Chapter 4
Regulating AI

Contents

Abstract Although the debate on AI regulation is still fluid at a global level and the European initiatives are in their early stages, three possible approaches to grounding AI regulation on human rights are emerging. One option is a principles-based approach, comprising guiding principles derived from existing binding and non-binding international human rights instruments, which could provide a comprehensive framework for AI. A different approach focuses more narrowly on the impacts of AI on individual rights and their safeguarding through rights-based risk assessment. This is the path followed by the Council of Europe in its ongoing work on AI regulation. Finally, as outlined in the EU proposal, greater emphasis can be placed on managing high-risk applications by focusing on product safety and conformity assessment. Despite the differences between these three models, they all share a core concern with protecting human rights, recognised as a key issue in all of them. However, in these proposals for AI regulation, the

A. Mantelero, *Beyond Data*, Information Technology and Law Series 36,
https://doi.org/10.1007/978-94-6265-531-7_4

emphasis on risk management is not accompanied by effective models for assessing the impact of AI on human rights. Analysis of the current debate therefore confirms that the HRESIA could not only be an effective response to human-rights oriented AI development that also encompasses societal values, but it could also bridge a gap in the current regulatory proposals.

Keywords Ad hoc Committee on Artificial Intelligence (CAHAI) · AI regulation · Artificial Intelligence Act · Conformity assessment · Co-regulation · Democracy · Technology assessment

4.1 Regulating AI: Three Different Approaches to Regulation

In its early stages, the regulatory debate on AI focused mainly on the ethical dimension of data use and the new challenges posed by data-intensive systems based on Big Data and AI. This approach was supported by several players of the AI industry, probably attracted by the flexibility of a self-regulation based on ethical principles, which is less onerous and easier to align with corporate values.[1]

As in the past, uncertainty about the potential impact of new technology and an existing legal framework not tailored to the new socio-technical scenarios was the main reason for rule makers to turn their gaze towards general principles and common ethical values.

The European Data Protection Supervisor (EDPS) was the first body to emphasise the ethical dimension of data use, pointing out how, in light of recent technological developments, data protection appeared insufficient to address all the challenges, while ethics "allows this return to the spirit of the [data protection] law and offers other insights for conducting an analysis of digital society, such as its collective ethos, its claims to social justice, democracy and personal freedom".[2]

This ethical turn was justified by the broader effects of data-intensive technologies in terms of social and ethical impacts, including the collective dimension of data use.[3] In the same vein, the European Commission set up a high-level group focusing on ethical issues.[4] This ethical wave later resulted in a flourishing of ethical principles, codes and ethical boards in private companies.[5]

[1] E.g., Center for Data Innovation 2021.

[2] European Data Protection Supervisor, Ethics Advisory Group 2018, 7. See also European Data Protection Supervisor 2018; European Data Protection Supervisor 2015.

[3] Mantelero 2016. See also Ferguson 2017; Goodman and Powles 2019.

[4] Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission 2019.

[5] See also Taylor and Dencik 2020.

This new focus, which also presented the danger of 'ethics-washing',[6] had the merit of shedding light on basic questions of the social acceptability of highly invasive predictive AI. Such systems may be legally compliant, while at the same time raising crucial questions about the society we want to create, in terms of technological determinism, distribution of power, inclusiveness and equality.

But the ethical debate frequently addressed challenging questions within a rather blurred theoretical framework, with the result that ethical principles were sometimes confused with fundamental rights and freedoms, or principles that were already part of the human rights framework were simply renamed.

A rebalancing of the debate has come from the different approach of the Council of Europe, which has remained focused on its traditional human rights-centred mission,[7] and the change of direction of the European Commission with a new bundle of proposals for AI regulation.[8] These bodies do not marginalise the role of ethics, but see moral and social values as complementary to a strategy based on legal provisions and centred on risk management and human rights.[9]

There are three possible approaches to grounding future AI regulation on human rights, which differ depending on the context in which they are placed – international or EU – and their focus.

The first is the principles-based approach, designed mainly for an international context characterised by a variety of national regulations. Here a set of key principles is clearly needed to provide a common framework for AI regulation at the regional or global level.

The second approach, also designed for the international context, is more focused on risk management and safeguarding individual rights. This approach taken by the Council of Europe, can be complementary to the first one, where the former sets out the key principles and the latter contextualises human rights and freedoms in relation to AI by adding rights-based risk management.

The third approach, embodied by the EU proposal on AI regulation, puts a greater emphasis on (high) risk management in terms of product safety and a conformity assessment. Here the regulatory strategy on AI is centred on a predefined risk classification, a combination of safety and rights protections and standardised processes.

These three models therefore offer a range of options, from a general principles-based approach to a more industry-friendly regulation centred on a

[6] Wagner 2018a.

[7] At its 1353rd meeting on 11 September 2019, the Committee of Ministers of the Council of Europe established an Ad Hoc Committee on Artificial Intelligence (CAHAI) to examine the feasibility and potential elements, on the basis of broad multi-stakeholder consultations, of a legal framework for the development, design and application of artificial intelligence, based on Council of Europe's standards on human rights, democracy and the rule of law.

[8] European Commission 2020d. See also European Commission 2020b.

[9] On the relationship between human right and fundamental rights, see Chap. 1, fn. 90.

conformity assessment of high-risk AI systems. Despite these differences, human rights remain a key element of all of them, though with significant distinctions in emphasis.

All these models also adopt the same co-regulation schema combining hard law provisions with soft-law instruments. This gives the framework flexibility in a field characterised by the rapid evolution of technology and emergence of new issues, while also giving space to sector-specific challenges and bottom-up initiatives.

The HRESIA framework can contribute to all three models by providing a human rights-centred perspective and bridging the two phases of the AI debate by combining a legal framework that takes into account ethical and societal issues with an operational focus that is often absent in the current proposals.

4.2 The Principles-Based Approach

The starting point in identifying the guiding principles that, from a human rights perspective, should underpin future AI regulation is to analyse the existing international legally binding instruments that necessarily represent the general framework in this field. This includes a gap analysis to ascertain the extent to which the current regulatory framework and its values properly address the new issues raised by AI.

Moreover, a principles-based approach focusing on human rights has to consider the state of the art with a view to preserving the harmonisation of the human rights framework, while introducing coherent new AI-specific provisions.

This principles-based approach consists in a targeted intervention, as it focuses on the changes AI will bring to society and not on reshaping every area where AI can be applied. The identification of key principles for AI builds on existing binding instruments and the contextualisation of their guiding principles.

Both the existing binding instruments and the related non-binding implementations – which in some cases already contemplate the new AI scenario – must be considered. This is based on the assumption that the general principles provided by international human rights instruments should underpin all human activities, including AI-based innovation.[10]

Defining key principles for the future regulation of AI through analysis of the existing legal framework requires a deductive methodology, extracting these principles from the range of regulations governing the fields in which AI solutions may be adopted. Two different approaches are possible to achieve this goal: a theoretical rights-focused approach and a field-focused approach based on the provisions set out in existing legal instruments.

In the first case, the various rights enshrined in human rights legal instruments are considered independently and in their abstract notion,[11] looking at how AI

[10] Council of Europe, Committee of Ministers 2020.

[11] Fjeld et al. 2020; Raso et al. 2018.

might affect their exercise. In the second, the focus shifts to the legal instruments themselves and areas they cover, to assess their adequacy in responding to the challenges that AI poses in each sector, from heath to justice.

From a regulatory perspective, and with a view to a future AI regulation, building on a theoretical elaboration of individual rights may be more difficult as it entails a potential overlap with the existing legal instruments and may not properly deal with the sectoral elaboration of such rights. On the other hand, a focus on legal instruments and their implementation can facilitate better harmonisation of new provisions on AI within the context of existing rules and binding instruments.

Once the guiding principles have been identified, they should be contextualised within the scenario transformed by AI, which in many cases requires their adaptation. The principles remain valid, but their implementation must be reconsidered in light of the social and technical changes due to AI.[12] This delivers a more precise and granular application of these principles so that they can provide a concrete contribution to the shape of future AI regulation.

This principles-based approach requires a vertical analysis of the key principles in each of the fields regulated by international instruments, followed by a second phase considering the similarities and common elements across all fields. Ultimately, such an approach should valorise the individual human rights, but departing from the existing legal framework and not from an abstract theoretical notion of each right and freedom.

As the existing international instruments are sector-specific and not rights-based, the focus of the initial analysis is on thematic areas and then a set of guiding principles common to all areas is developed. These shared principles can serve as the cornerstone for a common core of future AI provisions.

A key element in this process is the contextualisation of the guiding principles and legal values, taking advantage of the non-binding instruments which provide granular applications of the principles enshrined in the binding instruments.

AI technologies have an impact on a variety of sectors[13] and raise issues relating to a large body of regulatory instruments. However, from a methodological point of view, a possible principles-based approach to AI regulation can be validated by selecting a few key areas where the impact of AI on individuals and society is particularly marked and the challenges are significant. This is the case for data protection and healthcare.

The intersection between these two realms is interesting in view of future AI regulation, given the large number of AI applications concerning healthcare data

[12] This is the case, for example, with freedom of choice using so-called AI black boxes.

[13] See also UNESCO 2019.

and the common ground between the two fields. This is reflected in several provisions of international binding instruments,[14] as well as non-binding instruments.[15] Individual self-determination also plays a central role in both these fields, and the challenges of AI – in terms of the complexity and opacity of medical treatments and data processing operations – are therefore particularly relevant and share common concerns.

4.2.1 Key Principles from Personal Data Regulation

Over the past decade, the international regulatory framework in the field of data protection has seen significant renewal. Legal instruments shaped by principles defined in the 1970s and 1980s no longer responded to the changed socio-technical landscape created by the increasing availability of bandwidth for data transfer, data storage and computational resources (cloud computing), the progressive datafication of large parts of our life and environment (The Internet of Things, IoT), and large-scale and predictive data analysis based on Big Data and Machine Learning.

In Europe the main responses to this change have been the modernised version of Convention 108 (Convention 108+) and the GDPR. A similar redefinition of the regulatory framework has occurred, or is ongoing, in other international contexts – such as the OECD[16] – or in individual countries.

However, given the rapid development of the last wave of AI, these new measures fail to directly address some AI-specific challenges and several non-binding instruments have been adopted to bridge this gap, as well as future regulatory strategies under discussion.[17] This section examines the following data-related international non-binding legal instruments: Council of Europe, Guidelines on Artificial Intelligence and Data Protection [GAI];[18] Council of Europe, Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data [GBD];[19] Recommendation CM/Rec(2019)2 of the Committee of Ministers of the Council of Europe to member States on the

[14] E.g. the provisions of the Oviedo Convention (Council of Europe, Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, Oviedo, 4 April 1997) and Convention 108+ (Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data, adopted by the Committee of Ministers of the Council of Europe at its 128th Session of the Committee of Ministers, Elsinore, 18 May 2018).

[15] Council of Europe, Committee of Ministers 2019.

[16] OECD 2013.

[17] European Commission 2020c, d. See also European Commission 2020a.

[18] Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) 2019.

[19] Council of Europe, Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) 2017.

protection of health-related data [CM/Rec(2019)2];[20] Recommendation CM/Rec (2010)13 of the Committee of Ministers of the Council of Europe to member States on the protection of individuals with regard to automatic processing of personal data in the context of profiling [CM/Rec(2010)13]; UNESCO, Preliminary Study on a Possible Standard-Setting Instrument on the Ethics of Artificial Intelligence, 2019 [UNESCO 2019];[21] OECD, Recommendation of the Council on Artificial Intelligence, 2019 [OECD];[22] 40th International Conference of Data Protection and Privacy Commissioners, Declaration on Ethics and Data Protection in Artificial Intelligence, 2018 [ICDPPC].[23,24]

These instruments differ in nature: while some instruments define specific requirements and provisions, others are mainly principles-based instruments setting out certain guidelines but without, or only partially, providing more detailed rules.

Based on these instruments and focusing on those provisions that are most pertinent to AI issues,[25] it is possible to identify several general guiding principles which are then contextualised with respect to AI. Several of these principles can be extended to non-personal data, mainly in regard to the impact of its use (e.g. aggregated data) on individual and groups in decision-making processes.

A first group of principles (the primacy of the human being, human control and oversight, participation and democratic oversight) concerns the relationship between humans and technology, granting the former – either as individuals or social groups – control over technological development, in particular regarding AI.

To refine the key requirements enabling human control over AI and support human rights-oriented development, we can identify a second set of principles focussed on the following areas: transparency, risk management, accountability, data quality, the role of experts and algorithm vigilance.

Finally, the binding and non-binding international instruments reveal a further group of more general principles concerning AI development that go beyond data protection. These include rules on interoperability between AI systems,[26] as well as digital literacy, education and professional training.[27]

[20] This Recommendation has replaced Council of Europe, Committee of Ministers 1997. See also Council of Europe, Committee of Ministers 2016b and its Explanatory Memorandum.

