Regulating AI within the Human Rights Framework: A Roadmapping Methodology

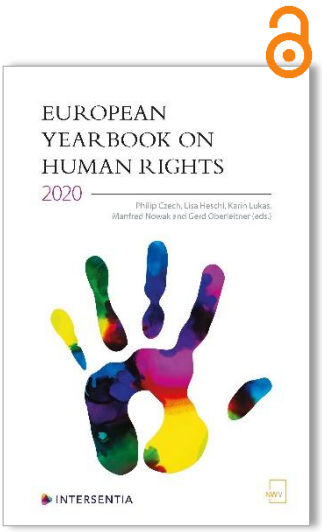
INTERSENTIA

'Regulating AI within the Human Rights Framework: A Roadmapping Methodology'

Alessandro Mantelero

From *European Yearbook on Human Rights 2020* by Philip Czech, Lisa Heschl, Karin Lukas, Manfred Nowak and Gerd Oberleitner (eds.)

The European Yearbook on Human Rights brings together renowned scholars, emerging voices and practitioners. Split into parts devoted to recent developments in the European Union, the Council of Europe and the OSCE as well as through reports from the field, the contributions engage with some of the most important human rights issues and developments in Europe.



For more information about this publication, visit:
https://intersentia.com/en/european-yearbook-on-human-rights-2020-41542.html

REGULATING AI WITHIN THE HUMAN RIGHTS FRAMEWORK

A Roadmapping Methodology

Alessandro Mantelero

ABSTRACT

The ongoing European debate on Artificial Intelligence (AI) is increasingly polarised between the initial ethics-based approach and the growing focus on human rights. The prevalence of one or the other of these two approaches is not neutral and entails consequences in terms of regulatory outcomes and underlying interests.

The basic assumption of this study is the need to consider the pivotal role of ethics as a complementary element of a regulatory strategy, which must have human rights principles at its core. Based on this premise, this contribution focuses on the role that the international human rights framework can play in defining common binding principles for AI regulation.

The first challenge in considering human rights as a frame of reference in AI regulation is to define the exact nature of the subject matter. Since a wide range of AI-based services and products have only emerged as a recent development of the digital economy, many of the existing international legal instruments are not tailored to the specific issues raised by AI. Moreover, certain binding principles and safeguards were shaped in a different technological era and social context.

Against this background, we need to examine the existing binding international human rights instruments and their non-binding implementations to extract the key principles that should underpin AI development and govern its ground-breaking applications.

However, the paradigm shift brought about by the latest wave of AI development means that the principles embodied in international legally binding instruments cannot be applied in their current form, and this contribution sets out to contextualise these guiding principles for the AI era.

Given the broad application of AI solutions in a variety of fields, we might look at the entire corpus of available international binding instruments. However, taking a methodological approach, this analysis focuses on two key areas – data protection and healthcare – to provide an initial assessment of the regulatory issues and a possible roadmap to addressing them.

1. INTRODUCTION

The last few years have seen a growing debate on the ethical dimension of data use and the new challenges and issues posed by data-intensive systems based on Big Data and Artificial Intelligence (AI). However, as in the past, uncertainty about the potential impact of new technology and an existing legal framework not tailored for the new socio-technical scenario have led policy-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 these 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'.[1]

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.[2] In the same vein, the European Commission set up a high-level group focusing on ethical issues.[3] This ethical wave later resulted in a flourishing of ethical principles, codes and ethical boards in private companies.[4]

This new focus, which also presented the danger of 'ethics-washing',[5] 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 (CoE), which has remained focused on its traditional human rights-centred mission,[6] and the recent change of direction of the

---

1 See EDPS Ethics Advisory Group, 'Towards a Digital Ethics', 2018, p. 7 available at https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf, last accessed 02.03.2020. See also EDPS, 'Public Consultation on Digital Ethics. Summary of Outcomes', 2018, available at https://edps.europa.eu/sites/edp/files/publication/18-09-25_edps_publicconsultationdigitalethicssummary_en.pdf, last accessed 02.03.2020.

2 See A. Mantelero, 'Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection', (2016) 32(2), *Computer Law & Sec.*, pp. 238–255. See also A.G. Ferguson, *The Rise of Big Data Policing : Surveillance, Race, and the Future of Law Enforcement*, New York University Press, New York 2017; E.P. Goodman and J. Powles, 'Urbanism Under Google: Lessons from Sidewalk Toronto', (2019) 88(2), *Fordham Law Review,* pp. 457–498.

3 See Independent High-level Expert Group on Artificial Intelligence set up by the European Commission, 'Ethics Guidelines for Trustworthy AI', 2019, available at https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, last accessed 02.03.2020.

4 See also L. Taylor and L. Dencik, 'Constructing Commercial Data Ethics', (2020) *Technology and Regulation* 1, available at https://techreg.org/index.php/techreg/article/view/35/9, last accessed 14.04.2020.

5 See B. Wagner, 'Ethics as an Escape from Regulation: From Ethics', in E. Bayamlioğlu, I. Baraliuc, L.A.W. Janssens et al. (eds.), *Being Profiling. Cogitas Ergo Sum*. Amsterdam University Press, Amsterdam 2018, pp. 84–89.

6 During its 1353rd meeting on 11 September 2019, the Committee of Ministers of the Council of Europe set up 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 the Council of Europe's standards on human rights, democracy and the rule of law.

