COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 15 June 2020 CAHAI(2020)08-fin

AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI)

Elaboration of the feasibility study

Analysis of the International legally binding instruments

Final report

Paper prepared by Alessandro Mantelero*

* Associate professor of Private Law and Data Ethics & Data Protection, Polytechnic University of Turin (Politecnico di Torino). The opinions expressed in this analysis do not necessarily reflect the position of the CAHAI or the Council of Europe.

Table of Contents

EXECUTIVE SUMMARY

The latest wave of Artificial Intelligence (AI) development is having a growing transformative impact on society and raises new questions in different fields, from predictive medicine to media content moderation, from the quantified self to judicial systems, without overlooking the issues of environmental impact.

An analysis of the international legally binding instruments is thus the obligatory starting point to define the existing legal framework, identify its guiding values and verify whether this framework and its principles properly address all the issues raised by AI.

With a view to preserving the harmonisation of the existing legal framework in the field of human rights, democracy and the rule of law, this study aims to contribute to the drafting of future AI regulation by building on the existing binding instruments, contextualising their principles and providing key regulatory guidelines for **a future legal framework**.

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. Moreover, only the human rights framework can provide **a universal reference for AI regulation**, while other realms (e.g. ethics) do not have the same global dimension, are more context-dependent and characterised by a variety of theoretical approaches.

The analysis of the existing binding legal instruments contained in this document is not limited to a harmonising study, extracting common values and principles from a given set of rules on AI. A more articulated investigation is carried out in different stages.

After an initial sector-specific analysis to map and identify key guiding principles in four core areas (data protection, health, democracy and justice), these principles are contextualised in the light of the changes to society produced by AI. In so doing, we benefit from the existing non-binding instruments that provide more granular applications of the principles enshrined in international legal instruments, in some cases also providing specific guidance on AI.

This **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. This makes it possible to formulate **an initial set of provisions for future AI regulation** focusing on the most challenging issues in each sector examined.

Considering the large number of documents adopted by several international and intergovernmental bodies and given the parallel ongoing study on ethical instruments carried out by CAHAI, this document focuses on the legally binding instruments, plus the non-binding instruments adopted to implement them.

The study is divided into two parts. The first one identifies the scope and methodology of this analysis, while the second presents the results of the sectoral analysis on guiding principles.

In the sector-specific analysis, the first two key areas examined are **health** and **data protection**. The intersection between these two realms is interesting in view of this study's focus, 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 the Oviedo Convention and Convention 108+, as well as by the non-binding instruments. Moreover, individual self-determination plays a central role in both the field of data protection and biomedicine, and the challenges of AI – in terms of the complexity and opacity of treatments and processing operations – are therefore particularly relevant and share common concerns.

The fourth and the fifth sections are centred on **democracy** and **justice**. Here the field of investigation is wider and there are no comprehensive legal instruments that can provide specific sectoral principles, such as Convention 108+ or the Oviedo Convention. The analysis is therefore more closely focused on high-level principles and their contextualisation with a more limited elaboration of key guiding provisions compared with the previous sections.

The last section provides an overview of the guiding principles identified and suggests a harmonisation framework pointing out the existing correlations and common ground between these principles and, at the same time, highlighting the unique contributions of each sector to future AI regulation.

The main objective of this study is not to add a new list of guiding principles to those already provided by a variety of bodies and entities, but to achieve a different result in methodological and substantive terms.

First, **the analysis carried out and the solution proposed have their roots and build on human rights and freedoms**, adopting a concrete approach centred on existing international legal instruments. Other studies are often sector-specific and have a different set of normative references (national or regional) or adopt a theoretical approach enunciating principles or referring to human rights in a general and abstract manner. Although these works do enhance the legal and ethical debate on AI, their impact in terms of contribution to the regulatory framework is limited and not specifically contextualised in the framework of the Council of Europe's standards on human rights, democracy and the rule of law.

Second, the result of this analysis of the legally binding instruments, including the non-binding instruments adopted to implement them, is not merely a list of principles however accurate that may be. **Identifying common guiding principles is important but not sufficient to provide a roadmap for future AI regulation**. Transparency, accountability, human oversight and many other principles already listed in several charters on AI are abstract concepts without a proper contextualisation.

The main contribution of this study is to furnish precisely this **contextualisation with regard to the legal framework and to AI challenges**. If this document succeeds in suggesting **concrete and effective ways to formulate and codify these guiding principles with regard to AI** and concretely **embed the Council of Europe's standards on human rights, democracy and the rule of law in the outline of the future AI regulation**, it will have achieved its goal in helping to frame the relationship between humans and AI from a legal standpoint.

PART I – SCOPE AND METHODOLOGY

Just as with the Internet, electricity and steam power, Artificial Intelligence (AI) comprise a range of different technologies having a broad impact on a variety of human activities and society.

In this context, many different legal instruments can assume importance in regulating AI applications. At the same time, these legal instruments were adopted in a pre-AI era and this might reduce their effectiveness in providing an adequate and specific response to the new challenges of AI.

An analysis of the international legally binding instruments is thus the obligatory starting point to define the existing legal framework, identify its guiding values and verify whether this framework and its principles properly address all the issues raised by AI, with the view to preserving the harmonisation of the existing legal framework in the field of human rights, democracy and the rule of law.

This approach does not set out to create a completely new and comprehensive reference framework, as the regulation should focus on what changes AI will bring to society, not on reshaping all areas where AI can be applied.[1] This targeted approach is made possible by building on the existing binding instruments, contextualising their guiding principles and providing key regulatory guidelines for a future legal framework for AI, which can cover areas that are not presently regulated by the existing binding instruments.

In this regard, it is important to highlight the difference between the existing legally binding instruments and other documents, such as soft law instruments or ethical charters on AI. Legally binding instruments pre-existed the current AI spring. They were not drafted with AI in mind and do not provide a specific set of rules for this field, while soft law and ethics documents on AI do provide a specific focus, albeit from different perspectives.

Analysis of the existing binding legal instruments is not therefore limited to a harmonising study (i.e. extracting common values and principles from a given set of rules on AI), but requires a more articulated process, in which harmonisation is just one of several stages. The process can be divided into three separate stages: (i) mapping and identification of key principles, (ii) contextualisation, and (iii) harmonisation.

I.1 The scenario

The latest wave of AI development is having a growing transformative impact on society and rises new question in different fields, from predictive medicine to media content moderation, from the quantified self to judicial systems, without overlooking the issues of environmental impact.

The rapid evolution of applied AI over the last few years has been incompatible with a specific legal response in terms of international legally binding instruments focused on AI. This is why we have seen the development of two different operating strategies to address these issues: (i) a significant effort in interpreting the existing legal framework in the light of AI related issues (see for example the ongoing debate on the GDPR provisions on transparency and automated decision-making); (ii) the use of

[1] See, for example, the EU approach to interstitial regulation of e-commerce.

non-binding rules to contextualise the principles provided by the existing binding instruments (e.g. T-PD(2019)01 Guidelines on Artificial Intelligence and Data Protection; CEPEJ. 2019. European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment[2]).

Future regulation of AI should therefore build on these efforts, focusing both on the guiding principles and values deriving from the existing binding instruments and on their related non-binding implementations, which in some cases already contemplate the new AI scenario.

I.2 Research focus and methodology

The main aim of this study is to define the key principles for the future regulation of AI through an analysis of the existing legal framework. The methodology is therefore necessarily deductive, extracting these principles from the variety of regulations concerning the fields where AI solutions can potentially be adopted.

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.[3] Moreover, only the human rights framework can provide a universal reference for AI regulation, while other realms (e.g. ethics) do not have the same global dimension, are more context-dependent and characterised by a variety of theoretical approaches.

Against this background, many questions arise, such as: when should an AI system make a decision? Which criteria should the system apply? Who is accountable for decisions that may negatively affect individuals and society? Around these and many other emerging questions, the existing regulations need to be reconsidered.

To provide a harmonised regulatory framework to address the challenges of AI, common and high-level guidance on the principles and values to be enshrined should be derived from international charters of human rights (e.g. Universal Declaration of Human Rights, International Covenant on Civil and Political Rights, Convention for the Protection of Human Rights and Fundamental Freedoms, Charter of Fundamental Rights of the European Union).

The guiding principles must be considered within the AI-driven transformative scenario, which in many cases will require their adaptation. These principles remain valid, but their operation should be reconsidered in the light of the social and technical changes induced by AI (e.g. freedom of choice in the event of so-called black boxes). This will deliver a more contextualised and granular application of the principles so that they can provide a concrete contribution to the shape of future AI regulation.

To conduct this study, we need to start by defining the main areas of investigation, considering both the potential impacts of AI and the fields of action of the Council of Europe. In this regard four key areas have been selected: data, health, democracy and justice.

[2] European Commission for the Efficiency of Justice (CEPEJ). 2018. European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment.

[3] See also Committee of Ministers. 2020. Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems.

I.3 Analysis and expected results

The study takes a top-down approach with a view to contributing to the future AI regulatory framework, to be implemented by additional binding and non-binding instruments, rather as happened in the field of biomedicine. The expected result is a set of provisions concerning the investigated areas and key common guiding principles, based on a comprehensive analysis of the entire corpus of the binding instruments, including the non-binding tools adopted.

First stage: Mapping and identification of key principles. Guiding principles will be identified in the different investigated areas. The first stage of the analysis is based on the different subjects, as binding instruments are sector-specific and not rights-based. The following two tables provide a first example of this mapping exercise based on a preliminary overview of the data protection and justice realms to identify the guiding principles for future regulation of AI.

Figure 1: Data protection

Binding instruments	Convention 108+ Convention on Cybercrime
Impacted areas	Decision-making systems Group privacy and collective dimension Profiling
Related non-binding instruments	CoE. 2019. Guidelines on the data protection implications of artificial intelligence CoE. 2017. Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data CoE. 2010. Recommendation on the protection of individuals with regard to automatic processing of personal data in the context of profiling [under revision] UNESCO. 2019. Preliminary Study on a Possible Standard-Setting Instrument on the Ethics of Artificial Intelligence OECD. 2019. Recommendation of the Council on Artificial Intelligence 40th International Conference of Data Protection and Privacy Commissioners. 2018
Guiding principles and legal values	Accountability Risk-based approach Precautionary principle Data quality & security Transparency Fairness

	Contextual approach Role of experts Participation/Inclusiveness Freedom of choice/Autonomy Human control/oversight Awareness Literacy Responsible innovation Cooperation between supervisory authorities

Figure 2: Justice

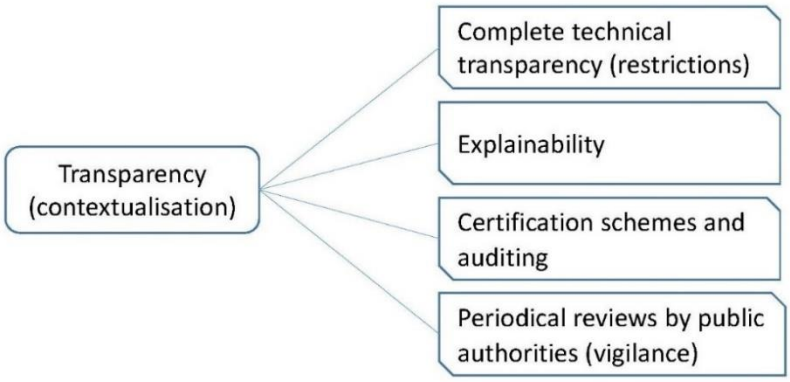
Binding instruments	Universal Declaration of Human Rights International Covenant on Civil and Political Rights International Convention on the Elimination of All Forms of Racial Discrimination Convention on the Elimination of All Forms of Discrimination against Women Convention for the Protection of Human Rights and Fundamental Freedoms Charter of Fundamental Rights of the European Union
Impacted areas	Processing of judicial decisions and data Predictive policing
Related non-binding instruments	CEPEJ. 2019. European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment
Guiding principles and legal values	Non-discrimination Data quality & security Transparency Impartiality Fairness Contextual approach Freedom of choice/ Independence of judges (decision-making process) Human control/oversight Guarantees of the right to a fair trial

Second stage: Contextualisation. The guiding values identified in the mapping exercise should be contextualised in the light of the changes to society produced by

AI. This phase will benefit from the existing non-binding instruments that provide more granular applications of the principles enshrined in the binding instruments, in some case also providing specific guidance on AI.

