Systems³⁵, contains a provision concerning the obligation to conduct a FRIA. In that regard, it is first to be noted that “*the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement*” is only permissible under certain circumstances pursuant to Article 5(1)(h) AI Act. Subsequently, Article 5 paragraphs 2 et seq. AI Act contain further requirements, procedures and specifications for the use of such systems in (principally) permissible use cases, also including the stipulation that the relevant law enforcement authority must complete a FRIA pursuant to Article 27 AI Act before a respective authorisation (see in detail Article 5(2) subparagraph 2 AI Act; and paragraph 3 on exemptions from prior authorisations).³⁶

[Explanations and analysis in the case at hand/concerning the given context; where appropriate with references to other sections, such as 2.5.2 on roles pursuant to the AI Act]

At last, it is to be mentioned that pursuant to Article 27(2) AI Act the obligation to perform a FRIA pursuant to Article 27(1) AI Act principally “*applies to the first use of the high-risk AI system*” (concerning updates see however particularly section 2.10.2 below).³⁷

[Explanations and analysis in the case at hand, where applicable]

2.3.3 Decision on Conducting Assessment(s) and Joint Performance

In light of the evaluations above, it is regarded [not] necessary/mandatory to conduct a FRIA/DPIA.

[concluding statement on respective applicability of provisions, as well as explanations on the decision to (not) conduct a FRIA/DPIA/Ethical IA (particularly also if that would not be mandatory)]

The AI Act bears some interconnections of the DPIA pursuant to Article 35 GDPR and the FRIA pursuant to Article 27 AI Act.³⁸ Of particular relevance here is Article 27(4) AI Act, stipulating that “[i]f any of the obligations laid down in [Article 27 concerning the FRIA] [...] is already met through the data protection

³⁵ See in detail sections 2.3.1.1 and 2.6.2.3.

³⁶ Cf. also *Fülöp/Poindl in Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International forthcoming) section 3.3.1 and 3.4.

³⁷ Cf. also *Fülöp/Poindl in Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International forthcoming) section 3.3.1.

³⁸ For further interconnections of Article 35 GDPR and the AI Act, e.g. concerning the use of instructions for use, see also below (particularly sections 2.4.3 and 2.6.2.2).

impact assessment conducted pursuant to Article 35 of [the GDPR] [...], the fundamental rights impact assessment [...] shall complement that data protection impact assessment". Therefore, it may also be advisable, or even necessary to perform those impact assessments jointly.³⁹

[Explanations and analysis in the case at hand and the given context; analysis as to how far a complementing or even joint performance is necessary (i.e. particularly, if any obligations pursuant to Article 27 AI Act are already met through a respective DPIA; if necessary, with references to sections below) or deemed useful]

2.4 Description of Facts

In this section, the relevant facts for a further assessment of the processes in scope shall be described and depicted. In this context, it is of particular relevance to gather all information necessary and relevant for further legal assessments, such as concerning the distribution of roles (see section 2.5) or compliance with relevant requirements of the GDPR and the AI Act (see section 2.6).

Where applicable, most likely the data processing system in scope comprises several data processing operations that should be distinguished, in particular different processing steps or functions of the system, each of which could have a specific purpose and may have a different legal basis. The description of data processing operations in this section as well as the assessment of GDPR-compliance below is structured according to these different processing operations. Where relevant, this already also concerns section 2.4.2. on the overall processes.

[With regard to section 2.4 on the description of facts it is to be noted that this task might require some flexibility concerning the structuring, layout and arrangement of the elements to be described, particularly considering the aim to make the explanations comprehensible, e.g., for data subjects.]

2.4.1 Architecture

[Description/depiction of the respective system's architecture and, where appropriate, depiction of the relevant processes in scope]

2.4.2 Processes in Scope

[Description of the object to be assessed/relevant processes in a general manner; depending on the circumstances of the case, this might require a wider description than just of the use of respective AI Systems or data processing operations, e.g. where other circumstances or processes are relevant for the understanding or assessing risks of the object of the assessment]

[Description of respective purpose(s) of the **specific** processes (not to be confused with the purpose of an AI System – e.g. pursuant to instructions for use – in a general manner)]

In particular, Article 27(1)(a) AI Act requires "*a description of the deployer's processes in which the high-risk AI system will be used in line with its intended purpose*". In that regard, the instructions for use pursuant to Article 13 AI Act are of particular relevance: In general, deployers of high-risk AI systems have to take "*appropriate technical and organisational measures to ensure they use such systems in accordance with the*

³⁹ Art. 27(4) is formulated very broadly in this regard, because *any obligation* pursuant to Art. 27 would suffice for its applicability, resulting in the *complementation* of respective DPIA; cf. e.g. *measures to be taken in the case of the materialisation of risks* pursuant to Art. 27(1)(f) AI Act and *measures to address risks* pursuant to Art. 35(7)(d) GDPR. Where the provision is applicable, perhaps also an *Addendum* to the DPIA could suffice, but joint performance seems in any case more advisable (e.g., for comprehensiveness and better transparency). Also, there might be cases where Art. 27(4) is not applicable (e.g. considering the different scope of the provisions) but joint performance of a DPIA and a FRIA would still be advisable; in that regard, Art. 27 AI Act does not seem to limit the content of the assessment principally (cf. paragraph 2 on other impact assessments or Recital 96 on other possible contents); see in total principally also *Fülöp/Poindl in Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.3.5.

instructions for use”, and those instructions pursuant to Article 13(3)(b)(i) also have to contain the AI system’s intended purpose (see in detail also section 2.6.2.4).⁴⁰

[Where applicable, respective explanations pursuant to Article 27(1)(a) AI Act]

Also, Article 27(1)(b) AI Act requires “*a description of the period of time within which, and the frequency with which, each high-risk AI system is intended to be used*”.

[Where applicable, respective explanations pursuant to Article 27(1)(b) AI Act]

2.4.3 Data Processing Operations in scope

With regard to the DPIA, and in light of the definitions concerning processing of personal data as outlined above (see section 2.3.1.2), Article 35(7)(a) GDPR requires “*a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller*”. Moreover, Article 26(9) AI Act stipulates that “*deployers of high-risk AI systems shall use the information provided under Article 13 of this Regulation to comply with their obligation to carry out a data protection impact assessment under Article 35 of [...] [the GDPR]*”. As mentioned above, Article 13 AI Act particularly comprises the instructions for use for high-risk AI systems, also including the intended purpose of the AI System (cf. in detail Article 13(2) and (3)). Such information should thus already be considered when describing the object of the assessment (respective data processing operations), which would furthermore also be in line with Article 27(1)(a) AI Act (see section 2.4.2 above).⁴¹

Resulting from the requirements of the GDPR, the processing operations should be described and depicted (see particularly 2.4.1 above) as precisely as possible, also involving data categories, categories of data subjects (see also section 2.7.3 below), location and duration of data storage and processing activities, purposes, and where applicable also recipients and legitimate interests pursuant to Article 6(1)(f) GDPR.⁴²

Consequently, below a respective description shall be given, taking into account those requirements:

[Description of data processing operations in scope (what data is being processed and how?), elaborating on respective personal data pursuant to definitions in relevant data protection legislation (which of the data processed are *personal data*?); particular explanations on processing of special categories of personal data (which of the personal data processed are to be regarded as *sensitive data*?); respective description of the purposes of the envisaged data processing operations (why are those conducted?); other above mentioned aspects to be addressed accordingly; in each case taking into account the information provided under Article 13 AI Act]

2.5 Identification of Stakeholders and Role Distribution

On the basis of the description of facts (see section 2.4 above), relevant stakeholders can be identified and their legal and organisational roles and accountabilities can be clarified pursuant to the pertinent provisions and requirements, particularly as set by the GDPR and the AI Act.⁴³

Furthermore, stakeholders, as a result of their role in the context at hand, can be regarded particularly as rights-holders or duty-bearers. While (specific and immediate) rights-holders are to be addressed in

⁴⁰ Cf. in total *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International forthcoming) section 3.3.4.1.

⁴¹ Cf. also *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International forthcoming) sections 3.3.4.1 and 3.3.5.

⁴² Cf. in total *Trieb* in *Knyrim* (ed.), *DatKomm Art.* 35 paras 108 et seq.

⁴³ Cf. *Vemou/Kadyra*, *Evaluating privacy impact assessment methods: guidelines and best practices* (2020) 28(1) *Information & Computer Security*, 35–53.

connection with the risk assessment in section 2.7.3 below, their views on the object of the assessment, as well as those of relevant representatives are already to be addressed in this section, as well as the identification of entities bearing respective obligations and other relevant stakeholders in the context at hand.

To systematically identify and classify relevant stakeholders, it should particularly be assessed to what extent the activities in question concern internal (e.g. data controllers, data processors, data protection officers, recipients; high-risk AI deployers, providers etc.) and external stakeholders (e.g. data subjects, third parties from the public and private sector representing affected persons), identifying the type and level of their involvement or concernment.⁴⁴ Based on that, these entities can then be addressed in the right place of the assessment.

2.5.1 Role Distribution pursuant to GDPR

The GDPR defines three fundamental roles in the context of processing of personal data, the data subject, the controller and the processor. Most importantly, the controller of the data processing has to be specified. The term controller, as defined by Art. 4(7) GDPR, is “*the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data*”. Article 4(8) GDPR defines the processor as “*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*”. The data subject is the natural person to which the information which is processed relates to. This definition is contained in the definition of the term personal data (see in detail above in section 2.3.1.2.1). Every data processing operation needs to be analysed in a separate manner regarding the question which stakeholders are involved in which role. There has to be one controller and there can be more than one (“*joint controllers*”). It is upon the controller to decide whether to involve one or more processors or none. In relation to a specific processing of personal data one entity can only fulfil one role at one point in time.

[respective explanations in the case at hand]

2.5.2 Role Distribution pursuant to AI Act

As was shown above in principle (2.3.1.1), the AI Act entails several roles in connection with AI systems, whereby the most important in the context at hand are those of the *provider* and the *deployer* of an AI system (defined in Article 3(3) and (4) AI Act). Both roles can be assumed by either “*a natural or legal person, public authority, agency or other body*”.

Subsequently, the essential characteristics that make up a **deployer** are (Article 3(4) AI Act):

- The **use** of an AI system
- **under the authority** of that relevant entity
- with the **exception** of the usage “*in the course of a personal non-professional activity*”

Whereas, the **provider** is the respective entity (Article 3(3) AI Act):

- **developing** an AI system [or a general-purpose AI model]
 - **or that has** an AI system [or a general-purpose AI model] **developed**
- and **placing** the AI system **on the market** under its **own name or trademark**
 - **or putting it into service** under its **own name or trademark**
- whether for payment or free of charge

⁴⁴ Cf. Kloza/Calvi/Casiraghi/Vazquez Maymir/Ioannidis/Tanas/Van Dijk, Data protection impact assessment in the European Union: developing a template for a report from the assessment process, Brussels Laboratory for Data Protection & Privacy Impact Assessments, Policy Brief 1/2020, VUB 2020, <https://doi.org/10.31228/osf.io/7qrfp>.

In that regard, “*placing on the market*” refers to “*the first making available of an AI system or a general-purpose AI model on the Union market*” (Article 3(9) AI Act), and “*putting into service*” to “*the supply of an AI system for first use directly to the deployer or for own use in the Union for its intended purpose*” (Article 3(11) AI Act).

These two concepts of the provider and the deployer therefore principally refer to somewhat different areas of operation, with the provider essentially developing (or having developed) and providing (or putting into service) an AI System and the deployer subsequently using it⁴⁵, but might also coincide (in contrast to the roles pursuant to the GDPR, see section 2.5.1 above). In that regard furthermore Art. 25 AI Act is to be mentioned, referring to situations in which (inter alia) a deployer would have to be considered as a provider of a high-risk AI system.⁴⁶ Recital 84 AI Act states some clarifications in that context:

- “*To ensure legal certainty, it is necessary to clarify that, under certain specific conditions, any distributor, importer, deployer or other third-party should be considered to be a provider of a high-risk AI system and therefore assume all the relevant obligations.*”
- “*This would be the case if that party puts its name or trademark on a high-risk AI system already placed on the market or put into service, without prejudice to contractual arrangements stipulating that the obligations are allocated otherwise.*”
- “*This would also be the case if that party makes a substantial modification to a high-risk AI system that has already been placed on the market or has already been put into service in a way that it remains a high-risk AI system in accordance with this Regulation, or if it modifies the intended purpose of an AI system, including a general-purpose AI system, which has not been classified as high-risk and has already been placed on the market or put into service, in a way that the AI system becomes a high-risk AI system in accordance with this Regulation.*”

Moreover, it is to be noted that the roles pursuant to the AI Act must not necessarily correlate with certain roles pursuant to the GDPR, as identified in section 2.5.1, presumably allowing different constellations. For instance, in case an AI system is used by an entity (deciding which particular system is used) in the course of that entity providing a service to another entity that essentially determines purposes and means of processing of personal data in the course of that service (cf. 2.5.1), the latter entity will then act as a *controller* pursuant to Article 4(7) GDPR. However, resulting from the fact that the *processor* pursuant to the GDPR is allowed some margin of manoeuvre concerning decision-making without becoming a *controller*,⁴⁷ it could be argued that the first entity is acting as the *deployer* pursuant to Article 3(4) AI Act, if its actions were still to be regarded as the use of the AI system “*under its authority*” (and not that of the data controller), while it may act as a *processor* pursuant to Article 4(8) GDPR, when processing personal data. In detail, this would therefore depend on the precise interpretation of the passage “*under its authority*” pursuant to Article 3(4) AI Act in comparison to *determining the purposes and means of the processing* pursuant to Article 4(7) GDPR, whereby the AI Act however indicates that there might indeed be respective differences⁴⁸.

⁴⁵ To the roles cf. in principle also: *Wendt/Wendt*, Das neue Recht der Künstlichen Intelligenz – Artificial Intelligence Act (AI Act) (2024) 52.

⁴⁶ See also *Wendt/Wendt*, AI Act, 52.

⁴⁷ Such as concerning the rather practical implementation of the processing and respective means, like regarding the selection of hard- or software or security measures: see in detail *Bogendorfer* in *Knyrim* (ed.), *DatKomm Art 28 DSGVO* para. 6 (status of 1 December 2022, rdb.at).

⁴⁸ Cf. in this regard particularly Recital 10 AI Act, which seems to acknowledge that deployers of AI Systems could be both data controllers as well as processors pursuant to the GDPR; also, respective provisions of the AI Act, particularly the obligations of deployers pursuant to Art. 26, indicate that the deployer has to be very close to the respective AI system and its use. For instance, deployers must, via appropriate technical and organisational measures, ensure that „*they use such systems in accordance with the instructions for use*“ (Art. 26(1)), furthermore assign human oversight to natural persons with certain characteristics (Art. 26(2)), and, where applicable, have to conduct the registration of the use of the systems pursuant to Art. 49 AI Act (see Art. 26(8)). Therefore,

[Explanations particularly on *deployer* and *provider* pursuant to AI Act in the case at hand; also, concerning other actors where appropriate]

[where appropriate, deeper analysis of Article 27(1) AI Act in light of addressed deployers (cf. 2.3.3.2)]

2.5.3 Views of Data Subjects or their Representatives (Article 35(9) GDPR)

Article 35(9) GDPR stipulates that “[w]here appropriate, the controller shall seek the views of data subjects or their representatives on the intended processing, without prejudice to the protection of commercial or public interests or the security of processing operations.” Unfortunately, the term “where appropriate” leaves open the circumstances under which the opinion must be obtained and when this can be omitted. From the wording it can be concluded that an unconditional obligation for controllers to obtain an opinion cannot be assumed.⁴⁹ However, it must in any way be documented and justified how this provision was dealt with in the particular case.⁵⁰

As an alternative to individual data subjects, their “representatives” can also be consulted, in particular various interest groups, workers’ councils or consumer protection associations. The position of these bodies should be taken into account in particular if the intended data processing involves a large number of data subjects whose interests are represented by the respective association or body.⁵¹

[respective analysis and explanations concerning the case at hand]

2.5.4 Involvement of the Data Protection Officers

Pursuant to Article 35(2) GDPR, the controller shall seek the advice of the data protection officer, where designated, when carrying out a DPIA. Whether seeking the advice of the data protection officer is mandatory and to what extent the advice obtained from the data protection officer must be followed is disputed in the literature: *Trieb*, for example, assumes that the GDPR does not impose such an obligation;⁵² *Jandt*, on the other hand, sees an obligation in the provision, but points out that the provision makes no statement as to whether the advice of the data protection officer must also be followed and does not provide for a right of veto or similar for the data protection officer.⁵³ However, if the controller does not agree with the advice (or parts thereof) obtained from the data protection officer, according to the Article 29 Working Party, a (reasonable) justification for the lack of compliance with the advice should be provided in the DPIA report.

[respective analysis and explanations concerning the case at hand]

2.5.5 Involvement of Other Relevant Stakeholders

Beyond the above-mentioned tendentially obligatory involvement of persons concerned/stakeholders, respective considerations are also valuable in a voluntary approach. In general, stakeholder-involvement is often regarded a central part of impact assessments⁵⁴ and can also support/extend collective aspects and

arguably situations might occur in which a respective data controller might not even be as close to the use of such systems to ensure respective compliance, but where respective actions and decisions (in the sense of „authority“) could rather fall to a contractor (arguably still acting as a data processor, cf. explanations above); cf. in total principally also *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*), section 3.3.5.

⁴⁹ Cf. *Trieb* in *Knyrim* (ed.), *DatKomm Art. 35 para. 131*.

⁵⁰ Cf. *Jandt*, in *Kühling/Buchner* (eds.), *DS-GVO/BDSG Art 35 para. 58*.

⁵¹ Cf. *Trieb* in *Knyrim* (ed.), *DatKomm Art. 35 para. 134*; Cf. in that regard also *Martin/Friedewald/Schierung/Mester/Hallinan/Jensen*, *Datenschutz-Folgenabschätzung nach Art 35 DSGVO*, *Fraunhofer-Institut für System- und Innovationsforschung, Karlsruhe* (2020) 38 et seq.

⁵² Cf. *Trieb* in *Knyrim* (ed.), *DatKomm Art. 35 para. 124*.

⁵³ Cf. *Jandt* in *Kühling/Buchner* (eds.), *DS-GVO/BDSG Art. 35 para. 18*.

⁵⁴ In relation to FRIA see *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*), section 3.3.2 with further references.

transparency of such, as well as the understanding of risks⁵⁵. Furthermore, Article 27 AI Act in its original draft (Article 29a in the proposal of the EP) also included the (mandatory) involvement of certain stakeholders. That part was deleted in the course of negotiations and a respective modified text passage is now only to be found in Recital 96 AI Act: ⁵⁶

“Where appropriate, to collect relevant information necessary to perform the impact assessment, deployers of high-risk AI system, in particular when AI systems are used in the public sector, could involve relevant stakeholders, including the representatives of groups of persons likely to be affected by the AI system, independent experts, and civil society organisations in conducting such impact assessments and designing measures to be taken in the case of materialisation of the risks.”

Therefore, considering the deletion from the text of Article 27 and the use of the term “could”, such stakeholder-involvement presumably is rather to be conducted voluntarily.⁵⁷ However, as is implied by recital 96 AI Act, such involvement could in any case be useful “to collect relevant information necessary” for the performance of the impact assessment.⁵⁸ Nevertheless, it should also be considered on a case-by-case-basis which form and manner of involving and consulting relevant parties is suitable in order to collect the respective information. Possible modes for respective engagement could, inter alia, be interviews, surveys/questionnaires and workshops.⁵⁹

[respective explanations in the case at hand concerning the involvement of relevant parties and respective considerations]

2.5.6 Role Distribution pursuant to Ethical Considerations

Specific to the ethical approach is to not have a clear distinction of controller and processor as in the GDPR or deployer and provider as in the AI Act. This allows to consider different roles that might overlap or roles that are not taken into account in the GDPR or AI Act. Ethical considerations are meant to bring all relevant persons and groups to the table and to balance their interests. While some persons and groups are able to take actions in view of the technology in question, others might not be able to do so to a full extent or not at all. Part of the EIA is to apply a concept of responsibility and to analyse to what extent actions towards others can be justified. Therefore, it makes sense to look at a broader scope of affected persons and groups as outlined in section 2.8.2 and of responsible persons in section 2.8.3.

⁵⁵ Furthermore, involving certain stakeholders can also provide advantages compared to involving affected persons directly, cf. in total: Mantelero, Computer Law & Security Review 34 (2018) 754 (758, 769-770) <https://doi.org/10.1016/j.clsr.2018.05.017>.

⁵⁶ See in detail Fülöp/Poindl in Pehlivan/Forgó/Valcke (eds.), Artificial Intelligence Act: A Commentary (Kluwer Law International forthcoming) sections 2.2 and 3.3.2.

⁵⁷ Cf. in principle (in comparison to the DPIA) also Müller/Poindl/Scheichenbauer, in Tagungsband der ÖFG-Tagung “KI-VO: Exekutive Rechtsetzung, Standardisierung, Zertifizierung und Grundrechte-Folgenabschätzung” (forthcoming) section 3.b.

⁵⁸ See also Fülöp/Poindl in Pehlivan/Forgó/Valcke (eds.), Artificial Intelligence Act: A Commentary (Kluwer Law International forthcoming) sections 3.3.2. and 3.1.2.

⁵⁹ Cf. e.g. in context of Human Rights Impact Assessments, also for further details concerning implementation: *The Danish Institute for Human Rights*, Crosscutting: Stakeholder Engagement – Human Rights Impact Assessment Guidance and Toolbox https://www.humanrights.dk/sites/humanrights.dk/files/media/document/HRIA%20Toolbox_Stakeholder%20Engagement_ENG_2_020.pdf (accessed 29. 8. 2024); see in context of the DPIA (and data subjects and their representatives) e.g. Trieb in Knyrim (ed.), *DatKomm*, Art. 35 DSGVO paras 130 et seq., citing further literature, and also section 2.5.3 above; in context of EIA and respective methods see Wright, a Framework for the Ethical Impact Assessment of Information Technology, Ethics and Information Technology (2011) 13/3, 199 (215 et seq.); see also Reijers/Brey/Jansen/Rodrigues/Koivisto/Tuominen, *Satori – A Common Framework for Ethical Impact Assessment – Annex 1 – A reasoned proposal for a set of shared ethical values, principles and approaches for ethics assessment in the European context – Deliverable D4.1 (2016) 11 et seq.*; UNESCO, *Ethical Impact Assessment – A Tool of the Recommendation on the Ethics of Artificial Intelligence*, <https://doi.org/10.54678/YTSA7796> 15-16 (accessed 25. 9. 2024); concerning AI in particular see also Kaminski/Maglieri, *Impacted Stakeholder Participation in AI and Data Governance*, Forthcoming on Yale Journal of Law and Technology (2024-2025), U of Colorado Law Legal Studies Research Paper No. 24-23, <https://ssrn.com/abstract=4836460> (accessed 25. 9. 2024).