European Commission, with its new bundle of proposals for AI regulation.[7] These bodies do not marginalise the role of ethics, but see moral and social values as complementary to a strategy based on human rights, rule of law and democracy.

As discussed elsewhere,[8] only a human rights-centred approach can benefit from a universal vision of common values, a corpus of existing sector-specific provisions and jurisprudence by *ad hoc* regional courts. On the other hand, the diversity of ethical approaches and the under-representation of non-Western ethics in the debate are intrinsic limits to any attempt to address AI challenges from a purely ethical perspective.

The growing interest in a human rights approach to AI needs to be better framed so as to avoid reduction to broad policy indications or a repetition of general principles (e.g. non-discrimination, transparency, solidarity, etc.) that are lacking in adequate contextualisation, which is crucial to any regulatory framework.

Thus, future AI regulation must be grounded on existing legal human rights instruments, but also go beyond it to provide a tailored and contextual application of these rights and bridge the gaps created by instruments drafted in a pre-AI era.

Having this regulatory goal in mind, this study hopes to contribute to the development of a dual process of contextualisation and integration of the human rights framework within the AI scenario. Since it would be too ambitious to cover the entire range of AI effects on human rights, this work focuses on the methodology, to provide an initial application of this approach in two key areas: data processing and healthcare.

By focusing on specific fields, the following sections aim to give a methodological contribution to the definition of a roadmap for the ongoing international regulatory debate on AI.

This work is made up of four sections. Following this introduction, Section 2 focuses on the methodological approach, while Section 3 examines the fields of data processing and healthcare. Based on this analysis, the last section draws some preliminary conclusions on the methodology adopted and its results.

---

7 See European Commission, 'White Paper on Artificial Intelligence – A European Approach to Excellence and Trust', COM(2020) 65 final, 2020, available at https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en, last accessed 02.03.2020. See also European Commission, 'Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee, COM(2020) 64 final, 2020, available at https://ec.europa.eu/info/files/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics_en, last accessed 02.03.2020.

8 See A. Mantelero, 'AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment', (2018) 34(4), *Computer Law & Security Review*, pp. 754–772.

2. METHODOLOGICAL APPROACH

The obligatory starting point in identifying the guiding legal values that should underpin future AI regulation within the framework of human rights is to analyse the existing international legally binding instruments. This includes a gap analysis to ascertain the extent to which the current regulatory framework and its values properly address the issues raised by AI.

Moreover, we need to analyse the state of the art with a view to preserving the harmonisation of the human rights framework, while introducing coherent new AI-specific provisions. From a methodological perspective, this approach does not set out to create a completely new and comprehensive reference framework, since AI regulation should focus on the changes AI will bring to society, not on reshaping every area where AI can be applied.

The methodology suggested is therefore based on a targeted approach, building on the existing binding instruments, contextualising their guiding principles and providing key regulatory guidelines for a future AI legal framework, which can cover areas that are not presently regulated by the existing legal instruments.

Analysis of the existing binding legal instruments cannot be limited to a merely harmonising effort (i.e. extracting common values and principles from a given set of rules), but requires a more articulated process in which harmonisation is just one of several stages. The process, described in the following sections, can be divided into four separate steps: mapping; identification of key principles; contextualisation; and harmonisation.

Regarding the existing binding instruments, we should note that the rapid evolution of applied AI over the last few years was incompatible with a specific response in terms of international legislation on AI. As a consequence, rule makers adopted two different strategies: (i) a widespread effort to interpret the existing legal framework in the light of AI-related issues; and (ii) use of non-binding instruments to contextualise the principles set forth in the existing binding instruments.[9]

In the mapping exercise, this study therefore takes into account the guiding principles and values deriving from both the existing binding instruments and the related non-binding implementations, which in some cases already contemplate the

[9] See e.g. Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108), 'Guidelines on Artificial Intelligence and Data Protection', T-PD(2019)01, 25 January 2019, available at https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8, last accessed 02.03.2020; European Commission for the Efficiency of Justice (CEPEJ), 'European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and Their Environment', 2018, available at https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment, last accessed 02.03.2020.

new AI scenario. The theoretical basis of this approach relies on the assumption that the general principles provided by international human rights instruments should underpin all human activities, including AI-based innovation.[10]

Since this contribution sets out to define the key principles for the future regulation of AI through analysis of the existing legal framework, the methodology is necessarily deductive, extracting these principles from the range of regulations governing the fields in which AI solutions may be adopted.

There are two different approaches to this analysis: 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 might affect their exercise. In the second approach, 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 health 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 these rights. On the other hand, a focus on these 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 will be contextualised within the scenario transformed by AI, which in many cases requires their adaptation. The principles remain valid, but their implementation will be reconsidered in light of the social and technical changes due to AI.[12] This will deliver a more precise and granular application of the principles so that they can provide a concrete contribution to the shape of future AI regulation.