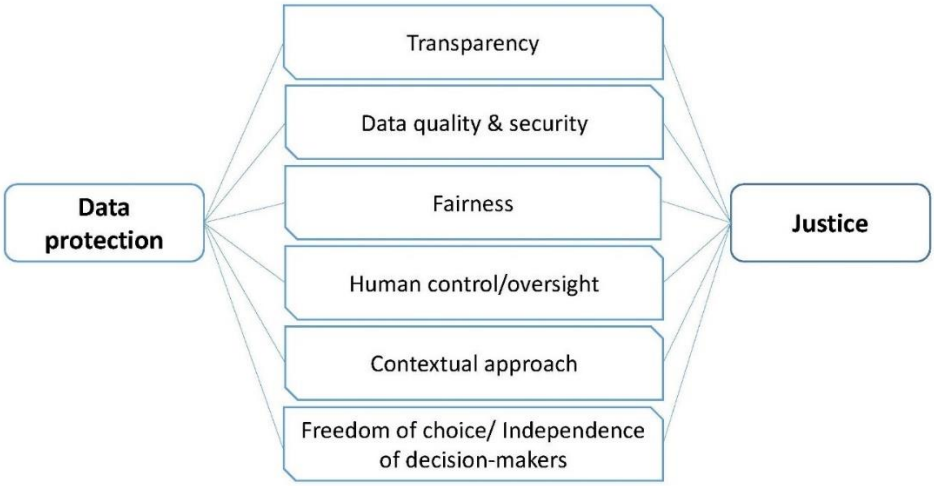
This contextualisation of the guiding principles and legal values will provide a more refined and elaborate formulation of them, considering the specific nature of AI products and services. At this stage, it will therefore be possible to formulate an initial set of provisions for future AI regulation focusing on the most challenging issues in each sector.

Figure 3: Context-specific implementation of the transparency principle



Third stage: Harmonisation (cross-sectoral analysis). Based on the sector-specific analysis carried out in this study, a list of key guiding principles common to the different realms will be drawn up in the last section (Figure 4). These shared principles will then be the cornerstone of the common core of the future provisions on AI.

Figure 4: Common guiding values in the field of data protection and justice



PART II – ANALYSIS

Considering the large number of documents adopted by several international and intergovernmental bodies and given the parallel ongoing study on ethical instruments carried out by CAHAI, this part focuses on the legally binding instruments, including the non-binding instruments adopted to implement them. Ethical guidelines are therefore not considered at this stage, and documents concerning future regulatory strategies (e.g. white papers) are only taken into account as background information.

This Part is divided into six sections followed by some concluding considerations. The first section presents a general overview of the existing instruments adopted by the Council of Europe and the main underlying principles/values. This helps to define the potential core principles of future AI regulation and its coherence with the existing framework.

The second and third sections focus on two key and related areas: health and data protection. The intersection between these two realms is interesting in view of this study's focus, 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 the Oviedo Convention and Convention 108+, as well as by the non-binding instruments.[4] Moreover, individual self-determination plays a central role in both the field of data protection and biomedicine, and the challenges of AI – in terms of the complexity and opacity of treatments and processing operations – are therefore particularly relevant and share common concerns.

The fourth and the fifth sections are centred on democracy and justice. Here the field of investigation is broader and there are no general legal instruments that can provide sector-specific principles, such as Convention 108+ or the Oviedo Convention. The analysis is therefore focused on high-level principles and their contextualisation, resulting in a more limited elaboration of key guiding provision than in previous sections.

Section 6 provides a general overview of the guiding principles identified and suggests a harmonisation framework that highlights both the existing correlations between these principles and the unique contribution of each sector to future AI regulation.

As highlighted by comments received during the monitoring exercise described in the next section, AI technologies impact on a variety of sectors and raise issues concerning a large body of regulatory instruments.[5] This initial study is therefore a starting point focused on the four core areas mentioned. However, despite its limited scope, the results validate the methodology proposed and provide a number of pointers towards future provisions in AI regulation.

[4] See Recommendation CM/Rec(2019)2 on the protection of health-related data.

[5] See Annex 1.

II.1 General overview

As AI impacts on a variety of situations[6] dealt with by different binding instruments covering several areas, we need to conduct an evidence-based analysis to identify key principles and common values to be considered for future regulation.

An initial monitoring exercise was carried out in this light between 12 and 28 February 2020, involving the different branches of the Council of Europe to benefit from the sector-specific expertise of the various units that have operated over the years in a range of fields relating to human rights, democracy, and the rule of law.

Using a survey based on open-ended questions, the different units interviewed were asked to provide information on the following areas: (i) binding instruments, (ii) impacted areas (applications), (iii) related non-binding instruments, (iv) guiding principles and legal values, and (v) missing principles/issues. Thanks to the positive commitment of the different areas, it was possible to collect a variety of different types of information.

From a methodological point of view, the structure of this preliminary survey based on open-ended questions necessarily affects the results of the quantitative analysis. The main limitations regard the use of different and partially overlapping general categories, as well as differing levels of granularity and specificity of the answers.

Nevertheless, by aggregation in macro-areas and focusing on similarities (i.e. frequency) in the principles and values identified, we were able to achieve some perspective in the results, and the exercise provided a more detailed map of the available non-binding instruments adopted by the Council of Europe that can help to establish a legal framework for future regulation (see Annex 1).

With regard to the impacted areas (see Annex 2), the exercise suggests focusing future AI regulation along two main axes: the use of AI and the development of AI. In both cases, different human rights and fundamental freedoms are potentially affected or can play an important role in shaping future AI scenarios.[7]

Regarding the use of AI, there are four main areas of application and consequent regulation: predictive analysis and decision support systems, automated decision-making systems, evidence collection/computer forensics, and content generation.

The first two areas are well known and debated, as they cover an extremely wide range of applications (see Annex 2). Nevertheless, the distinction between decision support and autonomous decision-making systems is crucial in terms of value oriented-design and the role of human beings in the decision-making process: the differing nature of these two types of systems will necessarily require different procedural and substantive safeguards in AI regulation.

The last two areas are sector-specific but should be considered separately since they do not concern the decision-making process directly but do provide the evidence that underpins it (evidence collection and computer forensics) or affect the creation processes (content generation). In these cases, the main issues seem to be different and more focused on the procedural aspects and their coherence with traditional (i.e. non-AI-based) approaches.

Although most of the existing literature and guidelines focus on AI systems and their potential consequences, an important impact of AI on human rights and fundamental freedoms is also related to the development of AI and the provision of AI services. In this respect, future AI regulation should carefully consider the issues relating to

[6] See also UNESCO, 2019.

[7] See Council of Europe-Committee of experts on internet intermediaries (MSI-NET), 2018.

working conditions of the people involved in the whole AI product and service supply chain.[8]

The second block of information provided by the monitoring exercise concerns the guiding principles and legal values that should underpin the future development and use of AI (see Annex 3). Here, the diversity of notions employed by the units surveyed suggests an aggregation of principles and values. The result of this analysis made it possible to group the guiding principles and values around a number of key elements which emerged in terms of distribution (frequency):

Non-discrimination (15)
Diversity, inclusion and pluralism (13)
Privacy and Data Protection (11)

Transparency (9)
Equality (8)
Access to justice, fair trial (7)
Human control (7)

Impartiality (6)
Access to information (5)
Security (5)
Fairness (5)
Participation (5)
Freedom of choice (5)
Freedom of expression and of creation (5)

Accountability (3)
Competence and capacity (2)
Independence (3)
Individual autonomy (3)
Cultural cooperation (2)
Sustainability and Long-term Orientation (2)

Despite the limitations of the analysis mentioned, it is clear that the first three principles are seen as key elements in the future regulation of AI and will therefore be its main focus. This is further confirmed by the second set of principles/values, which is closely related to the first: transparency and human control are important factors in non-discrimination and data protection, while access to justice is a general condition for addressing any potential infringement of these values. Similarly, though more substantively, equality is linked in various ways to the first three main values/principles. The other values/principles, addressing various specific concerns of AI implementation, differ more widely.

This exercise made it possible to identify a first list of guiding principles of AI regulation, already codified in binding and non-binding legal instruments, but in need of contextualisation in the field of AI. In the sector-specific analysis this contextualisation, based on an in-depth analysis of international legally binding instruments, will be achieved by assessing any potential gaps in the existing regulatory framework, sector-by-sector.

As AI is a cross-sector technology, it is expected that the results of this analysis may suggest similar regulatory interventions in other areas, as outlined in the part on methodology.[9] Once the sector-specific analysis is completed, all these potential

[8] See also Crawford and Joler, 2018.

[9] See above Part I.

interventions will be systematised to avoid overlaps and aggregating them into a coherent framework based on key values.

II.2 Data Protection

In the past decade, the international regulatory framework in the field of data protection has seen significant renewal. Legal instruments shaped on the basis of principles defined in the 1970s and 1980s[10] 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 (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 been, or is being, carried out in other international contexts – such as the OECD[11] – or by individual countries.

However, given the rapid development of the last wave of AI development, these new binding instruments 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.[12]

For the purposes of this study, the following non-binding legal instruments were therefore analysed:[13] T-PD(2019)01, Guidelines on Artificial Intelligence and Data Protection [GAI]; T-PD(2017)1, Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data; 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;[14] 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; UNESCO. 2019. Preliminary Study on a Possible Standard-Setting Instrument on the Ethics of Artificial Intelligence [UNESCO];[15] OECD. 2019. Recommendation of the Council on Artificial Intelligence [OECD]; 40th International Conference of Data Protection and Privacy Commissioners. 2018 [ICDPPC]. Declaration on Ethics and Data Protection in Artificial Intelligence.

[10] See also Mayer-Schönberger, 1997; González Fuster, 2014.

[11] See 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.

[12] See European Commission. 2020. Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM(2020) 64 final; European Commission. 2020. White Paper on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final; European Commission. 2020. A European strategy for data, COM(2020) 66 final.

[13] See Annex 4.

[14] This Recommendation has replaced Recommendation No. R(97)5 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.

[15] Despite the reference to ethics 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”.

These instruments differ in nature: while those adopted by the Council of Europe define different specific requirements and provisions, the others are mainly principles-based, setting out several principles but without, or only partially, providing more detailed guidance in terms of specific requirements. The following paragraphs illustrate the key principles derived from these different instruments and how they can be contextualised within the AI scenario.

Several of these principles classed in the field of personal data protection (e.g. data quality), can be extended to non-personal data, mainly in regard to the impact of the use of non-personal data (e.g. aggregated data) on individual and groups in the context of decision-making processes (e.g. mobility data or energy consumption data).

i) **Primacy of the human being**

AI systems shall be designed to serve mankind and any creation, development and use of AI systems shall fully respect human rights, democracy and the rule of law.[16]

ii) **Human control**

AI applications should allow meaningful control by human beings over their effects on individuals and society.[17]

iii) **Transparency and expandability**

Every individual shall have a right to be informed appropriately when she or he is interacting directly with an AI system, providing adequate and easy-to-understand information on the purpose and effects of this system, including the existence of automated decisions, in order to verify continuous alignment with the expectation of individuals, to enable overall human control on such systems and to enable those adversely affected by an AI system to challenge its outcome.[18]

Every individual shall also have a right to obtain, on request, knowledge of the reasoning underlying an AI-based decision process where the results of such process are applied to him or her.[19] Moreover, States shall promote scientific research on explainable artificial intelligence and best practices for transparency and auditability of AI systems.[20]

iv) **Precautionary approach**

When the potential risks of AI applications are unknown or uncertain, AI development shall be based on the precautionary principle.[21]

v) **Risk management**

AI developers, manufacturers and service providers should assess and document the possible adverse consequences of AI applications on human rights and fundamental freedoms, and adopt appropriate risk prevention and mitigation measures from the

[16] See CM/Rec(2019)2; ICDPPC; GAI, paras. I.1 and II.1; UNESCO. See also GDPR, Recital no. 4.

[17] See GAI, para. I.6.

[18] See ICDPPC, CM/Rec(2019)2, OECD, UNESCO. See also Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems.

[19] See Convention 108+; GAI, para. II.11.

[20] See ICDPPC.