2.6 Legal Admissibility

2.6.1 GDPR Compliance

Article 35(7) GDPR expressly mentions several minimum requirements which a DPIA shall contain, while the GDPR as a whole contains a lot more provisions which must be complied with in the course of the processing of personal data. It can be concluded from Article 5(2) and Article 24(1) GDPR that the compliance with all of these extensive provisions must be documented in writing. It is therefore advisable that, going beyond the requirements of Article 35, the DPIA report contains all this documentation regarding the data processing operations in scope because they have to exist anyway and it is in several ways practical that they can be found all in one document. Therefore, the following sections contain more aspects of documentation of the compliance with provisions of the GDPR than it would be strictly required by Article 35.

2.6.1.1 Lawfulness of Processing of Personal Data

The processing of personal data is subject to a prohibition with reservation of permission. Article 6(1) GDPR and Article 9(2) GDPR exhaustively and conclusively list the possible legal permissions for the processing of personal data. There is no hierarchical relationship between these permissive clauses but each of them is assigned an equal status.⁶⁰ Article 35(7) (a) GDPR explicitly requires that the DPIA has to document the legitimate interest pursued by the controller in case the controller bases the processing of personal data on such legitimate interest (Article 6(1) (f) GDPR).

2.6.1.1.1 Purposes of the Processing of Personal Data

The purpose of processing of personal data is the central component of the regulatory approach of the GDPR. In particular, the assessment of lawfulness and proportionality of the processing centre around the processing purpose. This is why Article 5(1) (b) requires the specification of the purposes prior to the processing⁶¹ and Article 35(7) (a) GDPR requires that the DPIA contains a systematic description of the purposes of the processing.

[It is advised that the purposes of the individual processing operations within the scope are already specified in the course of the description of these processing operations in section 2.4.]

2.6.1.1.2 Legal Bases for Processing of Personal Data

[respective explanations on legal bases of Articles 6(1) GDPR and 9(2) GDPR for processing of personal data within the scope]

2.6.1.1.3 Necessity and Proportionality of the Processing Operations

[respective explanations on processing of personal data within the scope in relation to the purposes pursuant to Article 35(7)(b) GDPR]

2.6.1.2 Transfer of Personal Data to Third Countries or IOs

[respective explanations on transfers pursuant to Chapter V GDPR, and particularly permissibility of such]

2.6.1.3 Measures to Ensure Compliance with the GDPR Principles

While specific measures for data protection and GDPR compliance are deduced and documented in the course of the risk assessment below, it is advisable to document the fundamental and general measures of

⁶⁰ Kastelitz/Hötzendorfer/Tschohl in Knyrim (ed.), DatKomm Art. 6 para. 14.

⁶¹ Hötendorfer/Tschohl/Kastelitz in Knyrim (ed.), DatKomm Art. 5 para. 22.

the system for compliance with the different data protection principles of Article 5(1) GDPR already in this section in a structured way along the different principles and where appropriate also along the different data protection operations identified within the scope of the Impact Assessment.

[respective explanations concerning the case at hand]

2.6.1.4 Measures to Safeguard the Rights of the Data Subject

In this section, the measures and processes for the compliance with the rights of the data subject of Chapter III of the GDPR are documented. The GDPR provisions regarding automated decisions (Article 22 GDPR) and the respective information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject (Article 13(2)(f), Article 14(2)(g) and Article 15(1)(h) GDPR) are covered separately in section 2.6.3 below together with the related Article 86 AIA.

[respective explanations concerning the case at hand]

2.6.2 Measures to Ensure Compliance with the AI Act (where relevant)

In this section, specifically relevant obligations and requirements set by the AI Act – particularly in light of Article 27 AI Act – shall be addressed. This shall in particular refer to obligations and requirements pertinent in view of the protection of fundamental rights.

[respective explanations in respective subsection; where appropriate (particularly when fundamental rights implications occur), also analysis of other relevant duties of deployers/requirements, as well as respective measures for compliance]

2.6.2.1 Implementation of Human Oversight Measures in Line with Instructions for Use

Human Oversight is a requirement of the AI Act for high-risk AI systems, which is particularly addressed in detail in Article 14 AI Act. Generally, Article 14(3) AI Act differentiates between two types of human oversight measures: “*measures identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service*” (point a) and “*measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the deployer*” (point b). The specific implementation of human oversight falls to the deployer, who specifically “*shall assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support*” (Article 26(2) AI Act) and in general also “*shall take appropriate technical and organisational measures to ensure they use such systems in accordance with the instructions for use accompanying the systems [...]*” (Article 26(1) AI Act; see also 2.6.2.4). Thereby, the instructions for use pursuant to Article 13 AI Act inter alia also comprise “*the human oversight measures referred to in Article 14, including the technical measures put in place to facilitate the interpretation of the outputs of the high-risk AI systems by the deployer*” (Article 13(3)(d) AI Act).

Subsequently in this regard, Article 27(1)(e) AI Act also requires “*a description of the implementation of human oversight measures, according to the instructions for use*” in the FRIA. Concerning the systematics of this stipulation see also 2.7.1.2 below. Here, it is only to be remarked – following the explanations above – that human oversight measures constitute an explicit requirement set by the AI Act independently of the FRIA. Subsequently, these measures are not necessarily only related to specific fundamental rights risks (cf. in that regard in term 2.7.1.2), but also to be considered in a more general manner (cf. e.g. Article 14(2) and (3) AI Act), which justifies a separate discussion besides such in connection with the specific risk assessment.⁶²

⁶² See in total principally Fülöp/Poindl in Pehlivan/Forgó/Valcke (eds.), Artificial Intelligence Act: A Commentary (Kluwer Law International forthcoming) section 3.3.4.5.

[specific explanations concerning the implementation in the case at hand]

2.6.2.2 Registration pursuant to Article 49 and Notification pursuant to Article 27(3)

In relation to certain deployers of certain high-risk AI Systems, particularly Article 49(3) and (4) stipulate requirements for the registration of the use of such systems in the EU database pursuant to Article 71 AI Act. In that regard, “*deployers that are public authorities, Union institutions, bodies, offices or agencies or persons acting on their behalf shall register themselves, select the system and register its use in the EU database [...] “[b]efore putting into service or using a high-risk AI system listed in Annex III, with the exception of high-risk AI systems listed in point 2 of Annex III” (Article 49(3) AI Act). Pursuant to Article 26(8) AI Act, the pertinent deployers must comply with those registration obligations, and furthermore shall not use such systems and inform respective providers or distributors, “[w]hen such deployers find that the high-risk AI system that they envisage using has not been registered in the EU database referred to in Article 71”.*

Consequently, not all deployers addressed by Article 27(1) AI Act are also required to register respective systems (e.g. excluding banking and insurance providers as addressed by Article 27(1) AI Act). Moreover, Article 49(4) AI Act restricts registration concerning “*high-risk AI systems referred to in points 1, 6 and 7 of Annex III, in the areas of law enforcement, migration, asylum and border control management*”, on the one hand regarding the content of the registration (see below), and on the other concerning public access to it, as “*the registration [...] shall be in a secure non-public section of the EU database referred to in Article 71*”. The content of the registration pursuant to Article 49(3) is principally stipulated in Annex VIII Section C, inter alia comprising “[a] *summary of the findings of the fundamental rights impact assessment conducted in accordance with Article 27*” (point 4) and “[a] *summary of the data protection impact assessment carried out in accordance with Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680 as specified in Article 26(8) of this Regulation, where applicable*” (point 5). However, the content of the registration regarding the use of systems referred to in Article 49(4) AI Act is limited only to points 1 to 3 of Section C of Annex VIII, thus excluding the aforementioned points 4 and 5 concerning FRIA and DPIA.⁶³ In regard to the summary of findings, see further section 2.10.1 below.

Ultimately, it is to be noted in that context that another prerequisite for the authorisation of “*the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement*” besides the completion of a FRIA pursuant to Article 27 AI Act principally is that the law enforcement authority “*has registered the system in the EU database according to Article 49*” (Article 5(2) AI Act; see 2.3.2.2 above). This obligation is however subject to certain restrictions (see particularly Article 5(2) subparagraph 2 and Article 5(3) AI Act).

Besides the registration obligation, respective deployers “*shall notify the market surveillance authority of its results, submitting the filled-out template referred to in paragraph 5 of this Article as part of the notification*” once they have performed a FRIA pursuant to Article 27 AI Act. Under certain circumstances pursuant to Article 46(1) AI Act, deployers of specific high-risk systems may however be excluded from this obligation (Article 27(3) AI Act). At this point, it can only be assumed that *results* of a FRIA do not refer to the documentation of every detail of the process of the FRIA, but only to its (essential) outcome (see also section 2.10.1). Furthermore, Article 27(5) in that regard refers to “*a template for a questionnaire, including through an automated tool, to facilitate deployers in complying with their obligations under this Article in a simplified manner*” that is to be developed by the AI Office.⁶⁴

⁶³ Cf. in total *Fülöp/Poindl in Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International forthcoming) section 3.3.2.

⁶⁴ The (English) wording of paragraph 3 might indicate that submitting the filled-out template alone may not be enough to fulfil the requirements of the provision (arg. „*as part of the notification*“); the German version, on the other hand using the term “*indem*” corresponding to “*by [submitting]*”, seems to be clearer about the filled-out template being the essential content of the submission;

[respective explanations concerning the case at hand on the implementation of those AI Act requirements, where appropriate]

2.6.2.3 Demarcation from Prohibited AI Practices pursuant to Article 5 AI Act

Certain AI Systems and use cases are prohibited pursuant to Article 5 AI Act due to unacceptable risks.⁶⁵ These essentially concern (cf. Article 5(1) AI Act; highlights added):

- *placing on the market, putting into service or the use of:*
 - *an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm; **(point a)***
 - *an AI system that exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm; **(point b)***
 - *AI systems for the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to either or both of the following **(point c)**:*
 - (i) *detrimental or unfavourable treatment of certain natural persons or groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected;*
 - (ii) *detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or its gravity;*
- *Placing on the market, putting into service for this specific purpose, or the use of:*
 - *an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics;⁶⁶ **(point d)***
 - *AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage; **(point e)***
 - *AI systems to infer emotions of a natural person in the areas of workplace and education institutions;⁶⁷ **(point f)***

cf. in total principally also Fülöp/Poindl in Pehlivan/Forgó/Valcke (eds.), Artificial Intelligence Act: A Commentary (Kluwer Law International forthcoming) section 3.3.2.

⁶⁵ Cf. e.g. Recitals 26, 31, 179; European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM(2021) 206 final, 2021/0106(COD), 5.2.2.

⁶⁶ „[...] this prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity“ (Art. 5(1)(d) AI Act).

⁶⁷ „[...] except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons“ (Art. 5(1)(f) AI Act).

- biometric categorisation systems that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation;⁶⁸ (point g)
- And, **principally**, the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement (point h).

Using 'real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement pursuant to point (h) may exceptionally be permissible provided that and "in so far as such use is strictly necessary for one of the following objectives":

- 1) "[T]he targeted search for specific victims of abduction, trafficking in human beings or sexual exploitation of human beings, as well as the search for missing persons" (Art. 5(1)(h)(i) AI Act)
- 2) "[T]he prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack" (Art. 5(1)(h)(ii) AI Act)
- 3) "[T]he localisation or identification of a person suspected of having committed a criminal offence, for the purpose of conducting a criminal investigation or prosecution or executing a criminal penalty for offences referred to in Annex II and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years." (Art. 5(1)(h)(iii) AI Act)

Moreover, Article 5 paragraphs 2 et seq. also contain several further requirements, specifications and procedures for the use of such systems in (principally) permissible use cases, inter alia including the requirement for the authorisation of such AI use that the relevant law enforcement authority has completed a FRIA pursuant to Article 27 AI Act (Art. 5(2) subparagraph 2; see already section 2.3.2.2).

In the case at hand the following measures, procedures, safeguards and considerations have been applied/taken into account to demarcate the processes in scope from AI practices prohibited pursuant to Article 5 AI Act:

[respective explanations concerning the case at hand, particularly on implementation of measures to comply with respective AI Act requirements]

2.6.2.4 Usage pursuant to Instructions for Use

As already mentioned above, Article 26(1) AI Act requires deployers of high-risk AI systems to "take appropriate technical and organisational measures to ensure they use such systems in accordance with the instructions for use accompanying the systems, pursuant to paragraphs 3 and 6 [of Article 26]". The instructions for use pursuant to Article 13 AI Act to a certain extent depict the specifications of usage foreseen by the provider, including the intended purpose of the AI system (Article 13(3)(b)(i) AI Act), and also involving "any known or foreseeable circumstance, related to the use of the high-risk AI system in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, which may lead to risks to the health and safety or fundamental rights", which have also been dealt with in the course of risk management pursuant to Article 9(2) AI Act (Article 13(3)(b)(iii) AI Act). Accordingly, the respective specifications of usage have been demarcated by the provider, particularly also through risk management pursuant to Article 9 AI Act, whereby the deployer shall subsequently not deviate from these specifications, which is evident from several passages of the AI Act, e.g. the above cited Article 26(1).⁶⁹

⁶⁸ „[...] this prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorizing of biometric data in the area of law enforcement" (Art. 5(1)(g) AI Act).

⁶⁹ Cf. in total detailed also Fülöp/Poindl in Pehlivan/Forgó/Valcke (eds.), Artificial Intelligence Act: A Commentary (Kluwer Law International forthcoming) sections 3.3.3 and 3.3.4.1.

[respective explanations concerning the case at hand, particularly on implementation of measures to adhere to instructions for use]

2.6.2.5 Other Considerations

In light of the specific case, also other requirements and obligations pursuant to the AI Act, such as post-market monitoring pursuant to Articles 26(5) and 72 AI Act should be taken into account (cf. introduction in 2.6.2 above).

[respective explanations, particularly on implementation of those AI Act requirements]

2.6.3 Automated Decisions (Article 22 GDPR and Article 86 AI Act)

Article 22 GDPR regulates the permissibility of automated decisions in individual cases, including profiling, as follows:

“Automated individual decision-making, including profiling

1. *The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.*

2. *Paragraph 1 shall not apply if the decision:*

(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

(b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests;

or

(c) is based on the data subject’s explicit consent.

3. *In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.*

4. *Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests are in place.”*

In addition, with respect to automated decisions, Article 13(2)(f) and Article 14(2)(g) GDPR stipulate the obligation of the controller to provide information regarding *“the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject”*, and Article 15(1) (h) GDPR stipulates a right of the data subject to obtain such information.

A related provision of the is Article 86 AI Act, which reads:

“Right to explanation of individual decision-making

1. *Any affected person subject to a decision which is taken by the deployer on the basis of the output from a high-risk AI system listed in Annex III, with the exception of systems listed under point 2 thereof, and which produces legal effects or similarly significantly affects that person in a way that they consider*

to have an adverse impact on their health, safety or fundamental rights shall have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken.

2. Paragraph 1 shall not apply to the use of AI systems for which exceptions from, or restrictions to, the obligation under that paragraph follow from Union or national law in compliance with Union law.

3. This Article shall apply only to the extent that the right referred to in paragraph 1 is not otherwise provided for under Union law.”

These provisions and their interplay are analysed in the following sections.

2.6.3.1 Permissibility of automated decisions (Article 22 GDPR)

The provision of Article 22 GDPR is structured in such a way that paragraph 1 establishes a prohibition, paragraph 2 contains exceptions to this prohibition and paragraph 4 contains counter-exceptions, which in turn lead to the applicability of the prohibition. Paragraph 3 specifies certain legal consequences for cases in which the prohibition does not apply due to certain exceptions and automated decision-making is therefore permitted.

Article 22 GDPR does not subject every automated individual decision to a legal consequence per se. An automated individual decision is only covered by Article 22 GDPR if it entails legal effects concerning the data subject or similarly significantly affects the data subject.

Therefore, the following three elements must be fulfilled for Article 22 GDPR to be applicable:

- There is a decision in the sense of Article 22 GDPR.
- The decision is based solely on automated processing.
- The decision has legal or other significant effects on the data subject.

2.6.3.1.1 Presence of an Automated Decision

Recital 71 GDPR gives the following typical examples of such decisions: “*automatic refusal of an online credit application*” or “*e-recruiting practices without any human intervention*”. These are processes that are traditionally decided without automated decision-making by a human being in the form of a more or less structured decision-making process involving several decision-making factors. This also applies to medical procedures such as anamnesis and diagnosis which traditionally aren’t automated. Data subjects whose data is used to train a machine learning model are not per se subject to an automated decision. However, persons whose personal data is processed in the course of the model’s application (inference) may very well be subject to an automated decision, such as whether to undergo a medical treatment or further examination.

According to *Haidinger*, the term “*decision*” in Article 22 GDPR is to be understood broadly⁷⁰ and includes “*measures*” according to Recital 71 GDPR. From Recital 71 GDPR it could be concluded that Article 22 GDPR is only applicable on a decision which involves the assessment of personal aspects of the data subject and would have to be teleologically reduced in this regard. This interpretation is conceivable,⁷¹ but by no means mandatory, in particular because it is not reflected in the wording of Article 22 GDPR. At least equally plausible is the interpretation that Recital 71 mentions the classic scenario that Article 22 GDPR is intended to regulate, which includes the assessment of personal aspects of the data subject, without wanting to reduce

⁷⁰ *Haidinger* in *Knyrim* (ed.), *DatKomm*, Art. 22 para. 18.

⁷¹ *Buchner* in *Kühling/Buchner* (eds.), *DS-GVO/BDSG*, Art 2 para. 19.

Article 22 GDPR to this scenario. This is supported by the fact that Article 22 GDPR still contained this element explicitly in earlier drafts – as did Article 15 GDPR – but this restriction of the provision has been removed in the final version.⁷²

[respective analysis regarding the application on the individual case]

2.6.3.1.2 Is the Decision Based Solely on Automated Processing?

This question can only be answered in the context of a specific system or specific data processing operation. From the *SCHUFA* judgement of the ECJ it becomes clear that the concept of a decision based solely on automated processing within the meaning of Article 22(1) GDPR also includes decisions of which one significant decision factor, basis or step of the decision was generated by automated processing.⁷³ This interpretation can be easily brought in line with the wording (“*solely*”) if the individual step, factor or decision basis generated by automated means is itself regarded as the decision within the meaning of Article 22(1) GDPR.

[respective analysis regarding the application on the individual case]

2.6.3.1.3 Is there a Decision with Legal or Other Significant Effect?

This element requires that a decision based solely on automated processing affects the rights of a natural person. It may also affect the legal status of the person or his or her rights under a contract.

Even if a decision-making process does not affect the rights of an individual, it may still fall within the scope of Article 22 GDPR if it has such an effect or significantly affects the individual in a similar way. In other words, even if the data subject's rights or obligations do not change, they may be sufficiently affected to require the protection of this provision.

[respective analysis regarding the application on the individual case]

2.6.3.1.4 Exemptions

Article 22(2) GDPR specifies the exemptions cited above. According to Article 22(3) GDPR, in the cases referred to in paragraph 2(a) and (c), the controller must “*implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.*” Article 22(4) GDPR standardises a restriction for automated decisions based on special categories of personal data within the meaning of Article 9(1) GDPR: Decisions based on such data are only permitted if

- a. the data subject has expressly consented to this processing of special categories of personal data (Art 9(2)(a) GDPR) or
- b. such “*processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject*” (Art 9(2)(g) GDPR).

[respective analysis regarding the application on the individual case]

⁷² Buchner in *Kühling/Buchner* (eds.), DS-GVO/BDSG, Art 2 para. 17.

⁷³ ECJ 7. 12. 2023, C-634/21, *SCHUFA Holding AG*, ECLI:EU:C:2023:957, para. 40 et seq.

2.6.3.2 Transparency of Automated Decisions

2.6.3.2.1 Right to Information about Automated Decision-Making (GDPR)

Article 13 GDPR stipulates an obligation of the controller to provide a range of information to the data subject when personal data are collected from the data subject, and Article 14 GDPR stipulates such obligation for the processing of data which have not been obtained from the data subject. Pursuant to Article 15 GDPR, data subjects have the right to obtain from the controller confirmation as to whether or not personal data concerning them are being processed. If this is the case, they have a right of access to this personal data and to a range of information. If the processing comprises automated decision-making, including profiling, as referred to in Article 22 GDPR, all three aforementioned provisions stipulate that in their respective context, three elements of information must be provided, the fact that automated decision-making is conducted, meaningful information about the logic involved, and meaningful information about the significance and the envisaged consequences of such processing for the data subject.

2.6.3.2.2 Right to Explanation of Individual Decision-Making (AI Act)

Article 86 AI Act stipulates that a person has the right to obtain from the deployer of a high-risk AI system “*clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken*” when a decision was taken on the basis of the output of such system which produces legal effects or similarly significantly affects that person “*in a way that they consider to have an adverse impact on their health, safety or fundamental rights*”. This is the only right in the AI Act that an affected person can exercise directly towards the operator.

That the provision refers to a decision with is taken “*on the basis of the output*” is a less ambiguous wording that is closer to practice than that of Article 22 GDPR, which refers to a decision “*based solely on automated processing*”. However, with reference to the analysis of Article 22 GDPR above it can be concluded that the interpretation of both is likely to be very similar, so that the scope of application of both provisions is already triggered if the output of the system is a significant decision factor, which is itself not questioned any more by a human actor.⁷⁴

The applicability of Article 86 AI Act is limited by several restrictions. In particular, as can be seen in detail in the full quote of the provision above, it does not apply to all high-risk AI systems. Also, the definition of the consequences the decision has to have, limits its applicability compared to the respective provisions of Articles 13 to 15 GDPR in conjunction with Article 22 GDPR: When a decision does not have a legal effect, Article 86 AI Act narrows down the relevant spheres of impact on a person to adverse effects on health, safety or fundamental rights. Moreover, Article 86 AI Act does not stipulate a prohibition but a subjective right, which is – in contrast to Article 22 GDPR – largely seen in the literature as an ex-post right, meaning that potentially affected persons cannot obtain information about the decision-making in a general manner before a decision is taken but only after the fact.⁷⁵

The restriction of applicability laid down in Article 86(3) AI Act, which stipulates the subsidiarity of the provision, directly concerns the relationship of Article 86 AI Act and Article 22 GDPR, among other provisions. If the information is already provided under Article 22 GDPR or another provision, Article 86 AI Act does not apply. It must be noted, however, that the wording in paragraph 3 refers to “*the right referred to in paragraph 1*”, which can be interpreted to the effect that Article 86 is only completely superseded if the effect of paragraph 1 is entirely provided for by another provision, while it can be questioned that, e.g. Article

⁷⁴ Cf. *Hornung*, Individualrechte in der KI-Verordnung. Die Rechte auf Beschwerde und auf Erläuterung der Entscheidungsfindung im Einzelfall, DuD 2024/8, 507 (510); ECJ 7. 12. 2023, C-634/21, *SCHUFA Holding AG*, ECLI:EU:C:2023:957.