What is more, the methodology requires a vertical analysis of the key principles in each of the fields regulated by these international instruments, followed by a second phase considering the similarities and common approaches

---

10 See Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, adopted by the Committee of Ministers on 8 April 2020.

11 See J. Fjeld, N. Achten, H. Hilligoss et al., 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI', Berkman Klein Center for Internet & Society, Cambridge, MA 2020, available at https://papers.ssrn.com/abstract=3518482, last accessed 02.03.2020; F. Raso, H. Hilligoss, V. Krishnamurthy et al., 'Artificial Intelligence & Human Rights Opportunities & Risks', Berkman Klein Center for Internet & Society, Cambridge, MA 2018, available at https://cyber.harvard.edu/sites/default/files/2018-09/2018-09_AIHumanRightsSmall.pdf?subscribe=Download+the+Report, last accessed 02.03.2020.

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

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 proposed model focuses on thematic areas, starting from an overview of the current legal framework, comprising both binding and non-binding instruments. On this basis, it is possible to draw up a list of guiding principles common to all realms in a second phase. These shared principles can then 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. This phase, which will be developed separately for each principle identified, is crucial to providing a more refined and elaborate formulation of the key principles, based on the nature of AI products and services.

From a more general perspective going beyond the data and healthcare sectors, this methodological approach can also reveal the existence of important areas affected by AI which are only partially covered by binding instruments. These areas will then be mapped to consider whether the guiding principles can be extended to them or if new values need to be developed in line with the existing framework of human rights, democracy and the rule of law.

3. FRAMEWORK ANALYSIS

AI technologies impact on a variety of sectors[13] and raise issues concerning a large body of regulatory instruments. Careful examination of the existing principles and provisions of these instruments and a gap analysis of AI-related issues necessarily requires a considerable research effort in terms of time and resources.

For this reason, in order to validate the methodological approach proposed, this study concentrates on just two key areas, data protection and healthcare, where the impact of AI on individuals and society is particularly marked and the challenges are significant.

The intersection between these two realms is interesting in view of the focus of this contribution on the core of future AI regulation, given the large number of AI applications concerning healthcare data and the common ground between the two fields. This is reflected in several provisions of international

[13] See also UNESCO, 'Preliminary Study on a Possible Standard-Setting Instrument on the Ethics of Artificial Intelligence', Paris 2019, available at https://unesdoc.unesco.org/ark:/48223/pf0000369455, last accessed 02.03.2020.

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.

The following sections consider relevant legally binging instruments, including the related non-binding instruments, adopted by international and intergovernmental organisations. Only in the case of data protection, the document adopted by the Conference of Data Protection and Privacy Commissioners (now Global Privacy Assembly)[16] has also been examined, given the role assigned to data protection authorities in international instruments.

3.1. DATA PROTECTION

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 General Data Protection Regulation (GDPR). A similar redefinition of the regulatory framework has occurred, or is ongoing, in other international contexts – such as the Organisation for Economic Co-operation and Development (OECD)[17] – and in individual countries.

However, given the rapid development of the most recent wave of AI, these new measures fail to directly address some AI-specific challenges, and several

14 See 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 See Recommendation CM/Rec(2019)2 of the Committee of Ministers to Member States on the protection of health-related data.

16 See https://globalprivacyassembly.org, last accessed 25.05.2020.

17 See OECD, 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.

non-binding instruments have been adopted to bridge this gap, as well as future regulatory strategies under discussion.[18]

For the purposes of this study, this section examines the following data-related, international, non-binding legal instruments: Council of Europe, Guidelines on Artificial Intelligence and Data Protection (GAI);[19] Council of Europe, Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data (GBD);[20] Recommendation CM/Rec(2019)2 of the Committee of Ministers of the Council of Europe to member States on the protection of health-related data (CM/Rec(2019)2);[21] 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); United Nations Educational, Scientific and Cultural Organisation (UNESCO), Preliminary Study on a Possible Standard-Setting Instrument on the Ethics of Artificial Intelligence, 2019 (UNESCO 2019);[22] OECD, Recommendation of the Council on Artificial Intelligence, 2019 (OECD);[23] and the 40th International Conference of Data Protection and Privacy Commissioners, Declaration on Ethics and Data Protection in Artificial Intelligence, 2018 (ICDPPC).[24,25]

---

18 European Commission, Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics, COM(2020) 64 final, Bruxelles, 2020; European Commission, White Paper on Artificial Intelligence – A European Approach to Excellence and Trust, COM(2020) 65 final, Bruxelles 2020. See also European Commission, 2020, A European strategy for data, COM(2020) 66 final.

19 Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108), 'Guidelines on Artificial Intelligence and Data Protection', *supra* note 9.

20 Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108), 'Guidelines on the protection of individuals with regard to the processing of personal data in a World of Big Data, T-PD(2017)01', 23.01.2017, available at https://rm.coe.int/t-pd-2017-1-bigdataguidelines-en/16806f06d0, last accessed 02.03.2020.

21 This Recommendation has replaced Recommendation No. R(97)5 of the Committee of Ministers to member States on the protection of medical data. See also Rec(2016)8 on the processing of personal health-related data for insurance purposes, including data resulting from genetic tests, and its Explanatory Memorandum.

22 Despite the reference to ethics only in the title, the purpose of the study 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'. The preliminary study is available at https://unesdoc.unesco.org/ark:/48223/pf0000369455, last accessed 02.03.2020.

23 Available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449, last accessed 02.03.2020.

24 Available at https://edps.europa.eu/sites/edp/files/publication/icdppc-40th_ai-declaration_adopted_en_0.pdf, last accessed 02.03.2020.