[21] See GAI, para. II.2.

design phase (human rights by-design approach) and during their entire lifecycle.[22] Adverse consequences include those due to the use of de-contextualised data and de-contextualised algorithmic models.[23]

AI developers, manufacturers, and service providers should consult competent supervisory authorities when AI applications have the potential to significantly impact the human rights and fundamental freedoms of individuals.[24]

vi) **Risk of re-identification**

Suitable measures should be introduced to guard against any possibility that anonymous and aggregated data may result in the re-identification of the data subjects.[25]

vii) **Data quality and minimisation**

AI developers should critically assess the quality, nature, origin and amount of personal data used, reducing unnecessary, redundant or marginal data during AI development and training phases, and monitoring the model's accuracy as it is fed with new data. The use of synthetic data may be considered as one possible solution to minimise the amount of personal data processed by AI applications.[26]

viii) **Role of experts**

AI developers, manufacturers and service providers are encouraged to set up and consult independent committees of experts from a range of fields, as well as engage with independent academic institutions, which can contribute to designing human rights-based, ethically and socially-oriented AI applications, and to detecting potential bias. Such committees may play an especially important role in areas where transparency and stakeholder engagement can be more difficult due to competing interests and rights, such as in the fields of predictive justice, crime prevention and detection.[27]

Appropriate mechanisms should be put in place to ensure the independence of these committees of experts.[28]

ix) **Participation and democratic oversight on AI development**

Participatory forms of risk assessment, based on the active engagement of the individuals and groups potentially affected by AI applications, shall be developed. Individuals, groups, and other stakeholders should be informed and actively involved in the debate on what role AI should play in shaping social dynamics, and in decision-making processes affecting them.[29]

---

[22] See GAI, paras II.2 and II.3; ICDPPC; OECD; UNESCO. See also Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems.

[23] See GAI, para, II.5.

[24] See GAI, para. III.5.

[25] See CM/Rec(2010)13.

[26] See GAI para. II.4; OECD. See also CM/Rec(2020)1.

[27] See also below Section II.3.

[28] See GAI, paras II.6 and II.7; ICDPPC. See also Article11, UNESCO. Declaration on the Human Genome and Human Rights (11 November 1997).

[29] See GAI, paras. II.7 and III.8; ICDPPC. See also CM/Rec(2020)1.

Derogations can be introduced for public interest, where proportionate in a democratic society and with adequate safeguards.

x) **Human oversight**

AI products and services shall be designed in a manner that ensures the right of individuals not to be subject to a decision significantly affecting them based solely on the automated processing of data, without having their views taken into consideration. AI products and services shall enable overall human control over them.[30]

In addition, 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.[31]

xi) **Algorithm vigilance**

AI developers, manufacturers, and service providers shall adopt forms of algorithm vigilance that promote the accountability of all relevant stakeholders by assessing and documenting the expected impacts on individuals and society in each phase of the AI system lifecycle on a continuous basis, to ensure compliance with human rights, the rule of law and democracy.[32] Governments should provide regular reports about their use of AI in policing, intelligence, and security.[33]

xii) **Freedom of choice**

In order to enhance users' trust, AI developers, manufacturers and service providers are encouraged to design their products and services in a manner that safeguards users' freedom of choice over the use of AI, by providing feasible alternatives to AI applications.[34]

xiii) **Right to object**

The right to object should be ensured in relation to AI systems based on technologies that influence the opinions and personal development of individuals.[35]

xiv) **Interoperability**

Interoperability between AI systems shall be implemented in full compliance with the principles of lawfulness, necessity and proportionality, putting in place appropriate safeguards for human rights, democracy and the rule of law.[36]

xv) **Cooperation**

Cooperation shall be encouraged between supervisory authorities with competence related to AI.[37]

xvi) **Digital literacy, education, and professional training**

[30] See Convention 108+; GAI para. II.8; ICDPPC; UNESCO.

[31] See GAI para. III. 4.

[32] See GAI para. II.10; OECD; ICDPPC. See also CM/Rec(2020)1.

[33] See UNESCO.

[34] See GAI, para. II.9.

[35] See GAI, para. II.12. See also below Section II.4.

[36] See CM/Rec(2019)2.

[37] See ICDPPC; GAI, para. III.6.

Policy makers should invest resources in digital literacy and education to increase data subjects' awareness and understanding of AI applications and their effects. They should also encourage professional training for AI developers to raise awareness and understanding of the potential effects of AI on individuals and society. They should support research in human rights-oriented AI.[38]

xvii) **Scientific research integrity**

Where a data subject withdraws from a scientific research project, the withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Personal data should be destroyed or anonymised in a manner which does not compromise the scientific validity of the research and the data subject should be informed accordingly.[39]

II.3 Health

The European regulatory framework for healthcare is characterised by a few Council of Europe binding instruments and a number of sector-specific instruments adopted at EU level, according to the different nature, scope and regulatory remit of these two entities.

The European Convention on Human Rights, as well as Convention 108+ and the European Social Charter, lay down several general provisions on health protection and related rights. However, these provisions and principles already set out in other general instruments at international level,[40] find a broader and more sector-specific contextualisation in the Oviedo Convention.

The Oviedo Convention – the only multilateral binding instrument entirely focused on biomedicine – and its additional protocols is therefore the main reference point to identify the key principles in this field,[41] which need further elaboration and, where necessary, amplification to regulate AI applications. Furthermore, the Convention is complemented by two non-binding instruments: the Recommendation on health data[42] and the Recommendation on research on biological materials of human origin.[43] The first of these two recommendations illustrates the close link between biomedicine (and healthcare more generally) and data processing, which will be discussed further below.

Most of the existing regulation on health focuses on medical treatment, research (including medical trials) and medical devices/products. AI has a potential impact on all these areas, given its application in precision medicine,[44] diagnosis, and medical devices and services.

---

[38] See ICDPPC; OECD; GAI, para. III.9; UNESCO. See also CM/Rec(2020)1.

[39] See Convention 108+; CM/Rec(2019)2.

[40] Some of the general principles enshrined in this convention have been affirmed in previous international human rights instruments, such as the International Covenant on Civil and Political Rights, the International Covenant on Economic, Social and Cultural Rights, and the Convention on the Rights of the Child of 20 November 1989.

[41] See Andorno, 2005; Seatzu, 2015.

[42] See Recommendation CM/Rec(2019)2 on the protection of health-related data.

[43] See Recommendation CM/Rec(2016)6 on research on biological materials of human origin.

[44] See Azencott, 2018; Ferryman and Pitcan, 2018.

Although the Oviedo Convention and 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.

Nevertheless, self-determination in the field of biomedicine faces the same challenges as already discussed in data processing. Notwithstanding the different nature of consent to medical treatments and consent to data processing, the high level of complexity and, often, a certain degree of obscurity of AI applications can undermine the effective exercise of individual autonomy in both cases.[45]

Against this background, the main contribution of the Oviedo Convention to future AI regulation does not concern the sector-specific safeguards it provides, but consists in the important set of general principles and values that can be extrapolated from it to form a building block of future AI regulation.

The Council of Europe's main contribution in the field of medicine concerns the following eight areas: human dignity, primacy of the human being, professional standards, general rule on informed consent, private life and the right to information, non-discrimination, protection of persons undergoing research, and public debate. The contribution of this Convention to the debate on the future regulation of AI goes beyond biomedicine since several provisions, centred on the right balance between technology and human rights, can be extended generally beyond the field of AI, as described in the following paragraphs.[46]

i) Primacy of the human being

In a geo-political and economic context characterised by competition in AI development, the primacy of the human being should generally be affirmed as a key element of the European approach: better performances of AI-based systems and their efficiency should not override the interests and welfare of human beings. The application of this principle should cover both the development (e.g. systems developed violating human rights and freedoms) and the use of AI systems.[47]

ii) Equitable access to health care

The equitable access principle can be extended to access to the benefits of AI. This entails the adoption of appropriate measures to tackle the risks concerning the digital divide, discrimination, marginalisation of vulnerable persons or cultural minorities, and limitations to the access to information.[48]

iii) Professional standards

AI development therefore embraces several areas of expertise and, where the development of AI systems can impact on individuals and society, it must be carried out in accordance with relevant professional obligations and standards of each area of expertise involved. The professional standards and skills required shall be based on the current state of the art.[49]

---

[45] See above Section II.2.

[46] 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 and informed consent is a principle that is also relevant in the context of data processing.

[47] See also Oviedo Convention, Article 2.

[48] See also Oviedo Convention, Article 3.

[49] See also Oviedo Convention, Article 4; Recommendation CM/Rec(2019)2 on the protection of health-related data.

States shall encourage professional training to raise awareness and understanding of AI and its potential effects on individuals and society. They should support research in human rights-oriented AI. States shall also cooperate in defining common educational programmes and common standards for professionals who deal with AI and society.

In using AI in the healthcare sector special attention shall be paid to the patient's confidence in his or her doctor and mutual trust, which shall not be compromised by the use of AI.

iv) **Protection of persons not able to consent and of persons not able to consent to research**

Respect for the principle of beneficence should be considered a requirement where, given the complexity or opacity of AI-based treatments, individual consent suffers from several limitations and cannot be the exclusive basis for treatment.[50]

v) **Private life and right to information**

According to Article 10 of the Oviedo Convention, AI health applications shall guarantee the right to information and respect the wishes of individuals not to be informed, except where compliance with an individual's wish not to be informed constitutes a serious risk for the health of others.[51]

vi) **Non-discrimination**

The principle of non-discrimination in the field of health should be complemented by forbidding any form of discrimination against a person or group based on predictions of future health conditions.[52]

vii) **Role of experts**

The experience of ethics committees in the field of biomedicine should be considered, introducing multidisciplinary committees of experts in the assessment of AI applications.[53]

viii) **Public debate**

Fundamental questions raised by the developments of AI shall be subject of appropriate public discussion in the light, in particular, of relevant social, economic, ethical and legal implications, and that their possible application is made the subject of appropriate consultation.[54]

***

These considerations show that the existing legal framework on biomedicine provides important principles and elements that can be extended to future AI regulation, even beyond the health sector. On the other hand, a series of shortcomings created by the impact of AI remain unresolved in the following areas.

a) Decision-making systems [Contextual approach, Fairness, Data quality, Human control/oversight]

In recent years a growing number of AI applications have been developed and used in the medical sector for diagnosis, using both analytics and ML solutions. Large-scale

[50] See also Oviedo Convention, Articles 6 and 17.

[51] See also Oviedo Convention, Article 10.

[52] See also Oviedo Convention, Article 11.

[53] See also Oviedo Convention, Article 16.

[54] See also Oviedo Convention, Article 28.

data pools are created, and predictive analytics is used to try and arrive at solutions for clinical cases based on existing knowledge and practices. Likewise, ML applications in image recognition look like they may provide increased cancer detection capability. In addition, in the field of the precision medicine, large-scale collection and analysis of multiple data sources (medical data but also non-medical data, such as air and housing quality) are used to develop individualised insights into health and disease.

The use of clinical data, medical knowledge 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 case opacity) of data processing, and the predictive nature of the results of analysis raise concerns about the potential weakness of decision-making systems.

Most of these issues are not limited to health sector, as potential biases (including lack of diversity and the exclusion of outliers and smaller populations), data quality, decontextualization, the context-based nature of data labelling and the re-use of data[55] are common to many cases of AI application and concern data in general.[56] In line with the methodology adopted,[57] the existing guidance in the field of data protection[58] can also be applied in this case and the data quality aspects extended to non-personal data.

b) Self-determination [Freedom of choice/Autonomy, Awareness]

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[59] and medical treatment, suggesting the adoption of new schemes – such as broad[60] or dynamic consent – which, however, could only contribute in part to solving this problem.

c) The doctor-patient relationship

Several factors concerning AI-based diagnosis – such as the loss of knowledge that cannot be encoded in data,[61] over-reliance on AI in medical decisions, effects of local

[55] Ferryman and Pitcan, 2018, 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").

[56] See above Section II.2.

[57] See e.g. above Figure 4: Common guiding values in the field of data protection and justice.

[58] See Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data. 2017. Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data. T-PD(2017)1; Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data. 2019. Guidelines on Artificial Intelligence and Data Protection. T-PD(2019)01. See also the related preliminary studies: Mantelero, A. 2019; Rouvroy, A. 2016.

[59] See also Recommendation CM/Rec(2019)2 on the protection of health-related data.

[60] See also Convention 108+. Explanatory Report, 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 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").

[61] See Caruana et al., 2015.

practices on training datasets, and potential deskilling in the medical sector[62] – may affect the care-patient relationship[63] and should be evaluated when adopting AI in this field.

d) Risk management [Risk-based approach, Accountability]

The field of medical devices[64] represents an interesting case study in terms of risk management, considering the significant consequences that the use of these devices can have on individuals. The European Union has already adopted a risk-based classification model[65] based on progressive safeguards according to the class of risk of each device (from conformity assessment procedures under the sole responsibility of the manufacturer or the intervention of a notified body, to inspection by a notified body and, in the cases of highest risk, the requirement of prior authorization before being placed 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 difficult, given their variety and different fields of application: several sector-specific classifications should be introduced, or general criteria adopted based on risk assessments procedures.