⁷⁵ Cf. *Radtke*, Das Verhältnis von KI-VO und Art. 22 DS-GVO unter besonderer Berücksichtigung der Schutzzwecke, RD 2024, 353 (358).

15(1)(h) GDPR entail the explanation of the “*role of the AI system in the decision-making procedure and the main elements of the decision*” pursuant to Article 86(1) AI Act. Therefore, in doubt it is recommended that in case the scope of applicability of both Article 22 GDPR and Article 86 AI Act is triggered, the content of the information shall fulfil both the respective provision of Articles 13 to 15 GDPR as well as Article 86 AI Act.

2.6.3.2.3 Content of the information obligation

The question, which information needs to be provided, leads to several aspects which are contested in the literature. What seems to be clear is that the obligations of Article 13(2)(f) and Article 14(2)(g) GDPR require the controller to provide ex-ante information before the decision is taken.⁷⁶ This clearly follows from the context and purpose of Articles 13 and 14 GDPR as a whole. In contrast, as mentioned above, Article 86 GDPR is largely seen as an ex-post right to information regarding a specific decision the right-holder was subject to. Whether Article 15(1)(h) GDPR adds an ex-post dimension to the ex-ante rights of Article 13(2)(f) and Article 14(2)(g) is contested in the literature.⁷⁷ This is indicated by its context as part of the right of access by the data subject, which can be exercised before or after such decision is taken, but the fact that the wording of the three provisions is exactly the same and refers to “*envisaged consequences*” speaks to the contrary.

Both regimes, the respective provisions of Articles 13 to 15 GDPR as well as Article 86 AI Act raise the question, whether the obliged entity has itself sufficient information about the functioning of an AI system in order to fulfil these obligations. In many cases, machine learning models function more or less as a black box, meaning that neither the provider nor the deployer of the system can explain their functioning in detail.⁷⁸ In particular, for machine learning models based on deep learning it is usually impossible to draw meaningful causal links between the input data and the output. This creates an obvious tension with the requirement of the respective provisions of Articles 13 to 15 GDPR described above, to provide the data subject “*meaningful information about the logic involved*”. Notably, the wording of Art. 86 AI Act does not explicitly contain an equivalent element referring to the functional dimension, but only requires explanations “*of the role of the AI system in the decision-making procedure and the main elements of the decision taken*”.

However, to what extent such tension actually exists, i.e. how the term “*meaningful information about the logic involved*” of the GDPR must be interpreted, is another aspect which is unclear. The Austrian Data Protection Authority and the Austrian Federal Administrative Court seem to interpret Article 15(1)(h) GDPR in a way that the information must explain which input parameters lead to the decision, whereby the Austrian Federal Administrative Court highlights that the “*algorithm itself*” is not to be explained, only the principle on which such a calculation is based, not the specific “*calculation formula*”.⁷⁹ It must be noted that on the one hand, these terms would not properly apply in the domain of machine learning, and on the other hand, as mentioned above, even the principle of which input leads to which output can be unknown. To conclude, it must be highlighted that the transparency provisions of the GDPR described in this section potentially cannot be completely fulfilled when using some particularly untransparent machine learning models within the scope of applicability of Art. 22(1) GDPR, whereby some important details remain to be clarified by the

⁷⁶ Cf. *Wachter/Mittelstadt/Floridi*, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, *International Data Privacy Law*, Volume 7, Issue 2, May 2017, 76–99, <https://doi.org/10.1093/idpl/ix005>.

⁷⁷ Cf. *Kaminski*, The Right to Explanation, Explained. *Berkeley Technology Law Journal* 34 (2019), 189, <https://doi.org/10.15779/Z38TD9N83H>; *Wachter/Mittelstadt/Floridi*, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, *International Data Privacy Law*, Volume 7, Issue 2, May 2017, 76–99, <https://doi.org/10.1093/idpl/ix005>.

⁷⁸ Cf. *Panigutti et al.*, The role of explainable AI in the context of the AI Act, *FACCT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1139–1150, <https://doi.org/10.1145/3593013.3594069>.

⁷⁹ DSB (Austria) 8. 9. 2020, 2020-0.436.002; BVwG (Austria) 23. 4. 2024, W292 2248672-1.

ECJ. If such a potential infringement of Articles 13, 14 or 15 amounts to an infringement of the transparency principle of Art. 5(1)(a) the whole processing activity would be rendered unlawful.⁸⁰

2.6.3.3 Conclusion regarding Automated Decision-Making

[respective analysis regarding the application of Article 22 and the respective provisions of Articles 13 to 15 GDPR and of Article 86 AI Act on the individual case]

2.6.4 General Fundamental Rights Considerations

First of all, it is to be noted that concerning the description of those rights that are **deemed as essential that they are granted in form of constitutional laws** the terms may vary from state to state, but nevertheless such an intention becomes evident by a respective analysis.⁸¹ In Austria, e.g. *human rights* are often deemed to describe those that are granted to all human beings irrespective of their nationality (even if not enforceable) – in opposition to *civilian rights* – while the term *fundamental rights* is particularly used as a synonym for the legally more relevant term of constitutionally guaranteed rights (*“verfassungsrechtlich gewährleistete Rechte”*).⁸² Also, on the one hand the ECHR refers to *human rights*, and the CFR, on the other hand refers to *fundamental rights*, while bearing a deep connection to the ECHR content-wise (cf. in detail 2.7.2 below). Therefore, the terms *fundamental rights* and *human rights* are to some degree to be regarded similar/to be used synonymous in the course of this document, while of course in detail there may be differences attributed to those terms, which have to be taken into account respectively.

For the analysis of fundamental/human rights and respective risks, some theoretical foundations are important to consider first.

While various fundamental rights exist that can also be classified in various ways,⁸³ it is important to internalise that initially fundamental rights primarily aimed at state actors, requiring them to refrain from interfering with those rights, which was subsequently developed further, consequently also requiring positive actions of states.⁸⁴

In particular, states are principally obliged to *respect* fundamental/human rights by refraining from interferences, except such interferences are foreseen by respective provisions and eventually meet certain criteria (see below). Furthermore, states must, via their duty to *protect*, also take action to safeguard rights holders against third parties violating their fundamental/human rights. Also, states must, via their *duty to fulfil*, create – through certain measures – circumstances in which it is possible to exercise fundamental/human rights as well as possible, such as by taking respective measures concerning the administration or courts.⁸⁵ Meanwhile, it e.g. is also common sense in Austria that the state is also bound by fundamental rights when acting in the private sector/commercially.⁸⁶

⁸⁰ Cf. Roßnagel/Richter in Spiecker gen Döhmman/Papakonstantinou/Hornung/De Hert (eds.), GDPR Art. 5 para 1.

⁸¹ See in detail on the understanding and genesis of human rights Nowak, Human Rights from a Legal Perspective, in Nowak/Januszewski/Hofstätter (eds.), All Human Rights for All – Vienna Manual on Human Rights (2012).

⁸² Cf. Kucsko-Stadlmayer, Allgemeine Strukturen der Grundrechte, in Marten/Papier/Kucsko-Stadlmayer, Handbuch der Grundrechte VII/1² (2014) § 3 para. 3 with further references; Heißl, Einführung – Grundlagen, in Heißl (ed.), Handbuch Menschenrechte (2009) para. 1/5 with further references.

⁸³ Cf. e.g. in detail (with reference in particular to Austria) Kucsko-Stadlmayer in Marten/Papier/Kucsko-Stadlmayer, Handbuch der Grundrechte VII/1² § 3 paras 16 et seq.; furthermore, the rights and their classification in the CFR.

⁸⁴ See e.g. Heißl in Heißl (ed.), Handbuch Menschenrechte, paras 1/7 et seq.

⁸⁵ Cf. in detail Nowak, Introduction to Human Rights Theory, in Nowak/Januszewski/Hofstätter (eds.), All Human Rights for All – Vienna Manual on Human Rights (2012) 270 et seq.; Kucsko-Stadlmayer in Marten/Papier/Kucsko-Stadlmayer, Handbuch der Grundrechte VII/1² § 3 paras 50 et seq., concerning the different state functions also paras 33 et seq.

⁸⁶ See Schulev-Steindl, Drittwirkung und Fiskalgeltung, in Marten/Papier/Kucsko-Stadlmayer, Handbuch der Grundrechte VII/1² (2014) § 6 para. 55.

Considering the fact that only a very limited number of fundamental/human rights is guaranteed in an absolute manner, meaning that they do not allow for interferences at all, and that the aforementioned fundamental/human rights obligations of states regarding positive actions hinge on certain circumstances of the case, it is essential to balance respective rights/interests. Concerning the achievement of a respectively fair balance of rights/interests, the principle of proportionality is crucial.⁸⁷

In this regard, it is therefore first to be examined whether a respective action or omission also constitutes an interference in the scope protection of the respective fundamental/human right.⁸⁸

Interferences in the respective scope of protection of fundamental/human rights principally require a law as a basis (while some rights are however absolute and do not allow for any interferences). Furthermore, they must fulfil certain criteria in order to be lawful, in particular serve a corresponding (legitimate) public interest and stand up to a proportionality test, which means in detail that the interference has to be suitable (even capable) and necessary to achieve its aim, as well as appropriate (proportional in a narrower sense), i.e. fairly balanced with the competing interests.⁸⁹ Article 52 CFR in this regard literally demands on the one hand that “[a]ny limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms” and on the other that “[s]ubject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others”.

In contrast to the comprehensive obligations of states as described above, it remains to some degree questionable to what extent also private actors are (directly) bound by fundamental/human rights,⁹⁰ whereby e.g. in Austria it is largely assumed that a binding of private actors principally only exists in an indirect manner, by way of interpretation of (other, non-constitutional, general) legal provisions⁹¹. However, certain rights and (non-constitutional) legislative acts also expressly aim at respective (direct) obligations *inter privatos* (such as concerning data protection)⁹², which are eventually to be analysed. At this point, also the GDPR and the AI Act shall be mentioned as a secondary EU law stipulating certain obligations of both public and private actors with regard to fundamental rights (cf. also the respective explanations in this document so far). With respect to the FRIA pursuant to Art. 27 AI Act, it is in this regard also to be mentioned with Recital 96 AI Act that this provision particularly also aims at “[s]ervices important for individuals that are of public nature” that “may also be provided by private entities” and that “[p]rivate entities providing such public services are linked to tasks in the public interest such as in the areas of education, healthcare, social services, housing, administration of justice”. It therefore specifically addresses such private entities/services with a certain link to public interests.

⁸⁷ Cf. Nowak in Nowak/Januszewski/Hofstätter (eds.), All Human Rights for All – Vienna Manual on Human Rights, 274 and 275; Kucsko-Stadlmayer in Marten/Papier/Kucsko-Stadlmayer, Handbuch der Grundrechte VII/1² § 3 paras 65 et seq.

⁸⁸ Cf. in detail Kucsko-Stadlmayer in Marten/Papier/Kucsko-Stadlmayer, Handbuch der Grundrechte VII/1² § 3 paras 79 et seq.; Nowak in Nowak/Januszewski/Hofstätter (eds.), All Human Rights for All – Vienna Manual on Human Rights 275.

⁸⁹ Cf. in detail Kucsko-Stadlmayer in Marten/Papier/Kucsko-Stadlmayer, Handbuch der Grundrechte VII/1² § 3 paras 65 et seq., 91-105; with regard to the EU/CFR/ECHR: Winkler, Grundrechte in der EU, in Heißl (ed.), Handbuch Menschenrechte (2009) paras 3/55 et seq.; Nowak in Nowak/Januszewski/Hofstätter (eds.), All Human Rights for All – Vienna Manual on Human Rights 274 et seq.; furthermore, section 2.9 below.

⁹⁰ See in detail e.g. Schulev-Steindl, in Marten/Papier/Kucsko-Stadlmayer, Handbuch der Grundrechte VII/1² § 6 paras 7 et seq.; with regard to the EU/CFR cf. also Winkler in Heißl (ed.), Handbuch Menschenrechte, para. 3/46; cf. also Nowak in Nowak/Januszewski/Hofstätter (eds.), All Human Rights for All – Vienna Manual on Human Rights 272 et seq.

⁹¹ See in detail Schulev-Steindl, in Marten/Papier/Kucsko-Stadlmayer, Handbuch der Grundrechte VII/1² § 6 paras 7 et seq., 19 et seq.

⁹² See in detail, particularly also concerning the differentiation of real third-party-effects *inter privatos* of fundamental rights and mere non-constitutional stipulations concerning the respective application of fundamental rights: Schulev-Steindl, in Marten/Papier/Kucsko-Stadlmayer, Handbuch der Grundrechte VII/1² § 6 paras 9 et seq.

Subsequently, the processes in scope of the object of the assessment shall be analysed with regard to fundamental/human rights in a general manner, in light of the explanations above.

[General considerations on legitimacy and proportionality of actions, particularly in case of public actors (as deployers)]

2.7 Fundamental Rights Risk Assessment

2.7.1 Methodology

The risk assessment is to some extent to be regarded as the core of the Combined Impact Assessment with respect to fundamental rights.⁹³ In this regard, for the DPIA, at first Article 35(7)(c) GDPR literally requires “an assessment of the risks to the rights and freedoms of data subjects referred to in [Article 35(1)]”, Article 27(1)(d) AI Act on the other hand addressing “the specific risks of harm likely to have an impact on the categories of natural persons or groups of persons identified pursuant to point (c) of this paragraph, taking into account the information given by the provider pursuant to Article 13” for the FRIA. Subsequently, also respective measures are required by the provisions: Article 35(7)(d) GDPR demands addressing “the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned”, while Article 27(1)(f) AI Act generally stipulates addressing “the measures to be taken in the case of the materialisation of those risks, including the arrangements for internal governance and complaint mechanisms”. Also to be mentioned again in this context is Article 27(1)(e) AI Act, requiring addressing “a description of the implementation of human oversight measures, according to the instructions for use” (see already section 2.6.2.1 above).

Principally, it is to be noted that certain indications for respective risk assessments can be drawn from both the GDPR and the AI Act (addressed subsequently), which should be taken into account respectively. However, it presumably may also be derived from the stipulation in Article 27(4) AI Act concerning the complementation of pertinent DPIA by respective FRIA (see already 2.3.3), that the law makers considered these assessments to be fundamentally similar, respectively, did not consider those to have fundamental differences (but only concerning specific aspects).⁹⁴

2.7.1.1 Identification and Analysis of Risks

Regarding the understanding of the concept of risks and of how to analyse them it can be deduced from Recitals 75 and 94 GDPR that a risk to the fundamental rights and freedoms of natural persons is conceptualised as the possibility of an event occurring, which itself represents damage or can lead to further damage to one or more natural persons.

The AI Act on the other hand refers to risk as “the combination of the probability of an occurrence of harm and the severity of that harm” (Article 3(2)), and in particular addresses “risks of harm” in Article 27.

⁹³ Cf. in regard to the DPIA in particular e.g. *Trieb in Knyrim* (ed.), *DatKomm Art 35 para. 113* citing further literature; from the perspective of the FRIA pursuant to Art. 27 AI Act cf. the (obligatory) elements listed in Art. 27(1) AI Act culminating in the respective analysis of risks and respective (counter) measures, while Recital 96 AI Act claims that “[t]he aim of the fundamental rights impact assessment is for the deployer to identify the specific risks to the rights of individuals or groups of individuals likely to be affected, identify measures to be taken in the case of a materialisation of those risks”.

⁹⁴ Cf. in principle in detail also *Fülöp/Poindl in Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International forthcoming) section 3.3.5.

It is to be noted that impact assessments in this regard should particularly relate to harm in the sense of unintentional adverse effects for individuals and groups caused by the operation of an AI system.⁹⁵

In general, the risk assessment is about an estimation and classification of the probability and severity of risks. To estimate the severity of risk scenarios several factors must be considered. These include among others:⁹⁶

- The number of people affected
- The characteristics of the impacted groups
- The geographical and demographical reach
- The extent of adverse effects and their reversibility
- The likelihood of exacerbating existing biases, stereotypes, discrimination and inequalities
- Possible cumulative impacts
- The effort required to minimise the risk (e.g. time spent amending information, extra costs, low or sufficient capacity to remediate the impact, long-term psychological or physical ailments, etc.)

With regard to the FRIA pursuant to Article 27 AI Act, the risk assessment must furthermore take “*into account the information given by the provider pursuant to Article 13*”, and should therefore also consider respective particularities of the AI system, such as its capabilities and limitations.⁹⁷ Moreover, as opposed to the risk-management system pursuant to Article 9 AI Act referring to *known and reasonably foreseeable risks*, the FRIA refers to “*the specific risks of harm **likely to have an impact** on the categories of natural persons or groups of persons [...]*” (Article 27(1)(d) AI Act, highlights added).

While the provision of Article 27 AI Act per se does not provide particular details in that regard, also concerning the FRIA, distinct steps of the impact assessment, also comprising risk identification, analysis and mitigation, can be deduced from literature.⁹⁸

As Recitals 84 and 90 GDPR point out, the risk assessment pursuant to the GDPR should consider the origin and nature of a risk as well as the scope, context and purposes of the respective data processing. The origin and nature of the risks can be distinguished by the following criteria:⁹⁹

- Internal/external human source or internal/external non-human source: e.g. internal or external employee, software error or hardware defect, environmental impact (natural forces), cybercriminal (attacker/malware), state institutions (intelligence service, law enforcement), management.
- Intentional, negligent, or unintentional: e.g. the damage to the affected person can be either condoned or intended and the goal of action, or due to individual or structural errors. In other words, a risk can emerge either from the intended functioning of the system, from an unintended malfunctioning or from an intended attack.

⁹⁵ See in total *Fülöp/Poindl* in in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International forthcoming) section 3.3.4.4.

⁹⁶ Cf. The Danish Institute for Human Rights, *Guidance on human rights impact assessment of digital activities*, 2020, <https://www.humanrights.dk/publications/human-rights-impact-assessment-digital-activities>, 22 et seq.

⁹⁷ Cf. in detail *Fülöp/Poindl* in in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International forthcoming) section 3.3.4.4 and Articles 13 and 27(1)(d) AI Act.

⁹⁸ Cf. *Fülöp/Poindl* in in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International forthcoming) sections 3.1.1 and 3.1.2.

⁹⁹ Cf. Bitkom, *Risk Assessment & Datenschutz-Folgenabschätzung – Leitfaden*, Bitkom e.V. 2017, <https://www.bitkom.org/sites/default/files/file/import/FirstSpirit-1496129138918170529-LF-Risk-Assessment-online.pdf>, 27.

From the perspective of the GDPR, a distinction can be made between physical, material and non-material types of damage (Recital 75 GDPR). Typical risk causes include unauthorised or unlawful processing, processing contrary to good faith, processing that is not transparent for the data subjects, unauthorised disclosure of and access to data, unintentional loss, destruction, or damage of data, denial of data subjects' rights, use of data by controllers for illegitimate purposes, not intended processing of data, processing of inaccurate data, incorrect processing (technical failures, human errors), processing beyond the retention period, processing itself when the harm lies in the performance of the pro-cessing (e.g. because it is illegitimate/lacks a legal basis) and processing contrary to the purpose limitation principle (Recitals 75 and 83 GDPR).

In context of fundamental rights per se, it is assumed that risks do not necessarily have to be of a material nature, which is also reflected by Recital 5 AI Act stating “AI may generate risks and cause harm to public interests and fundamental rights that are protected by Union law. Such harm might be material or immaterial, including physical, psychological, societal or economic harm”.¹⁰⁰ Furthermore, Recital 77 AI Act implies that risks to fundamental rights in context of high-risk AI systems might e.g. be connected to cybersecurity.

For the most comprehensive assessment of (potentially) all risks to all fundamental rights, at first a preliminary assessment concerning fundamental rights is to be conducted (see in detail section 2.7.2 below). The results of that preliminary assessment should then together with the analysis of affected categories of persons and groups serve as the primary basis for the subsequent risk analysis (section 2.7.5).

For the identification of AI-specific risks, apart from the preliminary fundamental rights assessment referred to in section 2.7.2, particularly also AI risk repositories or similar assessment tools may serve as a valuable tool. In this regard, e.g. the following can be mentioned:

- *Slattery/Saeri/Grundy/Graham/Noetel/Uuk/Dao/Pour/Casper/Thompson*, The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence, https://cdn.prod.website-files.com/669550d38372f33552d2516e/66bc918b580467717e194940_The%20AI%20Risk%20Repository_13_8_2024.pdf
- *Rosenthal*, Generative AI Risk Assessment (GAIRA) Tool, <https://www.rosenthal.ch>:
 - “Light” or “comprehensive”: https://www.rosenthal.ch/downloads/Rosenthal_GAIRA.xlsx
 - “One-pager”: https://www.rosenthal.ch/downloads/VISCHER_AI-Risk-Check.pdf
- *High-Level Expert Group on AI*, Assessment List for Trustworthy Artificial Intelligence (ALTAI), <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal> (introduction/instructions/links)

However, it is also to be emphasised that respective experts and/or professional literature might be required in order to properly conduct the fundamental rights risk assessment.¹⁰¹

¹⁰⁰ See in detail *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), Artificial Intelligence Act: A Commentary (Kluwer Law International forthcoming) section 3.3.4.4 with further references.