25 See also Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems [CM/Rec(2020)1], available at https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154, last accessed 10.04.2020.

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 providing, or only partially providing, more detailed rules.

Based on the mapping exercise of these instruments and focusing on those provisions that are most pertinent to AI issues,[26] we can identify several general guiding principles which are then contextualised in respect of AI in the following paragraphs. 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.

The 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 focused 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,[27] as well as digital literacy, education and professional training.[28]

3.1.1. Primacy of the Human Being

Although this principle is only explicitly enshrined in the Oviedo Convention and not in the binding instruments on data protection, such as Conventions 108 and 108+, the primacy of the human being is an implicit reference when data is

---

26 For a broader analysis of the issues concerning data protection and human rights in general, see Council of Europe-Committee of Experts on Internet Intermediaries (MSI-NET), 'Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications', 2018, available at https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5, last accessed 02.03.2020; A. Mantelero (2018), 'AI and Big Data: A Blueprint', *supra* note 8; F.Z. Borgesius, 'Discrimination, Artificial Intelligence, and Algorithmic Decision-Making', Anti-discrimination Department of the Council of Europe, 2018, available at https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73, last accessed 02.03.2020. See also J. Fjeld, N. Achten, H. Hilligoss et al., 'Principled Artificial Intelligence', *supra* note 11; F. Raso, H. Hilligoss, V. Krishnamurthy et al., 'Artificial Intelligence & Human Rights', *supra* note 11.

27 See also CM/Rec(2019)2, 1, para. 14.

28 See ICDPPC, OECD, GAI para. III.9, UNESCO 2019, and CM/Rec(2020)1, para. 7.

used in the context of innovative technologies.[29] This is reflected in the idea that data processing operations must 'serve the data subject'.[30] More generally, the primacy of the human being over science is 'a direct corollary'[31] of the principle of respect for human dignity. Dignity is a constitutive element of the European approach to data processing[32] and of the international approach to civil and political rights in general.[33] Wider reference to human dignity can also be found in the non-binding instruments focused on AI.[34]

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.

3.1.2. Human Control and Oversight

Since the notion of data protection originally rested on the idea of control over the use of information in information and communications technology (ICT), and the first data protection regulations were designed to give individuals some counter-control over the data that was collected,[35] human control plays a central role in this area. It is also related to the importance of self-determination[36] 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 technological applications ensures a degree of safeguarding against the possible adverse consequences for human rights and freedoms.

---

29 See Council of Europe – Parliamentary Assembly, 'Recommendation 2102 (2017)1 Technological Convergence, Artificial Intelligence and Human Rights', 2017. See also R. Strand and M. Kaiser, 'Report on Ethical Issues Raised by Emerging Sciences and Technologies', Council of Europe, Committee on Bioethics, Strasbourg, 2015, p. 6, available at https://www.coe.int/T/DG3/Healthbioethic/Activities/12_Emerging%20technologies/BergenStudy%20e.pdf, last accessed 02.03.2020.

30 See CM/Rec(2019)2, Preamble.

31 See H.A.M.J. ten Have and M.S. Jean, 'The UNESCO Universal Declaration on Bioethics and Human Rights: Background, Principles and Application', UNESCO, Paris, 2009, p. 93.

32 See 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').

33 See International Covenant on Civil and Political Rights, Preamble.

34 See GAI, paras. I.1 and II.1; UNESCO 2019, para. II.3, OECD, para. IV.1.2.

35 A.F. Westin, *Privacy and Freedom*, Atheneum, New York 1970, p. 7; D.J. Solove, *Understanding Privacy*, Harvard University Press, Cambridge, MA 2008, pp. 24–29. See also Secretary's Advisory Committee on Automated Personal Data Systems, 'Records, Computers and the Rights of Citizens', 1973, available at http://epic.org/privacy/hew1973report/, last accessed 02.03.2020.

36 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+[37] and the non-binding instruments on AI[38] – 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.

Contextualising human control and oversight with regard to artificial intelligence, AI applications should allow meaningful[39] control by human beings over their effects on individuals and society. Moreover, AI products and services must be designed in such a way as 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, AI products and services must allow general human control over them.[40]

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.[41]

3.1.3. Participation and Democratic Oversight of AI Development

Turning to the collective dimension of the use of data in AI,[42] human control and oversight cannot be limited to supervisory entities, data controllers or

---

37 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.

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

39 The adjective 'meaningful' was elaborated in the context of AWS, see R. Moyes, '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', Art. 36, 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 P. Asaro, 'Jus Nascendi, Robotic Weapons and the Martens Clause', in R. Calo, A Froomkin and I. Kerr (eds.), *Robot Law*, Edward Elgar Publishing, Cheltenham 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, 'Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679', Brussels, 06.02.2018, 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').

40 See Convention 108+; GAI, para. II.8; ICDPPC; UNESCO 2019.

41 See GAI, para. III. 4.

42 See A. Mantelero, 'Personal Data for Decisional Purposes in the Age of Analytics', *supra* note 2.

data subjects. Participatory and democratic oversight procedure should give voice to society at large, including various categories of people, minorities and under-represented groups.[43] This supports the notion that participation in decision-making serves to advance human rights and is crucially important in bringing specific issues to the attention of the public authorities.[44]

Since human control over potentially hazardous technology entails a risk assessment,[45] 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 the role that AI should play in shaping social dynamics and in the decision-making processes affecting them.[46]

Derogations may be introduced in the public interest, where proportionate in a democratic society and with adequate safeguards. In policing, intelligence and security, where public oversight is limited, governments should report regularly on their use of AI.[47]

3.1.4. Transparency and Intelligibility

Transparency is a challenging[48] and highly debated topic in the context of AI,[49] with several different interpretations, including the studies on 'Explainable AI'.