In addition, specific provisions on AI vigilance and the adoption of the precautionary principle in AI development, as discussed above,[66] can help to address these challenges.

II.4 Democracy

Democracy covers an extremely wide array of societal and legal issues,[67] most of them likely to be implemented with the support of ICT[68]. In this scenario, AI can play an important role in the present and future development of digital democracy in a information society.

Compared to the other areas examined (data protection and health), the broad dimension of this topic makes it difficult to identify a single binding sector-specific legal instrument for reference. Several international instruments deal with democracy and its different aspects, starting with the UN Declaration of Human Rights and the

[62] See Cabitza, Rasoini, and Gensini, 2017.

[63] See also WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects, 9th July 2018, https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/.

[64] See also European Commission, 2014.

[65] See Directive 93/42/EEC.

[66] See above Section II.2.

[67] See e.g. Council of Europe. Directorate General of Democracy – European Committee on Democracy and Governance. 2016. The Compendium of the most relevant Council of Europe texts in the area of democracy.

[68] See e.g. Directorate General of Democracy and Political Affairs – Directorate of Democratic Institutions. 2009. Project «Good Governance in the Information Society», CM(2009)9 Addendum 3. Indicatives Guides and Glossary relating to Recommendation Rec(2009) 1 of the Committee of Ministers to member states on electronic democracy (e-democracy), prepared by The Council of Europe's Ad hoc Committee on E-Democracy (CAHDE); Additional Protocol to the European Charter of Local Self-Government on the right to participate in the affairs of a local authority, 2009, Article 2.2.iii.

International Covenant on Civil and Political Rights. Similarly, in the European context, key principles for democracy are present in several international sources.

Based on Article 25 ICCPR, we can identify two main areas of intervention: (i) participation[69] and good governance, and (ii) elections. Undoubtedly, it is difficult or impossible to draw a red line between these fields as they are interconnected in various ways. AI can have an impact on all of them: participation (e.g. citizens engagement, participation platforms), good governance (e.g. e-government, decision-making processes, smart cities), pre-electoral phase (e.g. financing, targeting and profiling, propaganda), elections (e.g. prediction of election results, e-voting), and the post-election period (e.g. electoral dispute resolution).

As in any classification, this distinction is characterised by a margin of directionality. It is worth pointing here out that this is a functional classification based on different AI impacts, with no intention to provide a legal or political representation of democracy and its different key elements. The relationship between participation, good governance, and elections can therefore be considered from different angles and shaped in different ways, unifying certain areas or further subdividing them.

Participation is expressed both through taking part in the democratic debate and through the electoral process, but the way that AI tools interact with participation in these two cases differs and there are distinct international legal instruments specific to the electoral process.

II.4.1 Participation and good governance

The right to participate in public affairs (Article 25 Covenant) is based on a broad concept of "public affairs",[70] which includes public debate and dialogue between citizens and their representatives, with a close link to freedom of expression, assembly and association.[71] In this respect, AI is relevant from two different perspectives: as a means to participation and as the subject of participatory decisions.

Considering AI as a means, technical and educational barriers can undermine the exercise of the right to participate. Participation tools based on AI should therefore consider the risks of under-representation and lack of transparency in participative processes (e.g. platforms for the drafting of bills). At the same time, AI is also the subject of participatory decisions, as they include decisions on the development of AI in general and its use in public affairs.

[69] For a more detailed analysis see Faye Jacobsen, 2013. See also Maisley, 2017.

[70] See UN Office of the High Commissioner for Human Rights. 1996. General Comment No. 25: The right to participate in public affairs, voting rights and the right of equal access to public service (Art. 25). CCPR/C/21/Rev.1/Add.7.

[71] See also UN Office of the High Commissioner for Human Rights. 1981. CESCR General Comment No. 1: Reporting by States Parties, para 5 ("facilitate public scrutiny of government policies with respect to economic, social and cultural rights and to encourage the involvement of the various economic, social and cultural sectors of society in the formulation, implementation and review of the relevant policies").

AI-based participative platforms (e.g. Consul,[72] Citizenlab,[73] Decidim[74]) can make a significant contribution to the democratic process, facilitating citizen interaction, prioritising of objectives, and collaborative approaches in decision-making[75] on topics of general interests at different levels (neighbourhood, municipality, metropolitan area, region, country).[76] As these platforms are used in a social environment and collect information, the same aspects already discussed with regard to data protection, including security, can be recalled here by extending the guidelines discussed in the previous section on data to these applications.

However, other more specific issues arise in relation to AI tools for democratic participation (including those for preventing and fighting corruption[77]), which are associated with the following four main areas: transparency, **accountability**, **inclusiveness**, and **openness**. In this regard, the general principles set out in international binding instruments have an important implementation in the Recommendation CM/Rec(2009)1 of the Committee of Ministers to member states on electronic democracy (e-democracy), which provides a basis for further elaboration of the guiding principles in the field of AI with regard to democracy.

Transparency is a requirement for the use of technological applications for democratic purposes.[78] This principle is common to the fields analysed above, data and healthcare. However, transparency is a context-based notion. While in these fields transparency is closely related to self-determination, here it takes on a broader meaning. In a democratic process, transparency is not only a requirement for citizens' self-determination with respect to a technical tool, but is also a component of the democratic participatory process.[79] Transparency no longer has an individual dimension but assumes a collective dimension as a guarantee of the democratic process.

In this context, the use of AI-based solutions for e-democracy must be transparent in respect of their logic and functioning (e.g. content selection in participatory platforms) providing clear, easily accessible, intelligible and updated information about the AI tools used.[80]

[72] See <https://consulproject.org/en/>, accessed 29.12.2019.

[73] See <https://www.citizenlab.co/>, accessed 29.12.2019.

[74] See <https://decidim.org/>, accessed 29.12.2019.

[75] See also Council of Europe. Guidelines for civil participation in political decision making. CM(2017)83-final. Adopted by the Committee of Ministers on 27 September 2017 at the 1295th meeting of the Ministers' Deputies.

[76] See also Recommendation CM/Rec(2009)2 on the evaluation, auditing and monitoring of participation and participation policies at local and regional level.

[77] See United Nations Convention against Corruption, 2003, Article 13.

[78] See Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), para 6.

[79] See also Guidelines for civil participation in political decision making. CM(2017)83-final, IV.

[80] See Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), para. 6 ("facilitates and enhances access, accessibility […] by using, where feasible, transparent […] means") and Appendix to Recommendation CM/Rec(2009)1, para. P.57. See also Recommendation CM/Rec(2016)5 on Internet freedom. Appendix, paras 2.1.3 and 3.2.

Moreover, the implementation of this notion of transparency should also consider the range of different users of these tools, adopting an **accessible** approach[81] from the early stages of the design of AI tools. This is to ensure effective transparency with regard to vulnerable and impaired groups, giving added value to accessibility in this context.

Transparency and accessibility are closely related to the nature of the architecture used to build AI systems. **Open source and open standards**[82] can therefore contribute to democratic oversight of the most critical AI applications.[83] There are cases where openness is affected by limitations, due to the nature of the specific AI application (e.g. crime prevention). In these cases, auditability, as well as certification schemes, play a more important role than they already do in relation to AI systems in general.[84]

In the context of AI applications to foster democratic participation, an important role can be also played by **interoperability**[85] as it facilitates integration between different services/platforms for e-democracy and at different geographical levels. This aspect is already relevant for e-democracy in general,[86] and should therefore be extended to the design of AI-based systems.

Another key principle in e-democracy, as in the data and health sectors, is **accountability**. Unlike the previous principles examined, accountability does not take on a different meaning here, and therefore does not seem to require a sector-specific implementation in the context of AI, other than its general application.

Finally, given the role of media in the context of democratic participation and in line with Recommendation CM/Rec(2016)4 of the Committee of Ministers of the Council of Europe,[87] AI applications must not compromise the confidentiality and security of communications and protection of journalistic sources and whistle-blowers.[88]

In addressing the different aspects of developing AI solutions for democratic participation, a first consideration is that a democratic approach is incompatible with a techno-determinist approach. AI solutions to address societal problems should therefore be the result of an inclusive process. Hence, values such as the protection

[81] See also Recommendation CM/Rec(2018)4 on the participation of citizens in local public life, Appendix, para. B.IV.

[82] See also Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), para. 6 and Appendix, para P.54.

[83] See also Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), Appendix, para. G.58.

[84] It is worth to underline that auditing and certification schemes play an important role also in cases of open source AI architecture, as this nature does not imply per se absence of bias or any other shortcomings. See also Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), Appendix, paras P.55 and G.57 ("E-democracy software should either be open source software that can be inspected or, alternatively, be certified by an independent body").

[85] See also Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), Appendix, paras P. 56, G.56, 59 and 60.

[86] See also Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), para. 6.

[87] See Recommendation CM/Rec(2016)4 on the protection of journalism and safety of journalists and other media actors, Appendix, para. 2; Council of Europe, Parliamentary Assembly. 2019. Resolution 2254 (2019)1. Media freedom as a condition for democratic elections.

[88] See also Parliamentary Assembly, Resolution 2300 (2019)1, Improving the protection of whistle-blowers all over Europe; Recommendation CM/Rec(2014)7 on the protection of whistleblowers.

of minorities, pluralism and diversity should be a necessary consideration in the development of these solutions.

From a democratic perspective, the first question we should ask is: do we really need an AI-based solution to a given problem as opposed to other options,[89] considering the potential impact of AI on rights and freedoms? If the answer to this question is yes, the next step is to examine **value-embedding** in AI development.[90]

The proposed AI solutions must be designed from a human rights-oriented perspective, ensuring full respect for human rights and fundamental freedoms, including the adoption of **assessment tools and procedures** for this purpose.[91] In the case of AI applications with a high impact on human rights and freedoms, such as electoral processes, legal compliance should be **prior assessed**. In addition, AI systems for public tasks should be **auditable** and, where not excluded by competing prevailing interests, audits should be publicly available.

Another important aspect to be considered is the **public-private partnership** that frequently characterises AI services for citizens, weighing which is the best choice between in-house and third-party solutions, including the many different combinations of these two extremes. In this regard, when AI solutions are fully or partially developed by private companies, **transparency of contracts** and clear **rules on access and use of citizens' data** have a critical value in terms of democratic oversight.

Restrictions on access and use of citizens' data are not only relevant from a data protection perspective (principles of data minimisation and purpose limitation) but more generally with regard to the bulk of data generated by a community, which also includes non-personal data and aggregated data. This issue should be considered as a component of democracy in the digital environment, where the **collective dimension** of the digital resources generated by a community should entail forms of citizen control and oversight, as happens for the other resources of a territory/community (e.g. the environment).

The considerations already expressed above on openness as a key element of democratic participation tools should be recalled here, given their impact on the design of AI systems. Furthermore, the design, development and deployment of these systems should also consider the adoption of an environmentally friendly and sustainable strategy.[92]

Finally, it is worth noting that while AI-design is a key component of these systems, design is not neutral. Values can be embedded in technological artefacts,[93] including AI systems. These values can be chosen intentionally and, in the context of e-democracy, this must be based on a democratic process. But values may also be unintentionally embedded into AI solutions, due to the cultural, social and gender

[89] See also Recommendation CM/Rec(2020)1, Appendix, para. 5.7.

[90] See also Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (Adopted by the Committee of Ministers on 13 February 2019 at the 1337th meeting of the Ministers' Deputies, para. 7.

[91] See Recommendation CM/Rec(2009)1, paras 5 and 6, and Appendix to Recommendation CM/Rec(2009)1, para. G.67. See also above Section II.2 on data and the role of the committees of experts and A Mantelero, 'AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment' (2018) 34 Computer Law & Security Review 754.

[92] See also Recommendation CM/Rec(2009)1, Appendix, para. P. 58.

[93] See also P-P Verbeek, 2011, 41-65.

composition of AI developer teams. For this reason, **inclusiveness** has an added value here, in terms of inclusion and diversity[94] in AI development.

With regard to good governance,[95] the principles discussed for e-democracy can be repeated here.[96] This is the case with smart cities and sensor-based environmental management, where open, transparent and inclusive decision-making processes play a central role.[97] Similarly, the use of AI to supervise the activities of local authorities,[98] for auditing and anticorruption purposes,[99] should be based on **openness** (open source software), **transparency** and **auditability**.

More generally, AI can be used in government/citizen interaction to automate citizen' inquiries and information requests.[100] However, in these cases, it is important to guarantee the right to know we are interacting with a machine[101] and to have a human contact point. Moreover, access to public services must not depend on the provision of data that is unnecessary and not proportionate to the purpose.