¹⁰¹ Cf. e.g. concerning the composition of an impact assessment team: *Janssen/Lee/Singh*, Practical Fundamental Rights Impact Assessments, International Journal of Law and Information Technology, 2022, 30/2, 200 <https://doi.org/10.1093/ijlit/eaac018> (accessed 26. 9. 2024); *The Danish Institute for Human Rights*, Phase 1: Planing and Scoping – Guidance on HRIA of Digital Activities, 2020, https://www.humanrights.dk/sites/humanrights.dk/files/media/document/Phase%201_Planing%20and%20Scoping_ENG_accessible.pdf (accessed 26. 9. 2024); concerning a multidisciplinary approach in the team see also: *Ministry of the Interior and Kingdom Relations* (Government of the Netherlands), <https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms>

2.7.1.2 Assessment of likelihood and severity

The risk assessment is meant to be as objective as possible (Recitals 75 and 76 GDPR). This is, however, not always attainable in practice due to ambiguities about assignable likelihoods, possible types of damage, and the subjective perceptions of risk by the various stakeholders.

According to the GDPR, risks should be assessed by combining the likelihood (probability or chance) of occurrence (or happening) and the severity (or magnitude) of consequences (Recital 76 GDPR). As mentioned above, the AI Act in a similar manner defines risk as “*the combination of the probability of an occurrence of harm and the severity of that harm*” (Article 3(2) AI Act). Therefore, the risk evaluation is an estimation of the likelihood of a threat scenario materialising and an estimation of the impact severity of each risk scenario.¹⁰² These two variables are combined in the combinatorial matrix using an ordinal scaled system to estimate the overall risk level. The decision which risk level an expected impact corresponds to, lies with the evaluation of experts who possess knowledge of the system and processes in question, of case law, literature, and the relevant legal framework.

The exact methodology for assessing the impact varies from author to author. However, most models work with locating the respective likelihood and severity on a risk matrix, using an ordinal scale like (1) negligible, (2) limited, (3) substantial and (4) maximum.

In order to systematically determine the **probability**, the following statements can be attributed to each of the risk levels:¹⁰³

Negligible	It is very improbable that the damage/harm occurs and seems practically impossible for the selected risk sources to materialise the threat under the given circumstances (e.g. theft of paper documents stored in a room protected by a badge reader and access code, given the disproportionate effort necessary for an attacker).
Limited	It is rather improbable that the damage/harm occurs; it seems difficult for the selected risk sources to materialise the threat under the given circumstances (e.g. theft of paper documents stored in a room protected by a badge reader).
Substantial	It is rather probable that the damage/harm occurs; it seems possible for the selected risk sources to materialise the threat under the given circumstances (e.g. theft of paper documents stored in offices that cannot be accessed without first checking in at the reception).
Maximum	It is very probable or even inevitable that the damage/harm occurs; it seems easy for the selected risk sources to materialise the threat under the given circumstances (e.g. theft of paper documents stored in the public lobby).

¹⁰² Cf. *Vemou/Karyda*, Evaluating privacy impact assessment methods: guidelines and best practices, Information & Computer Security 28(1) (2020), 35–53, 48.

¹⁰³ Based on See *Mantelero*, Beyond Data – Human Rights, Ethical and Social Impact Assessment, Asser Press/Springer, Information Technology and Law Series, IT&LAW 36, 2022, 56.

The same goes for determining the level of **severity**:¹⁰⁴

Negligible	The severity of the damage/harm is very low; affected individuals and groups may encounter a few inconveniences, which they will overcome without any problem (e.g. time spent amending information, annoyances, irritations, etc.).
Limited	The severity of the damage/harm is rather low; affected individuals and groups may encounter inconveniences, which they will be able to overcome despite a few difficulties (e.g. extra costs, fear, lack of understanding, stress, minor physical ailments, etc.).
Substantial	The damage/harm is rather severe; affected individuals and groups may encounter consequences, which they should be able to overcome albeit with real and serious difficulties (e.g. economic loss, property damage, worsening of health, etc.).
Maximum	The damage/harm is very severe; affected individuals and groups may encounter serious or even irreversible consequences, which they may not overcome (e.g. long-term psychological or physical ailments, death, etc.).

The subsequent combination of these two variables results in the following risk matrix:

Risk assessment		Probability of occurrence			
		<i>negligible</i> (1)	<i>limited</i> (2)	<i>substantial</i> (3)	<i>maximum</i> (4)
Severity of damage/harm	<i>maximum</i> (4)	normal (4)	high (8)	high (12)	high (16)
	<i>substantial</i> (3)	normal (3)	normal (6)	high (9)	high (12)
	<i>limited</i> (2)	very low (2)	normal (4)	normal (6)	high (8)
	<i>negligible</i> (1)	very low (1)	very low (2)	normal (3)	normal (4)

¹⁰⁴ Commission Nationale de l'Informatique et des Libertés (CNIL), Privacy Impact Assessment (PIA): Knowledge Base, CNIL 2018, 4, <https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-3-en-knowledgebases.pdf> (accessed 26. 9. 2024).

With regard to AI, some indications concerning the assessment of risk levels in connection with fundamental rights in general may be drawn from the classifications of the AI Act per se, e.g. the specifications concerning high-risk AI in Articles 6 and 7 and Recitals 52, 53 and 58, the specifications concerning prohibited AI practises (related to unacceptable risks; see 2.6.2.3) in Article 5 and Recitals 26 and 28 et seq. Particularly, it is to be noted that the AI Act apparently considers AI systems principally to be “*high-risk*” in case “*in light of their intended purpose, they pose a high risk of harm to the health and safety or the fundamental rights of persons, taking into account both the severity of the possible harm and its probability of occurrence*”.¹⁰⁵

In accordance with Article 35(1) and Article 36(1) GDPR, the risk acceptance level is principally (see below) defined as normal or below. The point in the risk matrix where this level is exceeded cannot directly be deduced from the GDPR. For the purpose of the present methodology this is defined by the following principle: if the probability of occurrence of the maximum level of damage is not negligible then the resulting risk level must be considered as high. Consequently, with respect to the level of residual risks after the assessment and risk treatment, it is to be noted in the context of the DPIA pursuant to Article 35 GDPR that risks are particularly to be reduced to a level below *high*.¹⁰⁶ Risk mitigation measures must be, however, considered regarding all risks, not only regarding high risks. According to the risk assessment prescribed by Articles 24, 25 and 32 GDPR, depending on the available technical and/or organisational measures to mitigate a particular risk among other factors, risks with a level of normal, low, or very low can be considered to be acceptable. Risk scenarios located in the range of high or very high of the matrix, on the other hand, always require further risk treatment until the respective risks are sufficiently contained. If high risks would nonetheless remain, either the data protection supervisory authority must be consulted (see Article 36 GDPR) or the processing must not be carried out at all.

In context of the FRIA pursuant to Article 27 AI Act, at first it only seems to be required explicitly to address measures to be taken in case risks actually occur (“*measures to be taken in the case of the materialisation of those risks*”; Article 27(1)(f); highlights added). However, in the course of the interpretation of the provision and its purpose, as well as of impact assessments in general, this aspect should in our view (at the least as a precaution) be interpreted broadly, and in any case also involve measures aiming at the prevention of harm.¹⁰⁷

As regards specific measures, following Recital 96 AI Act these are to be selected in light of the **specific** use case/risks identified, whereby the variants for measures mentioned in Article 27(1)(f) appear to constitute minimum requirements rather than (as implied by Recital 96) mere examples, which concerns **arrangements for internal governance** and **complaint mechanisms**.¹⁰⁸ Furthermore, the **implementation of human oversight measures** (principally already addressed in section 2.6.2.1) is to be mentioned again in that context because Recital 96 AI Act also refers to such as example for/aspect of governance arrangements as example

¹⁰⁵ Cf. Recital 52 AI Act and *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.3.4.4.

¹⁰⁶ Cf. otherwise the applicability of the consultation obligation pursuant to Art. 36 GDPR (sections 2.9, 2.10.2): cf. e.g. *Trieb* in *Knyrim* (ed.), *DatKomm Art 35 para. 117*.

¹⁰⁷ See in detail *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.3.4.6.

¹⁰⁸ Cf. in detail *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.3.4.6; see also Recital 96 AI Act explicitly: “*In light of the risks identified, deployers should determine measures to be taken in the case of a materialisation of those risks, including for example governance arrangements in that specific context of use, such as arrangements for human oversight according to the instructions of use or, complaint handling and redress procedures, as they could be instrumental in mitigating risks to fundamental rights in concrete use-cases*”; whereas Art. 27(1)(f) AI Act explicitly: “*the measures to be taken in the case of the materialisation of those risks, including the arrangements for internal governance and complaint mechanisms*”.

for mitigation measures concerning risks.¹⁰⁹ This categorisation also seems to be supported by the systematics of Article 27(1) AI Act referring to the *implementation of human oversight measures* in paragraph 1(e) in-between the risk assessment pursuant to paragraph 1(d) and the (other) mitigation measures pursuant to paragraph 1(f). Therefore, (the implementation of) such measures should also be considered with respect to particular fundamental rights risks (cf. again 2.6.2.1).

Concerning the risk level to be achieved (particularly by the implementation of mitigation measures), Article 27 AI Act does not contain any clear specifications. In that context, it first is to be emphasised again that the AI systems addressed by the provision are principally already regarded to be **high-risk** per definition, which may also have some implications for respective risk mitigation processes. However, the AI Act in general requires that risks must not reach the threshold of Article 79(1) concerning risks that exceed what is “*considered reasonable and acceptable in relation to its intended purpose or under the normal or reasonably foreseeable conditions of use*” (cf. Article 3(19) Regulation (EU) 2019/1020¹¹⁰), which would particularly lead to the deployer having to suspend the use of the AI System and to fulfil information obligations pursuant to Article 26(5) AI Act. Beyond that, also considering general fundamental rights principles (see section 2.6.4 above), measures should particularly be selected in light of **proportionality** (see in particular section 2.9 below) and it should as well be ensured that the level of risk principally ensured by (the risk management of) the provider is not thwarted.¹¹¹

Finally, the risks identified and analysed in this section 2.7 are to be addressed from an overall perspective on the residual risk level in section 2.9.1 below, and subsequently to be subjected to a proportionality analysis, which should be conducted after taking into account all relevant risks as well as benefits (particularly also as described in section 2.8) of the object of the assessment, but nevertheless in light of the requirements described above as specifically as possible with respect to the specific risks (and benefits) an respective mitigation measures, where appropriate.

2.7.2 Preliminary Fundamental Rights Assessment

In order to conduct a comprehensive risk assessment, covering as many relevant risks to fundamental rights as possible, the pertinent rights should be identified and assessed preliminarily in light of the case at hand first. This approach shall ensure from the start that all relevant aspects and situations are shown from which possible risks may arise that have to be assessed consequently.¹¹²

While the AI Act essentially refers to the CFR as source of law for fundamental rights¹¹³, it is advisable to also analyse other sources for fundamental/human rights, such as in particular the ECHR (with a strong influence

¹⁰⁹ See *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.3.4.5.

¹¹⁰ Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011, OJ L 25. 6. 2019, 169, 1.

¹¹¹ See in total in detail *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.3.4.6.

¹¹² Cf. e.g. in this regard the application of a fundamental rights check-list by the EC in connection with Commission impact assessments to alleviate the understanding of the fundamental rights methodology and the performance of an initial screening of fundamental rights at the beginning of respective processes: *European Commission*, Commission Staff Working Paper – Operational Guidance on taking account of Fundamental Rights in Commission Impact Assessments, SEC(2011) 567 final; furthermore particularly the “fundamental rights schedule” (also serving “*a clear overview of the affected fundamental rights and their consequences*”) in: *Ministry of the Interior and Kingdom Relations* (Government of the Netherlands), Impact Assessment – Fundamental Rights and Algorithms (2022), <https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms> (accessed 26. 9. 2024; particularly pages 81 et seq.); cf. principally also, particularly concerning fundamental rights clusters: *Janssen/Lee/Singh*, *International Journal of Law and Information Technology*, 2022, 30/2, 200 (209-210) <https://doi.org/10.1093/ijlit/eaac018>.

¹¹³ Cf. e.g. Recitals 1, 6, 28, 48, 59, 176 and Art. 1(1).

on the CFR¹¹⁴) or relevant national laws on fundamental rights in the respective context, where appropriate.¹¹⁵

In this first step of the assessment, it should be examined how the processes in the case at hand, constituting the object of the assessment, might touch upon the scope of protection of the respective fundamental rights¹¹⁶, which should then be analysed further in the following sections, as outlined below, particularly in light of potential risks and infringements. Furthermore, in this step it should be examined from the perspective of the requirements for DPIAs pursuant to Article 35 GDPR in a similar manner, but particularly from a data protection point of view, which (other) “rights and freedoms” might be at risk.¹¹⁷

[respective explanations on preliminary assessment of relevant fundamental rights and their respective scope of protection and of other rights and freedoms in light of the case at hand and its impacts]

2.7.3 Affected Categories of Natural Persons and Groups

In this section, *categories of natural persons and groups likely to be affected* by the object of the assessment *in the specific context* shall be addressed. This particularly constitutes a requirement of the FRIA pursuant to the AI Act concerning the use of respective high-risk AI systems (Article 27(1)(c)). Given its context, this provision in essence relates to fundamental rights holders.¹¹⁸ Therefore, at first particularly the results of section 2.7.2. are to be consulted. Nevertheless, for the sake of working comprehensively and thoroughly, affected parties should not only be identified from the general perspective of (the scope of) fundamental rights, but additionally from a rather functional perspective of the particular use case (cf. also “[...] *in the specific context*” in Article 27(1)(c)),¹¹⁹ and in the course of that should particularly also involve the views of persons that apply respective AI. Furthermore, it is to be noted that also Article 27 AI Act presumably only aims at individuals and groups (*likely to be affected*) “that are located in the Union” (cf. Article 2(1)(g) AI Act; section 2.3.1.1 above). However, also considering the broad approach of the Combined Impact Assessment Methodology at hand in general and the application of the various relevant fundamental/human rights sources (section 2.7.2), it may be advisable to also involve a respectively wider circle of affected persons and groups, where appropriate.

Apart from that the DPIA, particularly the “*systematic description of the envisaged processing operations*” pursuant to Article 35(7)(a) GDPR to some extent requires addressing *data subjects*, which should therefore

¹¹⁴ See e.g. Art. 52(3) CFR.

¹¹⁵ Cf. in this regard again also *European Commission*, SEC(2011) 567 final on “*the ECHR and the constitutional traditions common to the Member States*” being “*a source from which the European Court of Justice deduces fundamental rights as general principles of the Union’s law (see Article 6(3) TEU)*”; cf. principally in detail also *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.2.

¹¹⁶ Cf. in this regard generally section 2.6.4 above and in light of impacts principally again also *European Commission*, SEC(2011) 567 final (sections “*What is the ‘Fundamental Rights Check-List’?*”, “*What is the difference between ‘absolute’ rights and rights ‘subject to limitations’? What is the difference between ‘absolute’ rights and rights ‘subject to limitations’?*” and “*What is the difference between assessing the impact on fundamental rights and verifying compliance with fundamental rights?*”).

¹¹⁷ Cf. with regard to the scope of Art. 35 GDPR section 2.7.1 above and in detail also *Müller/Poindl/Scheichenbauer*, in *Tagungsband der ÖFG-Tagung “KI-VO: Exekutive Rechtsetzung, Standardisierung, Zertifizierung und Grundrechte-Folgenabschätzung”* (*forthcoming*) section 2.

¹¹⁸ Cf. *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.3.4.3.

¹¹⁹ Cf. in this regard also the discussion of rights-based or risk-based approaches to assessing impacts on fundamental rights and the proposal of a new approach: *Malgieri/Santos*, *Assessing the (Severity of) Impacts on Fundamental Rights* (25. 6. 2024), <https://ssrn.com/abstract=4875937> (accessed on 26. 9. 2024).

be analysed in a specific manner in context of the personal data processing operations in scope (see also 2.4.3 above).¹²⁰

Lastly, it is questionable to which degree also legal persons are to be addressed mandatorily, considering that such could also be rights holders with regard to certain fundamental rights¹²¹. That is on the one hand because Article 27 AI Act only speaks of “*natural persons and groups*” (paragraph 1(c)) or “*natural persons or groups of persons*” (paragraph 1(d)),¹²² which could be questioned respectively, particularly with regard to *groups*, in light of that potential regulatory gap. Considering data protection, while the GDPR only protects natural persons (see section 2.3.1.2.1 above), according to some voices in the literature also Article 8 CFR protects legal persons (at least to some degree) as well, as according to prevailing doctrine and jurisprudence does Article 8 ECHR.¹²³ Therefore, it may be advisable to also involve legal persons in the assessment respectively.

[Elaboration on affected persons and groups particularly pursuant to Article 27(1)(c) AI Act, based on results of section 2.7.2 and assessed from the particular use of the object of the assessment, as well as on data subjects in context of respective data processing operations in particular in accordance with Article 35 GDPR]

2.7.4 Risk Table

The assessment of each identified risk structured according to the methodology outlined above is subsequently to be done in form of the following table.

2.7.4.1 Risk title

1) Risk identification	
	Risk description
	<i>Description and explanation of the possible risk scenario; naming of actors and persons involved</i>
	Risk source
	<p>What elements trigger the occurrence of the damage/harm? Is it a human or non-human risk source?</p> <p>Internal human source:</p> <p>Unintended actions: individual or structural errors</p> <p>Intended actions: harm to the person affected is either accepted or is intended by the perpetrator and is the aim of the action</p> <p>External human sources:</p> <p>Unintended actions: individual or structural errors</p>

¹²⁰ See Müller/Poindl/Scheichenbauer, in Tagungsband der ÖFG-Tagung “KI-VO: Exekutive Rechtsetzung, Standardisierung, Zertifizierung und Grundrechte-Folgenabschätzung” (*forthcoming*) sections 2 and 3.b with further references; Trieb in *Knyrim*, *DatKomm Art. 35 DSGVO* para. 109.

¹²¹ Cf. e.g. principally with regard to Art. 16 CFR: *Bezemek in Holoubek/Lienbacher*, *GRC-Kommentar*² Art. 16 para. 11 (status of 1. 4. 2019, rdb.at).

¹²² Cf. in this regard principally Fülöp/Poindl in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) sections 3.3.4.3 and 3.3.4.4, also concerning a former version of Art. 27(1)(d) AI Act.

¹²³ Cf. *Riesz in Holoubek/Lienbacher*, *GRC-Kommentar*² Art. 8 paras 48-56 (status of 1. 4. 2019, rdb.at).

	<p>Deliberate actions: Attacker or perpetrator with the aim of harming the system or those affected</p> <p>Internal / external technical sources:</p> <p>System errors (software/hardware)</p> <p>Examples (risk source):</p> <ul style="list-style-type: none"> • Internal personnel, • External personnel, • Data subjects, • Other third parties, • Software error, • Hardware defect (physical), • Environmental impact (force of nature), • Cybercriminals (attackers/malware), • Government institutions (intelligence services, law enforcement), • Management.
	Cause of damage/harm
	<p>What leads to the realisation of the risk (i.e. the damage or harm)?</p> <p>Examples (cause of damage/harm):</p> <ul style="list-style-type: none"> • Use of the system not known to the affected persons, • Limited, inadequate or lack of voluntariness of consent leading to unwanted data processing, • Lack of information of deployers or users leading to inadequate use of the system (e.g. overestimation of the system's capabilities or system unsuitable for the intended purpose), • Lack of information of the affected persons leading to wrong expectations or inadequate behaviour, • Market dynamics leading to rushed roll-out in the field.
	Possible damage/harm for the affected persons (with regard to the scope of application of relevant fundamental rights [and, with respect to Article 35 GDPR, other rights and freedoms of natural person])
	<p>Which damages/harms and infringements of the rights and freedoms of the affected persons can occur? Which (fundamental) rights (and freedoms) are affected, to what extent and how?</p>

2) Risk analysis and evaluation (before mitigation)			
	Probability of occurrence	Severity of damage/harm	Risk assessment
	<p>(1-4): Add probability level [negligible (1), limited (2), substantial (3), maximum (4)] and explain (see section 2.7.1.2 for guidance on determining the adequate probability).</p>	<p>(1-4): Add severity level [negligible (1), limited (2), substantial (3), maximum (4)] and explain (see section 2.7.1.2 for guidance on determining the adequate severity level).</p>	<p>(1-16): Calculate the overall risk level according to the risk matrix (2.7.1.2)</p>

	Explanation/justification	Explanation/justification	
	<i>Explanation/justification for/of the classification</i>	<i>Explanation/justification for/of the classification</i>	

3) Measures			
	Existing measures		
	<p>If it is envisaged to reduce/mitigate the identified risk: What regulatory, technical, organisational or behavioural measures need to be implemented to reduce the risk?</p> <p><i>Add list of selected envisaged/future mitigation measures, including a detailed description and explanation how they address the risks and disproportionality of the processing operations and other processes identified to protect the (fundamental) rights and freedoms of the affected persons/groups/data subjects and to demonstrate compliance with law.</i></p> <ul style="list-style-type: none"> • 		

4) Risk analysis and evaluation (after mitigation)			
	Probability of occurrence	Severity of damage/harm	Risk assessment
	<i>(1-4): Add probability level [negligible (1), limited (2), substantial (3), maximum (4)] and explain (see section 2.7.1.2 for guidance on determining the adequate probability).</i>	<i>(1-4): Add severity level [negligible (1), limited (2), substantial (3), maximum (4)] and explain (see section 2.7.1.2 for guidance on determining the adequate severity level).</i>	<i>(1-16): Calculate the overall risk level according to the risk matrix (2.7.1.2)</i>
	Explanation/justification	Explanation/justification	
	<i>Explanation/justification for/of the classification</i>	<i>Explanation/justification for/of the classification</i>	

2.7.5 Risk Analysis

2.7.5.1 [Insert Risk Tables of identified risks here]

2.7.6 Risk summary

2.7.6.1 Risks before mitigation

The identified risks with the according probability of occurrence and severity of damage or harm before the application of the identified mitigation measures can be plotted in the risk table as follows:

[enter risk numbers in the according table cells]

Risk assessment		Probability of occurrence			
		<i>negligible</i> (1)	<i>limited</i> (2)	<i>substantial</i> (3)	<i>maximum</i> (4)
Severity of damage/harm	<i>maximum</i> (4)				
	<i>substantial</i> (3)	RX, RY		RZ	
	<i>limited</i> (2)				
	<i>negligible</i> (1)				

2.7.6.2 Risks before mitigation

The identified risks with the according probability of occurrence and severity of damage or harm after the application of the identified mitigation measures can be plotted in the risk table as follows:

[enter risk numbers in the according table cells]

Risk assessment		Probability of occurrence			
		<i>negligible</i> (1)	<i>limited</i> (2)	<i>substantial</i> (3)	<i>maximum</i> (4)
Severity of damage/harm	<i>maximum</i> (4)				
	<i>substantial</i> (3)		RZ		
	<i>limited</i> (2)	RX, RY			
	<i>negligible</i> (1)				

2.8 Ethical Risk and Benefit Assessment

2.8.1 Introduction

2.8.1.1 Introductory explanations

The AI Act as a legal framework aims to ensure that AI systems respect not only fundamental rights, but also ethical principles.¹²⁴ Recital 27 of the AI Act recalls the non-binding principles listed in the 2019 Ethics guidelines for trustworthy AI developed by the independent AI HLEG appointed by the Commission, which are intended to help to ensure that AI is trustworthy and ethically sound. Because of the revolutionary character of laws on AI and the fact, that new technologies often develop faster than law can regulate them,¹²⁵ ethical principles play a huge role for a fair development and use of AI systems. Therefore, the goal of this ethical impact assessment is to go beyond legal necessities and evaluate ethical risks and benefits of AI systems.¹²⁶

¹²⁴ See the EC on the AI Act: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (accessed 14. 7. 2024).