---

43 See also CM/Rec(2020)1, para. 5.

44 See ICDPPC, para. 25. See also United Nations – Office of the High Commissioner for Human Rights, 'Guidelines for States on the Effective Implementation of the Right to Participate in Public Affairs', 2018, available at https://www.ohchr.org/EN/Issues/Pages/DraftGuidelinesRighttoParticipationPublicAffairs.aspx, last accessed 02.03.2020.

45 See below Section 3.1.5.

46 See GAI, paras. II.7 and III.8. See also United Nations – Office of the High Commissioner for Human Rights, 'Guidelines for States on the Effective Implementation of the Right to Participate in Public Affairs', *supra* note 44, para. 64.

47 See UNESCO 2019, para. 107.K.

48 See A. Mantelero, '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, available at https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6, last accessed 02.03.2020, pp. 11–13.

49 See e.g. A.D. Selbst and S. Barocas, 'The Intuitive Appeal of Explainable Machines', (2018) 87, *Fordham Law Review*, pp. 1085–1139; S. Wachter, B. Mittelstadt and L. Floridi, 'Why a right to explanation of automated decision – making does not exist in the General Data Protection Regulation', (2017) 7(2), *International Data Privacy Law*, pp. 76–99; A.D. Selbst and J. Powles, 'Meaningful Information and the Right to Explanation', (2017) 7(4), *International Data Privacy Law*, pp. 233–242; L. Edwards and M. Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For', (2017) 16(1), *Duke Law & Technology Review*, pp. 18–84.

In this sense, transparency is one of the data protection principles that is stressed most frequently.[50]

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 the collective use of AI at the group level[51] 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 she or he is interacting directly with an AI system and to be provided adequate and easy-to-understand information on its purpose and effects, including the existence of automated decisions. This information is necessary to enable overall human control on such systems, to verify alignment with individuals' expectations and to enable those adversely affected by an AI system to challenge its outcome.[52] Every individual must also have a right to obtain, on request, knowledge of the reasoning underlying any AI-based decision process where the results of such a process are applied to him or her.[53]

Finally, to foster transparency and intelligibility, governments should promote scientific research on explainable AI and best practices for transparency and auditability of AI systems.[54]

3.1.5. 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'.[55]

Where there is scientific uncertainty about the potential outcome, a precautionary approach[56] should be taken, rather than conducting a risk

50 See Convention 108+, Art. 8.

51 See L. Taylor, L. Floridi and B. Van der Sloot (eds.), *Group Privacy New Challenges of Data Technologies*, Springer International Publishing, Cham 2017.

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

53 See Convention 108+, Art. 9.1.c; GAI, para. II.11.

54 See ICDPPC, para. 3.a.

55 S. Ove Hansso, *The Ethics of Risk*, Palgrave Macmillan, New York 2013, p. 12.

56 See also J. Peel, 'Precaution – A Matter of Principle, Approach or Process?', (2004) 5(2), *Melbourne Journal of International Law* p. 483.

analysis.[57] The same conclusion can be drawn for AI where the potential risks of an AI application are unknown or uncertain.[58] 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.[59]

The development of AI also raises specific forms of risk in the field of data protection. One widely discussed example is that of re-identification,[60] while the risk of decontextualisation is less well known. In the latter case, data-intensive applications may ignore contextual information needed to understand and apply the proposed solution. Decontextualisation 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.[61]

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.[62] 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 (risk of re-identification).[63]

Convention 108+ (like the GDPR) adopts a two-stage approach to risk: an initial self-assessment is followed by a consultation with the competent

---

57 For a broader analysis of risk assessment in the field of AI, see also A. Mantelero (2018) 'AI and Big Data: A Blueprint', *supra* note 8.

58 See GAI, para. II.2. See also A. Mantelero, 'Regulating Big Data. The Guidelines of the Council of Europe in the Context of the European Data Protection Framework', (2017) 33(5), *Computer Law & Security Rev.*, pp. 584–602; ICDPPC ('Highlighting that those risks and challenges may affect individuals and society, and that the extent and nature of potential consequences are currently uncertain'); also CM/Rec(2020)1, Appendix, para. A.15.

59 See GAI, paras. II.2 and II.3; OECD, para. 1.4; UNESCO 2019. See also ICDPPC and OECD, 'Recommendation of the Council on Digital Security Risk Management for Economic and Social Prosperity', 2015, available at 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, last accessed 02.03.2020.

60 See e.g. A. Narayanan, J. Huey and E.W. Felten, 'A Precautionary Approach to Big Data Privacy', in S. Gutwirth, R. Leenes and P. De Hert (eds.), *Data Protection on the Move*, Springer, Dordrecht 2016, pp. 357–385; P. Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', (2010) 57, *UCLA Law Review*, pp. 1701–1777.