Special attention should also be paid to the potential use of AI in human-machine interaction to implement nudging strategies.[102] Here, due to the complexity and obscurity of the technical solutions adopted, AI can increase the passive role of citizens and negatively affect the democratic decision-making process. Otherwise, an active approach based on conscious and active participation in community goals should be preferred and better managed by AI participation tools. Where adopted, nudging strategies should still follow an evidence-based approach.

Finally, the use of AI systems in governance tasks raises challenging questions about the relationship between human decision-makers and the role of AI in the decision-making process.[103] These issues are more relevant with regard to the functions that have a high impact on individual rights and freedoms, as in the case of jurisdictional decisions. For this reason, concerns about transparency (including explainability) of AI reasoning and the relationship between the use of AI and the freedom of decision-makers will be analysed in Section 5.

---

[94] See also Recommendation CM/Rec(2020)1, Appendix, para. 3.5.

[95] See Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), Appendix, para. P.4 ("[…] good governance, which is the efficient, effective, participatory, transparent and accountable democratic exercise of power in electronic form, and includes informal politics and non-governmental players").

[96] See also Recommendation Rec(2004)15 on electronic governance ("e-governance"); Council of Europe. 2008. The 12 Principles of Good Governance enshrined in the Strategy on Innovation and Good Governance at local level, endorsed by a decision of the Committee of Ministers of the Council of Europe in 2008..

[97] See also Privacy International, 2017.

[98] See also Recommendation CM/Rec(2019)3 on supervision of local authorities' activities, Appendix, Guidelines on the improvement of the systems of supervision of local authorities' activities, paras 4 and 9.

[99] See also Savaget, Chiarini and Evans, 2019, discussing the Brazilian case of the 'Operação Serenata de Amor' (OSA).

[100] See Mehr. 2017.

[101] See also GAI 2.11.

[102] See, *ex multis*, Sunstein, 2015a; Sunstein, 2015a; Sunstein and Thaler, 2003; Thaler and Sunstein, 2008.

[103] See also Calo and Citron, 2020, Forthcoming.

II.4.2 Elections

As in other areas, the impact of AI on electoral processes is broad and concerns the pre-election, election, and post-election phases in different ways. However, an analysis focused on the stages of the electoral process does not adequately highlight the different ways in which AI solutions interact with it.

The influence of AI is therefore better represented by the following distinction: AI for the electoral process (e-voting, predictions of results, and electoral dispute resolution) and AI for electoral campaigns (micro-targeting and profiling, propaganda and fake news). While in the first area AI is mainly a technological improvement of an existing process, in the field of electoral campaigning AI-based profiling and propaganda raise new concerns that are only partially addressed by the existing legal framework. In addition, several documents have emphasised the active role of states in creating an enabling environment for freedom of expression.[104]

As regards the technological implementation of e-democracy (e-voting, prediction of results, and electoral dispute resolution), some of the key principles mentioned with regard to democratic participation are also relevant here. **Accessibility,**[105] **transparency,**[106] **openness,**[107] **risk management and accountability** (including the adoption of certification and auditing procedures)[108] are fundamental elements of the technological solutions adopted in these stages of the electoral process.

As regards AI for campaigning (micro-targeting and profiling, propaganda and fake news), some of the issues raised concern the processing of personal data in general. The principles set out in Convention 108+ can therefore be applied and properly contextualised.[109]

More specific and new responses are needed in the case of propaganda and disinformation.[110] Here the existing binding and non-binding instruments do not set

[104] See Recommendation CM/Rec(2018)1 on media pluralism and transparency of media ownership; Joint Declaration on "Fake News," Disinformation and Propaganda, The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information (3 March 2017). See also Recommendation CM/Rec(2016)5 on Internet freedom, Appendix, paras 1.5, 2.1 and 3; European commission for Democracy trough law (Venice Commission). 2019. Joint Report of the Venice Commission and of the Directorate of Information society and Actions Against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital Technologies and Elections, para. 151.E; OSCE, 2020. See also Bychawska-Siniarska, 2017.

[105] See Recommendation CM/Rec(2017)5 on standards for e-voting, Appendix I, E-voting Standards, paras 1 and 2.

[106] See Recommendation CM/Rec(2017)5, Appendix I, para. 32. See also Council of Europe. Directorate General of democracy and Political Affairs – Directorate of Democratic Institutions. 2011. Guidelines on transparency of e-enabled elections.

[107] See Recommendation CM/Rec(2017)5, Appendix I, para. 35.

[108] See Recommendation CM/Rec(2017)5, Appendix I, paras 36, 37, 38, 39 and 40.

[109] Recommendation CM/Rec(2010)13 on the protection of individuals with regard to automatic processing of personal data in the context of profiling and its ongoing review, see Council of Europe, Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data. 2019. Profiling and Convention 108+: Suggestions for an update. T-PD(2019)07BISrev.

[110] See Manheim and Kaplan, 2019; European Commission - Networks, Content and Technology-Directorate-General for Communication, 'A Multi-Dimensional Approach to Disinformation Report of the

specific provisions, given the novelty of the disinformation based on new forms of communication, such as social networks, which differ from traditional media[111] and often bypass the professional mediation of the journalists.

However, general principles, such as the **principle of non-interference** by public authorities on media activities to influence elections,[112] can be extended to these new forms of propaganda and disinformation. Considering the use of AI to automate propaganda, future AI regulation should extend the scope of the general principles of non-interference to AI-based systems used to provide false, misleading and harmful information. In addition, to prevent such interference, states[113] and social media providers should adopt a **by-design approach** to increase their resilience to disinformation and propaganda.

Similarly, the obligation to cover election campaigns in a **fair, balanced and impartial** manner[114] should entail obligations for media and social media operators regarding the transparency of the logic of the algorithms used for content selection,[115] ensuring pluralism and diversity of voices,[116] including critical ones.[117]

Moreover, states and intermediaries should promote and facilitate access to tools to detect disinformation and non-human agents, as well as support independent research on the impact of disinformation and projects offering fact-checking services to users.[118]

Given the important role played by advertising in disinformation and propaganda, the criteria used by AI-based solutions for political advertising should be **transparent**,[119] **auditable** and provide **equal conditions** to all the political parties and candidates.[120]

Independent High Level Group on Fake News and Online Disinformation' (2018). See also *Stoll v. Switzerlad* [GC], no.69698/01, § 104.

[111] See also Recommendation CM/Rec(2011)7 on a new notion of media.

[112] See Recommendation CM/Rec(2007)15 on measures concerning media coverage of election campaigns, para. I.1.

[113] See also Joint Declaration on "Fake News," Disinformation and Propaganda, para. 2.c.

[114] See Recommendation CM/Rec(2007)15, para. II.1.

[115] See also Joint Declaration on "Fake News," Disinformation and Propaganda; Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries, Appendix, paras. 2.1.3 and 2.3.5 ("Due to the current limited ability of automated means to assess context, intermediaries should carefully assess the human rights impact of automated content management, and should ensure human review where appropriate. They should take into account the risk of an overrestrictive or too lenient approach resulting from inexact algorithmic systems, and the effect these algorithms may have on the services that they provide for public debate").

[116] See also EU Code of Practice on Disinformation, 2018.

[117] See also Recommendation CM/Rec(2016)4, Appendix, para. 15.

[118] See also Joint Declaration on "Fake News," Disinformation and Propaganda, para. 4.e; European commission for Democracy trough law. 2019, para. 151.D.

[119] See also Council of Europe. Parliamentary Assembly. Resolution 2254 (2019)1. Media freedom as a condition for democratic elections, paras 9.2 and 11.1; European commission for Democracy trough law (Venice Commission). 2019. Joint Report of the Venice Commission and of the Directorate of Information society and Actions Against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital Technologies and Elections, paras 151.A and 151.B.

[120] See also Recommendation CM/Rec(2007)15, para. II.5.

In addition, intermediaries should review their advertising models to ensure that they do not adversely affect the **diversity of opinions and ideas**.[121]

II.5 Justice

As in the previous section, the field of justice is a broad domain and analysing the whole spectrum of the consequences of AI on justice and its related effects on democracy would be too ambitious. In line with the scope of this study, this section sets out to describe the main challenges associated with the use of AI and the principles which, based on international legally binding instruments, can contribute to its future regulation.

Justice differs from data protection and health in the absence of specific and dedicated binding instruments, such as Convention 108+ and the Oviedo Convention. This analysis is therefore more centred on the contextualisation of general guiding principles than on specific legal instruments.

This exercise is facilitated by the European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment, adopted by the CEPEJ in 2019, which directly addresses the relationship between justice and AI. Although this non-binding instrument is classed as an ethical charter, to a large extent it concerns legal principles enshrined in international instruments.

Guiding principles for the development of AI in the field of justice can be derived from the following binding instruments: the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, the Convention for the Protection of Human Rights and Fundamental Freedoms, the International Convention on the Elimination of All Forms of Racial Discrimination, the Convention on the Elimination of All Forms of Discrimination against Women, and the Convention for the Protection of Human Rights and Fundamental Freedoms.[122]

Given the range of types and purposes of operations in this field and the various professional figures and procedures involved, this section makes a functional distinction between two areas: (i) judicial decisions and alternative dispute resolutions (ADRs) and (ii) crime prevention/prediction. Before analysing and contextualising the key principles relating to these two areas, we should offer some general observation, which may also apply to the action of the public administration as a whole.[123]

First of all, it is worth noting that – compared to human decisions, and more specifically judicial decisions – the logic behind AI systems does not resemble legal reasoning. Instead they simply execute codes based on a data-centric and mathematical/statistical approach.

In addition, error rates for AI are close to, or lower than, the human brain in fields such as image labelling, but more complicated decision-making tasks have higher error

[121] See also Joint Declaration on "Fake News," Disinformation and Propaganda, para. 4.e.

[122] See also, with regard to the EU area, the Charter of Fundamental Rights of the European Union.

[123] See above Section II.4.

rates. This is the case with legal reasoning in problem solving.[124] At the same time, while a misclassification of an image of a cat may have limited adverse effects, an error rate in legal decisions has a high impact on rights and freedom of individuals.

It is worth pointing out that the difference between errors in human and machine decision-making has an important consequence in terms of scale: while human error affects only individual cases, poor design and bias in AI inevitably affect all people in the same or similar circumstances, with AI tools being applied to a whole series of cases. This may cause group discrimination, adversely affecting individuals belonging to different categories.

Given the textual nature of legal documents, natural language processing (NLP) plays an important role in AI applications for the justice sphere. This raises several critical issues surrounding commercial solutions developed with a focus on the English-speaking market, making them less effective in a legal environment that uses languages other than English.[125] Moreover, legal decisions are often characterised by implicit unexpressed reasoning, which may be amenable to expert systems, but not by language-based machine learning tools. Finally, the presence of general clauses requires a prior knowledge of the relevant legal interpretation and continual updates which cannot be derived from text mining.

All these constraints suggest a careful and more critical adoption of AI in the field of justice than in other domains and, with regard to court decisions and ARDs, suggest following a distinction between cases characterised by routinely and fact-based evaluations and cases characterised by a significant margin for legal reasoning and discretion.[126]

II.5.1 Court decisions and ADRs

Several so-called Legal Tech AI products do not have a direct impact on the decision-making processes in courts or alternative dispute resolutions (ADRs), but rather facilitate content and knowledge management, organisational management, and performance measurement.[127] These applications include, for example, tools for contracts categorisation, detection of divergent or incompatible contractual clauses, e-discovery, drafting assistance, law provision retrieval, assisted compliance review. In addition, some applications can provide basic problem-solving functions based on standard questions and standardised situations (e.g. legal chatbots).

---

[124] See Dupont et al., 2018, 148 ("Deep Learning has no natural way to deal with hierarchical structure, which means that all the available variables are considered on the same level, as 'flat' or non-hierarchical. This presents a major hurdle when decisions carry a heavy moral or legal weight that must supersede other features"). See also Osoba and Welser, 2017, 18 ("Another angle on the problem is that judgments in the space of social behavior are often fuzzy, rather than well-defined binary criteria […]. We are able to learn to navigate complex fuzzy relationships, such as governments and laws, often relying on subjective evaluations to do this. Systems that rely on quantified reasoning (such as most artificial agents) can mimic the effect but often require careful design to do so. Capturing this nuance may require more than just computer and data scientists."). See also Cummings et al., 2018, 13.

[125] See Council of Bars & Law Societies of Europe, 2020, 29.

[126] See the following Section on the distinction between codified justice and equitable justice.

[127] See European Commission for the Efficiency of Justice (CEPEJ). 2018. European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment, Appendix II.