¹²⁵ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 6 et seq., <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (accessed 26. 7. 2024).

¹²⁶ See *UNESCO*, Ethical impact assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence (2023), <https://unesdoc.unesco.org/ark:/48223/pf0000386276> (accessed 14. 7. 2024).

2.8.1.2 Relevance of the Ethical Impact Assessment

In ethics and in the legal theory some of the main questions are about what as humans we should do, what we have to do or what we are not allowed to do in specific situations. Therefore, we are looking at social norms, that tell us what to do. Ethics reflects on moral theories, which are concerned with whether an action is good or bad, whereas the law on the other hand will say an action is lawful or forbidden. A social norm consists of two parts. The first part describes a behaviour, and the second part imposes a sanction in case the addressed person doesn't follow the norm. The sanctions imposed for the violation of moral values may vary a lot, sometimes there is not even an appropriate sanction for moral misconduct. The characteristic what makes a social norm a legal norm, is that a legal norm can be lawfully executed with the power of a state.¹²⁷ This doesn't apply to other social norms, which points to the mutual relationship between law and state.¹²⁸

Often what is morally right and what is lawful go hand in hand, but this connection does not necessarily exist. It is therefore recommended to analyse law and ethics separately, although legal norms on fundamental rights and basic ethical principles often overlap. Moral concepts like human dignity and autonomy are supported by EU law as well as by morality, so the legal system in the EU and its member states represents at least a core of moral values. The ethical considerations of this EIA will mostly focus on risks and benefits that go beyond the legal framework of fundamental rights and data protection, because those frameworks are analysed under other parts of this combined method.

Ethical considerations are an important everyday mechanism to weigh each other's interests and find solutions, if interests get into conflict. This is brought to daylight especially when law is not able to offer satisfying mechanisms to find solutions as is seen often in moral dilemmas of the medical field. The law can set boundaries, but it cannot solve each case in a way to satisfy everyone. In medical ethics, ethical principles are applied to solve dilemmas, where not all persons and groups involved can be satisfied, but an acceptable decision has to be made. Just as much as medical ethics deals with human self-understanding in view of the beginning and the end of life or even in view of how to define a human being, the ethics of technology and in particular the ethics of AI looks at another form of human self-understanding when it comes to freedom, autonomy and decision making. AI reaches more and more a quality of decision making where human reasoning is pushed behind, which makes it necessary to have ethical considerations in an early stage of the process of implementing an AI system.

While this theoretical part served to show why the approach of a combined impact assessment can be of huge benefit, the following practical part will give a more concrete understanding of how to use the EIA.

2.8.1.3 How to Use the EIA

This EIA is considered a tool to help evaluate ethical risks and benefits. A person or group using this tool will take ethical considerations into account and should keep in mind the following question: *"Who's interests are at stake and what would they ask from a programmer, provider, controller etc. in view of an AI system?"* The EIA requires an identification and definition of the ethical impacts of the technology in question.¹²⁹ To do so, the following questions can be asked.

- What is your role and in which way do you have to take ethical considerations into account?

¹²⁷ See *Kelsen*, *Reine Rechtslehre* (1960) 2nd Edition, 64.

¹²⁸ See *Kelsen*, *Reine Rechtslehre* (1960) 2nd Edition, 289.

¹²⁹ See *Reijers et al.*, *Common Framework for Ethical Impact Assessment* (2016), 14, https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf.

- What is the current status of the technology in question and how does this affect ethical considerations?
- Who benefits from the technology in question? In which way?
- Who is at risk? In which way? In which way are risks not already covered by the DPIA and FRIA?
- In which way are the benefits of one person or group connected to the risks of another person or group?

For finding and analysing ethical impacts, it is useful to apply ethical values and principles.¹³⁰ The EIA offers a range of principles and moral values, that could be taken into account when looking at the technology in question. Identifying ethical problems to base a decision on how to apply a system in an ethical way is the main step to find solutions and recommendations.¹³¹ When ethical impacts are evaluated, different principles may conflict with each other, so it will be important to weigh and balance them.

The question of whether to start with the identification of ethical impacts or with the application of ethical principles or values may prove challenging to resolve. Finding impacts first and then using ethical principles to weigh them represents a valid approach, although some impacts might be first thought of when looking at the principles. Therefore, in this EIA, the approach is to first identify principles and then address ethical impacts by applying them. Should there come up impacts that cannot be dealt with the principles in the first place, it might be useful to think of additional principles to analyse the impacts. By using this method of implementing a loop, all relevant ethical impacts should be found and addressed with adequate principles. The EIA should provide an outcome in any way. There is either a solution for ethical problems by explaining why they are not too big of a threat for moral values. Additionally, recommendations may be proposed regarding the ethical use of the technology in question. It can also be concluded that the AI system in itself cannot be used in an ethical way. Therefore, all impacts should be covered, assessed and clearly presented in the end.

[if the previous sections do not provide sufficient coverage, include a description of the technology in question at this point, which, at the very least, includes: business function of technology at hand, impacts on critical functions and activities, including critical infrastructure, sector, conducting elections, maintaining supply chains, etc. and description of relevant phase of life cycle]

2.8.2 Affected Persons and Groups

From an ethical standpoint, the affected groups of persons can be broadly classified into two categories: “*vulnerable and marginalised groups*” and those who are “*other groups of people particularly affected in a specific context.*” It is imperative that both groups and the individuals listed therein be given special consideration in the development and utilisation of technology, given that their rights may be adversely affected at an earlier and more profound level by its deployment.¹³²

It is crucial to acknowledge that the notion of vulnerability is also inherently contextual. The vulnerability of an individual or group is often contingent upon the circumstances they are confronted with and may vary depending on whether a person finds themselves in a vulnerable position (“*universal vulnerability*”¹³³). Other

¹³⁰ See Reijers et al., Common Framework for Ethical Impact Assessment (2016), 14, https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf.

¹³¹ See Reijers et al., Common Framework for Ethical Impact Assessment (2016), 14, https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf.

¹³² See for further elaboration on this topic with regard to the climate crisis: *Hollaus*, Staatliche Schutzpflichten in der Klimakrise: Möglichkeiten, Herausforderungen und Grenzen im Rahmen der geltenden Grundrechtsdogmatik, ÖJA 2023, 377 (392).

¹³³ *Malgieri*, Vulnerability and Data Protection Law (2023) 73.

indicators of vulnerability frequently include the existence of a power imbalance between the individual or group affected and their counterpart, a lack of insight or knowledge regarding a process, and unequal treatment based on origin, gender, or social class.¹³⁴ That is to say, vulnerability is based on the specific characteristics of a person, on external influences such as power imbalances and on so-called “*vulnerability drivers*”, which together determine the classification of a particular person as vulnerable.¹³⁵ In combination, these factors serve to elevate the risk of adverse consequences for those individuals or groups that are affected. The following groups of persons are commonly considered “*vulnerable*” or “*marginalised*” or both. Please note that this list is not exhaustive and may change depending on the respective case in question.

- Children/minors
- Persons with disabilities
- Ethnic/religious minorities
- Elderly people
- Pregnant women
- Asylum seekers / unaccompanied minors
- Workers under precarious conditions
- Persons suffering from certain diseases
- Persons living in poverty/precarious economic or social situation
- Persons belonging to the LGBTQIA+-community
- Persons with addictions that could be exploited by the technology in question

Other persons or groups of persons, who are not regarded as particularly vulnerable in the strict sense, but may nevertheless be affected by the technology in question, include but are not limited to:

- Future generations
- Indigenous Peoples
- Workers in a specific professional field
- Certain (civil) associations, unions, organisations, communities, clubs
- The general public

[Introductory explanations on the assessment of involved persons/groups in Section 2.5 and why an ethical analysis beyond the identification of stakeholders/affected persons in Section 2.5 is necessary at this point]

[if not covered by previous sections: description of typical (end-)users of the technology at hand, their level of competency and the degree of optionality, including the possibility to opt out]

The ethical impact of the utilisation of [name the object of the assessment] thus goes beyond the elaborations in Section 2.5, as it additionally affects the following persons or groups of persons from an

¹³⁴ *Malgieri*, Vulnerability 54.

¹³⁵ *Malgieri*, Vulnerability 70.

ethical perspective [only analyse and describe the aforementioned categories in more detail where applicable/deemed necessary.]:

2.8.3 Responsibility

In contrast to the DPIA and FRIA, the EIA has a more expansive perspective on responsibility. From an ethical standpoint, any individual involved in the development, design, or deployment of the technology in question and engaged in a morally relevant action related to it should consider the corresponding ethical aspects. Nevertheless, when analysing the specific responsibility of an individual, it is essential to consider the phase of the technology's life cycle in which the individual is involved, as this may influence the type and extent of the individual's responsibility.

The ability to assume moral responsibility is contingent upon the moral theory or school of thought in question. In most cases, this ability is understood to necessitate a certain degree of intentionality, freedom of action, and knowledge. Furthermore, moral responsibility is often considered to be closely linked to moral agency, which is typically defined to comprise, at the very least, the abilities to communicate and answer for one's actions, autonomy, and (value) judgement.¹³⁶ If the aforementioned preconditions are met and an action that leads to a morally significant outcome is performed by a moral agent, moral responsibility may be ascribed. Causal connection between an agent's action and an outcome is often given.¹³⁷ Other theoretical frameworks conceptualise moral responsibility as a bundling of social duties of care and ascribe responsibility on the basis of social roles and the obligations one has towards others.¹³⁸ This can also be described as "*relational responsibility*", which refers to the fact that we do not only have responsibility *for* our actions, but also a responsibility *to others*.¹³⁹ In addition to the aforementioned affected groups or individuals, this may also encompass other sentient beings, such as animals, which are explicitly recognised as sentient beings in Article 13 of the Treaty on the Functioning of the European Union.

However, the fundamental prerequisites for attributing responsibility are not always fulfilled when technologies such as AI are involved. In particular, deep learning systems operate in ways that are not always transparent, and humans may not always be able to exert control over the system's functions.¹⁴⁰ The opacity, complexity and unpredictability of these systems can result in a responsibility gap, whereby no responsibility is attributed to humans due to a lack of control and knowledge.¹⁴¹ This has led to a debate in the field of AI ethics as to the extent to which humans should be held responsible for the behaviour of AI and at what stage of the system's life cycle responsibility should be ascribed. In addition, there have been numerous proposals for the consideration of AI systems as moral agents in themselves.¹⁴² At this time, the majority of the proposed characteristics or conditions that are thought to imply the moral responsibility of technical entities,

¹³⁶ See Loh, *Roboterethik. Eine Einführung* (2019) for a discussion of the most prominent theories on moral responsibility and agency with regard to AI and robots.

¹³⁷ Talbert, Moral Responsibility, in: *Stanford Encyclopedia of Philosophy* (2024), <https://plato.stanford.edu/entries/moral-responsibility/> (accessed 24. 7. 2024).

¹³⁸ Sullins, When Is a Robot a Moral Agent? *International Review of Information Ethics*, Vol. 6 (12/2006), 23 (28).

¹³⁹ See Coeckelbergh, Narrative responsibility and artificial intelligence. How AI challenges human responsibility and sense-making, in: *AI & SOCIETY* Vol. 38 (2023), 2437 (2440).

¹⁴⁰ See Coeckelbergh, Narrative responsibility 2437.

¹⁴¹ See Santoni de Sio and Mecacci, Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them, in: *Philosophy & Technology* Vol. 34 (2021), 1057 (1058) and Mittelstadt et al., The ethics of algorithms: Mapping the debate, in: *Big Data & Society* (2016), 1 (2).

¹⁴² See Mittelstadt et al., The ethics of algorithms 2 and for a more detailed discussion of this topic: Loh, *Roboterethik* or Noorman, Computing and Moral Responsibility, in: *Stanford Encyclopedia of Philosophy* (2023), <https://plato.stanford.edu/entries/computing-responsibility/> (accessed 25. 7. 2024).

including computers, robots, and AI systems, have not yet manifested. These include self-awareness¹⁴³, (intentionally exercised) autonomy¹⁴⁴, or sentience¹⁴⁵, which are all absent in the technology currently available. This is why the document at hand will not engage in discussions regarding the potential classification of [name object of the assessment] as a moral agent and, consequently, the possibility of holding it responsible. Nevertheless, as soon as there is a possibility that the technology in question may satisfy one of the aforementioned requirements, or as soon as technological progress is such that technological entities may be considered subjects of responsibility, the document at hand is to be revised and the EIA is to be repeated.

In the light of the above, the following parties can be identified as moral agents who bear a certain degree of responsibility that depends on the quality and quantity of the following conditions:

- **Persons with a more profound knowledge or understanding of the technology in question.** This is particularly relevant to developers and programmers, as they possess detailed knowledge of the system's characteristics, functions, properties, and so forth. Furthermore, this applies to all individuals who possess insight into the system and are involved in its development and deployment. This can be defined as "*design responsibility*", which holds that the individuals responsible for a technology's design and deployment are also responsible for its consequences.¹⁴⁶ With this, individuals who are not legally liable may be held morally responsible. This includes, for example, programmers or developers who design or develop a system on behalf of a legal entity. These individuals possess insight into the functions of a system and are therefore well placed to foresee, analyse and mitigate potential ethical risks. Such a moral responsibility entails that these individuals have a moral obligation to share their knowledge of the technology involved with staff members in a leading role or with those who are competent to take decisions. Moreover, the former should inform the latter as soon as they observe any ethically questionable aspects of the technology.
- **Persons with the ability to influence the development, design or deployment of a system.** Those persons who cause a morally significant outcome or who are able to influence the system in a manner that leads to a morally significant outcome are to be held responsible, provided that the causation is linked to a certain degree of intentionality.
- **Persons who act with a certain degree of intentionality.** This refers to persons who intentionally exercise control over a system's development, design, or deployment and are involved in the decision-making process. This could be a staff member in a leading role, but it could also be developers and programmers who decide on whether a certain function or property is to be integrated.
- **Persons who occupy a social role that entails certain obligations towards others with regard to the technology in question.** This includes deployers who are in most cases the direct points of contact for end-users. They have to ensure that the systems they use and the products and services they offer meet the required ethical standards.¹⁴⁷

It is strongly advised that the aforementioned persons ensure the technology's ethical compliance from the outset ("*ethics-by-design*" approach), encompassing the development and design phase and extending to the deployment and utilisation phase. In the event that ethical issues are identified, it is essential to implement

¹⁴³ See *Farina*, Artificial Intelligence Systems, Responsibility and Agential Self-Awareness, in: *Müller* (Ed.), *Philosophy and Theory of Artificial Intelligence* (2021), 15.

¹⁴⁴ See *Sullins*, When is a Robot a Moral Agent?

¹⁴⁵ See *Véliz*, Moral zombies: why algorithms are not moral agents, in: *AI & SOCIETY* Vol. 36, Iss. 02 (2021), 487.

¹⁴⁶ *Lokhorst and Van den Hoven*, Responsibility for Military Robots, in: *Lin et al* (Ed.), *Robot Ethics. The Ethical and Social Implications of Robotics* (2012), 145 (154).

¹⁴⁷ See Independent High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for trustworthy AI* (2019), 14, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

measures designed to eliminate or mitigate any associated ethical risks. Furthermore, it is of paramount importance to clearly designate the individuals responsible and to establish a clear framework for resolving value conflicts or ethical dilemmas (for further details, please refer to point 8.5.3). It is essential to differentiate between moral and legal responsibility of individuals, even if there may be some overlap between the two categories in one person, because the consequences most likely differ. Additionally, it is crucial to consider the impact of varying role distributions between controllers and processors, or providers and deployers on moral responsibility and the extent to which the persons in question bear responsibility.

It is imperative that those persons who are capable of bearing moral responsibility be made aware of their pivotal role throughout the technology's lifecycle. Furthermore, they must be provided with a comprehensive understanding of the responsibilities associated with this role. This is a management responsibility. Persons responsible for ethics in the respective project or ethical advisors shall actively provide the relevant information and carry out workshops and discussions to establish such understanding. The relevance of these information and activities must be effectively made clear by the management. It is, however, crucial to understand that the presence of persons responsible for ethics in the respective project or ethical advisors does not in any way reduce the moral responsibility of the actors in the roles described above, as they take the decisions, may they be larger or smaller, that actually shape the outcome.

Annex A, which, among others, contains the principle of accountability, outlines the potential strategies for mitigating and eliminating ethical risks in more detail (see Annex A point 4). This is because accountability encompasses responsibility and complements it with the aspects of answerability, authority recognition, interrogation, and limitation of power.¹⁴⁸ The ethical principle will be described in more detail in Annex A point 4 and its practical relevance shown by formulating key questions to determine if the preconditions of accountability are met.

2.8.4 Theoretical Framework of Ethical Principles

The fields of computer ethics, the ethics of technology, and AI ethics are experiencing significant growth, which serves to illustrate the considerable ethical implications that all forms of technology have for individuals and society all large. These ramifications extend to the evolving relationship between humans and machines, prompting crucial inquiries into human autonomy, dignity, justice, and transparency as well as the wider societal and environmental effects of a technology.¹⁴⁹ Consequently, the ethical impact of technology manifests itself in the potential for both ethical risks and benefits to emerge from the utilisation of specific systems, whether in particular sectors, towards designated individuals or groups, or within defined contexts. In order to assess and mitigate the impact of technology (especially AI), numerous international ethics frameworks and guidelines advocate the realisation of trustworthy AI, which is largely contingent upon the adherence to fundamental ethical principles and values.¹⁵⁰ These principles and values serve as the foundation for the ethical development, design, deployment, and utilisation of AI systems and technology in general.

¹⁴⁸ See *Novelli et al.*, Accountability in artificial intelligence: what it is and how it works, in: AI & SOCIETY (2023), <https://doi.org/10.1007/s00146-023-01635-y>.

¹⁴⁹ See *Mantelero*, Beyond Data – Human Rights, Ethical and Social Impact Assessment, Asser Press/Springer, Information Technology and Law Series, IT&LAW 36, 2022, 93.

¹⁵⁰ E.g. *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

Annex A will provide an overview of the fundamental ethical principles that should be considered when assessing the technology in question.¹⁵¹ Each ethical principle is followed by a list of guiding questions that assist in determining the impact of the technology in question.¹⁵²

2.8.5 Ethical Impact Analysis

After identifying ethical impacts, they should be analysed by applying the ethical principles and values. To have a positive outcome, two basic moral values should be fulfilled. The technology in question should be developed in a way that benefits humans, with the assumption that the greater the benefit to humans, the more ethical the technology. The consequences of the use of the technology in question should be good in a moral sense. While often a utilitarian view will represent the consequentialist approach best, balancing interests is allowed and freedom and autonomy of humans in the process allow them to put their own interests on top. The second principle is of a deontological nature and means to ensure that individual interests are not overruling basic legal and moral rights. Although it seems to be legitimate to have one's interests fulfilled, taking everyone's involved dignity and freedom into account is necessary. If basic moral rights are attacked, a technology cannot be ethical. The following assessments should always take into account, who benefits from the use of the system and which principles direct to a positive use of the system. On the other hand, ethical risks have to be taken into account. Who might be at risk of the use of the technology in question and which principles direct to a negative use of the system?

To identify benefits and risks, ethical impacts should be described in the following table and assessed with the given categories. In the text section, additional information can be presented and the main aspects of the problem should be made explicit. If possible, it is useful to already present strategies how to deal with ethical risks.

The goal then should be to have a list of ethical impacts in the table of benefits and a list of ethical impacts in the table of risks. Often some benefits can be assigned to risks. For example, if an AI system helps a car drive by itself, the autonomous decision of the driver will be taken away, but the potential benefit is that (statistically) there will be less car accidents.

Especially if benefits of one person or group seem to outweigh the risks of another person or group, it should be argued and justified, why it is considered still in the boundaries of fairness to use the technology in question. This must be done even more carefully, if vulnerable or marginalised persons and groups are at risk.

In order to identify potential risks and benefits, in addition to responding to the guiding questions presented in Annex A, the following methods, which are not exhaustive, are recommended. It should be noted that depending on the scale, impact level, number and kind of affected persons, the likelihood of impact, and resources available, different methods or a combination of multiple methods may be appropriate.

¹⁵¹ It should be noted that this list is not exhaustive and may be subject to change at any time.

¹⁵² The questions have been derived from the following sources: *Independent High-Level Expert Group on Artificial Intelligence*, The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment (2020), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342; UNESCO, Ethical Impact Assessment. A Tool of the Recommendation on the Ethics of Artificial Intelligence (2023), https://unesdoc.unesco.org/ark:/48223/pf0000386276_eng; Bundesministerium Kunst, Kultur, öffentlicher Dienst und Sport, Leitfaden Digitale Verwaltung und Ethik. Praxisleitfaden für KI in der Verwaltung, Version 1.0 (2023), <https://oeffentlicherdienst.gv.at/wp-content/uploads/2023/11/Leitfaden-Digitale-Verwaltung-Ethik.pdf>.