61 See R. Caplan, J. Donovan, L. Hanson et al., 'Algorithmic Accountability: A Primer', 2018, available at https://datasociety.net/output/algorithmic-accountability-a-primer/, last accessed 02.03.2020, p. 7; AI Now Institute, 'Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems', 2018, available at https://ainowinstitute.org/litigatingalgorithms.pdf, last accessed 02.03.2020.

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

63 See also CM/Rec(2010)13, para. 8.5.

supervisory authority if there is residual high risk. A similar model can be extended to AI-related risks.[64] 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.[65]

3.1.6. Accountability

The principle of accountability is recognised in Convention 108+[66] and is more generally considered as a key element of risk management policy. In the context of AI,[67] it is important to stress that human accountability cannot be hidden behind the machine. Although AI generates more complicated scenarios,[68] this does not exclude accountability and responsibility of the various human actors involved in the design, development, deployment and use of AI.[69]

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.

3.1.7. 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.[70] 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.[71]

In the context of AI, this means that developers must assess the nature and amount of data used (data quality), minimise the presence of redundant or

64 See 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, 'Opinion of the Data Ethics Commission', 2019, available at https://www.bmjv.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission_EN_node.html, last accessed 02.03.2020, p. 42, which also suggests the introduction of licensing and oversight procedures.

65 See GAI, para. III.4.

66 See Convention 108+, Art. 10.1.

67 See OECD para. IV.1.5; GAI paras. I.2 and III.1.

68 See also European Commission – Expert Group on Liability, 'Liability for Artificial Intelligence and Other Emerging Digital Technologies', 2019.

69 See also Council of Europe – Parliamentary Assembly, 2017, Recommendation 2102 (2017)1 Technological Convergence, Artificial Intelligence and Human Rights, para. 9.1.1.

70 See Convention 108+, Art. 5.

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

marginal data[72] during the development and training phases and then monitor the model's accuracy as it is fed with new data.[73]

AI development and deployment should avoid any potential bias, including unintentional or hidden bias, 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.[74]

3.1.8. Role of Experts and Participation

The complex potential impacts of AI solutions on individuals and society demand that the AI development process cannot be delegated to technicians alone. Experts from various domains can therefore play an important role in this regard and in discerning the potentially adverse consequences of AI applications.[75] Where AI solutions have a significant and extensive impact on society, such vigilance is ineffective without engaging the target communities or groups.

For these reasons, AI developers, manufacturers and service providers should set up and consult independent committees of experts from a range of fields and also engage with independent academic institutions, which can help in the design of human rights-based AI applications.[76] Participatory forms of AI development, based on the active engagement of the individuals and groups potentially affected by AI applications, should also be encouraged.[77]

3.1.9. Algorithm Vigilance

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

---

72 Synthetic data can make a contribution to this end; see also The Norwegian Data Protection Authority, 'Artificial Intelligence and Privacy Report', 2018, available at https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf, last accessed 02.03.2020.

73 See also GBD, paras. IV.4.2 and IV.4.3.

74 See GAI, para. II.4; OECD; UNESCO 2019.

75 Committees of experts play an important role in areas where transparency and stakeholder engagement are made more difficult by competing interests and rights, such as in predictive justice, crime prevention and detection.

76 See GAI, para. II.6, ICDPPC. See also UNESCO, 'Declaration on the Human Genome and Human Rights', 11 November 1997, Article11. See also CM/Rec(2020)1, Appendix, para. B.5.3.

77 See GAI, para. II.7.

should both adopt forms of algorithm vigilance analogous to pharmacovigilance to react quickly in the event of unexpected and hazardous outcomes.[78]

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, the rule of law and democracy.[79] Cooperation should be encouraged in this regard between different supervisory authorities having competence for AI.[80]

3.2. HEALTH PROTECTION

Compared with data protection, the legal instruments in health protection provide a more limited and sector-specific contribution to the road-mapping 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; and (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.[81] Among them, the International Covenant on Economic, Social

---

78 See also Commission Nationale de l'Informatique et des Libertes – LINC, 'La Plateforme d'une Ville Les Données Personnelles Au Cœur de La Fabrique de La Smart City', available at https://www.cnil.fr/sites/default/files/atoms/files/cnil_cahiers_ip5.pdf, last accessed 02.03.2020; The Public Voice, 'Universal Guidelines for Artificial Intelligence', 2018, available at https://thepublicvoice.org/AI-universal-guidelines/, last accessed 02.03.2020.

79 See GAI, para. II.10; OECD; ICDPPC.

80 See ICDPPC; GAI, para. III.6.

81 See e.g. Office of the High Commissioner for Human Rights, '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)', p. 21; A.E. Yamin, 'The Right to Health Under International Law and Its Relevance to the United States', (2005) 95, *American Journal of Public Health,* p. 1156. At 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 C.A. Azencott, 'Machine Learning and Genomics: Precision Medicine versus Patient Privacy', (2018) 376, *Philosophical Transactions Royal Society A,* 20170350; K. Ferryman and M. Pitcan, 'Fairness in Precision Medicine', *Data & Society*, 2018, available at https://datasociety.net/wp-content/uploads/2018/02/Data.Society.Fairness.In_.Precision.Medicine.Feb2018.FINAL-2.26.18.pdf, last accessed 02.03.2020.

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.[82] 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[83] (Council of Europe).