[21] Despite the reference to ethics only in the title, the purpose of the study UNESCO 2019 is described as follows: "This document contains the preliminary study on the technical and legal aspects of the desirability of a standard-setting instrument on the ethics of artificial intelligence and the comments and observations of the Executive Board thereon".

[22] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Accessed 2 March 2020.

[23] The text of the Declaration is available at https://edps.europa.eu/sites/edp/files/publication/icdppc-40th_ai-declaration_adopted_en_0.pdf. Accessed 2 March 2020.

[24] See also Council of Europe, Committee of Ministers 2020.

[25] For a broader analysis of the issues related to data protection and human rights in general, Council of Europe-Committee of experts on internet intermediaries (MSI-NET) 2018; Mantelero 2018a; Zuiderveen Borgesius 2018. See also Fjeld et al. 2020; Raso et al. 2018.

[26] See also CM/Rec(2019)2, 1, para 14.

[27] ICDPPC, OECD, GAI para III.9, UNESCO 2019, and CM/Rec(2020)1, para 7.

4.2.1.1 Primacy of the Human Being

Although this principle is only explicitly enshrined in the Oviedo Convention and not in the binding international instruments on data protection, such as Convention 108 and 108+, the primacy of the human being is an implicit reference when data is used in the context of innovative technologies.[28] This is reflected in the idea that data processing operations must "serve the data subject".[29] More generally, the primacy of the human being over science is a direct corollary of the principle of respect for human dignity.[30] Dignity is a constitutive element of the European approach to data processing,[31] and of the international approach to civil and political rights in general.[32] Wider reference to human dignity can also be found in the non-binding instruments focused on AI.[33]

In affirming the primacy of the human being within the context of artificial intelligence, AI systems must be designed to serve mankind and the creation, development and use of these systems must fully respect human rights, democracy and the rule of law.

4.2.1.2 Human Control and Oversight

Since the notion of data protection originally rested on the idea of control over use of information in information and communication technology and the first data protection regulations were designed to give individuals some counter-control over the data that was collected,[34] human control plays a central role in this area. It is also related to the importance of self-determination[35] in the general theory of personality rights and the importance of human oversight in automated data processing.

Moreover, in the field of law and technology, human control plays an important role in terms of risk management and liability. Human control over potentially harmful technology applications ensures a degree of safeguard against the possible adverse consequences for human rights and freedoms.

[28] Council of Europe, Parliamentary Assembly 2017. See also Strand and Kaiser 2015, 6.

[29] CM/Rec(2019)2, Preamble.

[30] ten Have and Jean 2009, 93.

[31] Convention 108+, Preamble. See also Explanatory Report, para 10 ("Human dignity requires that safeguards be put in place when processing personal data, in order for individuals not to be treated as mere objects").

[32] International Covenant on Civil and Political Rights, Preamble.

[33] GAI, paras I.1 and II.1; UNESCO 2019, para II.3, OECD, para IV.1.2.

[34] See Chap. 1, Sect. 1.2.

[35] See also ICDPPC, para 1.1; Universal Declaration of Human Rights.

Human control is thus seen as critical from a variety of perspectives – as borne out by both Convention 108+[36] and the non-binding instruments on AI[37] – and it also encompasses human oversight on decision-making processes delegated to AI systems. Several guiding principles for future AI regulation can therefore be discerned in the instruments examined.

By contextualising human control and oversight with regard to AI applications, these applications should allow meaningful[38] control by human beings over their effects on individuals and society. Moreover, AI products and services must be designed in such a way to grant individuals the right not to be subject to a decision which significantly affects them taken solely on the basis of automated data processing, without having their views taken into consideration. In short, Al products and services must allow general human control over them.[39]

Finally, the role of human intervention in AI-based decision-making processes and the freedom of human decision makers not to rely on the result of the recommendations provided using AI should be preserved.[40]

4.2.1.3 Participation and Democratic Oversight on AI Development

Turning to the collective dimension of the use of data in AI,[41] human control and oversight cannot be limited to supervisory entities, data controllers or data subjects. Participatory and democratic oversight procedure should give voice to society at large, including various categories of people, minorities and underrepresented groups.[42] This supports the notion that participation in decision-making serves to

[36] Convention 108+, Preamble ("[Considering that it is necessary to secure] personal autonomy based on a person's right to control of his or her personal data and the processing of such data"). See also Explanatory Report, para 10.

[37] Council of Europe, Parliamentary Assembly 2017, para 9.3 ("the need for any machine, any robot or any artificial intelligence artefact to remain under human control") and GAI, para I.6.

[38] The adjective meaningful was discussed in the context of AWS, Moyes 2016. The author explains his preference for the adjective thus: "it is broad, it is general rather than context specific (e.g. appropriate), derives from an overarching principle rather being outcome driven (e.g. effective, sufficient), and it implies human meaning rather than something administrative, technical or bureaucratic". See also Asaro 2016, pp. 384–385. The term has been used to insist that automated tools cannot relegate humans to mere approval mechanisms. The same reasoning underpins human oversight in data processing in Europe, see Article 29 Data Protection Working Party 2018, p. 21 ("To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the relevant data").

[39] See Convention 108+; GAI, para II.8; ICDPPC; UNESCO 2019.

[40] GAI, para III. 4.

[41] Mantelero 2016.

[42] See also CM/Rec(2020)1, para 5.

advance human rights and is crucially important in bringing specific issues to the attention of the public authorities.[43]

Since human control over potentially hazardous technology entails a risk assessment,[44] this assessment should also adopt a participatory approach. Adopting this approach in the context of AI, participatory forms of risk assessment should be developed with the active engagement of the individuals and groups potentially affected. Individuals, groups, and other stakeholders should therefore be informed and actively involved in the debate on what role AI should play in shaping social dynamics, and in the decision-making processes affecting them.[45]

Derogations may be introduced in the public interest, where proportionate in a democratic society and with adequate safeguards. In this regard, in policing, intelligence, and security, where public oversight is limited, governments should report regularly on their use of AI.[46]

4.2.1.4 Transparency and Intelligibility

Transparency is a challenging[47] and highly debated topic in the context of AI,[48] with several different interpretations, including the studies on 'Explainable AI'. In this sense, it is one of the data protection principles that is stressed most frequently.[49]

But effective transparency is mired by complex analysis processes, non-deterministic models, and the dynamic nature of many algorithms. Furthermore, solutions such as the right to explanation focus on decisions affecting specific persons, while the problems of collective use of AI at group level[50] remain unaddressed.

In any case, none of these points diminishes the argument for the central role of transparency and AI intelligibility in safeguarding individual and collective self-determination. This is truer still in the public sector, where the limited variability of algorithms (ensuring equality of treatment and uniform public procurement procedures) can afford greater transparency levels.

In the AI context, every individual must therefore have the right to be properly informed when interacting directly with an AI system and to receive adequate and

[43] ICDPPC, para 25. See also United Nations, Office of the High Commissioner for Human Rights 2018.

[44] See below in Sect. 4.2.1.5.

[45] GAI, paras II.7 and III.8. See also United Nations, Office of the High Commissioner for Human Rights 2018, para 64.

[46] UNESCO 2019, para 107.K.

[47] Mantelero 2018a, pp. 11–13.

[48] E.g. Selbst and Barocas 2018; Wachter et al. 2017; Selbst and Powles 2017; Edwards and Veale 2017.

[49] Convention 108+, Article 8.

[50] Taylor et al. 2017.

easy-to-understand information on its purpose and effects, including the existence of automated decisions. This information is necessary to enable overall human control over such systems, to verify alignment with individuals' expectations and to enable those adversely affected by an AI system to challenge its outcome.[51] Every individual should also have a right to obtain, on request, knowledge of the reasoning underlying any AI-based decision-making process where the results of such process are applied to him or her.[52]

Finally, to foster transparency and intelligibility, governments should promote scientific research on explainable AI and best practices for transparency and auditability of AI systems.[53]

4.2.1.5 Precautionary Approach and Risk Management

Regarding the potentially adverse consequences of technology in general, it is important to make a distinction between cases in which the outcome is known with a certain probability and those where it is unknown (uncertainty). Since building prediction models for uncertain consequences is difficult, we must assume that "uncertainty and risk are defined as two mutually exclusive concepts".[54]

Where there is scientific uncertainty about the potential outcome, a precautionary approach[55] should be taken, rather than conducting a risk analysis.[56] The same conclusion can be drawn for AI where the potential risks of an AI application are unknown or uncertain.[57] In all other cases, AI developers, manufacturers and service providers should assess and document the possible adverse consequences of their work for human rights and fundamental freedoms, and adopt appropriate risk prevention and mitigation measures from the design phase (human rights by-design approach) and throughout the lifecycle of AI products and services.[58]

The development of AI raises specific forms of risk in the field of data protection. One widely discussed example is that of re-identification,[59] while the risk of de-contextualisation is less well known. In the latter case, data-intensive AI applications may ignore contextual information needed to understand and apply the

[51] Convention 108+, Article 8; CM/Rec(2019)2, para 11.3; OECD, para 1.3; UNESCO 2019, Annex I, p. 28. See also ICDPPC, para 3; CM/Rec(2020)1, Appendix, para C.4.1.

[52] Convention 108+, Article 9.1.c; GAI, para II.11.

[53] ICDPPC, para 3.a.

[54] Hansson 2013, p. 12.

[55] See also Peel 2004.

[56] See also Chap. 2, Sect. 2.2. For a broader analysis of risk assessment in the field of AI, see also Mantelero 2018b.

[57] GAI, para II.2. See also Mantelero 2017; ICDPPC ("Highlighting that those risks and challenges may affect individuals and society, and that the extent and nature of potential consequences are currently uncertain"); CM/Rec(2020)1, Appendix, para A.15.

[58] GAI, paras II.2 and II.3; OECD, para 1.4; UNESCO 2019. See also ICDPPC and OECD 2015.

[59] E.g., Narayanan et al. 2016; Ohm 2010.

proposed solution. De-contextualisation can also impact the choice of algorithmic models, re-using them without prior assessment in different contexts and for different purposes, or using models trained on historical data of a different population.[60]

The adverse consequences of AI development and deployment should therefore include those that are due to the use of de-contextualised data and de-contextualised algorithmic models.[61] Suitable measures should also be introduced to guard against the possibility that anonymous and aggregated data may result in the re-identification of the data subjects.[62]

Finally, Convention 108+ (like the GDPR) adopts a two-stage approach to risk: an initial self-assessment is followed by a consultation with the competent supervisory authority if there is residual high risk. A similar model can be extended to AI-related risks.[63] AI developers, manufacturers, and service providers should consult a competent supervisory authority where AI applications have the potential to significantly impact the human rights and fundamental freedoms of individuals.[64]

4.2.1.6 Accountability

The principle of accountability is recognised in Convention 108+[65] and is more generally considered as a key element of risk management policy. In the context of AI,[66] it is important to stress that human accountability cannot be hidden behind the machine. Although AI generates more complicated scenarios,[67] this does not exclude accountability and responsibility of the various human actors involved in the design, development, deployment and use of AI.[68]

From this follows the principle that the automated nature of any decision made by an AI system does not exempt its developers, manufacturers, service providers, owners and managers from responsibility and accountability for the effects and consequences of the decision.

[60] Caplan et al. 2018, 7; AI Now Institute 2018.

[61] GAI, para II.5. This principle is also repeated in CM/Rec(2020)1, Appendix, para B3.4.

[62] See also CM/Rec(2010)13, para 8.5.

[63] GAI, para III.5. See also Data Ethics Commission of the Federal Government, Federal Ministry of the Interior Building and Community and Data Ethics Commission 2019, 42, which also suggests the introduction of licensing and oversight procedures.

[64] GAI, para III.4.

[65] Convention 108+, Article 10.1.

[66] OECD para IV.1.5; GAI paras I.2 and III.1.

[67] See also European Commission, Expert Group on Liability 2019.

[68] See also Council of Europe, Parliamentary Assembly 2017, para 9.1.1.