- **Exploration of existing work:** The process of analysing ethical impacts should commence with an evaluation of existing work, studies, and other pertinent materials within the field in question. This primarily entails desk research and a systematic literature review.¹⁵³
- **Stakeholder consultation:** In order to adequately assess the potential impact of the technology in question, it is recommended that open discussion and the involvement of stakeholders be considered, so that they can contribute their views and, ideally, submit reports.¹⁵⁴ It is also possible to organise panels including various members, such as affected persons, representatives of a specific group, etc. Actively seeking participation and dialogue supports the evaluation of results and approaches, and can particularly be helpful in complex cases.¹⁵⁵ This method is open and transparent, but the response rate is often low.¹⁵⁶
- **Expertise-based methods:** This encompasses the organisation of expert panels, expert consultations and expert surveys, which may take the form of interviews or the conduction of small workshops.¹⁵⁷ It is strongly recommended that experts from a range of disciplines participate in discussions and the development of solutions together.¹⁵⁸ Expertise-based methods also include the ethical Delphi, which is an iterative process in which experts repeatedly and anonymously exchange opinions and arguments.¹⁵⁹ The underlying concept of the Delphi method is that this feedback loop will allow for better judgements to be made without there being undue influence from forceful or high-status advocates.¹⁶⁰
- **Trend analysis:** The EIA may be employed to identify a general tendency or direction of technological development, which may be discerned from past events and thus suggest a pattern.¹⁶¹
- **Ethical matrix:** Another approach is the development of an ethical matrix. This is a tool based on ethical principles that are applied to specific interest groups with the objective of supporting decision-making and facilitating the structured analysis of ethical impacts.¹⁶²

¹⁵³ Reijers *et al.*, A Common Framework for Ethical Impact Assessment. Annex I. A reasoned proposal for a set of shared ethical values, principles, approaches for ethics assessment in the European Context (2016), 32, https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf (accessed 12. 8. 2024).

¹⁵⁴ Wright, A framework for the ethical impact assessment of information technology, *Ethics and Information Technology* 13 (2011), 199 (216).

¹⁵⁵ See Independent High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for trustworthy AI (2019), 23, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

¹⁵⁶ Wright, A framework for the ethical impact assessment of information technology, *Ethics and Information Technology* 13 (2011), 199 (216).

¹⁵⁷ Reijers *et al.*, A Common Framework for Ethical Impact Assessment. Annex I. A reasoned proposal for a set of shared ethical values, principles, approaches for ethics assessment in the European Context (2016), 32, https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf.

¹⁵⁸ Wright, A framework for the ethical impact assessment of information technology, *Ethics and Information Technology* 13 (2011), 199 (pp216).

¹⁵⁹ Wright, A framework for the ethical impact assessment of information technology, *Ethics and Information Technology* 13 (2011), 199 (pp217).

¹⁶⁰ Reijers *et al.*, A Common Framework for Ethical Impact Assessment. Annex I. A reasoned proposal for a set of shared ethical values, principles, approaches for ethics assessment in the European Context (2016), 33, https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf.

¹⁶¹ Reijers *et al.*, A Common Framework for Ethical Impact Assessment. Annex I. A reasoned proposal for a set of shared ethical values, principles, approaches for ethics assessment in the European Context (2016), 33, https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf.

¹⁶² Wright, A framework for the ethical impact assessment of information technology, *Ethics and Information Technology* 13 (2011), 199 (217). For an example of an ethical matrix, kindly consult: https://www.researchgate.net/figure/A-generic-ethical-matrix-example-Mephram-et-al-2006_fig1_363155443, or <https://core.ac.uk/download/pdf/29269684.pdf>, or (in German): https://www.vetmeduni.ac.at/fileadmin/v/messerli/ethik/Projekte/VETHICS/Scan_Matrix.pdf.

- **Other methods** include the conducting of trend analyses, situational approaches, road mapping, foresight analyses, ethical checklist approaches and scenario writing.¹⁶³

2.8.5.1 Ethical Benefits

The following table was created using the template provided in the UNESCO recommendations for conducting an ethical impact assessment.¹⁶⁴

Benefit description	Positive impact level	Extent of impact	Groups of affected persons	Likelihood of occurrence
B1	Transformative	Short-term	Humankind as a whole	Very high
	Significant	Medium-term		High
	Moderate	Long-term	Specific groups	Medium
	Minor	Intergenerational		Low
BN...				

2.8.5.2 Ethical Risks

The following table was created using the template provided in the UNESCO recommendations for conducting an ethical impact assessment.¹⁶⁵

Description of the risk	Gravity level	Extent of impact	Groups of affected persons	Likelihood of occurrence	Possible mitigation measures
R1 (insert risk description)	Catastrophic	Short-term		Very high	
	Critical	Medium-term		High	
	Serious	Long-term		Medium	
	Moderate minor	Intergenerational		Low	
RN...					

¹⁶³ For more details, kindly consult: *Reijers et al.*, A Common Framework for Ethical Impact Assessment. Annex I. A reasoned proposal for a set of shared ethical values, principles, approaches for ethics assessment in the European Context (2016), 32, https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf.

¹⁶⁴ UNESCO, Recommendations on the Ethics of Artificial Intelligence (2021), https://unesdoc.unesco.org/ark:/48223/pf0000381137_eng (accessed 7. 8. 2024).

¹⁶⁵ UNESCO, Recommendations on the Ethics of Artificial Intelligence (2021), https://unesdoc.unesco.org/ark:/48223/pf0000381137_eng.

2.8.5.3 Value Conflicts

Once ethical impacts are found, they should be analysed in accordance with the relevant ethical principles and values. In general, the ethical principles mentioned in 8.4 should be fulfilled as far as possible.¹⁶⁶ This is not always possible, because principles are sometimes pointing in contrary directions and, therefore, have to be weighed against each other or it has to be examined as to what extent a principle has to be fulfilled to have an ethically positive outlook.

As a first approximation, it is recommended to identify principles or values that are fundamental.¹⁶⁷ The best example of an absolute or fundamental value is human dignity:¹⁶⁸ Kant describes it as follows:

“So act as to treat humanity, whether in thine own person or in that of any other, in every case as an end withal, never as means only.”¹⁶⁹

New technologies and in particular AI systems have the potential to lose an anthropocentric view, if they just treat humans as a means to an end. This would violate human dignity and therefore infringe human rights, because the latter are based on human dignity. Therefore, the main purpose of all AI systems should be to make human life better or at least not worse by following fundamental moral values.

For other principles it might not be so simple to define, if they should be seen as fundamental.¹⁷⁰ Of course, one would think of the core of transparency as a necessity in the use of AI systems. The minimum would be for humans to be able to trace back how the system was built, which data was used to train it, who built it for which purpose, etc. However, in contrast to human dignity, transparency is not an absolute necessity, since disclosing such information could violate other rights (e.g. confidential business information). Hence, even if some principles are not seen as absolute, balancing them against other principles would have to be justified very well.

For most principles it might be hard to define, if they have a core that is fundamental,¹⁷¹ but it does not seem necessary to discuss this at this point, as most moral principles are already represented in human rights. The reason why these values enjoy protection of the law might be, that they are of a high not only legal, but also moral value. For the purpose of the combined method given, we can refer to the corresponding sections and this section will often have to weigh principles against each other.

For ethical impacts that cannot be addressed with a fundamental moral value, there will often come up a moral dilemma. This can also be the case, if two principals have or seem to have fundamental moral value. A moral dilemma is a situation, where not all interests of all the people involved can be fulfilled by taking different actions, so there will always be some form of dissatisfaction.¹⁷² The given ethical principles represent those different interests and by applying them, users of this EIA should seek compromises or the

¹⁶⁶ See Prem, From ethical AI frameworks to tools: a review of approaches (2022), https://www.researchgate.net/journal/AI-and-Ethics-2730-5961/publication/368391699_From_ethical_AI_frameworks_to_tools_a_review_of_approaches/links/63e5bdc964252375639f88d0/From-ethical-AI-frameworks-to-tools-a-review-of-approaches.pdf (accessed 26. 7. 2024), 702.

¹⁶⁷ See Reijers et al., Common Framework for Ethical Impact Assessment (2016), 43, https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf.

¹⁶⁸ See Independent High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for trustworthy AI (2019), 13, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

¹⁶⁹ Kant, FUNDAMENTAL PRINCIPLES OF THE METAPHYSIC OF MORALS (1785), translated by Thomas Kingsmill Abbott (updated 2021) <https://www.gutenberg.org/cache/epub/5682/pg5682-images.html> (accessed 26. 7. 2024).

¹⁷⁰ See Reijers et al., Common Framework for Ethical Impact Assessment (2016), https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf, 43.

¹⁷¹ See Beauchamp/Childress, Principles of biomedical ethics (2009) 7th edition, 19; they find for example prohibition of cruelty and unnecessary inflicting of pain and suffering to be absolute virtues.

¹⁷² See Beauchamp/Childress, Principles of biomedical ethics (2009) 7th edition, 11.

best possible solution in a specific case. Applying moral values and ethical principles should not only be understood as balancing interests, as interests are not moral values.¹⁷³ While *Prem* points out that it is easier to identify ethical principles than to provide clear instructions,¹⁷⁴ the goal here is to give a tool to do so.

In a first step, ethical impacts should be identified and addressed with regard to ethical principles that are at risk or that might be of any benefit and be written in the corresponding table. In a second step, some ethical principles could be in conflict.¹⁷⁵ If so, then the conflict should be outlined in the table. Benefits and risks should be identified for each principle itself and for each conflict of principles. Subsequently, it should be described in more detail in which way principles in conflict pose risks and benefits so they can be weighed against each other.¹⁷⁶

As the ethical principles pointed out can be abstract, it makes sense to specify them and find rules for concrete actions. Outlaying concrete rules and adding content brings principles closer to the specific case.¹⁷⁷ This might be a first step to work with the principles. A main goal of specification should be to identify where there are boundaries in the process of weighing the principles against each other. There can be conflicting norms and the question is, how to define which norm or principle to value more.¹⁷⁸ Weighing and balancing principles is one of the main tasks of the EIA, because at this point, there is no universal moral theory or system that can be applied to solve ethical problems.¹⁷⁹ Balancing should not be spontaneous or unreflective, instead it should be based on reasoning.¹⁸⁰ In analogy to fundamental rights cases, the principle of proportionality or risk minimalization could be used to identify the main ethical principles applicable.¹⁸¹ *Beauchamp and Childress* suggest six conditions that have to be met to justify choosing one norm over another.¹⁸² While they wrote on medical ethics, these conditions apply to balancing and weighing in the ethics of technology and AI ethics just as well.

1. *Good reasons can be offered to act on the overriding norm rather than on the infringed norm.*
2. *The moral objective justifying the infringement has a realistic prospect of achievement.*
3. *No morally preferable alternative actions are available.*
4. *The lowest level of infringement, commensurate with achieving the primary goal of the action, has been selected.*
5. *All negative effects of the infringement have been minimized.*
6. *All affected parties have been treated impartially.*¹⁸³

¹⁷³ See *Reijers et al.*, Common Framework for Ethical Impact Assessment (2016), 12 et seq., https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf (accessed 26. 7. 2024).

¹⁷⁴ See *Prem*, From ethical AI frameworks to tools: a review of approaches (2022), https://www.researchgate.net/journal/AI-and-Ethics-2730-5961/publication/368391699_From_ethical_AI_frameworks_to_tools_a_review_of_approaches/links/63e5bdc964252375639f88d0/From-ethical-AI-frameworks-to-tools-a-review-of-approaches.pdf (accessed 26. 7. 2024), 701

¹⁷⁵ See *Reijers et al.*, Common Framework for Ethical Impact Assessment (2016), https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf (accessed 26. 7. 2024), 14.

¹⁷⁶ See *Reijers et al.*, Common Framework for Ethical Impact Assessment (2016), https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf, 14.

¹⁷⁷ See *Beauchamp/Childress*, Principles of biomedical ethics (2009) 7th edition, 17.

¹⁷⁸ See *Beauchamp/Childress*, Principles of biomedical ethics (2009) 7th edition, 20.

¹⁷⁹ See *Beauchamp/Childress*, Principles of biomedical ethics (2009) 7th edition, 16.

¹⁸⁰ See *Beauchamp/Childress*, Principles of biomedical ethics (2009) 7th edition, 22.

¹⁸¹ See *Ernst*, Am Anfang und Ende des Lebens. Grundfragen medizinischer Ethik (2020), 44 and 47.

¹⁸² See *Beauchamp/Childress*, Principles of biomedical ethics (2009) 7th edition, 23.

¹⁸³ See *Beauchamp/Childress*, Principles of biomedical ethics (2009) 7th edition, 23.

If there are still conflicts of principles, another possibility is to think out of the box to come up with solutions for a moral dilemma. It could also be an option to consult different stakeholders to resolve moral conflicts by deliberation and negotiation. Maybe the dilemma can be avoided by setting up the technology in question in a different way.¹⁸⁴ Maybe there are non-algorithmic options or different methods?¹⁸⁵

Even if all specification, weighing and balancing are done, there can still be moral disagreement. Moral diversity is nothing bad and it is normal for different persons to come to different conclusions.¹⁸⁶ If moral reasoning and good ethical arguments are done, a solution does not have to be perfect.

Not only can ethical values conflict with each other, but they can also be in conflict with the law. Especially if you look at the benefits of an AI system, it might seem that data protection goes too far by restricting the use of potential training data. One may also argue that small violations of some fundamental rights are justified, if the technology in question could bring huge advantages for society. As mentioned above, ethics and law do not necessarily have to come to the same conclusions. Still, it makes sense to point out ethical benefits and show how they would outweigh ethical risks.

2.8.6 Mitigation of Ethical Risks and Weighing with Benefits

Based on the theoretical considerations and guiding questions in Annex A as well as the ethical impact analysis in section 2.8.5, this section lists recommendations to mitigate ethical risks that go beyond legal recommendations.

[The specific recommendations will vary depending upon the case in question and will have to be specified accordingly. However, common recommendations are listed in Annex B]¹⁸⁷

In conclusion, the preceding considerations, including the ethical risks, benefits, their gravity level, extent and likelihood of occurrence, as well as the selected recommendations on how to mitigate ethical risks and balance them with ethical benefits, can be summarised in the table below.

¹⁸⁴ See *Reijers et al., Common Framework for Ethical Impact Assessment* (2016), 43 et seq., https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf.

¹⁸⁵ See *UNESCO, Ethical impact assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence* (2023), 12, <https://unesdoc.unesco.org/ark:/48223/pf0000386276>.

¹⁸⁶ See *Beauchamp/Childress, Principles of biomedical ethics* (2009) 7th edition, 24f.

¹⁸⁷ Sources: *Algemene Rekenkamer, An audit for algorithms. Nine algorithms used by the Dutch Government* (2022), (<https://english.rekenkamer.nl/binaries/rekenkamer-english/documenten/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government/An+Audit+of+Algorithms.pdf>), *NSW Government, Mandatory Ethical Principles for the use of AI* (<https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-ethics-policy/mandatory-ethical-principles>), *OECD, Advancing accountability in AI. Governing and managing risks throughout the lifecycle for trustworthy AI*, no. 349 (2023), https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/advancing-accountability-in-ai_753bf8c8/2448f04b-en.pdf (accessed 7. 8. 2024), *UNESCO, Recommendations on the Ethics of Artificial Intelligence* (2021), https://unesdoc.unesco.org/ark:/48223/pf0000381137_eng.

Risks	Benefits weighed with risks	Description of weighing and balancing risks and benefits	Recommendations	Gravity level if recommendations are followed	Extent of impact if recommendations are followed	Likelihood of occurrence if recommendations are followed
R1	BN 1	Reason 1	Rec 1	Catastrophic	Short-term	Very high
	BN ...	Reason ...	Rec ...	Critical	Medium-term	High
				Serious	Long-term	Medium
				Moderate	Intergenerational	Low
				Minor		
RN...						

2.8.7 Conclusion concerning Ethical Impacts

It follows from this, that from an ethical perspective **[short summary of findings]**

Similar to the DPIA and FRIA, the EIA should be conducted prior to the first use of a technology, tool, system, etc. Ideally, the assessment or parts thereof is/are carried out already during the development phase, to ensure an “ethics-by-design”-approach. If, during the use of the technology in question, the responsible parties consider that any of the elements described in this section have changed or are no longer up to date, necessary steps should be taken to update the information or repeat the assessment. With regard to the EIA, it is recommended that the assessment be updated as soon as an update of the DPIA and/or FRIA is required in accordance with the respective provisions of the GDPR and/or the AI Act (see 2.10.2 below).

2.9 Residual Risk Level and Proportionality

Based on the residual risks the overall risk level (see 2.9.1), the Combined Impact Assessment should then examine the proportionality of potential interferences with fundamental rights and other rights and freedoms. The principle of proportionality (and the related step of examining the necessity) can be understood as a doctrinal tool for the resolution of conflicts between two competing rights or interests¹⁸⁸ and has in principle been discussed already in sections 2.6.4 and 2.7.1 above in connection with fundamental/human rights and respective impact assessments.

While e.g. with regard to impact assessments conducted by the EC (in light of fundamental rights) in particular in connection with legislative proposals, deeper fundamental rights compliance checks are subject to later, final analyses, also here the impact assessment aims at an identifying analysis of inter alia “necessity and proportionality of the interference in terms of policy options and objectives”¹⁸⁹, and taking into account sorts of such proportionality considerations plays a rather important role in other approaches to impact assessments in connection with fundamental rights¹⁹⁰. Furthermore, such considerations presumably

¹⁸⁸ Cf. Möller, Proportionality: Challenging the Critics, 2012 10(3) International Journal of Constitutional Law 709–731; <https://doi.org/10.1093/icon/mos024>, 10.

¹⁸⁹ Cf. European Commission, SEC(2011) 567 final (section “What is the difference between assessing the impact on fundamental rights and verifying compliance with fundamental rights?”).

¹⁹⁰ Cf. e.g. Ministry of the Interior and Kingdom Relations (Government of the Netherlands), <https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms>;

constitute an essential indication with respect to measures to be taken against fundamental rights risks in connection with the FRIA (besides Article 79(1) AI Act, essentially referring to unacceptable risks, and risk management by the provider; see section 2.7.1 above), which apart from general fundamental rights principles (see already section 2.6.4) might also be deduced from the understanding of the AI Act of risks and respective counter/mitigation measures.¹⁹¹

Regarding the DPIA in particular, some further general remarks are to be made: The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights.¹⁹² In this spirit, Article 35 (7) (b) GDPR stipulates an assessment of the necessity and proportionality of the processing operations in relation to the purpose, in case the operations interfere with other fundamental rights.¹⁹³ Also, according to the European Data Protection Supervisor (EDPS), the appraisal of necessity and proportionality is an essential requirement with which any proposed measure that involves processing of personal data must comply.¹⁹⁴ Contrary to the several existing methodologies for risk assessments, approaches for assessing the proportionality and necessity in the context of personal data protection are rather scarce.¹⁹⁵

In its related guidelines the EDPS points out that the examination of necessity (just as the entire DPIA) is a facts-based process, rather than a merely abstract legal notion. It must therefore be considered in light of the specific circumstances surrounding the use-case in question as well as the concrete purpose it aims to achieve. This means that the respective data protection operation should be genuinely effective to achieve the pursued objective of general interest.¹⁹⁶

In addition, the envisaged data protection operation should be the least intrusive for the fundamental right at stake. Consequently, the assessor needs to consider alternative measures which are comparably effective but with less impact on e.g. the protection of personal data or the right to respect of private life.¹⁹⁷ Only if existing or less intrusive measures are not available and the envisaged data processing operation is essential

Janssen/Lee/Singh, International Journal of Law and Information Technology, 2022, 30, 200 <https://doi.org/10.1093/ijlit/eaac018>; Malgieri/Santos, Assessing the (Severity of) Impacts on Fundamental Rights (25. 6. 2024), <https://ssrn.com/abstract=4875937>; in connection with data processing cf. also, however rather sceptical in that regard: Mantelero, Computer Law & Security Review 34 (2018) 754 (particularly 762) <https://doi.org/10.1016/j.clsr.2018.05.017>.

¹⁹¹ Cf. e.g. Recital 96 referring to measures to be determined *in light of the risks identified* (also implying the importance of the specific context of use/use-case); furthermore already the risk-based approach of the regulation itself, e.g. referred to explicitly in Recitals 26 and 27; concerning (risk mitigation) measures adopted by the providers furthermore Article 9 (particularly concerning “appropriate and targeted risk management measures” pursuant to paragraph 2(d); as specified in paragraphs 4 and 5) and e.g. Recitals 64, 65, 70, 76 and 118; concerning deployers in advance also Art. 26(1) and Recital 91; in connection with human oversight measures Art. 14(2) and (4) and Recital 73; Recital 52 concerning the classification of stand-alone AI systems as high-risk itself (as well considering probability and severity of harms); moreover e.g. 5(2) concerning the permissibility of the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement; and the aforementioned principle acceptance of applying AI systems classified as high-risk but ultimately particularly restricting such meeting the threshold of Art. 79(1) (see particularly section 2.7.1.2 above and respective explanations in the sections before); also in principle Fülöp/Poindl in Pehlivan/Forgó/Valcke (eds.), Artificial Intelligence Act: A Commentary (Kluwer Law International *forthcoming*) section 3.3.4.6.

¹⁹² European Data Protection Supervisor (EDPS), Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit, EDPS, 2017, https://edps.europa.eu/sites/edp/files/publication/17-04-11_necessity_toolkit_en_0.pdf, 4.

¹⁹³ Cf. Jandt in Kühling/Buchner (eds.), DS-GVO/BDSG, Art. 35 para. 39 et seq.

¹⁹⁴ Cf. European Data Protection Supervisor (EDPS), Guidelines on assessing the proportionality of measures that limit the fundamental rights to privacy and to the protection of personal data, 2019, 3, https://edps.europa.eu/sites/edp/files/publication/19-12-19_edps_proportionality_guidelines2_en.pdf.

¹⁹⁵ Cf. Kloza/Calvi/Casiraghi/Vazquez Maymir/Ioannidis/Tanas/Van Dijk, Data protection impact assessment in the European Union: developing a template for a report from the assessment process, Brussels Laboratory for Data Protection & Privacy Impact Assessments, Policy Brief 1/2020, VUB 2020, 29, <https://doi.org/10.31228/osf.io/7qrfp>.

¹⁹⁶ Cf. European Data Protection Supervisor (EDPS), Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit, EDPS, 2017, 8, https://edps.europa.eu/sites/edp/files/publication/17-04-11_necessity_toolkit_en_0.pdf.