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,[84] which require further elaboration to be applied to AI regulation. The Convention is complemented by two non-binding instruments: the Recommendation on health data[85] and the Recommendation on research on biological materials of human origin.[86] 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, the obscurity of AI applications can often undermine the effective exercise of individual autonomy in both cases.[87]

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

---

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

83 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.

84 See R. Andorno, 'The Oviedo Convention: A European Legal Framework at the Intersection of Human Rights and Health Law', (2005) 2(1), *Journal of International Biotechnology Law*, pp. 133–143; F. Seatzu, 'The Experience of the European Court of Human Rights with the European Convention on Human Rights and Biomedicine', (2015) 31(81), *Utrecht Journal of International and European Law,* p. 5.

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

86 See Recommendation CM/Rec(2016)6 of the Committee of Ministers to member States on research on biological materials of human origin.

87 See above Sections 3 and 3.1.

These key principles regard the following nine areas: primacy of the human being; equitable access; acceptability; the principle of beneficence; private life and the 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.[88]

3.2.1. Primacy of the Human Being

In a geopolitical and economic context characterised by competitive AI development, the primacy of the human being must be affirmed as a key element in the human rights-oriented approach:[89] 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).

3.2.2. Equitable Access to Healthcare

The principle of equitable access to healthcare[90] should be extended to the benefits of AI,[91] 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 and cultural minorities and limited access to information.

3.2.3. 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'.[92]

---

88 Human dignity and informed consent are not included in the table, as the former 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 Section 3.1) and informed consent is a principle that is also relevant in the context of data processing.

89 See also Oviedo Convention, Art. 2, and GAI.

90 See Oviedo Convention, Art. 3.

91 See also UNESCO. Universal Declaration on Bioethics and Human Rights, Art. 2.f.

92 See Office of the High Commissioner for Human Rights, 'CESCR General Comment No. 14', *supra* note 81. See also UNESCO, Universal Declaration on Bioethics and Human Rights, Art. 12; GBD, paras. IV.1 and IV.2.

Given the potentially high impact of AI-based solutions on society and groups,[93] 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.[94]

3.2.4. Principle of Beneficence

Respect for the principle of beneficence in biomedicine and bioethics and human rights[95] should be seen as a requirement where, as mentioned above, the complexity or opacity of AI-based treatments places limitations on individual consent which 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.[96]

3.2.5. Private Life and Right to Information

In line with the remarks made 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[97] 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 pose a serious risk to the health of others.[98]

---

93 See T. Linnet, L. Floridi and B. van der Sloot (eds.), *Group Privacy, supra* note 51.

94 See GAI paras. I.4 and II.6; CM/Rec(2020)1.

95 See UNESCO, Universal Declaration on Bioethics and Human Rights, Art. 4. See also Oviedo Convention, Art. 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'), 16, 17.

96 See also T.L. Beauchamp, 'Promise of the Beneficence Model for Medical Ethics' (1990) 6, *J. Contemp. Health L. & Pol'y*, pp. 145, 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'); E.D. Pellegrino and D.C. Thomasma, 'The Conflict between Autonomy and Beneficence in Medical Ethics: Proposal for a Resolution' (1987) 3, *The Journal of Contemporary Health Law and Policy*, pp. 23, 42 ('[in the beneficent model] No ethical stance, other than acting for the patient's best interests, is applied beforehand').

97 See Oviedo Convention, Art. 10. See also UNESCO, Universal Declaration on Bioethics and Human Rights, Art. 10.

98 See also CM/Rec(2019)2, 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, Art. 5.c.

3.2.6. Professional Standards

Professional standards are a key factor in biomedicine,[99] 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.

3.2.7. Non-Discrimination

The principle of non-discrimination[100] and non-stigmatisation in the field of biomedicine and bioethics[101] should be complemented by ruling out any form of discrimination against a person or group based on predictions of future health conditions.[102]

3.2.8. Role of Experts

The expertise of ethics committees in the field of biomedicine[103] should be called upon to provide independent, multidisciplinary and pluralist committees of experts in the assessment of AI applications.[104]

3.2.9. Public Debate

As with biomedicine,[105] 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 should be subject to consultation.

3.2.10. Unresolved and Partially Addressed Issues

Examination of these nine key areas (Sections 3.2.1 to 3.2.9) demonstrates that the current legal framework on biomedicine can provide important principles

99 See Oviedo Convention, Art. 4. See also CM/Rec(2019)2.
100 See Oviedo Convention, Art. 11.
101 See UNESCO, Universal Declaration on Bioethics and Human Rights, Art. 11.
102 See also, CM/Rec(2016)6, Art. 5.
103 See Oviedo Convention, Art. 16. See also UNESCO, Universal Declaration on Bioethics and Human Rights, Art. 19.
104 See also GBD; A. Mantelero (2018), 'AI and Big Data: A Blueprint', *supra* note 8.
105 See Oviedo Convention, Art. 28.

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. These shortcomings are discussed below.

3.2.10.1. Decision-Making Systems

In recent years, a growing number of AI applications have been developed for medical diagnosis, using data analytics and machine learning (ML) solutions. Large-scale data pools and predictive analytics are used to try to 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 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, decontextualisation, context-based data labelling and the re-use of data[106] are common to many AI applications and concern data in general. In coherence with the methodology adopted, existing guidance in the field of data protection[107] can be applied here, too, and the data quality aspects extended to non-personal data.