4.2.1.7 Data Minimisation and Data Quality

Data-intensive applications, such as Big Data analytics and AI, require a large amount of data to produce useful results, and this poses significant challenges for the data minimisation principle.[69] Furthermore, the data must be gathered according to effective data quality criteria to prevent potential bias, since the consequences for rights and freedoms can be critical.[70]

In the context of AI, this means that developers are required to assess the nature and amount of data used (data quality) and minimise the presence of redundant or marginal data[71] during the development and training phases, then monitoring the model's accuracy as it is fed with new data.[72]

AI development and deployment should avoid any potential bias, including unintentional or hidden, and critically assess the quality, nature, origin and amount of personal data used, limiting unnecessary, redundant or marginal data, and monitoring the model's accuracy.[73]

4.2.1.8 Role of Experts and Participation

The complex potential impacts of AI solutions on individuals and society demand that AI development process cannot be delegated to technicians alone. The role of experts from various domains was highlighted in the first non-binding document on AI and data protection, suggesting AI developers, manufacturers and service providers set up and consult independent committees of experts from a range of fields, and engage with independent academic institutions, which can help in the design of human rights-based AI applications.[74] Participatory forms of AI development, based on the active engagement of the individuals and groups potentially affected by AI applications, should also been encouraged.[75]

4.2.1.9 Algorithm Vigilance

The existing supervisory authorities (e.g. data protection authorities, communication authorities, antitrust authorities, etc.) and the various stakeholders involved in

[69] Convention 108+, Article 5.

[70] GAI paras II.2 and II.6. See also CM/Rec(2020)1, Appendix, para B.2.2.

[71] Synthetic data can make a contribution to this end; see also The Norwegian Data Protection Authority 2018.

[72] See also GBD, paras IV.4.2 and IV.4.3.

[73] GAI, para II.4; OECD; UNESCO 2019.

[74] GAI, para II.6, ICDPPC. See also UNESCO, Declaration on the Human Genome and Human Rights, 11 November 1997, Article 11; CM/Rec(2020)1, Appendix, para B.5.3.

[75] GAI, para II.7.

the development and deployment of AI solutions should both adopt forms of algorithm vigilance to react quickly in the event of unexpected and hazardous outcomes.[76]

AI developers, manufacturers, and service providers should therefore implement algorithm vigilance by promoting the accountability of all relevant stakeholders, assessing and documenting the expected impacts on individuals and society in each phase of the AI system lifecycle on a continuous basis, so as to ensure compliance with human rights.[77] Cooperation should be encouraged in this regard between different supervisory authorities having competence for AI.[78]

4.2.2 Key Principles from Biomedicine Regulation

Compared with data protection, international legal instruments on health protection provide a more limited and sector-specific contribution to the draft of future AI regulation. While data is a core component of AI, such that several principles can be derived from international instruments of data protection, healthcare is simply one of many sectors in which AI can be applied. This entails a dual process of contextualisation: (i) some principles stated in the field of data protection can be further elaborated upon with regard to biomedicine; (ii) new principles must be introduced to better address the specific challenges of AI in the sector.

Starting with the Universal Declaration of Human Rights, several international binding instruments include provisions concerning health protection.[79] Among them, the International Covenant on Economic, Social and Cultural Rights, the European Convention on Human Rights, Convention 108+ and the European Social Charter, all lay down several general provisions on health protection and related rights.[80] Provisions and principles already set out in other general instruments have a more sector-specific contextualisation in the Universal Declaration on Bioethics and Human Rights (UNESCO) and the Oviedo Convention[81] (Council of Europe).

[76] See also Commission Nationale de l'Informatique et des Libertés – LINC 2017; The Public Voice 2018.

[77] GAI, para II.10; OECD; ICDPPC.

[78] ICDPPC; GAI, para III.6

[79] E.g. Office of the High Commissioner for Human Rights 2000, p. 21; Yamin 2005. At a national and EU level, most of the existing regulation on health focuses on medical treatment, research (including clinical trials) and medical devices/products. AI has a potential impact on all these areas, given its application in precision medicine, diagnosis, and medical devices and services. See also Azencott 2018; Ferryman and Pitcan 2018.

[80] See also the International Covenant on Civil and Political Rights, and the Convention on the Rights of the Child of 20 November 1989.

[81] Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, Oviedo, 4 April 1997.

The Oviedo Convention – the only multilateral binding instrument entirely focused on biomedicine – and its additional protocols is the main source to identify the key principles in this field,[82] which require further elaboration to be applied to AI regulation. The Convention is complemented by two non-binding instruments: the Recommendation on health data[83] and the Recommendation on research on biological materials of human origin.[84] The former illustrates the close links between biomedicine (and healthcare more generally) and data processing.

Although the Universal Declaration on Bioethics and Human Rights and the Oviedo Convention – including the related non-binding instruments –, were adopted in a pre-AI era, they provide specific safeguards regarding self-determination, human genome treatments, and research involving human beings, which are unaffected by AI application in this field and require no changes.

However, self-determination in the area of biomedicine faces the same challenges as already discussed for data processing. Notwithstanding the different nature of consent to medical treatment and to data processing, the high degree of complexity and, in several cases, obscurity in AI applications can often undermine the effective exercise of individual autonomy in both cases.[85]

Against this background, the main contribution of the binding international instruments in the field of biomedicine does not concern the sector-specific safeguards they provide, but consists in the important set of general principles and values that can be extrapolated from them to form a building block of future AI regulation.

The key principles can be identified in relation to the following nine areas: primacy of the human being, equitable access, acceptability, the principle of beneficence, private life and right to information, professional standards, non-discrimination, the role of experts, and public debate. This contribution goes beyond biomedicine since several provisions, centred on an appropriate balance between technology and human rights, can be extended to AI in general and contextualised in this field, as explained in the following analysis.[86]

4.2.2.1 Primacy of the Human Being

In a geo-political and economic context characterised by competitive AI development, the primacy of the human being must be affirmed as a key element in the

[82] Andorno 2005; Seatzu 2015.

[83] Council of Europe, Committee of Ministers 2019.

[84] Council of Europe, Committee of Ministers 2016a.

[85] See above Sect. 4.2.1.

[86] Human dignity and informed consent are not included in the table as the first is a value common to the instruments adopted by the Council of Europe in the area of human rights, democracy and the rule of law (see Sect. 3.1) and informed consent is a principle that is also relevant in the context of data processing.

human rights-oriented approach:[87] the drive for better performance and efficiency in AI-based systems cannot override the interests and welfare of human beings.

This principle must apply to both the development and use of AI systems (e.g. ruling out systems that violate human rights and freedoms or that have been developed in violation of them).

4.2.2.2 Equitable Access to Health Care

The principle of equitable access to healthcare,[88] should be extended to the benefits of AI,[89] especially considering the increasing use of AI in the healthcare sector. This means taking appropriate measures to combat the digital divide, discrimination, marginalisation of vulnerable persons or cultural minorities, and limited access to information.

4.2.2.3 Acceptability

Based on Article 12 of the International Covenant on Economic, Social and Cultural Rights, the Committee on Economic, Social and Cultural Rights clarified the notion of acceptability, declaring that all health facilities, goods and services must “be respectful of medical ethics and culturally appropriate”.[90] Given the potentially high impact of AI-based solutions on society and groups,[91] acceptability is also a key factor in AI development, as demonstrated by the emphasis on the ethical and cultural dimension found in some non-binding instruments.[92]

4.2.2.4 Principle of Beneficence

Respect for the principle of beneficence in biomedicine and bioethics and human rights[93] should be seen as a requirement where, as mentioned above, the complexity or opacity of AI-based treatments places limitations on individual consent which

[87] See also Oviedo Convention, Article 2, and GAI.

[88] Oviedo Convention, Article 3.

[89] See also UNESCO, Universal Declaration on Bioethics and Human Rights, Article 2.f.

[90] Office of the High Commissioner for Human Rights 2000. See also UNESCO, Universal Declaration on Bioethics and Human Rights, Article 12; GBD, paras IV.1 and IV.2.

[91] Taylor et al. 2017.

[92] GAI paras I.4 and II.6; CM/Rec(2020)1.

[93] UNESCO, Universal Declaration on Bioethics and Human Rights, Article 4. See also Oviedo Convention, Article 6 (“an intervention may only be carried out on a person who does not have the capacity to consent, for his or her direct benefit”), and Articles 16 and 17.

cannot therefore be the exclusive basis for intervention. In such cases, the best interest of the person concerned should be the main criterion in the use of AI applications.[94]

4.2.2.5 Private Life and Right to Information

In line with the considerations expressed earlier on data protection, the safeguards concerning self-determination with regard to private life and the right to information already recognised in the field of medicine[95] could be extended to AI regulation.

With specific reference to the bidirectional right to information about health, AI health applications must guarantee the right to information and respect the wishes of individuals not to be informed, unless compliance with an individual's wish not to be informed entails a serious risk to the health of others.[96]

4.2.2.6 Professional Standards

Professional standards are a key factor in biomedicine,[97] given the potential impacts on individual rights and freedoms. Similarly, AI development involves several areas of expertise, each with its own professional obligations and standards, which must be met where the development of AI systems can affect individuals and society.

Professional skills requirements must be based on the current state of the art. Governments should encourage professional training to raise awareness and understanding of AI and its potential effects on individuals and society, as well as supporting research into human rights-oriented AI.

[94] See also Beauchamp 1990, p. 153 ("virtually everyone acknowledges-under any model-that a person who is nonautonomous or significantly defective in autonomy is highly dependent on others, does not properly fall under the autonomy model, and therefore should be protected under the beneficence model"); Pellegrino and Thomasma 1987, 42 ("[in the beneficent model] No ethical stance, other than acting for the patient's best interests, is applied beforehand").

[95] Oviedo Convention, Article 10. See also UNESCO, Universal Declaration on Bioethics and Human Rights, Article 10.

[96] See also Council of Europe, Committee of Ministers 2019, para 7.6 "The data subject is entitled to know any information relating to their genetic data, subject to the provisions of principles 11.8 and 12.7. Nevertheless, the data subject may have their own reasons for not wishing to know about certain health aspects and everyone should be aware, prior to any analysis, of the possibility of not being informed of the results, including of unexpected findings. Their wish not to know may, in exceptional circumstances, have to be restricted, as foreseen by law, notably in the data subject's own interest or in light of the doctors' duty to provide care"); UNESCO, Declaration on the Human Genome and Human Rights, 11 November 1997, Article 5.c.

[97] Oviedo Convention, Article 4. See also Council of Europe, Committee of Ministers 2019.

4.2.2.7 Non-discrimination

The principle of non-discrimination[98] and non-stigmatisation in the field of biomedicine and bioethics[99] should be complemented by ruling out any form of discrimination against a person or group based on predictions of future health conditions.[100]

4.2.2.8 Role of Experts

The expertise of ethics committees in the field of biomedicine[101] should be called upon to provide independent, multidisciplinary and pluralist committees of experts in the assessment of AI applications.[102]

4.2.2.9 Public Debate

As with biomedicine,[103] fundamental questions raised by AI development should be exposed to proper public scrutiny as to the crucial social, economic, ethical and legal implications, and their application subject to consultation.

Examination of the above key areas demonstrates that the current legal framework on biomedicine can provide important principles and elements to be extended to future AI regulation, beyond the biomedicine sector. However, four particular shortcomings created by the impact of AI remain unresolved, or only partially addressed, and should be further discussed:

(a) Decision-making Systems

In recent years a growing number of AI applications have been developed for medical diagnosis, using data analytics and ML solutions. Large-scale data pools and predictive analytics are used to try and arrive at clinical solutions based on available knowledge and practices. ML applications in image recognition may provide increased cancer detection capability. Likewise, in precision medicine, large-scale collection and analysis of multiple data sources (medical as well as non-medical data, such as air and housing quality) are used to develop personalised responses to health and disease.

The use of clinical data, medical records and practices, as well as non-medical data, is not in itself new in medicine and public health studies. However, the scale

[98] Oviedo Convention, Article 11.

[99] UNESCO. Universal Declaration on Bioethics and Human Rights, Article 11.

[100] See also Council of Europe, Committee of Ministers 2016a, Article 5.

[101] Oviedo Convention, Article 16. See also UNESCO, Universal Declaration on Bioethics and Human Rights, Article 19.

[102] See Chap. 3. See also GBD.

[103] Oviedo Convention, Article 28.

of data collection, the granularity of the information gathered, the complexity (and in some cases opacity) of data processing, and the predictive nature of the results raise concerns about the potential fragility of decision-making systems. Most of these issues are not limited to the health sector, as potential biases (including lack of diversity and the exclusion of outliers and smaller populations), data quality, de-contextualisation, context-based data labelling and the re-use of data[104] are common to many AI applications and concern data in general. Existing guidance in the field of data protection[105] can therefore be applied here too and the data quality aspects extended to non-personal data.

(b) Self-determination
The opacity of AI applications and the transformative use of data in large-scale data analysis undermine the traditional notion of consent in both data processing[106] and medical treatment. New schemes could be adopted, such as broad[107] or dynamic consent,[108] which however – at the present state of the art – would only partially address this problem.

(c) The Doctor-Patient Relationship
There are several factors in AI-based diagnosis – such as the loss of knowledge that cannot be encoded in data,[109] over-reliance on AI in medical decisions, the effects of local practices on training datasets, and potential deskilling in the healthcare sector[110] – that might affect the doctor-patient relationship[111] and need to be evaluated carefully before adoption.