Although AI has an impact in such cases on legal practice and legal knowledge that raises various ethical issues,[128] the potential adverse consequences for human rights, democracy and the rule of law are limited. To a large extent, they are related to inefficiencies or flaws of these systems.

In the case of content and knowledge management, including research and document analysis, these flaws can generate incomplete or inaccurate representations of facts or situations, but this affects the meta-products, the results of a research tool that need to be interpreted and adequately motivated when used in court. Liability rules, in the context of product liability, for instance, can address these issues.

In addition, bias (poor case selection, misclassification etc.) affecting standard text-based computer-assisted search tools for the analysis of legislation, case-law and literature,[129] ca be countered by suitable **education and training** of legal professionals and the **transparency** of AI systems (i.e. description of their logic, potential bias and limitations) can reduce the negative consequences.

Transparency should also characterise the use by courts of AI for legal research and document analysis. Judges must be transparent as to which decisions depend on AI and how the results provided by AI are used to contribute to the arguments, in line with the **principles of fair trial and equality of arms**.[130]

Finally, transparency can play an important role with regard to legal chatbots based on AI, making users aware of their logic and the resources used (e.g. list of cases analysed). Full transparency should also include the sources used to train these algorithms and access to the database used to provide answers. Where these databases are private, third party **audits** should be available to assess the quality of datasets and how potential biases have been addressed, including the risk of under- or over-representation of certain categories (**non-discrimination**).

Further critical issues affect AI applications designed to automate alternative dispute resolution or to support judicial decision. Here, the distinction between codified justice and equitable justice[131] suggests that AI should be circumscribed for decision-making purposes to cases characterised by routine and fact-based evaluations. This entails the importance to carry out further research on the classification of the different kind of decisional processes to identify those routinised applications of legal reasoning that can be demanded to AI, preserving in any case human overview that also guarantees legal creativity of decision-makers.[132]

---

[128] See also Nunez, 2017.

[129] See the notion of e-justice in Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), Appendix, para. 38.

[130] See also European Commission for the Efficiency of Justice (CEPEJ). 2018. European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment.

[131] See Re and Solow-Niederman, 2019, 252-254 ("Equitable justice entails both reflection on the values set in place by the legal system and the reasoned application of those values, in context […] Codified justice refers to the routinized application of standardized procedures to a set of facts […] In short, codified justice sees the vices of discretion, whereas equitable justice sees its virtues").

[132] See also Clay, 2019. In this regard, for example, a legal system that provides compensation for physical injuries on the basis of the effective patrimonial damages could be automatised, but it will not be able to reconsidered the foundation of the legal reasoning and extend compensation to non-personal and existential damages.

Regarding equitable justice, as the literature points out,[133] its logic is more complicated than the simple outcome of individual cases. Expressed and unexpressed values and considerations, both legal and non-legal, characterise the reasoning of the courts and are not replicable by the logic of AI. ML-based systems are not able to perform a legal reasoning. They extract inferences by identifying patterns in legal datasets, which is not the same as the elaboration of legal reasoning.

Considering the wider context of the social role of courts, jurisprudence is an evolving system, open to new societal and political issues. AI path-dependent tools could therefore stymie this evolutive process: the deductive and path-dependent nature of certain AI-ML (Machine Learning) solutions can undermine the important role of human decision-makers in the evolution of law in practice and legal reasoning.

Moreover, at the individual level, path-dependency may also entail the risk of "deterministic analyses",[134] prompting the resurgence of deterministic doctrines to the detriment of doctrines of individualisation of the sanction and with prejudice to the principle of rehabilitation and individualisation in sentencing.

In addition, in several cases, including ADR, both the mediation between the parties' demands and the analysis of the psychological component of human actions (fault, intentionality) require emotional intelligence that AI systems do not have.

These concerns are reflected in the existing legal framework provided by the international legal instruments. The Universal Declaration of Human Rights (Articles 7 and 10), the ICCPR (Article 14), the Convention for the Protection of Human Rights and Fundamental Freedoms (Article 6) and also the Charter of Fundamental Rights of the European Union (Article 47) stress the following key requirements with regard to the exercise of judicial power: equal treatment before the law, impartiality, independence and competency. AI tools do not possess these qualities and this limits their contribution to the decision-making process as carried out by courts.

As stated by the European Commission for the Efficiency of Justice, "the neutrality of algorithms is a myth, as their creators consciously or unintentionally transfer their own value systems into them". Many cases of biases regarding AI applications confirm that these systems too often – albeit in many cases unintentionally – provide a partial representation of society and individual cases, which is not compatible with the principles of **equal treatment before the law** and **non-discrimination**.[135] **Data quality** and other forms of quality **assessment** (impact assessment, audits, etc.) can reduce this risk[136] but, given the degree of potentially affected interests in the event of biased decisions, the risks remain high in the case of equitable justice and seem disproportionate to the benefits largely in terms of efficiency for the justice system.[137]

Further concerns affect the **principles of fair trial and of equality of arms**,[138] when court decisions are based on the results of proprietary algorithms whose training data

[133] See Re and Solow-Niederman, 2019.

[134] See European Commission for the Efficiency of Justice (CEPEJ). 2018. European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment, 9.

[135] See also CEPEJ, 2018.

[136] See also CEPEJ, 2018.

[137] See also Recommendation CM/Rec(2020)1, Appendix, para. 11.

[138] See also CEPEJ, 2018, Appendix I, para. 138.

and structure are not publicly available.[139] A broad notion of **transparency** might address these issues in relation to the use of AI in judicial decisions, but the transparency of AI – a challenging goal in itself – cannot address the other structural and functional objections cited above.

In addition, data scientists can shape AI tools in different ways in the design and training phases, so that were AI tools to became an obligatory part of the decision-making process, governments selecting the tools to be used by the courts could potentially indirectly interfere with the **independence** of the judges.[140]

This risk is not eliminated by the fact that the judge remains free to disregard AI decisions, providing a specific motivation. Although **human oversight** is an important element,[141] its effective impact may be undermined by the psychological or utilitarian (cost-efficient) propensity of the human decision-maker to take advantage of the solution provided by AI.[142]

II.5.2 Crime prevention

The complexity of crime detection and prevention has stimulated research in AI applications to facilitate human activities. In recent years, several solutions[143] and a growing literature have been developed in the field of predictive policing, which is a proactive data-driven approach to crime prevention. Essentially, the available solutions pursue two different goals: to predict where and when crimes might occur or to predict who might commit a crime.[144]

These two purposes have a distinct potential impact on human rights and freedom, which is more pronounced when AI is used for individual predictions. However, in both cases, we can repeat here the considerations about the general challenges related to AI (obscurity, intellectual property rights, large-scale data collection[145], etc.) discussed in the previous sections and partially addressed by transparency, **data quality**, **data protection**, **auditing** and the other measures. It is worth noting that the role of

[139] See also CEPEJ, Appendix I, para. 131 ("the lack of transparency in the algorithm operation processes designed by private companies (which claim intellectual property) was another cause for concern. If we take into account the fact that they take their source data from the state authorities themselves, their lack of accountability to citizens poses a major democratic problem […] an example of this is when ProPublica revealed the flaws in the COMPAS algorithm following the owner company's refusal to share it").

[140] See also CEPEJ, 2018.

[141] See also CEPEJ, 2018.

[142] See also Mantelero, 2019 ("the supposedly reliable nature of AI mathematics-based solutions can induce those taking decisions on the basis of algorithms to place trust in the picture of individuals and society that analytics suggest").

[143] See Završnik, 2019; European Union Agency for Fundamental Rights, 2018, 98-100; Osoba and Welser, 2017.

[144] For a taxonomy of predictive methods, see Perry et al., 2013, who identifies the following four categories: methods for predict crimes (focused on places and times of crimes), method for predicting offenders (focused on individuals), methods for predicting perpetrators' identities (focused on individuals), and methods for prediction victims of crimes (focused on groups and, in some cases, on individuals).

[145] See also Recommendation Rec(2001)10 on the European Code of Police Ethics, Appendix, para. 42.

transparency[146] in the judicial context could be limited so as not to frustrate the deterrent effect of these tools. Full transparency could therefore be replaced by auditing and oversight by independent authorities.

Leaving aside the organisational aspects regarding the limitation of police officers' self-determination in the performance of their duties, the main issues with regard to the use of AI to predict crime on geographic and temporal basis concern the impact of these tools on the **right to non-discrimination**.[147] Self-fulfilling bias, community bias[148] and historical bias[149] can produce forms of stigmatisation for certain groups and the areas where they typically live.

Where data analysis is used to classify crimes and infer evidence on criminal networks, proprietary solutions raise issues in terms of respect for the **principles of fair trial and of equality of arms** with regard to the collection and use of evidence. Moreover, if the daily operations of policy departments are guided by predictive software, this raises a problem of **accountability** of the strategies adopted, as they are partially determined by software and hence by software developer companies, rather than the police.

A sharper conflict with human rights arises in the area of predictive policing tools that use profiling to support individual forecasting. Quite apart from the question of data processing and profiling,[150] these solutions can also adversely affect the principle of **presumption of innocence**,[151] procedural **fairness**, and the right to **non-discrimination**.[152]

While non-discrimination issues could be partially addressed, the remaining conflicts seem to be more difficult to resolve. From a human rights standpoint and in terms of proportionality (including the right to respect for private and family life[153]), the risk of prejudice to these principles seems high and not adequately countered by the evidence of benefits for individual and collective rights and freedoms.[154] In the light of future AI regulation, this should urge careful consideration of these issues, taking into account the distinction between the technical possibilities of AI solutions and their concrete benefits in safeguarding and enhancing human rights and freedoms.

Finally, from a wider and comprehensive human rights perspective, the focus on crime by data-driven AI tools drives a short-term factual approach that underrates the social

[146] See also Barrett, 2017, 361-62.

[147] See European Union Agency for Fundamental Rights, 2018, 10.

[148] See also Barrett, 2017, 358-59 ("For some, the goal of collective safety merits a unilateral sacrifice of some degree of individual rights in this particular context. But that calculus must change if the sacrifice is not collective, but instead confined to minority groups, or becomes fundamentally arbitrary by virtue of an unacceptable degree of error.")

[149] See Bennett Moses and Chan, 2018.

[150] See above Section II.2.

[151] See also Recommendation Rec(2001)10 on the European Code of Police Ethics, Appendix, para. 47.

[152] See also Recommendation Rec(2001)10 on the European Code of Police Ethics, Appendix, para. 49.

[153] See van Brakel and De Hert, 2011, 183. See also *Szabó and Vissy v Hungary* [2016] European Court of Human Rights Fourth Section. Application no. 37138/14.

[154] See Meijer and Wessels, 2019.

issues that are often crime-related and require long-term social strategies involving the effective enhancement of individual and social rights and freedoms.[155]

II.6 Harmonisation of the principles identified

The previous sections identified **several guiding principles for the future regulation of AI**. These principles were **contextualised with regard to the challenges associated with AI** in the various areas examined, but it is worth looking at the existing level of harmonisation between these principles.

The findings of this study indicate that in a limited number of cases **there are common principles** (the primacy of the human being, individual self-determination, non-discrimination, human oversight). This is due to several factors.

First, **some principles are sector specific**. This is the case, for instance, of the independence of the judges or the principles of fair trial and of equality of arms, which concern justice alone.[156]

Second, some guiding principles are the same in different areas, but with **different nuances** in each context. This is true for transparency, which is often regarded as pivotal in AI regulation, but takes on different meanings in different regulatory contexts.

In the fields of health and personal data, transparency relates to the information given to individuals about the treatment concerning them, with particular attention to the process and related risks and with a strong connotation of individual self-determination. But transparency is also relevant in data protection to control the exercise of power over data in the hands of public and private entities. This different face of transparency is then considered with regard to AI applications for democratic participation and good governance. Then again, in the context of justice, transparency has a more complex significance being vital to safeguard fundamental rights and freedoms (e.g. use of AI in the courts), but also requiring limitations to avoid prejudicing competing interests (e.g. crime detection and prevention in predictive policing).

We can therefore conclude that transparency is a guiding value, but we must go beyond a mere claim to transparency as a key principle for AI regulation. As with other key principles (such as participation, inclusion, democratic oversight, and openness), a proper contextualisation is necessary, adopting provisions that take into account the different contexts in which they operate.

Third, some principles are different, but belong to the **same conceptual area**, assuming various nuances in the different contexts. This is the case with accountability and guiding principles on risk management in general. Here the level of detail and related requirements can be more or less elaborate. For instance, in the field of data protection there are several provisions implementing these principles with a significant degree of detail, whereas in the case of democracy and justice these principles are less developed with regard to data-intensive applications such as AI.