¹⁹⁷ Cf. Jandt in Kühling/Buchner (eds.), DS-GVO/BDSG, Art. 35 para. 39.

and limited to what is absolutely necessary to achieve the objective of general interest, the criterion of necessity is met.¹⁹⁸

At the core of the appraisal process lies the concept of a balancing. This final step is a procedure of weighing up the intensity of the interference against the legitimacy (or importance) of the objective pursued in the given context. The balancing (of advantages/disadvantages and benefits/costs) should lead to the decision whether the data processing operation in question is proportionate or not. If the conclusion is that it is not proportionate, the assessor should make sure to take all factors which determine the appraisal as disproportionate into account and determine and introduce (if possible) safeguards which render the data processing activity proportionate.¹⁹⁹

In general, as well as in accordance with Article 52 CFR and the relevant legal literature on the methodological implementation of the proportionality principle, the following steps/criteria can be differentiated after implementing the principles of legality²⁰⁰ and legitimacy²⁰¹ (see in general already section 2.6.4):²⁰²

- **Suitability**

- Are the means of the interference suited and capable to achieve its aspired aim?
- The scrutiny of which may also depend on the (kind of) interference²⁰³
- e.g.: Is the envisaged data processing operation even capable to achieve the given legitimate aim?

- **Necessity**

- Is the interference necessary to achieve its aspired aim?
- This should also include assessing alternative (less invasive) means/measures suitable to achieve the aspired aim, especially concerning particularly intrusive interferences/means.²⁰⁴
- e.g.: Is the envisaged data processing operation necessary to achieve its (legitimate) aim?

- **Appropriateness (Proportionality *sensu stricto*):**

- Is the interference also appropriate (proportional in a narrower sense) in light of its aspired aim, i.e. fairly balanced with competing interests?

¹⁹⁸ Cf. *European Data Protection Supervisor (EDPS)*, Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit, EDPS, 2017, 17, https://edps.europa.eu/sites/edp/files/publication/17-04-11_necessity_toolkit_en_0.pdf.

¹⁹⁹ Cf. *European Data Protection Supervisor (EDPS)*, Guidelines on assessing the proportionality of measures that limit the fundamental rights to privacy and to the protection of personal data, 2019, 11, https://edps.europa.eu/sites/edp/files/publication/19-12-19_edps_proportionality_guidelines2_en.pdf.

²⁰⁰ E.g. referring to the question if e.g. the legal basis for data processing is provided for by a law of a sufficient quality and if this legal basis respects the essence of the fundamental rights and freedoms.

²⁰¹ E.g. referring to the question if e.g. the envisaged data processing operation serves a legitimate aim or meets objectives of general interest to protect the fundamental rights and freedoms of others.

²⁰² Cf. furthermore Möller, Proportionality: Challenging the Critics, 2012 10(3) *International Journal of Constitutional Law* 709–731, <https://doi.org/10.1093/icon/mos024>, 711 et seq.; *European Data Protection Supervisor (EDPS)*, Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit, EDPS, 2017, 4 et seq., https://edps.europa.eu/sites/edp/files/publication/17-04-11_necessity_toolkit_en_0.pdf; *European Data Protection Supervisor (EDPS)*, Guidelines on assessing the proportionality of measures that limit the fundamental rights to privacy and to the protection of personal data, 2019, 6 et seq., https://edps.europa.eu/sites/edp/files/publication/19-12-19_edps_proportionality_guidelines2_en.pdf.

²⁰³ Cf. in total in detail, in particular regarding expropriations: *Kucsko-Stadlmayer* in *Marten/Papier/Kucsko-Stadlmayer*, *Handbuch der Grundrechte VII/1² § 3* paras 99-100; cf. in principle also *Nowak* in *Nowak/Januszewski/Hofstätter* (eds.), *All Human Rights for All – Vienna Manual on Human Rights*, 275; *Winkler* in *Heißl* (ed.), *Handbuch Menschenrechte*, paras 3/58-59.

²⁰⁴ Cf. in total in detail *Kucsko-Stadlmayer* in *Marten/Papier/Kucsko-Stadlmayer*, *Handbuch der Grundrechte VII/1² § 3* paras 101-102; cf. principally also *Winkler* in *Heißl* (ed.), *Handbuch Menschenrechte*, paras 3/58-59.

- Suitable and necessary measures may still be inappropriate considering the relation of the of the respective measure and its aim as well as competing interests, which is particularly important concerning particularly intrusive interferences/means.²⁰⁵

2.9.1 Discussion and Overall Assessment of Residual Risks

[Particularly elaboration on the existence of high risks after the implementation of respective mitigation measures (see particularly explanations in section 2.7.1)]

[Particular explanations on risks concerning which the risk level after measures does not only result in acceptance, but where additional steps such as an insurance are to be implemented]

[Also, assessing residual risks from an overall view, considering the total risk level and amount/variety of such risks (cf. also section 2.10.2 on potential next steps in light of overall residual risk level)]

2.9.2 Proportionality Considerations in the Case at Hand

[Particular explanations on proportionality considerations concerning the object of the assessment and related risks that have been identified and analysed based on the assessments in sections 2.7 and 2.8 and the explanations in section 2.9.1., as well as on general fundamental rights considerations as explained in section 2.6.4, involving specific risks and benefits concerning the object of the assessment as well as the overall perspective]

2.10 General Conclusion of the Impact Assessment

2.10.1 Summary of the Findings

In this section, the findings of the Combined Impact Assessment shall be summarised. As explained in detail above (section 2.6.2.2), certain deployers of certain high-risk AI systems are also obliged to register summarised findings of FRIA and summaries of respective DPIA in the (principally public) database pursuant to Article 71 AI Act. In exact wording, this refers to “[a] *summary of the findings of the fundamental rights impact assessment conducted in accordance with Article 27*” (Annex VIII, Section C, point 4) and “[a] *summary of the data protection impact assessment carried out in accordance with Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680 as specified in Article 26(8) of this Regulation, where applicable*” (Annex VIII, Section C, point 5). Presumably, such a “*summary of the findings*” of the FRIA might refer to a condensed version of its *results* (cf. Article 27(3) AI Act; see section 2.6.2.2).²⁰⁶

As a result of the impact assessment process, it is also advisable to compile a guidance document, comprising and explaining all measures identified to mitigate the risks identified in the risk assessment. The measures identified in the risk assessment can be distinguished into two different categories, those that can be implemented into the system “by design”, i.e. already in the development process, and those which need to be implemented at a later stage, typically when the respective system is put into operation or during operation. Depending on at which point of the development process the first version for the impact assessment report is finalised, the guidance document should contain all measures to be implemented in the remaining development process, if any, and the measures to be implemented at a later stage.

[respective explanations concerning the case at hand]

²⁰⁵ Cf. in total in detail *Kucsko-Stadlmayer* in *Marten/Papier/Kucsko-Stadlmayer*, *Handbuch der Grundrechte VII/1² § 3* paras 103-105; cf. principally also *Winkler* in *Heißl* (ed.), *Handbuch Menschenrechte*, paras 3/58-59.

²⁰⁶ *Fülöp/Poindl* in *Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.3.2; see also the German version of the AI Act, translating both *results* and *findings* in Art. 27(3) and Annex VIII Section C point 4 in the same way as “*Ergebnisse*”.

2.10.2 Following Decisions

Here at last, subsequent decisions based on the process and the results of the Combined Impact Assessment shall be discussed.

[particularly explanations on consequences of results of the assessment and residual risks. In this regard, the focus should be placed on the risks that require further actions (like informing of authorities and further risk handling such as insurance measures; depending on the results in section 2.9), or, concerning which such further actions are considered in any case, as well as the overall view on all (residual) risks. Apart from the case that high or even unacceptable risks remain, where further actions may be necessary, particularly also in case this overall view reveals that exceptionally many risks remain, further actions should be considered from a proportionality perspective (cf. 2.9). Such actions may range from the consultation of competent authorities, further risk handling (e.g. insurance measures) or mitigation, to the suspension of the processes at hand until a lower risk level has been reached]

Furthermore, as partially explained above, responsible entities may also be obliged to inform competent authorities, particularly concerning results of respective impact assessments. Apart from the above-mentioned (public) registration of summaries of impact assessment (findings) pursuant to Article 49(3) in connection with Annex VIII AI Act, this on the one hand concerns the notification of the market surveillance authority of the results of the FRIA pursuant to Article 27(3) AI Act (cf. 2.6.2.2 above), and on the other hand the consultation of the supervisory authority prior to processing pursuant to Article 36 GDPR (also including providing the authority with the DPIA: paragraph 3(e)). The latter must be conducted in case a DPIA pursuant to Article 35 GDPR “*indicates that the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk*”. While the wording of that provision suggests that the consultation may already have to take place if high risks are just identified initially (without considering any potential mitigation measures by the controller), it presumably only refers to situations where the controller cannot (in an economically justifiable manner) mitigate such high risks sufficiently, thus would principally leave the consultation obligation respectively at the discretion of the controller.²⁰⁷

[respective explanations on decisions on submissions and information of competent authorities pursuant to Articles 36 GDPR, 27 and 49 (Annex VIII) AI Act, as well as concerning the further (voluntary) publication, as this is often regarded crucial in respective impact assessment processes²⁰⁸ and is essential to promoting the ethical principle of transparency, which often serves as the foundation for ethical compliance with other principles (such as accountability, safety, fairness, and so forth)]

Moreover, both the AI Act and the GDPR require responsible entities under certain circumstances to undergo respective revision processes concerning corresponding impact assessments. Regarding the FRIA, Article in this regard 27(2) is to be quoted, requiring obliged deployers in case they consider “*during the use of the high-risk AI system*” „*that any of the elements listed in paragraph 1 has changed or is no longer up to date*“ to „*take the necessary steps to update the information*“. Concerning the DPIA, Article 35(11) is of relevance here, principally requiring respective controllers to „*carry out a review to assess if processing is performed in accordance with the data protection impact assessment*“. According to the provision this applies „*[w]here necessary*“ and „*at least when there is a change of the risk represented by processing operations*“, which particularly refers to changes concerning factual or legal circumstances, as well as false assumptions in the

²⁰⁷ See *Trieb in Knyrim*, DatKomm Art. 36 DSGVO para. 12 (status of first September 2019, rdb.at).

²⁰⁸ Cf. *Fülöp/Poindl in Pehlivan/Forgó/Valcke* (eds.), *Artificial Intelligence Act: A Commentary* (Kluwer Law International *forthcoming*) section 3.1.2.

original assessments.²⁰⁹ Respective changes, particularly when required by the aforementioned legal provisions, should furthermore also be visible in the History of Changes at the beginning of this report.

2.10.3 Outlook

Concerning the FRIA pursuant to Article 27 AI Act, it is once more to be noted that the AI Office is yet to publish a template for a questionnaire (including through an automated tool) for the facilitation of the FRIA obligations, it shall develop pursuant to Article 27(5) AI Act. That template should then be considered particularly in respect of the methodology for the FRIA, which are to be amended where appropriate subsequently (cf. section 2.2.1 above).

[Other explanations concerning (foreseeable) developments with relevance for the object of the assessment]

²⁰⁹ See in detail *Trieb in Knyrim*, DatKomm Art. 35 DSGVO paras 89 et seq., particularly also taking into account in a slightly different context respectively if essential changes to the processing are made, such as regarding the nature, scope, circumstances and purposes of the processing.

3 Conclusion, next steps

This report sets out the Ethical and Human Rights Impact Assessment Framework of the Microb-AI-ome project, which is the methodological basis for carrying out the ongoing (impact) assessment of the activities in the project and its results. The methodology combines the necessary elements of a Data Protection Impact Assessment (DPIA) pursuant to Article 35 GDPR and of a Fundamental Rights Impact Assessment (FRIA) pursuant to Article 27 AI Act with an Ethical Impact Assessment (EIA). As a next step, after the relevant design decisions are taken in Microb-AI-ome – a process which is in turn influenced by security, data protection, human rights and ethical considerations – the relevant description of facts has to be written (see section 2.4 above). After that, or in fact rather in parallel, the relevant stakeholders need to be analysed, permissibility, necessity and proportionality of the processes in scope need to be assessed, as well as, in particular, their relevant risks, including ethical considerations. Finally, corresponding mitigation measures need to be identified and described. These activities, which are in fact not strictly following each-other but are carried out partly in parallel, will be the backbone of the work in WP 6 of the Microb-AI-ome project in the remaining years.

Annex A: Ethical Principles

1 Transparency

A main component of the principle of transparency is explainability. This means that the capabilities and purpose of AI systems need to be openly communicated, and decisions explained to those directly and indirectly affected, including the mechanisms behind the decision-making process and their design.²¹⁰ This can also be referred to as “*procedural transparency*”.²¹¹ Moreover, providers and deployers must be able to explain how the system, model, algorithm, etc. operates, which is also referred to as “*technical transparency*”.²¹² Users should be given the knowledge and tools to comprehend and interact with AI systems as well as to understand the systems limitations.²¹³ This also applies to providers and deployers of AI systems who have to ensure that their staff has a sufficient level of AI literacy (Article 4 AI Act). In the event that an explanation as to why a model has generated a particular output or decision is not possible, other measures may be required, including traceability, auditability and transparent communication.²¹⁴ Traceability describes the need to maintain a complete account of the provenance of data, processes, code, and other elements in the development of an AI system in order to capture granular information.²¹⁵ It moreover refers to the documentation of the data sets and processes that yield an AI system’s decision as well as the algorithms used in order to enable identification of the reasons for an AI system’s decision and to facilitate auditability.²¹⁶ Furthermore, users have the right to be informed when they interact with an AI system, where this is not obvious from the outset.

Guiding questions:

- Is the technology in question designed to interact with end-users?
- Is the technology in question designed to guide or take decisions that affect humans and if so, can the decision-making process be traced back?
- In cases where an interaction with the technology in question (particularly an AI- based system) is not obvious to end-users, are they made fully aware when they are interacting with an AI system, as opposed to a human being?
- Are individuals (directly or indirectly) impacted by the AI system made fully aware of when a decision, outcome, content or advice was informed by or made on the basis of an AI system or AI algorithms?

²¹⁰ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 13, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²¹¹ *Algemene Rekenkamer*, An audit for algorithms. Nine algorithms used by the Dutch Government (2022), 34 (<https://english.rekenkamer.nl/binaries/rekenkamer-english/documenten/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government/An+Audit+of+Algorithms.pdf>).

²¹² *Algemene Rekenkamer*, An audit for algorithms. Nine algorithms used by the Dutch Government (2022), 34 (<https://english.rekenkamer.nl/binaries/rekenkamer-english/documenten/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government/An+Audit+of+Algorithms.pdf>).

²¹³ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 16 and 18, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²¹⁴ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 13, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²¹⁵ *OECD*, Advancing accountability in AI. Governing and managing risks throughout the lifecycle for trustworthy AI, no. 349 (2023), 33, https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/advancing-accountability-in-ai_753bf8c8/2448f04b-en.pdf.

²¹⁶ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 18 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

- Are they made aware of the extent to which they are impacted, including the rationale, benefits and limitations of the decision(s)?
- Have appropriate explanations been put in place to help users and other impacted individuals understand the decision-making process and how the system works as well as its purpose, criteria and limitations? Has appropriate training material been provided to end-users?
- Have measures been put in place to continuously assess the quality of the input data and the quality of the outputs?
- Has the decision to adopt the AI system been documented and communicated online?
- Is the algorithm, including its inner-working logic, open to the public or any oversight authority?
- Is the code of the AI system in an open-source format?
- Can public authorities request a copy of the code?
- Are the datasets used for training the system known and traceable?
- Will the system be used by the public or only internally? If the system will only be used internally, what is the level of competency of those who will interact with it?

2 Fairness

There are numerous different interpretations of fairness, and the term involves different dimensions that must be equally taken into account. Firstly, fairness relates to justice and refers to ensuring equal and just distribution of benefits and costs alike. This includes equal opportunity in terms of access to education, goods, services, and technology.²¹⁷ The principle of fairness also includes avoiding unfair bias, discrimination, and stigmatisation, thereby encompassing societal fairness.²¹⁸ This includes that the technology in question should also take account of diversity in the population and does not discriminate.²¹⁹ This encompasses the avoidance of bias, which can be defined as “*the action of supporting or opposing a particular person or thing in an unfair way, because of personal opinions to influence your judgment*”²²⁰ in a way that is inaccurate, closed-minded and prejudicial. With regard to algorithms, the term bias refers to “*systematic and repeatable errors in a computer system that create ‘unfair’ outcomes, such as ‘privileging’ one category over another [...]*”²²¹ It is crucial to differentiate between “*bias in the algorithm*” and “*bias in data*”.²²² Moreover, it is possible to distinguish between several additional types of bias: These include “*historical bias*” (pre-existing biases and patterns in the training data), “*representation bias*” (incomplete information due to a lack of parameters, sample sizes or the underrepresentation of specific sub-groups), “*measurement bias*” (certain variables or characteristics are either disproportionately included or excluded), “*methodological and*

²¹⁷ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 12, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²¹⁸ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 12, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²¹⁹ *Algemene Rekenkamer*, An audit for algorithms. Nine algorithms used by the Dutch Government (2022), 32 (<https://english.rekenkamer.nl/binaries/rekenkamer-english/documenten/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government/An+Audit+of+Algorithms.pdf>).

²²⁰ See the definition in the Cambridge Dictionary at: <https://dictionary.cambridge.org/dictionary/english/bias> (accessed 7. 8. 2024).

²²¹ https://en.wikipedia.org/wiki/Algorithmic_bias (accessed 7. 8. 2024).

²²² *Algemene Rekenkamer*, An audit for algorithms. Nine algorithms used by the Dutch Government (2022), 32-33 (<https://english.rekenkamer.nl/binaries/rekenkamer-english/documenten/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government/An+Audit+of+Algorithms.pdf>).

evaluation bias” (errors in model validation, calibration or monitoring of outcomes), and “*monitoring bias and skewed samples*” (inappropriate interpretation of results during the monitoring process and pre-existing biases that pollute the training data).²²³ The perpetuation of such biases has the potential to result in discriminatory practices against specific individuals or groups. Therefore, it is imperative to address and eliminate these biases at the earliest possible stage of the data collection process. It is also possible that certain biases are deliberately exploited in order to manipulate users, consumers, etc. into a certain behaviour or to engage in unfair competition.²²⁴ Apart from that, fairness encompasses accessibility, diversity, plurality, intergenerational justice, and stakeholder participation.

Guiding questions:

- Has the technology been tested with different groups and was there any difference detectable that could produce discriminatory outcomes or lead to a different performance for different groups?
- What definition of fairness do you use and is it implemented in any phase of the system’s life cycle?
- What criteria are used to determine whether the application is fair?
- Are processes in place to test data against biases?
- Is the data well-balanced and does it reflect the diversity of the targeted end-user population?
- Did you put in place educational and awareness initiatives to help AI designers and developers gain awareness of the possible bias they can introduce through the design and development of the system?
- What mechanisms have been put in place that allow for the flagging of issues related to bias, discrimination or poor performance of the AI system?
- Does the technology in question correspond to the variety of preferences and abilities to society?
- Is the technology in question able to take or inform decisions on the distribution of resources, goods, services, etc. that might lead to injustice and inequality among affected persons?

3 Non-maleficence, Safety, Robustness

In principle, the digital world should be a secure environment for all users and parties, regardless of age, gender, ethnicity, etc. It is therefore essential that the physical and mental integrity of individuals is safeguarded from potential harm or exploitation by technology. This means that systems should neither cause nor exacerbate harm or otherwise adversely affect human beings.²²⁵ This includes individual, collective, social, cultural or political harm. Furthermore, systems need to be safe and secure, as do the environments in which they operate. They need to be secure from malicious use and possess resilience to attack. Another major component of the principle of safety is robustness, referring to the ability to endure or overcome

²²³ See OECD, Advancing accountability in AI. Governing and managing risks throughout the lifecycle for trustworthy AI, no. 349 (2023), 27-28, https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/advancing-accountability-in-ai_753bf8c8/2448f04b-en.pdf.

²²⁴ See Independent High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for trustworthy AI (2019), 18, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²²⁵ Independent High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for trustworthy AI (2019), 12, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

adverse conditions, including digital security risks, and maintain its level of performance.²²⁶ The technology in question should thus have a technically robust foundation, should be developed in a manner such that it reliably behaves as intended and that its behaviour and outputs are reproducible.²²⁷ Their outputs, predictions and judgments need to be as correct and precise as possible, and the system should be able to indicate how likely certain errors are and where they could occur.²²⁸

Guiding questions:

- Could the technology in question have adversarial, critical or damaging effects?
- Are the expected impacts irreversible or difficult to reverse or could they involve life and death decisions? (e.g., setting prison sentences or determining medical treatments)
- Is the technology in question intended to be used or could be adapted for social scoring or mass surveillance? If so, have measures been put in place to safeguard against this?
- What measures were put in place to ensure the safety and security of the technology in question and protect it from system manipulation?
- Is the technology in question compliant with specific security standards?
- What measures were put in place to ensure the safety and security of the system's training data and of the data processed from data poisoning, corruption, and malicious use?
- How often will the technology in question be tested in the future and which components will be tested?
- What security measures are in place to protect against the inappropriate or malevolent use of technology in question?
- Is the system vulnerable to cyber-attacks and if so, what measures have been put in place to ensure the integrity, robustness and overall security of the AI system against potential attacks over its lifecycle?
- Have end-users been informed about existing and potential risks and threats to the technology in question?
- Have measures been put in place to ensure that the data (including training data) used to develop the system is up-to-date, of high quality, complete and representative of the environment the system will be deployed in?
- Have verification and validation methods and documentation (e.g. logging) been put in place to evaluate and ensure different aspects of the system's reliability and reproducibility?
- Have tested failsafe fallback plans been defined to address system errors of whatever origin and governance procedures been put in place?

²²⁶ OECD, Advancing accountability in AI. Governing and managing risks throughout the lifecycle for trustworthy AI, no. 349 (2023), 34, https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/advancing-accountability-in-ai_753bf8c8/2448f04b-en.pdf.