[104] Ferryman and Pitcan 2018, pp. 19–20 (“Because disease labels, such as sepsis, are not clear cut, individual labels may be used to describe very different clinical realities” and “these records were not designed for research, but for billing purposes, which could be a source of systematic error and bias”).

[105] GBD and the related preliminary studies: Mantelero 2018a, and Rouvroy 2015.

[106] See Chap. 1; see also Council of Europe, Committee of Ministers 2019.

[107] Sheehan 2011. See also Convention 108+, Explanatory Report, p. 43 (“In the context of scientific research it is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose”) and Council of Europe, Committee of Ministers 2019, 15.6 (“As it is not always possible to determine beforehand the purposes of different research projects at the time of the collection of data, data subjects should be able to express consent for certain areas of research or certain parts of research projects, to the extent allowed by the intended purpose, with due regard for recognised ethical standards”).

[108] Kaye et al. 2015.

[109] Caruana et al. 2015.

[110] Cabitza et al. 2017.

[111] See also, UNESCO, Universal Declaration on Bioethics and Human Rights, Article 20; WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects, 9 July 2018. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/, accessed 6 March 2020.

(d) Risk Management

The medical device industry has already developed risk-based regulatory models, such as Regulation (EU) 2017/745 – based on progressive safeguards according to the class of risk of each device –, which could be generalised for the future AI regulation focusing on the impact on human rights and fundamental freedoms. However, a risk-based classification of AI by law is complicated, given its variety and different fields of application.[112]

4.2.3 A Contribution to a Future Principles-Based Regulation of AI

Based on the analysis of two key areas of AI application, the principles-based approach has revealed how it is possible to define future AI regulation by focusing on a set of guiding principles developed in a way consistent with the existing international human rights framework and reaffirming the central role of human dignity and human rights in AI, where machine-driven solutions risk dehumanising individuals.[113]

The principle-based methodological process, consisting of analysis (mapping and identification of key principles) and contextualisation, has proven its merit in the areas examined, with the development of several key principles. Correlations and a common ground between these principles have been identified facilitating their harmonisation, while other principles represent the unique contributions of each sector to future AI regulation.

The table below (Table 4.1) summarises these findings and the level of harmonisation in these two areas and, notwithstanding the limitations of the scope of this analysis, shows how its results validate the principles-based methodology as a possible scenario for future AI regulation.

[112] See in this regard the considerations expressed in Sect. 4.3.2.

[113] See also UNESCO, Declaration on the Human Genome and Human Rights (11 November 1997), Article 2. This may also include the adoption of bans on specific AI technologies developed in a manner inconsistent with human dignity, human rights, democracy and the rule of law. See also UNESCO, Declaration on the Human Genome and Human Rights (11 November 1997), Article 11; Data Ethics Commission of the Federal Government, Federal Ministry of the Interior 2019; Access Now 2019.

Table 4.1 Key principles in Data and Health (AI regulation)

Data	Health
Primacy of human being	Primacy of the human being
Data protection and right to information on data processing	Private life and right to information
Digital literacy, education and professional training Accountability	Professional standards
Transparency and intelligibility	Right to information
Precautionary approach and risk management Algorithm vigilance	Principle of beneficence Non-discrimination Equitable access
Role of experts	Role of experts
Participation and democratic oversight on AI development	Public debate
	Acceptability
Data minimisation and data quality	

Source The author

4.3 From Design to Law – The European Approaches and the Regulatory Paradox

In previous sections we have seen how the future regulation of AI could be based on existing international principles. We can carry out a similar exercise with respect to EU law, where similar principles are recognised, though in the presence of a wider variety of binding instruments, owing to the EU's broader field of action.

Rather than adopt the principles-based methodology described, neither the EU legislator nor the Council of Europe decided to follow this path. Both European legislators abandoned the idea of setting common funding principles for AI development and opted for a different and more minimalist approach with a greater emphasis on risk prevention.

While the focus on risk is crucial and in line with the HRESIA, there is something of a regulatory paradox in Europe's approach to AI. An attempt to provide guiding principles was made through ethical guidelines – such as those drafted by the HLEGAI[114] –, vesting legal principles in ethical requirements. On the other hand, recent regulatory proposals based on binding instruments have preferred not to provide a framework of principles but focus on specific issues such as banning applications, risk management and conformity assessment.

This is a regulatory paradox, where general legal principles are set out in ethical guidelines while the actual legal provisions lack a comprehensive framework. Although this is more pronounced in Brussels than in Strasbourg, concerns at a

[114] See Chap. 3, Sect. 3.1.2.

European level about the impact of AI regulation on competition and the weakness of the AI industry in Europe appear to take precedence over far-reaching regulation.

Such concerns have restricted measures to high-risk applications,[115] leaving aside a broader discussion of the role of AI in society and citizen participation in AI project development. This bears similarities with what we witnessed with the first generation of data protection law in Europe in the 1960's, where the principle concern was risk and the need to provide safeguards against the danger of a database society.[116] Only in later waves of legislation was a more sophisticated framework established with reference to general principles, fundamental rights, and comprehensive regulation of data processing. A similar path could be foreseen for AI and here a principles-based methodology described above might figure in more extensive regulation to resolve the present paradox.

The two European legislators also display further similarities in their approach to co-regulation – combining hard and soft law –, setting red lines on the most harmful AI applications, and oversight procedures.

Finally, neither of the proposals seem oriented towards the creation of a new set of rights specifically tailored to AI. This decision is important since the contextualisation of existing rights and freedoms can often provide adequate safeguards, while some proposals for new generic rights – such as the right to digital identity – rest on notions that are still in their infancy, and not mature enough to be enshrined in a legal instrument.

Against these similarities between the two European initiatives, differences necessarily remain, given the distinct institutional and political remits of the Council of Europe and the European Union: the Council's more variable political and regulatory situation, compared with the EU; the different goals of the two entities, one focused on human rights, democracy and the rule of law, and the other on the internal market and more detailed regulation; the different status of the Council of Europe's international instruments, which are addressed to Member States, and the EU's regulations which are directly applicable in all Member States; and – not least – the business interests and pressures which are inevitably more acute for the European Union given the immediate impact of EU regulation on business.

Having described the key features of Europe's approach, we can go on to discuss the main ways in which it deals with AI risk. After looking at the framing of the relationship between the perceived risks of AI and the safeguarding of human rights in Strasbourg and Brussels, we will examine the possible contribution of the HRESIA model to future regulation.

[115] See the subject matter of the European Commission, Proposal for Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) as set in its Article 1: (i) prohibition of certain artificial intelligence practices; (ii) specific requirements for high-risk AI systems; (iii) transparency rules for certain AI systems; (iv) market monitoring and surveillance.

[116] Westin and Baker 1972, p. 346.

4.3.1 The Council of Europe's Risk-Based Approach Centred on Human Rights, Democracy and Rule of Law

On 11 September 2019, during its 1353rd meeting, the Committee of Ministers of the Council of Europe set up the Ad hoc Committee on Artificial Intelligence (CAHAI), mandated to examine the feasibility and potential elements of a legal framework for the development, design and application of AI based on the Council of Europe's standards on human rights, democracy and the rule of law.[117] This was the fruit of several ongoing AI initiatives in different branches of the Council of Europe, which had already led to the adoption of important documents in specific sectors.[118]

The CAHAI mandate also confirmed the Council of Europe's focus on legal instruments and its disinclination to regulate AI on the basis of ethical principles.[119] In this sense, the Council of Europe anticipated the EU's turn towards legislation.

After a preliminary study of the most important international and national legal frameworks and ethical guidelines, and an analysis of the risks and opportunities of AI for human rights, democracy and the rule of law,[120] the CAHAI conducted a Feasibility Study on the development of a horizontal cross-cutting regulatory framework[121] on the use and effects of AI (plus policy tools, such as impact

[117] This author served as an independent scientific expert to the CAHAI for the preliminary study of the existing legally binding instruments on AI and was a member of the CAHAI as scientific expert to the Council of Europe's Consultative Committee of the Convention for the protection of individuals with regard to automatic processing (Convention 108). The views and opinions expressed in this chapter are those of the author and do not necessarily reflect the Council of Europe's official policy or position. They are based solely on publicly available documents and do not rely on, or refer to, confidential information or internal procedures and exchanges of opinions.

[118] Council of Europe, Committee of Ministers 2020; Council of Europe, Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) 2019; Council of Europe, Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) 2021.

[119] This is also evident in Council of Europe, European Commission for the Efficiency of Justice (CEPEJ) 2018 which, despite the reference to ethics, focuses on fundamental rights and the principle of non-discrimination.

[120] Council of Europe 2020.

[121] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, para 76 ("it was noted that ethics guidelines are useful tools to exert some influence on public decision making over AI and to steer its development towards social good. However, it was also underlined that soft law approaches cannot substitute mandatory governance. […] there is a particular risk that self-regulation by private actors can bypass or avoid mandatory governance by (inter)governmental authorities. Soft law instruments and self-regulation initiatives can however play an important role in complementing mandatory governance").

assessment models) which might also include a sectoral approach.[122] The Feasibility Study gives a general overview of the key issues and describes the CAHAI's main directions of travel towards a legal framework and policy instruments.

The approach outlined in the Feasibility Study is based on recognition that the existing human rights legal framework already provides guiding principles and provisions that can be applied to AI.[123] These need to be better contextualised in light of the changes to society brought by AI[124] to fill three perceived gaps in the legal landscape: (i) the need to move from general principles to AI-centred implementation; (ii) the adoption of specific provisions on key aspects of AI (e.g. human control and oversight, transparency, explicability); (iii) the societal impact of AI.[125]

Thus, the Feasibility Study refers to human dignity, the right to non-discrimination, the right to effective remedy and other rights and freedoms enshrined in international human rights law. But it also makes new claims, such as: the right to be informed that one is interacting with an AI system rather than with a human being (especially where there is a risk of confusion which can affect human dignity);[126] the right to challenge decisions informed and/or made by an AI system and demand that such decisions be reviewed by a human being; the right to freely refuse AI-enabled manipulation, individualised profiling and predictions, even in the case of non-personal data processing; the right to interact with a human being

[122] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, para 89, p. 176 ("The study has noted that no international legal instrument specifically tailored to the challenges posed by AI exists, and that there are gaps in the current level of protection provided by existing international and national instruments. The study has identified the principles, rights and obligations which could become the main elements of a future legal framework for the design, development and application of AI, based on Council of Europe standards, which the CAHAI has been entrusted to develop. An appropriate legal framework will likely consist of a combination of binding and non-binding legal instruments, that complement each other"). This approach is in line with the conclusion of the preliminary study on the legal framework, Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020b ("A binding instrument establishing the legal framework for AI, including both general common principles and granular provisions addressing specific issues, could therefore be combined with detailed rules set out in additional non-binding sectoral instruments. This model would provide both a clear regulatory framework and the flexibility required to address technological development.").

[123] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, para 83 ("The CAHAI therefore notes that, while there is no legal vacuum as regards AI regulation, a number of substantive and procedural legal gaps nevertheless exist").

[124] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020b ("contextualisation of the guiding principles and legal values provides a more refined and elaborate formulation of them, considering the specific nature of AI products and services, and helps better address the challenges arising from AI").

[125] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, paras 84–86.

[126] See also Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data (Convention 108) 2019, para 2.11.

rather than a robot (unless ruled out on legitimate overriding and competing grounds).[127]

In considering these proposals, it is worth noting that the Feasibly Study is not a legal document and uses the language of a policy document rather than the technical language of a legal text like regulation. Many of these rights are not therefore new stand-alone rights, but intended (including through creative interpretation) to complement already existing rights and freedoms, as part of the Council of Europe's contextualisation and concretisation of the law to deal with AI and human rights.

Along with these proposals for the future legal framework, the Feasibility Study also suggests several policy initiatives to be further developed by non-binding instruments or industrial policy, such as those on auditing processes, diversity and gender balance in the AI workforce or environmental-friendly AI development policies.[128]

In line with the CAHAI mandate and the Council of Europe's field of action, the path marked out by the Feasibility Study also includes two sections on democracy and the rule of law.[129] While extension of the proposed rights and obligations to these fields is significantly narrower than those on human rights, this move is atypical in the global scenario of AI regulation, which tends to exclude holistic solutions comprising democracy and the rule of law, or rely on sector-specific guidelines to address these questions.[130]

Regarding democracy, the most important rights with regard to AI are those concerning democratic participation and the electoral process, diverse information, free discourse and access to a plurality of ideas, and good governance. They also entail the adoption of specific policies on public procurement, public sector oversight, access to relevant information on AI systems, and fostering digital literacy and skills.