[155] See also Rosenbaum, 2006, 245–266.

[156] See also the principles of equitable access and of beneficence in health sector, or the principles of non-interference by public authorities on media activities to influence elections and the obligation to offer equal conditions to all the political parties and candidates in electoral advertising.

Finally, there are certain components of an AI regulatory strategy that are not principles, but **operational approaches and solutions,** common to the different areas though requiring context-based development. This is the case with the important role played by education and training, interoperability and expert committees.

Such considerations suggest only partial harmonisation is achievable. The regulatory approach to AI should therefore be based on **a legally binding instrument that includes both general provisions** – focusing on common principles and operational solutions – and **more specific and sectoral provisions**, covering those principles that are only relevant in a given field or cases where the same principle is contextualised differently in the different fields.

II.7 Conclusions

This analysis has confirmed the validity of the methodological approach adopted, which focuses on the **contextualisation** of guiding principles extracted from legally binding and non-binding intentional instruments. At the same time, it also highlighted the complexity of systematising the provisions of a wide variety of instruments, which differ not only in their binding nature, but also in their specific focus and approach, as well as their structure.

The results have also confirmed that the existing framework based on human rights, democracy and the rule of law can provide an appropriate and common context for the elaboration of **a more specific binding instrument to regulate AI in line with the principles and values enshrined in the international legal instruments, capable of addressing more effectively the issues raised by AI**.

This international framework necessarily leads us to reaffirm the central role of human dignity in the context of AI, where machine-driven solutions cannot be allowed to dehumanise individuals. This may also suggest the introduction of specific limitations to AI when developed or used in a way that is not consistent with respect for human dignity,[157] human rights, democracy and the rule of law.

With a view to future AI regulation, this positive methodological and substantive outcome does not exclude the existence of some gaps. These mainly concern broad areas, such as democracy and justice, where different options and interpretations are available, depending on the political and societal vision of the future relationship between humans and machines.

Further investigation in the field of human rights and AI, as well as the ongoing debate at international and regional level, will contribute to bridging these gaps. However, given the evolving nature of AI, a **co-regulatory approach** is desirable.

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.

[157] See also UNESCO. 1997. Declaration on the Human Genome and Human Rights, Article11.

References

Andorno, R. 2005. The Oviedo Convention: A European Legal Framework at the Intersection of Human Rights and Health Law. Journal of International Biotechnology Law, January 2005. https://doi.org/10.1515/jibl.2005.2.4.133, accessed 20.02.2020.

Azencott, C.-A. 2018. Machine Learning and Genomics: Precision Medicine versus Patient Privacy. Phil. Trans. R. Soc. A 376, no. 2128 (13 September 2018): 20170350. https://doi.org/10.1098/rsta.2017.0350, accessed 14.01.2020.

Barrett, L. 2017. Reasoably Suspicious Algorithms: Predictive Policing at the United States Border. 41 N.Y.U. Rev. Law & Social Change 327.

Bennett Moses, L., Chan, J. 2018. Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability. 28 Policing and Society 806.

Bychawska-Siniarska, D. 2017. Protection the Right to Freedom of Expression under the European Convention on Human Rights. Council of Europe.

Cabitza, F., Rasoini, R., and Gensini, G.F. 2017. Unintended Consequences of Machine Learning in Medicine. JAMA 318, no. 6 (8 August 2017): 517. https://doi.org/10.1001/jama.2017.7797, accessed 18.12.2019.

Calo, R., Citron, D.K. 2020. The Automated Administrative State: A Crisis of Legitimacy. Emory Law Journal, Forthcoming. https://ssrn.com/abstract=3553590, accessed 20.04.2020.

Caruana, R. et al. 2015. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Cham, Switzerland: Springer International Publishing AG.

Clay, T. (ed). 2019. L'arbitrage en ligne. Rapport du Club des Juristes. Paris. 58 https://www.leclubdesjuristes.com/les-commissions/larbitrage-en-ligne/, accessed 30.05.2020.

Committee of Ministers. 1999. Recommendation No. R (99) 5 for the protection of privacy on the internet. Adopted by the Committee of Ministers on 23 February 1999 at the 660th meeting of the Ministers' Deputies.

Committee of Ministers. 2001. Recommendation Rec(2001)10 of the Committee of Ministers to member states on the European Code of Police Ethics. Adopted by the Committee of Ministers on 19 September 2001 at the 765th meeting of the Ministers' Deputies.

Committee of Ministers. 2003. Recommendation CM/Rec(2003)4 on common rules against corruption in the funding of political parties and electoral campaigns. Adopted by the Committee of Ministers on 8 April 2003 at the 835th meeting of the Ministers' Deputies.

Committee of Ministers. 2004. Recommendation CM/Rec(2004)15 on Electroic Governance ("E-Governance"). Adopted by the Committee of Ministers on 15 December 2004 at the 909th meeting of the Ministers' Deputies.

Committee of Ministers. 2007. Recommendation CM/Rec(2007)15 of the Committee of Ministers to member states on measures concerning media coverage of election campaigns. Adopted by the Committee of Ministers on 7 November 2007 at the 1010th meeting of the Ministers' Deputies.

Committee of Ministers. 2009. Recommendation CM/Rec(2009)1 of the Committee of Ministers to member states on electronic democracy (e-democracy). Adopted by the

Committee of Ministers on 18 February 2009 at the 1049th meeting of the Ministers' Deputies.

Committee of Ministers. 2009. Recommendation CM/Rec(2009)2 of the Committee of Ministers to member states on the evaluation, auditing and monitoring of participation and participation policies at local and regional level. Adopted by the Committee of Ministers on 11 March 2009 at the 1050th meeting of the Ministers' Deputies.

Committee of Ministers. 2010. Recommendation CM/Rec(2010)13 on the protection of individuals with regard to automatic processing of personal data in the context of profiling. Adopted by the Committee of Ministers on 23 November 2010 at the 1099th meeting of the Ministers' Deputies.

Committee of Ministers. 2011. Recommendation CM/Rec(2011)7 of the Committee of Ministers to member states on a new notion of media. Adopted by the Committee of Ministers on 21 September 2011 at the 1121st meeting of the Ministers' Deputies.

Committee of Ministers. 2012. Recommendation CM/Rec(2012)3 on the protection of human rights with regard to search engines. Adopted by the Committee of Ministers on 4 April 2012 at the 1139th meeting of the Ministers' Deputies.

Committee of Ministers. 2012. Recommendation CM/Rec(2012)4 on the protection of human rights with regard to social networking services. Adopted by the Committee of Ministers on 4 April 2012 at the 1139th meeting of the Ministers' Deputies.

Committee of Ministers. 2014. Recommendation CM/Rec(2014)7 on the protection of whistleblowers. Adopted by the Committee of Ministers on 30 April 2014, at the 1198th meeting of the Ministers' Deputies.

Committee of Ministers. 2016. Recommendation CM/Rec(2016)1 on protecting and promoting the right to freedom of expression and the right to private life with regard to network neutrality. Adopted by the Committee of Ministers on 13 January 2016, at the 1244th meeting of the Ministers' Deputies.

Committee of Ministers. 2016. Recommendation CM/Rec(2016)4 of the Committee of Ministers to member States on the protection of journalism and safety of journalists and other media actors. Adopted by the Committee of Ministers on 13 April 2016 at the 1253rd meeting of the Ministers' Deputies.

Committee of Ministers. 2016. Recommendation CM/Rec(2016)5 of the Committee of Ministers to member States on Internet freedom. Adopted by the Committee of Ministers on 13 April 2016 at the 1253rd meeting of the Ministers' Deputies.

Committee of Ministers. 2017. Recommendation CM/Rec(2017)5 of the Committee of Ministers to member States on standards for e-voting. Adopted by the Committee of Ministers on 14 June 2017 at the 1289th meeting of the Ministers' Deputies.

Committee of Ministers. 2018. Recommendation CM/Rec(2018)1 of the Committee of Ministers to member States on media pluralism and transparency of media ownership. Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers' Deputies.

Committee of Ministers. 2018. Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries. Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers' Deputies.

Committee of Ministers. 2018. Recommendation CM/Rec(2018)4 of the Committee of Ministers to member States on the participation of citizens in local public life. Adopted by the Committee of Ministers on 21 March 2018 at the 1311th meeting of the Ministers' Deputies.

Committee of Ministers. 2019. Recommendation CM/Rec(2019)2 of the Committee of Ministers to member States on the protection of health-related data. Adopted by the Committee of Ministers on 27 March 2019 at the 1342nd meeting of the Ministers' Deputies.

Committee of Ministers. 2019. Recommendation CM/Rec(2019)3 of the Committee of Ministers to member States on supervision of local authorities' activities. Adopted by the Committee of Ministers on 4 April 2019 at the 1343rd meeting of the Ministers' Deputies.

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. Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies.

Council of Bars & Law Societies of Europe. 2020. CCBE Considerations on the Legal Aspects of Artificial Intelligence. Brussels.

Council of Europe - Venice Commission, OSCE/ODIHR. 2011. Joint Guidelines on Political Party Regulation. https://www.osce.org/odihr/77812, accessed 20.12.2019.

Council of Europe, Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data. 2019. Guidelines on Artificial Intelligence and Data Protection. T-PD(2019)01. https://rm.coe.int/2018-lignes-directrices-sur-l-intelligence-artificielle-et-la-protecti/168098e1b7, accessed 16.11.2019.

Council of Europe, Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data. 2017. Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data. T-PD(2017)1. http://rm.coe.int/t-pd-2017-1-bigdataguidelines-en/16806f06d0, accessed 16.11.2019

Council of Europe, Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data. 2019. Profiling and Convention 108+: Suggestions for an update. T-PD(2019)07BISrev.

Council of Europe, Directorate General of Democracy – European Committee on Democracy and Governance. 2016. The Compendium of the most relevant Council of Europe texts in the area of democracy http://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806b5f2c, accessed 18.11.2019.

Council of Europe, Directorate General of democracy and Political Affairs – Directorate of Democratic Institutions. 2011. Guidelines on transparency of e-enabled elections.

Council of Europe, Directorate General of Democracy and Political Affairs – Directorate of Democratic Institutions. 2009. Project «Good Governance in the Information Society», CM(2009)9.

Council of Europe, Parliamentary Assembly. 2019. Resolution 2254 (2019)1. Media freedom as a condition for democratic elections.

Council of Europe. 2008. The 12 Principles of Good Governance https://www.coe.int/en/web/good-governance/12-principles#{%2225565951%22:[0]}, accessed 16.03.2020.

Council of Europe. 2009. Additional Protocol to the European Charter of Local Self-Government on the right to participate in the affairs of a local authority, Utrecht, 16.XI.2009.

Council of Europe. 2017. Guidelines for civil participation in political decision making, CM(2017)83-final. Adopted by the Committee of Ministers on 27 September 2017 at the 1295th meeting of the Ministers' Deputies.

Council of Europe. 2019. Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes. Adopted by the Committee of Ministers on 13 February 2019 at the 1337th meeting of the Ministers' Deputies.

Council of Europe. 2019. Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (Adopted by the Committee of Ministers on 13 February 2019 at the 1337th meeting of the Ministers' Deputies).

Council of Europe. Guidelines for civil participation in political decision making. CM(2017)83-final. https://www.coe.int/en/web/youth/-/guidelines-for-civil-participation-in-political-decision-making, accessed 15.03.2020.

Council of Europe-Committee of experts on internet intermediaries (MSI-NET). 2018. Algorithms and Human Rights. 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 28.11.2019.

Crawford, K., and Joler, V. 2018. Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources. AI Now Institute and Share Lab. http://www.anatomyof.ai, accessed 27.12.2019.

Cummings, M. L. , Roff H. M., Cukier K., Parakilas J. and Bryce H. 2018. Chatham House Report. Artificial Intelligence and International Affairs Disruption Anticipated. London: Chatham House. The Royal Institute of International Affairs. https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial-intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf, accessed 21.03.2020.

Dupont, B. et al. 2018. Artificial Intelligence in the Context of Crime and Criminal Justice. Korean Institute of Criminology 2018. https://www.cyberjustice.ca/publications/lintelligence-artificielle-dans-le-contexte-de-la-criminalite-et-de-la-justice-penale/, accessed 30.05.2020.

EU Code of Practice on Disinformation, 2018 https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation, 23.03.2020.

European Commission for Democracy Through Law (Venice Commission). 2002. Code of Good Practice in Electoral Matters. Guidelines and Explanatory Report. Adopted by the Venice Commission at its 51st and 52nd sessions (Venice, 5-6 July and 18-19 October 2002).