²²⁷ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 16-17, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²²⁸ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 17, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

4 Accountability

The principle of accountability necessitates that mechanisms be put in place to ensure responsibility for systems and their outcomes.²²⁹ This firstly includes auditability, entailing the enablement of the assessment of algorithms, data and design processes, including the evaluation by internal and external auditors. Furthermore, both the ability to report on actions or decisions that contribute to a certain outcome, and to respond to the consequences of such an outcome, must be ensured.²³⁰ Accountability also refers to the documentation of trade-off decisions due to tensions between ethical principles or fundamental rights and to a clearly defined and communicated responsibility of the decision-maker. This particularly applies in cases, where legal provisions do not provide sufficient guidance. This includes that oversight mechanisms be put in place. Additionally, when unjust adverse impact occurs, accessible mechanisms should be foreseen to ensure adequate redress.²³¹

Guiding questions:

- Is it always possible to attribute ethical and legal responsibility for any stage of the lifecycle of the AI system to physical persons or to existing legal entities?
- Who has ultimate decision-making authority within the entity responsible for the technology in question?
- What technical and institutional designs have been put in place to ensure the accountability, auditability and traceability of the technology in question?
- Has there been foreseen any kind of external guidance or third-party auditing processes to oversee ethical concerns and accountability measures?
- Is there a protocol regarding liability allocation in case of adverse effects and damages caused by the technology in question?
- Is there a designated staff member or public sector institution who can review complaints, inform impacted individuals of the explanation(s), and correct the decision if needed?
- Is there a procedure in place to investigate claims raised about the system by the general public, researchers or the media?
- What measures have been put in place for whistle-blower protection?
- Can individuals appeal a decision made by an AI system? If so, are details on how to do so provided to them?
- Has there been organised a risk training for persons working with the technology in question or persons involved in its development and design?

²²⁹ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 19, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²³⁰ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 20, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²³¹ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 20, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

5 Privacy, Data Protection

The need to safeguard people's privacy and to ensure that personal data are handled with care is widely acknowledged in ethics and law.²³² The right to privacy and protection of personal data is a fundamental right enshrined in Article 8 CFR and Article 8 ECHR, which is why most of its scope is already protected by legally binding norms. Beyond that, from an ethical perspective, the principle of privacy necessitates adequate data governance that covers the quality and integrity of the data used. Risks to privacy and data governance can arise at the data and the model levels, at their intersection, as well as during the interaction between humans and the technology in question.²³³ For example, when data is gathered, it may contain socially constructed biases, inaccuracies, errors, or malicious data, which needs to be addressed prior to and while training with any given data set.²³⁴ Moreover, it should be outlined, who has access to the data of individual persons and what data protocols are in place.²³⁵ It is imperative that the security of technologies and the protection of personal data be an integral part of the design process, with the objective of ensuring the privacy and security of users. In this context, informed, freely given and unambiguous consent plays a pivotal role. It also important to adhere to the principles of privacy-by-design and privacy-by-default pursuant to Article 25 GDPR.

Guiding questions:

- Is the system being trained, or was it developed, by using or processing personal data (including special categories of personal data)? If so, what categories of personal data?
- Are the data and input collected by humans, automated sensors or both?
- Is the data being stored at a level of security commensurate to its sensitivity?
- Is the data minimisation principle being applied? In other words, is there an ex-ante assessment of the relevance and necessity of including each one of the data types in the system?
- Has the quality of the training data been evaluated in terms of fairness and non-discrimination?
- Have measures to achieve privacy-by-design and privacy-by-default been put in place, including encryption, pseudonymisation, aggregation, and anonymisation?
- Have the privacy implications of the system's non-personal training-data or other processed non-personal data been considered?
- Was the technology in question aligned with relevant standards or widely adopted protocols for (daily) data management and governance?
- Are there less invasive and less "data hungry" technologies/solutions available?

²³² *Algemene Rekenkamer*, An audit for algorithms. Nine algorithms used by the Dutch Government (2022), 32 (<https://english.rekenkamer.nl/binaries/rekenkamer-english/documenten/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government/An+Audit+of+Algorithms.pdf>).

²³³ *OECD*, Advancing accountability in AI. Governing and managing risks throughout the lifecycle for trustworthy AI, no. 349 (2023), 29, https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/advancing-accountability-in-ai_753bf8c8/2448f04b-en.pdf.

²³⁴ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 17, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²³⁵ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 17, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

6 Beneficence

It should be the ultimate goal of technology to enhance overall well-being and fostering a positive social impact. Especially AI systems are and increase to be prominent in many areas of our lives (education, work, care, entertainment, creative professions, etc.) and may alter our lifestyle. They have the potential to both increase and deteriorate our social relationships, which is why their social impact must be constantly monitored.²³⁶ The technology in question should not be an obstacle to peaceful and just societies, but rather contribute to peace and well-being.²³⁷ The value of living in peaceful and just societies points to the potential of AI systems to contribute throughout their life cycle to the interconnectedness of all living creatures with each other and with the natural environment.

Guiding questions:

- Does the system in question serve people or the public good? If so, to what extent does the use of the system meet their needs and serve their interests?
- What are the objectives of using the technology in question and why is its utilisation important and to whom?
- Could the technology in question have a negative impact on society at large? What measures have been taken to mitigate or eliminate the potential harm?
- Has careful consideration been given to alternative, less invasive options which may be used to achieve the same goal? If so, why is the option involving the technology in question favoured?
- Has the scope of the application of the technology in question been clearly defined? What limitations have been placed on the scope to ensure it remains proportional to the stated objective?
- If the technology in question has already been in operation, how effective has it been in achieving its stated aim?
- Does the technology in question impact human work and work arrangements? If so, have impacted workers and their representatives been informed in advance and have the adopted measures been communicated to the affected (groups of) persons?
- Are whole professions affected by the development of this technology and what would this mean for workers in this field?

7 Autonomy, Freedom, Human Agency

The requirement that people using technology maintain complete and effective self-determination over themselves, as well as the human-centric allocation of functions between humans and technology, is known as the principle of autonomy. The technology in question should not force, control, subjugate, deceive, or herd people in an unjustified manner.²³⁸ This includes that human behaviour and thinking are not shaped or manipulated through mechanisms that are hard to detect or influence sub-conscious process, ensuring the fundamental right to freedom of opinion and thought. This is closely linked to human agency, meaning that

²³⁶ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 19, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²³⁷ UNESCO, Recommendations on the Ethics of Artificial Intelligence (2021), 19-20, https://unesdoc.unesco.org/ark:/48223/pf0000381137_eng.

²³⁸ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 12, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

users should be able to make informed autonomous decisions regarding technology, and digital sovereignty.²³⁹ Human oversight mechanisms help ensuring that a system does not undermine human autonomy or causes other adverse effects and can be achieved through governance mechanisms, including *human-in-the-loop*, *human-on-the-loop*, or *human-in-command*.²⁴⁰ Closely linked to autonomy and freedom is the so-called “*right to opt out*”. This refers to the requirement, that humans must be provided with the option to decide against the interaction with a technology in favour of human interaction.²⁴¹

Guiding questions:

- Is the system (a) replacing an existing computer system; (b) replacing human beings; (c) adding new functionality or supplementing existing functionality?
- If the system took over a task that was previously conducted by humans, how was the knowledge transfer preserved? How involved were the humans who were conducting the task previously in the development and training of the system?
- Does the technology in question have the authority to make a decision that would impact people? If yes, is the decision subjected to meaningful human oversight before it takes effect?
- Which detection, response, and control mechanisms for undesirable adverse effects of the technology in question were established?
- What recourse is available to those affected by the technology in question, and how can they opt out of AI-based decision-making processes?
- Can the technology in question take any decisions which the physical persons or legal entities in charge of the system lack expertise or competence to critique, modify or override?
- Does the system in question risk creating human attachment, dependency, manipulation or a simulation of social interaction?
- Is the system a self-learning or autonomous system?
- Is the system overseen by a *Human-in-the-loop*, *Human-on-the-loop* or *Human-in-command*?
- Is it possible to “*pull the emergency brake*” on a process in order to safely abort an operation when needed?
- Are the persons to whom human oversight has been delegated aware of a possible tendency towards overconfidence in the outcome produced by the technology in question (“*automation bias*”) and are they able to correctly interpret the output of the system? Do they adequately understand the capabilities and limitations of the system in question and adequately monitor its operation?
- Is the technology in question capable of exploiting addictions and vulnerabilities?

²³⁹ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 16, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²⁴⁰ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 16, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²⁴¹ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 16, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

8 Sustainability

The development, design, deployment, and utilisation of an AI system should be assessed with regard on its impact on the environment, its usage of resources, its impact on the achievement of the Sustainable Development Goals, its energy consumption and its environmental friendliness.²⁴² Furthermore, environmental and ecosystem flourishing should be recognised, promoted and protected by the technology in question.²⁴³

Guiding questions:

- Are there potential negative impacts of the technology in question on the environment, including the high use of resources and energy consumption?
- Where possible, were mechanisms established to evaluate the environmental impact of the system's development, deployment and/or use (for example, the amount of energy used and carbon emissions)?
- Does the technology in question comply with applicable environmental laws and policies?
- Are accountability metrics for responsible innovation (SDGs, ESGs) used to project how the system can increase environmental flourishing (long-term sustainability) versus just avoiding immediate, likely regional and short-term harms?
- Was there an estimation of the environmental impact of raw material extraction, processing and transportation involved in manufacturing the hardware of the system?
- Once the system is decommissioned, how will the process of dismantling, recycling and/or disposing of obsolete IT hardware be handled?

9 Human Dignity

The development, design, deployment, and use of technologies should be oriented towards the objective of respecting and protecting human dignity. The concept of human dignity is based on the idea that every person has an “*inherent*” value simply by virtue of being human, which must never be compromised by others. As an “*inescapable premise of normative obligations*”²⁴⁴, human dignity is understood as the foundation of mutual respect and especially applies in cases where new technologies and AI systems are used. This implies that technologies must be designed in a manner that preclude any infringement upon fundamental rights and freedoms. It is imperative that new technologies enhance the quality of life for those affected and facilitate the realisation of fundamental rights and freedoms, rather than impeding their fulfilment. Moreover, individuals should be regarded as moral agents, rather than as mere objects. Humans must never be solely treated as a means to an end, but rather as an end in themselves and shall not be manipulated through algorithms. This encompasses the right not to be subjected to automated decisions

²⁴² See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 19, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²⁴³ UNESCO, Recommendations on the Ethics of Artificial Intelligence (2021), 19, https://unesdoc.unesco.org/ark:/48223/pf0000381137_eng.

²⁴⁴ Bielefeldt, Universalität und Gleichheit, in: *Pollmann/Lohmann* (ed.), *Menschenrechte. Ein interdisziplinäres Handbuch* (2012), 162.

without one's knowledge (Art 22 GDPR) and it is of great importance to ascertain whether the system is engaged in tasks that are relevant to decision-making processes.²⁴⁵

Guiding questions:

- Is there any possibility that the technology in question is able to manipulate, subjugate, coerce or objectify human beings?
- Have appropriate oversight mechanisms been put in place?
- Does the operation of the technology in question involve the treatment of humans as an object or as a means to an end?
- Is the system capable of taking automated decisions without human interference?

10 Solidarity, Inclusion, Accessibility

Universal access to the internet, digital skills and fair working conditions should be guaranteed. It is of utmost importance to ensure that technologies are accessible to people with different abilities, backgrounds and cultures, that they are interactive and that the population benefits from the use of the technology in question. Particular attention must be paid to vulnerable and marginalised groups and they should be involved in the development, deployment, and utilisation of the technology in question. Additionally, special consideration needs to be given to scenarios in which asymmetries in knowledge or power, such as those between companies and employees, governments and citizens, might result in or worsen negative effects caused by the system.²⁴⁶ The technology should also take into account cultural differences, cultural identities and different ways of life.

Guiding questions:

- What stakeholder groups are most likely to be impacted by the deployment of the AI system?
- Who has the greatest needs for this tool?
- Who has the least power to influence the development of this tool?
- Does the design allow all people, especially vulnerable marginalised groups, to access and interact with the system? Is the user interface usable by those with disabilities (e.g., accessible to screen readers, including alt text for images, colour-blind friendly palettes, etc.) or those at risk of exclusion?
- Has there been an assessment whether the system is usable by those with a precarious economic situation or by the elderly?
- To which segment of the population will the system be applied? Is the population affected vulnerable or marginalised or particularly affected in the given context?

11 Participation

The principle of participation encompasses different dimensions. Firstly, it refers to a direct form of participation of relevant stakeholders in the development, design, deployment, and use of the technology in

²⁴⁵ See *Bundesministerium für Kunst, Kultur, öffentlicher Dienst und Sport (BMKÖS)*, Leitfaden Digitale Verwaltung und Ethik. Praxisleitfaden für KI in der Verwaltung, Version 1.0 (2023), 18.

²⁴⁶ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 12, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

question through stakeholder participation. This is because stakeholders who may directly or indirectly be affected by the system throughout the system's life cycle should be able to give feedback and participate in a consultation process.²⁴⁷ On the other hand, the system in question should not undermine but facilitate participation of society as a whole in decision-making processes, such as elections. Additionally, technology should aim at closing the digital divide rather than exacerbating it.

Guiding questions:

- Were stakeholders engaged in any phase of the system's life cycle? If so, explain the criteria of the selection of stakeholders. If not, please explain why no stakeholder participation was considered.
- What mechanisms were considered to include participation of the widest range of possible stakeholders?
- Has there been a public announcement regarding the intention to design this technology and its possible impact? Can information be found online regarding the system, its capabilities, its purpose and functionality? If not, is there a plan to publish this information at a particular stage of the project lifecycle?
- Is the language used to present the system appropriate for the general public?
- If the system will be used by the public, can people report their experience interacting with the system and concerns related to its impacts? Is the process for doing so simple, accessible and clearly advertised?
- Have any schemes been put in place to help educate users and impacted groups about this system and the reason behind its deployment? For example, an educational media campaign or workshops involving community leaders?

12 Democracy

The assessment of the development, deployment and use of the technology in question should take into account its effect on institutions, democracy and society at large. Especially the use of AI systems should be given careful consideration particularly in situations relating to the democratic process, including not only political decision-making but also electoral contexts.²⁴⁸ It is imperative that inclusive review mechanisms are put in place to ensure that democratic decision-making processes, pluralism, access to information and economic and social rights are safeguarded in relation to the development and use of technology. It is crucial that technology in question does not compromise the independence of the judiciary, due process, and impartiality. Additionally, technology must not be developed, deployed and used in a way that it is able to undermine the rule of law, particularly with regard to its use in public administration or judiciary. Additionally, adherence to fundamental and human rights frameworks is of the utmost importance.²⁴⁹

Guiding questions:

²⁴⁷ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 19, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²⁴⁸ See *Independent High-Level Expert Group on Artificial Intelligence*, Ethics Guidelines for trustworthy AI (2019), 19, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

²⁴⁹ OECD, Advancing accountability in AI. Governing and managing risks throughout the lifecycle for trustworthy AI, no. 349 (2023), 31, https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/advancing-accountability-in-ai_753bf8c8/2448f04b-en.pdf.

- Could the technology in question have negative impacts on democracy? What measures have been taken to mitigate or eliminate the potential harm?
- Is there a risk that the technology in question negatively impacts the conduction of elections or other principles of democracy?
- Is there a risk that the technology in question negatively impacts the right to a fair trial or other fundamental legal principles?
- Does the technology comply with applicable legal norms, especially with fundamental and human rights instruments?

13 Efficiency

The deployment of technology in a specific sector must be conducted in a manner that enhances the effectiveness and efficiency of the sector's processes without adversely impacting the work environment of those employed in this sector and without impairment of the overall situation. It is essential to define clear criteria that can be used to distinguish between instances where the utilisation of technology results in improved efficacy, and instances where it has a detrimental impact on the work situation of the human situation as a whole.

Guiding questions:

- What criteria are used to determine when and whether the use of the technology in question improves the effectiveness and efficiency?
- What criteria are used to determine whether the work situation of people working in the affected sector is worsened by the use of a technology?
- Does the technology in question relieve staff of monotonous and stressful routine tasks, reduce waiting times or improve the quality of services and decisions?

14 Trust

The compliance with the aforementioned principles is of utmost importance to ensure trust in technology. This includes trustworthy systems, researchers, developers, organisations, compliance tools, design principles and trustworthy technology in general. Suggestions for building or sustaining trust include education, reliability, accountability, processes to monitor and evaluate the integrity of technology over time, and tools and techniques ensuring compliance with norms and standards.²⁵⁰ Nevertheless, while trust in the technology in question should be encouraged, it is also important to avoid placing excessive trust in it.

Guiding questions:

- Which measures have been put in place to ensure trust in the technology in question among its end-users and the general public?
- With which norms and standards does the technology in question comply?

²⁵⁰ Jobin et al., The global landscape of AI ethics guidelines, nature machine intelligence (2019), 389 (395).

Annex B: Ethical Recommendations

- AI literacy should be facilitated among staff to foster a basic understanding for AI-based technologies, application and impacts. Public bodies should train the general population in the usage of AI-based technologies. →transparency
- Prior to deploying an AI-based technology or prior to starting a project that involves the utilisation of AI, public bodies should inform citizens about the project goals, planned results, the use of data and the fundamental methodology. →transparency
- The public should be able to question and seek reviews of the technology in question →transparency
- It must be made clear to users that they are interacting with an AI system—especially for systems that simulate human communication, such as chatbots. →transparency
- The purpose, capabilities, limitations, benefits and risks of the AI system and the decisions it makes must be openly communicated to all stakeholders. →transparency
- AI systems must be constructed so that people can audit, query, dispute or seek to change its activities. This includes organizational processes by which the operators can receive and assess requests from third parties. →transparency
- Transparency requires that development processes and tools record these ethical design decisions so that it is possible to understand how ethical obligations were met. This information may be required for audits, for disputing decisions made by the system or for correcting any ethical issues which arise after deployment. →transparency
- Whenever relevant, AI decisions should be explainable to users. Where possible this should include the reasons why the system made a particular decision. We recognize that this may not be possible with some systems. Nevertheless, the system (or those deploying it) should always have a mechanism by which to explain what the decision was and what data were used to make that decision. Explainability is especially important for systems that make decisions or perform actions for which accountability may be required, such as decisions that can cause harm or restrict an individual's rights. →transparency
- Training data must be diverse and representative, particularly with regard to detecting and avoiding bias. →fairness
- The data must be regularly – both during the design and during the use of the technology in question – checked for the effect of bias →fairness
- Measures to guarantee cybersecurity and avoid adversarial attacks must be taken in cooperation with the relevant departments (e.g. IT-unit). (Higher standards have to be maintained for critical infrastructure) →safety
- Potential risks need to be assessed, evaluated, mitigated and eliminated where possible →safety
- Development of a security concept, which protects the system from misuse and guarantees the protection of the mental and physical integrity of end-users. →safety
- AI systems should be safe to use and should not have a propensity to harm or significantly reduce the health and physical or psychological well-being of any stakeholders (users, clients, data subjects, and other affected parties). →safety

- Those who are harmed must have access to effective legal remedies to redress the harm. Moreover, the specific parties responsible must take appropriate measures to ensure that any potential harm caused by the use of the technology is adequately compensated. →accountability, rule of law, trust
- AI systems should not negatively impact the quality of communication, social interaction, information, social relations or democratic processes; for example, by amplifying fake news or segregating people into filter bubbles. →rule of law, beneficence
- Appropriate review mechanisms, audits and due diligence obligations need to be developed and implemented. →accountability
- Ensure that end-users have the possibility to report back on errors or problems of the system →accountability
- Where relevant and practical, the system should generate human accessible logs of the AI system's internal processes. →accountability, transparency
- Privacy-by-design-approach pursuant to Article 25 GDPR has to be implemented. →privacy
- Less invasive and "*data hungry*" technologies should be given priority over the system at hand →privacy
- Data should be acquired, stored and processed in a manner which can be audited by humans →privacy
- Measures must be in place to safeguard the rights of data subjects through technical measures, such as anonymization, as well as through organisational measures, such as access control systems. →privacy
- It should be outlined in protocols who can access and delete data under which circumstances. →privacy
- AI projects should clearly demonstrate the clear benefits for the community and provide a significant gain in knowledge that could not have been realised without the use of AI. →beneficence
- Implementation of "*Self-Sovereign Identity*" and guarantee of the possibility to "*opt-out*". →autonomy
- Utilisation of the least environment-invasive and most energy saving technology. →sustainability
- Environmental and ecosystem flourishing should be recognized, protected and promoted through the life cycle of AI systems. →sustainability
- AI development should be mindful of the principles of environmental sustainability, both regarding the system itself and the supply chain to which it connects. →sustainability
- It should be avoided to subject individuals to automated decision-making where possible and it should be clear, where there is a human-in-the-loop, human-on-the-loop and human-in-command. It should always be possible to "*pull the emergency brake*" on a process. →autonomy and human oversight, human dignity
- Stakeholder consultation and Co-Creation-processes →solidarity, inclusion, participation
- The scope of lifestyle choices, beliefs, opinions, expressions or personal experiences, including the optional use of the technology in question should not be restricted during any phase of the life cycle of AI systems. →solidarity
- Clear definition of purpose of the technology and how this purpose is justified →trust

- Measures to ensure compliance with the rights of the persons affected →accountability, rule of law, trust
- An AI system should not be designed or used in a manner which deprives people of the ability to make decisions which they should be able to make for themselves. →autonomy
- AI applications should be designed to give system operators and (as much as possible), end-users the ability to control, direct and intervene in operations of the system. →autonomy
- AI systems should never make the final decision about important issues of a personal, moral or political nature. They may recommend, but the final decision must always be made by a human. →autonomy
- An AI should not be designed or used in a way that results in the reduction of basic human freedoms, including freedom of movement, assembly, speech and information →human dignity
- AI systems should not be designed or used to subordinate, coerce, deceive, manipulate, objectify or dehumanize people. →Human Dignity
- AI systems should not be designed or used to create addiction to the system or to the services which it provides. →Human Dignity
- The processes of the life cycle of AI systems should not objectify human beings or divide and turn individuals and groups against each other. →human dignity
- AI systems should, to the extent relevant and possible, be universally accessible and offer the same functionality and benefits to end-users irrespective of their different abilities, beliefs, preferences or interests. →solidarity, accessibility, inclusion
- Demonstration of a clear benefit or insight of the technology's development, deployment, or use →beneficence
- It should be communicated how the technology in question aligns with core values and principles of the community, organisation, society, population, etc. affected by it. →beneficence
- The use of the technology in question should be appropriate and proportional to achieve a given legitimate aim →beneficence

Microb-AI-ome consortium partners

- Universitaet Hamburg (UHAM), DE
- University College Cork – National University Of Ireland, Cork (UCC), IE
- GNOME Design SRL (GND), RO
- tp21 GmbH (TP21), DE
- Research Institute AG & CO KG (RI), AT
- Institut National De Recherche Pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), FR
- Assistance Publique Hopitaux de Paris (APHP), FR
- Mater Misericordiae University Hospital (MMUH), IE



The Microb-AI-ome project has received funding from the European Union's Horizon Europe research and innovation programme under the grant agreement Nr. 101079777.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.