As for the rule of law, the main risks concern the use of AI in the field of justice. Here the Feasibility Study refers to the right to judicial independence and impartiality, the right to legal assistance, and the right to effective remedy. In policy terms, Member States are encouraged to provide meaningful information to individuals on the AI systems used in justice and law enforcement, and to ensure these systems do not interfere with the judicial independence of the court.[131]

[127] These and other proposed rights are discussed in Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, Section 7.

[128] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, Section 7.

[129] See also Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, Sections 7.8 and 7.9.

[130] E.g. Council of Europe, European Commission for the Efficiency of Justice (CEPEJ) 2018.

[131] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, 42–43. See also Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020b, Section 2.5.1.

The Council of Europe thus takes a risk-based approach to AI[132] including introducing risk assessment criteria, 'red lines' for AI compatibility with human rights, and mechanisms for periodic review and audits.[133]

More specifically, the Feasibility Study considers risk assessment and management as part of the wider human rights due diligence process and as an ongoing assessment process rather than a static exercise.[134] For the future development of its impact assessment approach, the study takes as a reference framework the "factors that are commonly used in risk-impact assessments". It explicitly mentions the following main parameters: (i) the potential extent of the adverse effects on human rights, democracy and the rule of law; (ii) the likelihood that an adverse impact might occur; (iii) the scale and ubiquity of such impact, its geographical reach, its temporal extension; and (iv) the extent to which the potential adverse effects are reversible.[135]

On the basis of this Feasibility Study, the CAHAI created three working groups:[136] the Policy Development Group (CAHAI-PDG) focused on policies for AI development (soft law component); the Consultations and Outreach Group (CAHAI-COG) tasked with developing consultations with various stakeholders on key areas of the Feasibility Study and the CAHAI's ongoing activity; and the Legal Frameworks Group (CAHAI-LFG) centred on drafting proposals for the future legal framework (hard law component). Though from different angles, these three working groups all adopt the Council of Europe's risk-based approach and its implementation through impact assessment tools and provisions.

The main outcomes are expected to come from the CAHAI-LFG, in the form of binding provisions on impact assessment, and the CAHAI-PDG, with the development of an impact assessment model centred on human rights, democracy, and the rule of law. The CAHAI-COG multi-stakeholder consultations found clear expectations of the impact assessment in AI regulation, and stakeholders saw this as the most important mechanism in the Council of Europe's new framework.[137]

[132] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, paras 87 ("A comprehensive legal framework for AI systems, guided by a risk-based approach") and 125 ("As noted above, when member States take measures to safeguard the listed principles, rights and requirements in the context of AI, a risk-based approach – complemented with a precautionary approach where needed – is recommended").

[133] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, paras 42, 43, 44 ("A contextual and periodical assessment of the risks arising from the development and use of AI is necessary").

[134] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, para 169.

[135] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020a, para 126, where the CAHAI also notes that "in specific contexts, 'integrated impact assessments' might be deemed more appropriate to reduce the administrative burden on development teams (bringing together, for example, human rights, data protection, transparency, accountability, competence, and equalities considerations)".

[136] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2020c.

[137] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2021a.

Based on the CAHAI's work, and the more specific contribution of the CAHAI-LFG working group, the Council of Europe's risk-based AI model will introduce an assessment of the impact of AI applications on human rights, democracy, and the rule of law.[138] While the HRIA is not new, as discussed above, the inclusion of democracy and the rule of law is innovative and challenging.

The democratic process, and democracy in its different expressions, covers a range of topics and it is not easy, from a methodological perspective, to assess the impact on it of a technology or its applications, particularly since it is hard to assess the level of democracy itself.

This does not mean that it is impossible to carry out an impact assessment on specific fields of democratic life, such as the right to participation or access to pluralist information, but this remains a HRIA, albeit one centred on civil and political rights.[139] Evaluation of the impact of AI on democracy and its dynamics in general is still quite difficult.[140]

Different considerations regard the rule of law, where the more structured field of justice plus the limited application of AI make it easier to envisage uses and foresee their impact on a more uniform and regulated set of principles and procedures than democracy. Here again however, the specificity of the field and the interests involved may raise some doubts about the need for an integrated risk assessment model – including human rights, democracy, and the rule of law – as opposed to a more circumscribed assessment of the impact of certain AI applications on the rule of law.

The HUDERIA (HUman rights, DEmocracy and the Rule of law Impact Assessment)[141] proposed by the CAHAI therefore seems much more challenging in its transition from theoretical formulation to concrete implementation than the HRESIA, given the latter's modular structure and its distinction between human rights assessment (entrusted to the customised HRIA) and the social and ethical assessment (entrusted to committees of experts). The HUDERIA's difficulties

[138] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2021b, p. 2.

[139] Council of Europe – Ad hoc Committee on Artificial Intelligence (CAHAI) 2021c, p. 3 seems to be aware of this challenge when it "agreed to use human rights as proxies to democracy and the rule of law. The idea is to explore if the magnitude of certain individual human rights violations closely linked to the good functioning of democratic institutions and processes, as well as rule of law core elements, could undermine democracy and the rule of law". However, using human rights as proxies for democracy and the rule of law means that the proposed model is *de facto* a HRIA.

[140] This is the case with the overall impact of AI-based solutions for smart cities. The case study discussed in Chap. 2 shows that the use of AI in a smart city can foster citizen engagement and interaction, public interest data sharing etc. But at the same time this environment can be captured by big private players and result in a shift in powers traditional exercised by public bodies, on the basis of democratic rules, towards private companies who can privatise and contractualise public tasks and interaction with citizens. It is difficult therefore to define overall impact on democracy as a stand-alone item of the impact assessment. A more feasible solution might be to perform a HRIA but consider the results for the democratic process as an issue for discussion and analysis (see Chap. 3).

[141] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2021c.

appear to be confirmed by the slower progress of the CAHAI-PDG's work on this model compared with the rest of the CAHAI's activities.

Looking at the criteria proposed by the CAHAI-LFG for the impact assessment, they are largely those commonly used in impact assessment theory, i.e. likelihood and severity. Several factors are considered in relation to the severity of the impact (gravity, number of people affected, characteristics of impacted groups, geographical and demographical reach, territorial extension, extent of adverse effects and their reversibility, cumulative impact, likelihood of exacerbating existing biases, stereotypes, discrimination and inequalities). The assessment model should also consider further concurring factors, such as AI-specific risk increasing factors, the context and purpose of AI use, possible mitigation measures, and the dependence of potentially affected persons on decisions based on AI.[142]

The model envisaged is based on the traditional five risk levels (no risk, low, medium, high, extreme). The proposed provisions also leave room for the precautionary principle when it is impossible to assess the envisaged negative impact.

Finally, the level of transparency of the results of the assessment – in terms of their publicly availability –, accountability, auditability and transparency of the process are also considered in the CAHAI-LFG proposal.

At the time of writing, the proposed HUDERIA model adopts a four-stage iterative and participatory model – identification of relevant rights, assessment of the impact on those rights, governance mechanisms, continuous evaluation – which are common to all impact assessments. Its distinguishing feature is "that it includes specific analysis of impact on fundamental rights proxies which are directed towards the Rule of Law and Democracy".[143] In this the CAHAI documents do not limit the impact assessment obligations to specific AI applications in certain fields, a (high) level of risk or the nature and purpose of the technology adopted.

4.3.2 The European Commission's Proposal (AIA) and Its Conformity-Oriented Approach

After an initial approach centred on ethics[144] and the White Paper on Artificial Intelligence,[145] in April 2021 the European Commission proposed an EU regulation

[142] See also Council of Europe – Ad hoc Committee on Artificial Intelligence (CAHAI) 2021d, p. 3 ("the CAHAI-LFG has considered, besides the likelihood and severity of the negative impact, also contextual factors, such as the sector and area of use; the complexity of the AI-system and the level of automation; the quality, type and nature of data used, or the level of compliance with regulation in other fields").

[143] Council of Europe – Ad hoc Committee on Artificial Intelligence (CAHAI) 2021e.

[144] See also Chap. 2, Sect. 2.1.

[145] European Commission 2020d.

on AI (hereinafter the AIA Proposal).[146] This proposal introduces two new elements: the departure from more uncertain ethical grounds towards the adoption of a hard law instrument, albeit within the familiar framework of co-regulation;[147] the adoption of a regulation in the absence of national laws on AI or differing approaches among EU Member States.

The latter aspect highlights the EU legislator's concerns about the rapid development of AI, the EU's limited competitive power in this area in terms of market share, and the need to address the public's increasing worries about AI which might hamper its development.[148] The typical harmonisation goal of EU regulations – not applicable here in the absence of national laws on AI – is therefore replaced by a clear industrial strategy objective embodying a stronger and more centralised regulatory approach by the Commission which is reflected in the AIA Proposal.

As in the case of data protection, the EU proposal therefore stands within the framework of internal market interests, while protecting fundamental rights.[149] This focus on the market and competition appears to be the main rationale behind regulating an as yet unregulated field, designed to encourage AI investment in the EU.[150] It also emerged clearly from the four objectives of the proposed regulation: (i) ensure that AI systems marketed and used in the Union are safe and respect existing law on fundamental rights and Union values; (ii) guarantee legal certainty to facilitate investment and innovation in AI; (iii) enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems; (iv) facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.[151]

In this context a central role is necessarily played by risk regulation, as in the first generation of data protection law where citizens were concerned about the potential misuse of their data and public and (some) private entities were aware of

[146] European Commission, Proposal for Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending legislative acts, COM(2021) 206 final, Brussels, 21 April 2021.

[147] European Commission, Proposal for Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending legislative acts, COM(2021) 206 final, Brussels, 21 April 2021, Explanatory Memorandum (hereinafter AIA Explanatory Memorandum), 9.

[148] European Commission, AIA Explanatory Memorandum, 6 ("This proposal constitutes a core part of the EU digital single market strategy. The primary objective of this proposal is to ensure the proper functioning of the internal market by setting harmonised rules in particular on the development, placing on the Union market and the use of products and services making use of AI technologies or provided as stand-alone AI systems").

[149] This is clearly evident in Article 1 (Subject matter) of the Proposal where there is no explicit or direct reference to the safeguarding of fundamental rights and freedoms and AI's potential impact on them, but only general references to "certain artificial intelligence practices" and "high-risk AI systems". For a different approach, see Article 1 of the General Data Protection Regulation.

[150] European Commission, AIA Explanatory Memorandum. ("It is in the Union interest to preserve the EU's technological leadership"). See also Recital No. 6 AIA Proposal.

[151] European Commission, AIA Explanatory Memorandum, p. 3.

the value of personal data in enabling them to carry out their work. For this reason, the EU proposal wishes to limit itself to the "minimum necessary requirements to address the risks and problems linked to AI, without unduly constraining or hindering technological development or otherwise disproportionately increasing the cost of placing AI solutions on the market".[152]

These goals and the framing of the risk-based approach reveal how the EU differs from the Council of Europe, which places greater emphasis on the safeguarding of human rights and fundamental freedoms. This inevitably impacts on the risk management solutions outlined in the AIA Proposal.

The European Commission's 'proportionate'[153] risk-based approach addresses four level of risks: (i) extreme risk applications, which are prohibited;[154] (ii) high risk applications, dealt with by a conformity assessment (where HRIA is only one of its components); (iii) a limited number of applications that have a significant potential to manipulate persons, which must comply with certain transparency obligations; (iv) no high-risk uses, dealt with by codes of conduct designed to foster compliance with AIA main requirements.[155] Of these, the most important from a human rights impact assessment perspective are the provisions on high risk applications.

The first aspect emerging from these provisions is the combination, under the category of high-risk applications, of AI solutions impacting on two different categories of protected interests: physical integrity, where AI systems are safety components of products/systems or are themselves products/systems regulated under the New Legislative Framework legislation (e.g. machinery, toys, medical devices, etc.),[156] and human rights in the case of so-called stand-alone AI systems.[157]

Safety and human rights are two distinct realms. An AI-equipped toy may raise concerns around its safety, but have no or only limited impact on human rights (e.g. partially automated children's cars). Meanwhile another may raise concerns largely in relation to human rights (e.g. the smart doll discussed in Chap. 2). AI may have a negative impact and entail new risks for both safety and human rights, but the fields, and related risks, are separate and require different remedies. This does not mean that an integrated model is impossible or even undesirable, but that different assessments and specific requirements are essential.

[152] European Commission, AIA Explanatory Memorandum, p. 3.

[153] European Commission, AIA Explanatory Memorandum, p. 3.

[154] AIA Proposal, Article 5, and AIA Explanatory Memorandum, para 5.2.2, which refers to unacceptable risks (prohibited practices) as "contravening Union values, for instance by violating fundamental rights".

[155] AIA Proposal, Article 69.

[156] AIA Proposal, Annex II. The AIA Proposal is not applicable to products/systems regulated under the Old Approach legislation (e.g. aviation, cars), see AIA Proposal, Article 2.2.