3.2.10.2. 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[108] and medical treatment. New schemes could be adopted, such as

---

106 K. Ferryman and M. Pitcan, 'Fairness in Precision Medicine', *supra* note 81, 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').

107 See GBD. See also the related preliminary studies: A. Mantelero, 'Artificial Intelligence and Data Protection: Challenges and Possible Remedies. Report on Artificial Intelligence', *supra* note 48; A. Rouvroy, '"Of Data and Men" – Fundamental rights and freedoms in a world of Big Data', T-PD-BUR(2015)09Rev, Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of personal data, Strasbourg, available at http://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020, last accessed 02.03.2020.

108 See also CM/Rec(2019)2.

broad[109] or dynamic consent,[110] which however – at the present state of the art – would only partially address this problem.

3.2.10.3. The Doctor-Patient Relationship

There are several factors in AI-based diagnosis – such as the loss of knowledge that cannot be encoded in data,[111] over-reliance on AI in medical decisions, the effects of local practices on training datasets and potential deskilling in the healthcare sector[112] – that might affect the doctor-patient relationship[113] and need to be evaluated carefully before adoption.

3.2.10.4. Risk Management

The medical device industry represents an interesting case study in terms of risk management, considering the significant consequences that these devices can have for individuals. The European Union has already adopted a risk-based classification (Directive 93/42/EEC) based on progressive safeguards according to the class of risk of each device (from conformity assessments under the sole responsibility of the manufacturer or the intervention of a notified body, to

---

109 See M. Sheehan, 'Can Broad Consent be Informed Consent?', (2011) (3), *Public Health Ethics*, pp. 226–235. 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 Recommendation CM/Rec(2019)2 of the Committee of Ministers of the Council of Europe to member States on the protection of health-related data, 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').

110 See J. Kaye, E.A. Whitley, D. Lund et al., 'Dynamic Consent: A Patient Interface for Twenty-first Century Research Networks', (2015) 23(2), *European Journal of Human Genetics,* p. 141.

111 See R. Caruana, P. Koch, Y. Lou et al., '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, available at http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf, last accessed 02.03.2020, pp. 1721–1730.

112 F. Cabitza, R. Rasoini and G.F. Gensini, 'Unintended Consequences of Machine Learning in Medicine', (2017) 318, *JAMA*, p. 517.

113 See also, UNESCO. Universal Declaration on Bioethics and Human Rights, Art. 20; WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects, 9th July 2018, accessible at https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/, last accessed 02.03.2020.

inspection by a notified body and, in cases of highest risk, a requirement for prior authorisation before being allowed on the market).

A model based on such progressive safeguards could be generalised for future AI regulation and also adopted outside the field of medical devices, focusing on the impact on human rights and fundamental freedoms. However, the classification of AI products/services is more complicated given their variety and different fields of application. Several sector-specific classifications would be needed, or general criteria based on risk assessments.

Finally, specific provisions on AI vigilance and the adoption of the precautionary principle in AI development, as discussed above, could help to address these challenges.

4. CONCLUSIONS

In view of the ongoing debate on AI regulation and the limited scope of this contribution, which does not cover all the different areas of AI, only certain provisional conclusions can be drawn at this stage. This is in line with the goal of this research which aimed to outline a methodological approach to the question and not to set out a comprehensive series of regulatory principles and provisions.

Through an initial analysis of the binding and non-binding legal instruments in the fields of healthcare and data protection, several principles have been identified as the basis for future AI regulation. This framework and the instruments themselves[114] reaffirm the central role of human dignity and human rights in the application of AI, where machine-driven solutions may dehumanise individuals.

This also suggests introducing bans on specific AI technologies[115] that are developed in such a way that is inconsistent with human dignity,[116] human rights, democracy and the rule of law.

The methodological process – consisting of analysis (mapping and identification of key principles) and contextualisation – has proven useful in the areas examined, with the elaboration of several key principles and procedural approaches. Numerous correlations and a common ground between these

---

114 See also UNESCO, Declaration on the Human Genome and Human Rights (11 November 1997), Art. 2.

115 See e.g. Data Ethics Commission of the Federal Government – Federal Ministry of the Interior, 'Opinion of the Data Ethics Commission' (December 2019), available at https://datenethikkommission.de/en/, last accessed 06.04.2020; Access Now, 'The European Human Rights Agenda in the Digital Age' (November 2019), available at https://www.accessnow.org/access-now_the-european-human-rights-agenda-in-the-digital-age_final1/, last accessed 01.06.2020.

116 See also UNESCO, Declaration on the Human Genome and Human Rights (11 November 1997), Art. 11.

principles and approaches have been identified facilitating their harmonisation, while other principles represent the unique contributions of each sector to future AI regulation. The table below summarises these findings and the level of harmonisation in these two realms.

Table 1. Common ground

Data	**Health**
Primacy of the 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: Compiled by the author.

Notwithstanding the limitations of the scope of this analysis, the results appear to validate the methodology proposed for a road-mapping of AI regulation based on a four-step process, consisting of mapping legal instruments, identification of key principles, contextualisation and harmonisation. At the same time, this contribution highlights the significant effort required to systematise the provisions of a wide variety of instruments, with differing binding forces, focuses, approaches and structures.