[157] AIA Proposal, Annex III; see also rec. 64 ("different nature of risks involved").

Looking at the risk model outlined by the AIA Proposal, its structure is based on Article 9. The chief obligations on providers of high-risk AI systems,[158] as set out in Article 16, regard the performance of a conformity assessment (Articles 19 and 43, Annexes VI and VII) and the adoption of a quality management system (Article 17). The conformity assessment – except for the AI regime for biometric identification and the categorisation of natural persons[159] and the AI applications regulated under the New Legislative Framework (NLF) legislation – is an internal self-assessment process based on the requirements set out in Annex VI. This Annex requires an established quality management system in compliance with Article 17 whose main components include the risk management system referred to in Article 9.[160]

In this rather convoluted structure of the AIA Proposal, Article 9 and its risk management system is the key component of a combined conformity assessment and quality management system. Indeed, the quality management system comprises a range of elements which play a complementary role in risk management. However, the risk assessment and management model defined by Article 9 is based on three traditional stages: risk identification, estimation/evaluation, and mitigation.

The peculiarity of the AIA model consists in the fact that the risk assessment is performed in situations that are already classified by the AIA as high-risk cases. In the EU's proposal, the risk-based approach consists mainly of risk mitigation rather than risk estimation.

The proposal makes a distinction between use of AI in products already regulated under safety provisions, with some significant exceptions,[161] and the rest. In the first group, AI is either a safety component of these products[162] or itself a product in this category. The second group consists of stand-alone AI systems not covered by the safety regulations but which, according to the European Commission, carry a high-risk.

This classification emphasises the importance of the high-risk evaluation set out in the AIA Proposal. With regulated safety applications, risk analysis is only broadened from safety to the HRIA.[163] For stand-alone AI systems, on the other hand, it introduces the completely new regulation based on a comprehensive conformity assessment, which includes the impact on fundamental rights.

[158] An AI provider is "a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge". See AIA Proposal, Article 3.2.

[159] AIA Proposal, Article 43.1.

[160] AIA Proposal, Article 17.1.g.

[161] AIA Proposal, Article 2.2.

[162] On the notion of safety component of a product or system, AIA Proposal, Article 2, No. 14 ("a component of a product or of a system which fulfils a safety function for that product or system or the failure or malfunctioning of which endangers the health and safety of persons or property").

[163] But AIA Proposal, rec. 31.

However, the approach adopted raises questions concerning the following issues: (i) a top-down and more rigid system of high-risk assessment; (ii) a critical barrier between high risk and lower risk; (iii) opaque regulation of technology assessment (Annex III) and risk assessment carried out by providers (Article 9); (iv) use of the notion of acceptability; (v) marginalisation of the role of AI system users. These elements, discussed below, all reveal the distinction between the AIA Proposal's complicated model of risk management and the HRIA's cleaner model based on a general risk assessment.[164]

Given the variety of fields of application of AI and the level of innovation in this area, dividing high-risk applications into eight categories and several sub-fields seems to underestimate the evolving complexity of the technology scenario.

Considering how rapidly AI technology is evolving and the unexpected discoveries regarding its abilities,[165] a closed list of typical high-risk applications may not be easy to keep up-to-date properly or promptly.[166] In addition, the decision to delegate such a key aspect to the Commission, the EU's executive body,[167] is likely to raise concerns in terms of power allocation.

A closed list approach (albeit using broad definitions and open to updating) appears to be reactive rather than preventive in anticipating technology development. By contrast, a general obligation of an AI impact assessment (HRIA) does not suffer from this shortcoming and can act more swiftly in detecting critical new applications. Moreover, a general risk assessment removes the burden of rapidly updating the list of stand-alone high-risk applications, which can remain an open list of presumed high-risk cases, as in Article 35.3 of the GDPR.

The focus on a list of high-risk cases also introduces a barrier between them and the rest where risks are lower. This sharp dichotomy contrasts with the more nuanced consideration of risk and its variability depending on the different technology solutions, contexts, etc. Furthermore, a rigid classification of high-risk applications leaves room for operators wishing to circumvent the regulation by denying that their system falls into one of the listed categories.[168]

[164] A similar general risk assessment, not based on a predefined close list of high-risk cases, was also adopted by the GDPR. See GDPR, Article 35.

[165] E.g., Simonite 2021.

[166] The Commission can add new cases, but on the basis of those listed as a benchmark, AIA Proposal, Article 7.2 ("an AI system poses a risk of harm to the health and safety or a risk of adverse impact on fundamental rights that is equivalent to or greater than the risk of harm posed by the high-risk AI systems already referred to in Annex III").

[167] Regarding the power of the Commission to update the list in Annex III with the addition of new high-risk AI systems, it was also pointed out that "it remains nebulous when the threshold of high-risk, as defined in Article 7(2), will be reached, i.e., when a system's risk count[s] as 'equivalent to or greater' than those of other systems already on the list", AlgorithmWatch 2021.

[168] AlgorithmWatch 2021, p. 3.

Finally, as pointed out in Chap. 2, this cumulative quantification of the level of risk of a given application (described as a high-risk use of AI) contradicts the necessarily multifaced impact of AI applications, which usually concerns different rights and freedoms. The impact may therefore be high with respect to some rights and medium or law with respect to others. The different nature of the impacted rights does not make it possible to define an overall risk level.

The only possible conclusion is that if there is a high risk of a negative impact on even one right or freedom, the overall risk of AI application is high. This is in line with the idea that all human rights must be protected and the indivisible, interdependent and interrelated nature of human rights.

The categories of high-risk application set out in Annex III are defined on the basis of a technology assessment resting on four key elements: (i) AI system characteristics (purpose of the system and extent of its use or likely use); (ii) harm/impact (caused or foreseen harm to health and safety or adverse impacts on fundamental rights; potential extent of such harm or such adverse impacts; reversibility); (iii) condition of affected people (dependency or vulnerability); (iv) legal protection (measures of redress[169] or to prevent or substantially minimise those risks).

This is necessarily an abstract exercise by the legislator (and in future by the Commission) which uses a future scenario approach or, when referring to existing practices, generalises or aggregates several cases. The assessment required by Article 9 on the other hand is a context-specific evaluation based on the nature of the particular case of AI application. These different types of assessment suggest that the applications listed in Annex III, in their context-specific use, may not entail the high level of risk presumed by the Regulation.

In addition, the Proposal fails to explain how and on the basis of which parameters, and method of evaluation, these risks should be assessed in relation to specific AI applications, according to Article 9. Nor, with regard to the general technology assessment used for the Annex III list, does the Commission's Proposal provide transparency on the methodology and criteria adopted.[170]

Another aspect that requires attention is the relationship between high-risk, residual risk and acceptability.[171] Risk assessment and mitigation measures should act in such a way that the risk "associated with each hazard as well as the overall residual risk of the high-risk AI systems is judged acceptable". But the AIA Proposal fails to provide a definition of acceptable risk.

[169] It is worth emphasising that these measures are not directly related to risk assessment.

[170] European Center for Not-for-Profit Law 2021, 9 ("there is currently no provision nor clearly identified procedure allowing for adding new categories to annex III related to the list of high-risk uses of AI systems"). Another major shortcoming is the lack of public debate on the cases listed, AlgorithmWatch 2021 ("many of these sensitive applications have not yet been the object of public debate. Before they are put to use, citizens should have the opportunity to discuss whether there are limits to what decisions should be automated in the first place").

[171] AIA Proposal, Article 9.4.

The notion of acceptable risk comes from product safety regulation, while in the field of fundamental rights the main risk factor is proportionality. While acceptability is largely a social criterion,[172] Article 2(b) of Directive 2001/95/EC on general product safety define a safe product as one that "does not present any risk or only the minimum risks compatible with the product's use, considered to be acceptable". Here acceptability results from an absence of risk or "minimum risks", which is necessarily context-dependent[173] and suggests a case-specific application of the criteria set out in Article 7.2 of the AIA Proposal. What is more, these criteria – like the focus on the product characteristics, the categories of consumers at risk and the measures to prevent or substantially minimise the risks – are coherent with those considered by Article 2(b) of Directive 2001/95/EC.

If we accept this interpretation, acceptability is incompatible with AI's high risk of adverse impacts on fundamental rights and any impact assessment based on a quantification of risk levels will play a crucial part in risk management.

Finally, the AIA Proposal marginalises the role of the AI users. They play no part in the risk management process and have no obligations in this regard, even though AI providers market solutions that are customisable by users. AI users[174] may independently increase or alter the risks of harm to health and safety by their particular use of the systems, especially in terms of impact on individual and collective rights, given their variety and context dependence.

For example, an AI company can offer a platform for participatory democracy, but its implementation can be affected by exclusion biases depending on the user's choice of settings and the specific context. AI providers cannot fully take into account such contextual variables or foresee the potentially affected categories, so

[172] Nordlander et al. 2010, pp. 241–42 ("Determining the acceptable level of risk is not the function of the risk assessment itself, which simply attempts to identify the ranges of risk. The decision as to what constitutes acceptable risk is a socio-political, not a scientific, decision"); Whipple 1988, 85–86. See also Muhlbauer 2004, p. 335 ("In general, society decides what is an acceptable level of risk for any particular endeavor"); Bergkamp 2015.

[173] See also Commission Implementing Decision (EU) 2019/417 of 8 November 2018 laying down guidelines for the management of the European Union Rapid Information System 'RAPEX' established under Article 12 of Directive 2001/95/EC on general product safety and its notification system (notified under document C(2018) 7334) 2019 (OJ L), p. 171 ("Taking action to counteract a risk may also depend on the product itself and the 'minimum risks compatible with the product's use, considered to be acceptable and consistent with a high level of protection'. This minimum risk will probably be much lower for toys, where children are involved, than for a chain-saw, which is known to be so high-risk that solid protective equipment is required to keep the risk at a manageable level") and 183 ("Any injury harm that could easily have been avoided will be difficult to accept for a consumer").

[174] An AI user is "'any 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", AIA Proposal, Article 3.4.

their adoption of a general mitigation strategy will have only a limited effect.[175] Risk management and risk assessment duties should therefore also apply to AI users in proportion to their role in system design and deployment.

In line with risk management theory and practice, a circular iterative approach is adopted by the AIA Proposal, including post-market monitoring.[176] This is crucial since lifelong monitoring, learning and re-assessment are essential elements in an evolving technology scenario where risk levels may change over time.[177]

Considering the AIA Proposal as a whole, the legislator's rationale is to largely exempt AI users (i.e. entities using AI systems under their own authority) from risk management duties and to avoid creating extensive obligations for the AI producers, limiting the regulatory impact only to specific sectors, characterised by potential new AI-related risks or the use of AI in already regulated product safety areas.

While this is effective in terms of policy impact and acceptability, it is a weak form of risk prevention. The Proposal makes a quite rigid distinction between high-level risk and the rest, providing no methodology to assess the former, and largely exempting the latter from any mitigation (with the limited exception of transparency obligations in certain cases).

In addition, two large elements are missing from the EU's Proposal: integration between law and ethical/societal issues and the role of participation. As for the first, following several years of discussion of the ethical dimension of AI, the prevailing vision seems to be to delegate ethical issues to other initiatives[178] not integrated with the legal assessment. In the same way that focusing exclusively on ethics was critical,[179] this lack of integration between the legal and societal impacts of AI is problematic. An integrated assessment model, like the HRESIA, could overcome this limitation in line with the proposed risk-based model.

Equally, introducing a participatory dimension to the assessment model, covering both legal and societal issues, would bridge the second gap, related to the lack of participation, and align the AIA proposal with the emphasis on civic engagement of other EU initiatives and a vision of AI use for the benefit of citizens.[180]

[175] In addition, different AI systems can be combined by the user to achieve a specific goal.

[176] AIA Proposal, Article 61. See also Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) 2019, Section 2.10.

[177] AIA Proposal, Recital No. 66 ("as regards AI systems which continue to 'learn' after being placed on the market or put into service (i.e. they automatically adapt how functions are carried out), it is necessary to provide rules establishing that changes to the algorithm and its performance that have been pre-determined by the provider and assessed at the moment of the conformity assessment should not constitute a substantial modification").

[178] E.g. Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission 2020.

[179] See Chap. 3.

[180] See also European Center for Not-for-Profit Law 2021, pp. 11 and 15–16.

4.4 The HRESIA Model's Contribution to the Different Approaches

Looking at the three approaches to AI regulation described at the beginning of this chapter, neither the Council of Europe nor the European Commission decided to adopt a principles-based approach. This, even though several of the key principles enshrined in binding and non-binding human rights instruments can be valuable – with due contextualisation – to AI regulation and are also partially reflected in the proposals of both bodies.

The predominant focus on risk and accountability is probably due to the reductive and incremental approach of this first stage of AI regulation, as was the case with data protection in the 1970s or with regard to product safety in the first phase of industrial mass production.[181] As with the early data protection regulations, the priority is to establish specific procedural, technical and organisational safeguards against the most serious risks rather than building a clear and complete set of governing principles.

The EU's closed list of high-risk systems, and the Council of Europe's key guiding principles for AI development and use reflect the fact that these proposals represent the first generation of AI regulation.

As with data protection,[182] further regulation will probably follow, broader in scope and establishing a stronger set of guiding principles. In regard to the EU initiative, a fuller consideration of the potential widespread impact of non-high-risk applications and the challenges of rigid pre-determined risk evaluation systems could provide more effective protection of individual rights and collective interests.

Both proposals are also characterised by a focus on the legal dimension at the expense of a more holistic approach covering ethical and societal issues, which are either ignored or delegated to non-legal instruments.

This gap could be bridged by a hybrid model, such as the HRESIA, combining human rights and ethical and societal assessments to give a more complete view of the consequences of AI applications and affect their design. This is even more important in the case of large-scale projects or those with significant effects on social communities.

In addition, the key notion of acceptability in the AIA Proposal,[183] discussed in the previous section, necessarily implies the value of the HRIA to assess the impact on fundamental rights covered by Article 9. But it would also benefit from the broader HRESIA model given the societal dimension of acceptability[184] which should be paid greater attention with regard to each context-specific AI application and addressed by expert committees, as described in Chap. 3.

[181] Gregory 1951, p. 385; Traynor 1965; McMahon 1968; Oliphant 2005.

[182] Mayer-Schönberger 1997.

[183] AIA Proposal, Article 9.

[184] See above fn. 172.

Regarding the costs and resources involved in extending the HRESIA, we should recall the considerations expressed above about the model's modularity and scalability.[185] Based on a HRIA and adopting internal advisors for the societal issues, the burdens are proportional to the impact of the technology and minimum or negligible in the case of low risk. Moreover, the experience gained by the HRESIA experts would further reduce the costs in relation to the frequency of the assessments.

Both the Council of Europe and the European Commission suggest a self-assessment procedure in line with the HRESIA model. The latter also includes a layer of participation, which is mentioned by the Council of Europe[186] and one of the recognised shortcomings of the AIA Proposal.

The EU Proposal limits the obligation to perform an impact assessment to AI providers, in line with the thinking behind product safety regulation. However, a more nuanced approach is required, given the part played by providers and users in the development, deployment and use of AI applications, and the potential impacts of each stage on human rights and freedoms.

It is worth remembering that AI differs from data protection in the greater role that AI providers play in the complicated and often obscure AI processing operations.[187] This makes it inappropriate to recreate the controller/provider distinction, albeit with different nuances,[188] regardless of the criticisms expressed about the distinction itself.[189] Still, the effective role played by AI users[190] in system design and deployment should be addressed by their involvement in risk management and assessment duties.

This can be achieved for most of the AI systems in use, excepting those cases where the user has little ability to customise or train the system for a specific context, and a HRESIA should be performed by all entities that use third-party AI services for their own purposes. This does not mean that the HRESIA cannot be used by producers in the design of their systems. but suggests a model – already

[185] See Chap. 2.

[186] Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) 2021f, pp. 26–27.

[187] Mantelero 2018b.

[188] Microsoft Corporation 2021, 5 ("we recommend creating a new designation of "deployer," defined as the entity that takes the specific decision to implement an AI system for one of the high-risk scenarios detailed in Annex III. We also recommend that this entity be responsible for ensuring that any such Annex III deployment satisfies the requirements set out in Article 16. This approach has the virtue of ensuring that regulatory responsibilities fall in the first instance on the entity that has the greatest control over, and visibility into, the operation of the specific deployment that brings it within scope of Annex III (and thus subject to the requirements of Articles 9–17). It is, however, contingent on "technology suppliers" also assuming responsibilities that they are well-placed to bear, as described below.)").

[189] de Hert and Papakonstantinou 2016, p. 184.

[190] AIA Proposal, Article 3(4) ("'user' means any 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").

proposed in data protection regulation[191] –, in which the providers perform the HRESIA on their products, but AI users perform their own HRESIA with regard to specific implementation.

Finally, both the Council of Europe and the European Commission base their approaches on risk assessment and a series of variables to be considered but fail to specify a method of assessing the level of risk, making them difficult to put into practice.[192] In contrast, the HRESIA not only identifies the assessment criteria but also explains a how to go about defining the risk levels and evaluating the systems.

With its nature, scope, and methodology the HRESIA model not only responds to AI impact assessment requirements of the European proposals, but it could also address the shortcomings of the proposed provisions and serve as a model that is as yet absent in the ongoing work of these regulatory bodies.

4.5 Summary

The ongoing debate on AI in Europe has been characterised by a shift in focus, from the identification of guiding ethical principles to a first generation of legal obligations on AI providers.

Although the debate on AI regulation is still fluid at a global level and the European initiatives are in their early stages, three possible approaches are emerging to ground AI regulation on human rights.

One option is a principles-based approach, comprising guiding principles derived from existing international binding and non-binding human rights instruments, which could provide a comprehensive framework for AI, in line with previous models such as Convention 108 or the Oviedo Convention.

A different approach focuses more narrowly on the impacts of AI on individual rights and their safeguarding through rights-based risk assessment. This is the path followed by the Council of Europe in its ongoing work on AI regulation.

Finally, as outlined in the EU proposal, greater emphasis can be placed on managing high-risk applications focusing on product safety and conformity assessment, combining safety and rights protection with a predefined risk classification.

[191] Article 29 Data Protection Working Party 2017, 8 (“A DPIA can also be useful for assessing the data protection impact of a technology product, for example a piece of hardware or software, where this is likely to be used by different data controllers to carry out different processing operations. Of course, the data controller deploying the product remains obliged to carry out its own DPIA with regard to the specific implementation, but this can be informed by a DPIA prepared by the product provider, if appropriate”).

[192] As demonstrated in the field of Corporate Social Responsibility, the lack or vagueness of specific operational implementation of general law requirements can hamper the effectiveness of value-oriented regulations; Wagner 2018b.

Despite the differences between these three models, they each share a core concern with protecting human rights, recognised as a key issue in all of them. Moreover, while this first generation of AI regulation reveals a pragmatic approach with a focus on risk management at the expense of a framework of guiding principles and a broader consideration of the role of AI in society, this does not rule out a greater emphasis on these aspects in future regulation, as happened with data protection.

Identifying a common core of principles can be of help for this second stage of AI regulation. In the end, therefore, all three approaches can contribute in different ways and probably with different timescales to posing the building blocks of AI regulation.

In these early proposals for AI regulation, the emphasis on risk management is not accompanied by effective models to assess the impact of AI on human rights. Following the turn from ethical guidelines to legal provisions, there are no specific instruments to assess not just the legal compliance of AI solutions, but their social acceptability, including a participatory evaluation of their coherence with the values of the target communities.

Analysis of the current debate confirms that the HRESIA may not only be an effective response to human-rights oriented AI development which also encompasses societal values, but it may also bridge a gap in the present regulatory proposals. Furthermore, a general risk assessment methodology is better suited to the variety of AI and technology developments than regulatory models based on a predefined list of high-risk applications or, at any rate, might represent a better guide to rule-makers in their definition.

References

40th International Conference of Data Protection and Privacy Commissioners, Declaration on Ethics and Data Protection in Artificial Intelligence, 2018.

Access Now (2019) The European Human Rights Agenda in the Digital Age. https://www.accessnow.org/access-now_the-european-human-rights-agenda-in-the-digital-age_final1/. Accessed 4 July 2020.

AI Now Institute (2018) Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems. https://ainowinstitute.org/litigatingalgorithms.pdf. Accessed 5 February 2020.

AlgorithmWatch (2021) Draft AI Act: EU Needs to Live up to Its Own Ambitions in Terms of Governance and Enforcement. https://algorithmwatch.org/en/wp-content/uploads/2021/08/EU-AI-Act-Consultation-Submission-by-AlgorithmWatch-August-2021.pdf. Accessed 6 August 2021.

Andorno R (2005) The Oviedo Convention: A European Legal Framework at the Intersection of Human Rights and Health Law. Journal of International Biotechnology Law 2(1):133–143.

Article 29 Data Protection Working Party (2017) Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679’, WP 248 rev.01.

Article 29 Data Protection Working Party (2018) Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679.

Asaro P (2016) Jus Nascendi, Robotic Weapons and the Martens Clause. In: Calo R, Froomkin A, Kerr I (eds) Robot Law. Edward Elgar Publishing, Cheltenham, pp 367–386.

Azencott CA (2018) Machine Learning and Genomics: Precision Medicine versus Patient Privacy. Phil. Trans. R. Soc. A 376:20170350.

Beauchamp TL (1990) Promise of the Beneficence Model for Medical Ethics. J. Contemp. Health L. & Pol'y 6:145–155.

Bergkamp L (2015) Is There a Defect in the European Court's Defect Test? Musings about Acceptable Risk. European Journal of Risk Regulation 6:309–322.

Cabitza F, Rasoini R, Gensini GF (2017) Unintended Consequences of Machine Learning in Medicine. JAMA 318:517–518.

Caplan R, Donovan J, Hanson L, Matthews J (2018) Algorithmic Accountability: A Primer. https://datasociety.net/output/algorithmic-accountability-a-primer/. Accessed 24 May 2019.

Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N (2015) Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the 21st Annual SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721–1730. http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf. Accessed 14 January 2020.

Center for Data Innovation (2021) How Much Will the Artificial Intelligence Act Cost Europe? https://www2.datainnovation.org/2021-aia-costs.pdf. Accessed 16 August 2021.

Commission Nationale de l'Informatique et des Libertés – LINC (2017) La Plateforme d'une Ville Les Données Personnelles Au Cœur de La Fabrique de La Smart City. https://www.cnil.fr/sites/default/files/atoms/files/cnil_cahiers_ip5.pdf. Accessed 18 November 2019.

Council of Europe (2020) Towards regulation for AI systems. Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law, Compilation of contributions prepared by the CAHAI Secretariat, DGI (2020)16. https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a. Accessed 5 January 2021.

Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) (2020a) Feasibility Study, CAHAI(2020)23. https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da. Accessed 29 July 2021.

Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) (2020b) Elaboration of the feasibility study. Analysis of the International legally binding instruments. Final report. Paper prepared by Alessandro Mantelero, CAHAI(2020)08-fin. https://rm.coe.int/cahai-2020-08-fin-mantelero-binding-instruments-report-2020-def/16809eca33. Accessed 29 July 2021.

Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) (2020c) Working methods of the CAHAI: functioning of the working groups, CAHAI(2020)10 ADD REV1. https://rm.coe.int/cahai-2020-10-add-rev1-en/16809ee918. Accessed 2nd August 2021.

Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) (2021a) Analysis of the Multi-Stakeholder Consultation, Strasbourg, CAHAI(2021)07. https://rm.coe.int/cahai-2021-07-analysis-msc-23-06-21-2749-8656-4611-v-1/1680a2f228. Accessed 4 August 2021.

Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) (2021b) 5th meeting. Strasbourg, 5–7 July 2021. Abridged meeting report and list of decisions, CAHAI(2021)10. https://rm.coe.int/cahai-2021-10-5th-plenary-abridged-report-2776-1003-8532-v-2/1680a31d48. Accessed 5 August 2021.

Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) (2021c) Progress Report by the Co-Chairs of the CAHAI-PDG, CAHAI(2021)09. https://rm.coe.int/cahai-2021-09-pdg-progress-report-2784-0682-4452-v-1/1680a2fd49. Accessed 4 August 2021.

Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) (2021d) Progress Report by the Co-Chairs of the CAHAI-LFG, CAHAI(2021)08. https://rm.coe.int/cahai-2021-08-eng-cahai-lfg-progress-report-june-2021-2770-4668-9539-v/1680a2f5cc. Accessed 5 August 2021.

Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) (2021e) Human Rights, Democracy and Rule of Law Impact. Assessment of AI systems, CAHAI-PDG(2021)05.

https://rm.coe.int/cahai-pdg-2021-05-2768-0229-3507-v-1/1680a291a3. Accessed 5 August 2021.

Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) (2021f) – Policy development Group (CAHAI-PDG) Human Rights, Democracy and Rule of Law Impact. Assessment of AI systems, CAHAI-PDG(2021)05.

Council of Europe-Committee of experts on internet intermediaries (MSI-NET) (2018) Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications. https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5. Accessed 10 March 2020.

Council of Europe, Committee of Ministers (1997) Recommendation No. R(97)5 of the Committee of Ministers to member States on the protection of medical data.

Council of Europe, Committee of Ministers (2010) Recommendation CM/Rec(2010)13 of the Committee of Ministers of the Council of Europe to member States on the protection of individuals with regard to automatic processing of personal data in the context of profiling.

Council of Europe, Committee of Ministers (2016a) Recommendation CM/Rec(2016)6 of the Committee of Ministers to member States on research on biological materials of human origin.

Council of Europe, Committee of Ministers (2016b) Recommendation CM/Rec(2016)8 on the processing of personal health-related data for insurance purposes, including data resulting from genetic tests and its Explanatory Memorandum.

Council of Europe, Committee of Ministers (2019) Recommendation CM/Rec(2019)2 of the Committee of Ministers to member States on the protection of health-related data.

Council of Europe, Committee of Ministers (2020) Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems. https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154. Accessed 10 March 2020.

Council of Europe, Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) (2017) Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data, T-PD(2017)01. https://rm.coe.int/t-pd-2017-1-bigdataguidelines-en/16806f06d0. Accessed 15 April 2020.

Council of Europe, Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) (2019) Guidelines on Artificial Intelligence and Data Protection, T-PD(2019)01. https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8. Accessed 15 April 2020.

Council of Europe, Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data (Convention 108) (2021) Guidelines on Facial Recognition, 28 January 2021, T-PD(2020)03rev4.

Council of Europe, European Commission for the Efficiency of Justice (CEPEJ) (2018) European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment. https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment. Accessed 4 March 2019.

Council of Europe, Parliamentary Assembly (2017) Recommendation 2102 (2017)1 Technological Convergence, Artificial Intelligence and Human Rights.

Data Ethics Commission of the Federal Government, Federal Ministry of the Interior Building and Community and Data Ethics Commission (2019) Opinion of the Data Ethics Commission. https://www.bmjv.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission_EN_node.html. Accessed 16 June 2020.

De Hert P, Papakonstantinou V (2016) The New General Data Protection Regulation: Still a Sound System for the Protection of Individuals? Computer Law & Security Review 32(2):179–194.

Edwards L, Veale M (2017) Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For. Duke Law & Technology Review 16(1):18–84.

European Center for Not-for-Profit Law (2021) ECNL Position Statement on the EU AI Act. https://ecnl.org/sites/default/files/2021-07/ECNL%20EU%20AI%20Act%20Position%20Paper.pdf. Accessed 7 August 2021.

European Commission (2020a) A European strategy for data, COM(2020) 66 final. https://ec.europa.eu/info/sites/default/files/communication-european-strategy-data-19feb2020_en.pdf. Accessed 15 March 2020.

European Commission (2020b) Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee, COM/2020/64 final. https://ec.europa.eu/info/files/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics_en. Accessed 12 March 2020.

European Commission (2020c) Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics, COM/2020/64 final. https://ec.europa.eu/info/files/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics_en. Accessed 12 March 2020.

European Commission (2020d) White Paper on Artificial Intelligence – A European Approach to Excellence and Trust, COM(2020) 65 final. https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en. Accessed 12 March 2020.

European Commission, Expert Group on Liability (2019) Liability for Artificial Intelligence and Other Emerging Digital Technologies. https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2020/01-09/AI-report_EN.pdf. Accessed 13 January 2020.

European Data Protection Supervisor (2015) Towards a new digital ethics: Data, Dignity and Technology. https://edps.europa.eu/data-protection/our-work/publications/opinions/towards-new-digital-ethics-data-dignity-and_en. Accessed 4 October 2021.

European Data Protection Supervisor (2018) Public Consultation on Digital Ethics. Summary of Outcomes. https://edps.europa.eu/sites/edp/files/publication/18-09-25_edps_publicconsultation-digitalethicssummary_en.pdf. Accessed 12 March 2020.

European Data Protection Supervisor, Ethics Advisory Group (2018) Towards a Digital Ethics. https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf. Accessed 12 March 2020.

Ferguson AG (2017) The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement. New York University Press, New York.

Ferryman K, Pitcan M (2018) Fairness in Precision Medicine. Data & Society. https://datasociety.net/wp-content/uploads/2018/02/Data.Society.Fairness.In_.Precision.Medicine.Feb2018.FINAL-2.26.18.pdf. Accessed 25 June 2020.

Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M (2020) Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center for Internet & Society, Cambridge, MA, https://papers.ssrn.com/abstract=3518482. Accessed 12 April 2020.

Goodman EP, Powles J (2019) Urbanism Under Google: Lessons from Sidewalk Toronto. Fordham L. Rev. 88(2):457–498.

Gregory CO (1951) Trespass to Negligence to Absolute Liability. Virginia Law Review 37 (3):359–397.

Hansson SO (2013) The Ethics of Risk. Palgrave Macmillan, New York.

https://www.sciencedirect.com/science/article/pii/B9780750675796500182. Accessed 5 August 2021.

Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission (2019) Ethics Guidelines for Trustworthy AI. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html. Accessed 12 March 2020.

Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission (2020) The Assessment List For Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment. Accessed 17 September 2021.

Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K (2015) Dynamic Consent: A Patient Interface for Twenty-first Century Research Networks. European Journal of Human Genetics 23(2):141–146.

Mantelero A (2016) Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection. Computer Law & Sec. 32(2):238–255.

Mantelero A (2017) Regulating Big Data. The Guidelines of the Council of Europe in the Context of the European Data Protection Framework. Computer Law & Security Rev. 33(5):584–602.

Mantelero A (2018a) Artificial Intelligence and Data Protection: Challenges and Possible Remedies. Report on Artificial Intelligence, T-PD(2018)09Rev, Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of personal data: Strasbourg, 2019. https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6. Accessed 20 July 2020.

Mantelero A (2018b) AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment. Computer Law & Sec. Rev. 34(4):754–772.

Mayer-Schönberger V (1997) Generational Development of Data Protection in Europe. In: Agre PE, Rotenberg M (eds) Technology and Privacy: The New Landscape. The MIT Press, Cambridge, pp 219–241.

McMahon BME (1968) The Reactions of Tortious Liability to Industrial Revolution: A Comparison: I. Irish Jurist 3(1):18–32.

Microsoft Corporation (2021) Feedback from: Microsoft Corporation [to the European Commission's Proposal for a Regulation on Artificial Intelligence (AI) Systems]. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665556_en. Accessed 9 August 2021.

Moyes R (2016) Key Elements of Meaningful Human Control. Background Paper to Comments. Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) Geneva, 11–15 April 2016. Article 36. https://article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf. Accessed 24 May 2021.

Muhlbauer WK (2004) Risk Management. In: Muhlbauer WK (ed) Pipeline Risk Management Manual. Gulf Professional Publishing, Amsterdam, pp 331–355.

Narayanan A, Huey J, Felten EW (2016) A Precautionary Approach to Big Data Privacy. In: Gutwirth S, Leenes R, De Hert P (eds) Data Protection on the Move. Springer, Dordrecht, pp 357–385.

Nordlander K, Simon C-M, Pearson H (2010) Hazard v. Risk in EU Chemicals Regulation. European Journal of Risk Regulation 1 (3):239 239–250.

OECD (2013) Recommendation of the Council concerning Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data, C(80)58/FINAL, as amended on 11 July 2013 by C(2013)79.

OECD (2015) Recommendation of the Council on Digital Security Risk Management for Economic and Social Prosperity. https://www.oecd-ilibrary.org/science-and-technology/digital-security-risk-management-for-economic-and-social-prosperity/recommendation-of-the-council-on-digital-security-risk-management-for-economic-and-social-prosperity_9789264245471-1-en. Accessed 18 March 2019.

OECD (2019) Recommendation of the Council on Artificial Intelligence. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Accessed 29 July 2019.

Office of the High Commissioner for Human Rights (2000) CESCR General Comment No. 14: The Right to the Highest Attainable Standard of Health (Art. 12) Adopted at the Twenty-Second Session of the Committee on Economic, Social and Cultural Rights, on 11 August 2000 (Contained in Document E/C.12/2000/4).

Ohm P (2010) Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA L. Rev. 57:1701–1777.

Oliphant K (2005) Rylands v Fletcher and the Emergence of Enterprise Liability in the Common Law. In: Koziol H, Steininger BC (eds) European Tort Law, Vol. 2004. Tort and Insurance Law Yearbook. New York, Springer, pp 81–120.

Peel J (2004) Precaution – A Matter of Principle, Approach or Process? Melb. J. Int. Law 5 (2):483–501.

Pellegrino ED, Thomasma DC (1987) The Conflict between Autonomy and Beneficence in Medical Ethics: Proposal for a Resolution. The Journal of Contemporary Health Law and Policy 3:23–46.

Raso F, Hilligoss H, Krishnamurthy V, Bavitz C, Kim L (2018) Artificial Intelligence & Human Rights Opportunities & Risks. Berkman Klein Center for Internet & Society. https://cyber.harvard.edu/sites/default/files/2018-09/2018-09_AIHumanRightsSmall.pdf?subscribe=Download+the+Report. Accessed 12 April 2020.

Rouvroy A (2015) "Of Data and Men" – Fundamental rights and freedoms in a world of Big Data. Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of personal data, T-PD-BUR(2015)09Rev. http://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020. Accessed 24 June 2020.

Seatzu F (2015) The Experience of the European Court of Human Rights with the European Convention on Human Rights and Biomedicine. Utrecht Journal of International and European Law 31(81):5–16.

Selbst AD, Barocas S (2018) The Intuitive Appeal of Explainable Machines. Fordham L. Rev. 87:1085–1139.

Selbst AD, Powles J (2017) Meaningful Information and the Right to Explanation. International Data Privacy Law 7(4):233–242.

Sheehan M (2011) Can Broad Consent be Informed Consent? Public Health Ethics 3:226–235.

Simonite T (2021) These Algorithms Look at X-Rays—and Somehow Detect Your Race. Wired, May 5. https://www.wired.com/story/these-algorithms-look-x-rays-detect-your-race/. Accessed 7 August 2021.

Strand R, Kaiser M (2015) Report on Ethical Issues Raised by Emerging Sciences and Technologies. Council of Europe, Committee on Bioethics, Strasbourg. https://www.coe.int/T/DG3/Healthbioethic/Activities/12_Emerging%20technologies/BergenStudy%20e.pdf. Accessed 12 May 2020.

Taylor L, Dencik L (2020) Constructing Commercial Data Ethics. Technology and Regulation. https://techreg.org/index.php/techreg/article/view/35/9. Accessed 14 April 2020.

Taylor L, Floridi L, van der Sloot B (eds) (2017) Group Privacy New Challenges of Data Technologies. Springer International Publishing, Cham.

ten Have HAMJ, Jean MS (2009) The UNESCO Universal Declaration on Bioethics and Human Rights: Background, Principles and Application. UNESCO, Paris.

The Norwegian Data Protection Authority (2018) Artificial Intelligence and Privacy Report. https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf. Accessed 15 July 2019.

The Public Voice (2018) Universal Guidelines for Artificial Intelligence. https://thepublicvoice.org/AI-universal-guidelines/. Accessed 5 May 2019.

Traynor RJ (1965) The Ways and Meanings of Defective Products And Strict Liability. Tenn. L. Rev. 32(3):363–376.

UNESCO (2019) Preliminary Study on a Possible Standard-Setting Instrument on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000369455. Accessed 8 March 2020.

United Nations, Office of the High Commissioner for Human Rights (2018) Guidelines for States on the Effective Implementation of the Right to Participate in Public Affairs. https://www.ohchr.org/EN/Issues/Pages/DraftGuidelinesRighttoParticipationPublicAffairs.aspx. Accessed 20 November 2019.

Wachter S, Mittelstadt B, Floridi L (2017) Why a right to explanation of automated decision – making does not exist in the General Data Protection Regulation. International Data Privacy Law 7(2):76–99.

Wagner B (2018a) Ethics as an Escape from Regulation: From Ethics. In: Bayamlioğlu E, Baraliuc I, Janssens LAW, Hildebrandt M (eds) Being Profiling. Cogitas Ergo Sum. Amsterdam University Press, Amsterdam, pp 84–89.

Wagner CZ (2018b) Evolving Norms of Corporate Social Responsibility: Lessons Learned from the European Union Directive On Non-Financial Reporting. Transactions: The Tennessee Journal of Business Law 19:619–708.

Westin AF, Baker MA (1972) Databanks in a Free Society. Computers, Record-Keeping and Privacy. Quadrangle/The New York Time Book Co., New York.

Whipple C (1988) Acceptable Risk. In: Travis CC (ed) Carcinogen Risk Assessment. Springer, Boston, pp 157–170.

Yamin AE (2005) The Right to Health Under International Law and Its Relevance to the United States. American Journal of Public Health 95(7):1156–1161.

Zuiderveen Borgesius F (2018) Discrimination, Artificial Intelligence, and Algorithmic Decision-Making. Anti-discrimination department of the Council of Europe. https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73. Accessed 16 May 2020.