European commission for Democracy trough law (Venice Commission). 2019. Joint Report of the Venice Commission and of the Directorate of Information society and Actions Against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital Technologies and Elections.

European Commission for the Efficiency of Justice (CEPEJ). 2018. European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment. Adopted by the CEPEJ during Its 31st Plenary Meeting (Strasbourg, 3-4 December 2018) https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c, accessed 04.12.2018.

European Commission, Networks, Content and Technology- Directorate-General for Communication. 2018. A Multi-Dimensional Approach to Disinformation Report of the Independent High Level Group on Fake News and Online Disinformation.

https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation, accessed 22.03.2018.

European Commission. 2014. Green paper on mobile health. http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=5147, accessed 12.01.2020.

European Commission. 2020. A European strategy for data, COM(2020) 66 final. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en, accessed 20.02.2020.

European Commission. 2020. 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/publications/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics-0_en, accessed 20.02.2020.

European Commission. 2020. White Paper on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en, accessed 20.02.2020.

European Union Agency for Fundamental Rights. 2018. #BigData: Discrimination in Data-Supported Decision Making. https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making, accessed 20.05.2020.

European Union Agency for Fundamental Rights. 2018. Preventing Unlawful Profiling Today and in the Future: A Guide. https://fra.europa.eu/en/publication/2018/preventing-unlawful-profiling-today-and-future-guide, accessed 20.05.2020.

Faye Jacobsen, A. 2013. The Right to Public Participation. A Human Rights Law Update. Issue Paper. The Danish Institute for Human Rights.

Ferryman, K. and Pitcan, M. 2018. Fairness in Precision Medicine. Data & Society, February, https://datasociety.net/wp-content/uploads/2018/02/Data.Society.Fairness.In_.Precision.Medicine.Feb2018.FINAL-2.26.18.pdf, accessed 20.12.2019.

González Fuster, G. 2014. The Emergence of Personal Data Protection as a Fundamental Right of the EU. Cham-New York: Springer International Publishing.

Maisley, N. 2017. The International Right of Rights? Article 25(a) of the ICCPR as a Human Right to Take Part in International Law-Making. 28 European Journal of International Law 89.

Manheim, K., Kaplan, L. 2019. Artificial Intelligence: Risks to Privacy and Democracy. 21 Yale J.L. & Tech. 106.

Mantelero, A. 2018. AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment. 34 Computer Law & Security Review 754.

Mantelero, A. 2019. Artificial Intelligence and Data Protection: Challenges and Possible Remedies. Report on Artificial Intelligence. Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of personal data: Strasbourg. T-PD(2018)09Rev, https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6, accessed 20.02.2020.

Mayer-Schönberger, V. 1997. Generational development of data protection in Europe? In Agre, P.E., Rotenberg, M. (eds). Technology and privacy: The new landscape. Cambridge, MA: MIT Press.

Mehr, H. 2017. Artificial Intelligence for Citizen Services and Government. Harvard Kennedy School. Ash Center for Democracy and Innovation.

Meijer, A., Wessels, M. 2019. Predictive Policing: Review of Benefits and Drawbacks. 42 International Journal of Public Administration 1031.

Nunez, C. 2017. Artificial Intelligence and Legal Ethics: Whether AI Lawyers Can Make Ethical Decisions. 20 Tul. J. Tech. & Intell. Prop. 189-204.

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.

OSCE. 2020. Non-Paper on the Impact of Artificial Intelligence on Freedom of Expression. https://www.osce.org/representative-on-freedom-of-media/447829, accessed 11.06.2020.

Osoba, O.A., Welser, W. 2017. An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1744.html, accessed 20.05.2020.

Parliamentary Assembly. 2019. Resolution 2254 (2019)1. Media freedom as a condition for democratic elections.

Parliamentary Assembly. 2019. Resolution 2300 (2019)1, Improving the protection of whistle-blowers all over Europe

Perry, W.L. et al. 2013. Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations. RAND Corporation 2013. https://www.rand.org/pubs/research_reports/RR233.html, accessed 30.03.2020.

Privacy International. 2017. Smart Cities: Utopian Vision, Dystopian Reality. https://privacyinternational.org/report/638/smart-cities-utopian-vision-dystopian-reality, accessed 21.03.2020.

Re, R.M., Solow-Niederman, A. 2019. Developing Artificially Intelligent Justice. 22 Stan. Tech. L. Rev. 242.

Rosenbaum, D., 2006. The limits of hot spots policing. In: D. Weisburd and A. Braga, eds. Police innovation: contrasting perspectives. New York, NY: Cambridge University Press, 245–266.

Rouvroy, A. 2016. “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: Strasbourg. T-PD-BUR(2015)09Rev, http://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020, accessed 04.11.2019.

Savaget, P., Chiarini, T., Evans, S. 2019. Empowering Political Participation through Artificial Intelligence. 46 Science and Public Policy 369.

Seatzu, F. 2015. The Experience of the European Court of Human Rights with the European Convention on Human Rights and Biomedicine. 31(81) Utrecht Journal of International and European Law 5, DOI: http://dx.doi.org/10.5334/ujiel.da, accessed 07.12.2019.

Sunstein, C.R. 2015a. The Ethics of Nudging. Yale Journal on Regulation 32: 413

Sunstein, C.R. 2015b. Why Nudge? The Politics of Libertarian Paternalism. New Haven: Yale University Press.

Sunstein, C.R., Thaler, R. 2003. Libertarian Paternalism in Not an Oxymoron. University of Chicago Law Review 70 (4): 1159.

Thaler, R., Sunstein, C.R. 2008. Nudge. New Haven, CT: Yale University Press.

The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. 2017. Joint Declaration on "Fake News," Disinformation and Propaganda. http://www.osce.org/fom/302796?download=true, accessed 02.02.2020.

T-PD(2019)07BISrev. Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data. 2019. Profiling and Convention 108+: Suggestions for an update

UN Office of the High Commissioner for Human Rights. 1981. CESCR General Comment No. 1: Reporting by States Parties. Adopted at the Thirteenth Session of the Committee on Economic, Social and Cultural Rights, on 27 July 1981 (Contained in Document E/1989/22).

UN Office of the High Commissioner for Human Rights. 1996. General Comment No. 25: The right to participate in public affairs, voting rights and the right of equal access to public service (Art. 25). CCPR/C/21/Rev.1/Add.7.

UNESCO. 2019. Preliminary Study on a Possible Standard-Setting Instrument on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000369455, accessed 21.11.2019.

UNESCO. Declaration on the Human Genome and Human Rights (11 November 1997).

United Nations Convention against Corruption, 2003.

van Brakel, R., De Hert, P. 2011. Policing, Surveillance and Law in a Pre-Crime Society: Understanding the Consequences of Technology Based Strategies. 3 Cahiers Politiestudies, Jaargang 163.

Verbeek, P-P. 2011. Understanding and Designing the Morality of Things. Chicago-London: The University of Chicago Press.

Završnik, A. 2019. Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings. European Journal of Criminology, 1-20, https://doi.org/10.1177/1477370819876762, accessed 20.02.2020.

Annex 1 – Legal instruments

Binding instruments	Related non-binding instruments
Biomedicine	
Council of Europe Convention on Human Rights and Biomedicine ('Oviedo Convention') Additional Protocol concerning Genetic Testing for Health Purposes Additional Protocol concerning Biomedical Research	Rec(2016)8 on the processing of personal health-related data for insurance purposes, including data resulting from genetic tests and its Explanatory Memorandum Recommendation CM/Rec(2016)6 of the Committee of Ministers to member States on research on biological materials of human origin Strategic Action Plan on Human Rights and Technologies in Biomedicine 2020-2025
Antidiscrimination	
- Universal Declaration of Human Rights - International Covenant on Civil and Political Rights International Covenant on Economic, Social and Cultural Rights - International Convention on the Elimination of All Forms of Racial Discrimination - Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) - Convention on the Rights of Persons with Disabilities - European Convention on Human Rights (ECHR) and its Protocols (No.12 in particular) - European Social Charter - Convention on Cybercrime and its Additional Protocol concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems - Convention on preventing and combating violence against women and domestic violence (Istanbul Convention) - Charter of the Fundamental Rights of the European Union	ECRI's General Policy Recommendations, no. 2 (on equality bodies), 11 (on combating racial discrimination in policing) and 15 (on hate speech) in particular. PACE Recommendation 2098 (2017) on Ending cyberdiscrimination and online hate CM Recommendation (2019)1 on Preventing and Combating Sexism
Cybercrime and electronic evidence	
Convention on Cybercrime	Guidance Notes by the Cybercrime Convention Committee on DDOS attacks, Critical information infrastructure attacks, Malware, Spam, Identity theft etc.
Justice	
- Universal Declaration of Human Rights - International Covenant on Civil and Political Rights	CEPEJ. 2019. European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment

- International Convention on the Elimination of All Forms of Racial Discrimination - Convention on the Elimination of All Forms of Discrimination against Women - Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR) - Charter of Fundamental Rights of the European Union	
Congress of Local and Regional Authorities	
The European Charter of Local Self-Government	Congress Resolution 435 (2018) and Recommendation 424 (2018) "Transparency and open government." Congress Resolution 417 (2017) and Recommendation 398 (2017) "Open data for better public services. Congress Resolution 394 (2015) E-media: game changer for local and regional politicians. Congress Resolution 290 (2009) E-democracy: opportunities and risks for local authorities.
Democracy and participation	
- Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR) - Convention on the protection of individuals with regard to automatic processing of personal data ETS No. 108 of 1981 and the 2018 Protocol modernising the Convention	- Committee of Ministers Recommendation Rec(2003)4 on common rules against corruption in the funding of political parties and electoral campaigns - Code of Good Practice in Electoral Matters (Venice Commission) - Joint Guidelines on Political Party Regulation (Venice Commission and OSCE/ODIHR) - Recommendation CM/Rec(2007)15 of the Committee of Ministers to member states on measures concerning media coverage of electoral campaigns - see also Recommendation CM/Rec(2018)1 on media pluralism and transparency of media ownership, Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries, Recommendation CM/Rec(2016)1 on protecting and promoting the right to freedom of expression and the right to private life with regard to network - 1999 Committee of Ministers Recommendation No. R (99) 5 for the protection of privacy on the internet, 2010 Recommendation CM/Rec(2010)13 on the protection of individuals with regard to automatic processing of personal data in the context of profiling, Recommendation CM/Rec(2012)3 on the protection of human rights with regard to search engines,

	Recommendation CM/Rec(2012)4 on the protection of human rights with regard to social networking services
Freedom of expression	
- European Convention on Human Rights - International Covenant on Civil and Political Rights - Charter of Fundamental Rights of the European Union	UDHR CM/Rec(2018)2 on roles and responsibilities of internet intermediaries CM/Rec(2020)x on the human rights impacts of algorithmic systems Decl(13/02/2019) on the manipulative capabilities of algorithmic processes CM/Rec(2018)1 on media pluralism and transparency of media ownership CM/Rec(2020)x on promoting a favourable environment for quality journalism in the digital age
Elections	
Universal Declaration on Human Rights International Covenant on Civil and Political Rights United Nations Convention on the Elimination of All Forms of Racial Discrimination United Nations Convention on the Elimination of All Forms of Discrimination against Women United Nations Convention on the Rights of Persons with Disabilities United Nations Convention against Corruption Convention for the Protection of Human Rights and Fundamental Freedoms (ETS No. 5) Protocol to the Convention for the Protection of Human Rights and Fundamental Freedoms (ETS No. 9) European Charter of Local Self-Government (ETS No. 122) European Charter for Regional or Minority Languages (ETS No. 148) Convention on Cybercrime (ETS No. 185) Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (ETS No. 108) Additional Protocol to the Convention for the Protection of Individuals with Regard to	Code of Good Practice in Electoral Matters, adopted by the Council for Democratic Elections of the Council of Europe and the European Commission for Democracy through Law (Venice Commission) Recommendation Rec(2003)3 of the Committee of Ministers to member states on balanced participation of women and men in political and public decision making Convention on the Standards of Democratic Elections, Electoral Rights and Freedoms in the Member States of the Commonwealth of Independent States (CDL-EL(2006)031rev) Recommendation Rec(99)5 of the Committee of Ministers to member States on the protection of privacy on the Internet Recommendation Rec(2004)15 of the Committee of Ministers to member States on electronic governance (e-governance) Recommendation CM/Rec(2007)15 of the Committee of Ministers to member states on measures concerning media coverage of election campaigns Recommendation CM/Rec(2009)1 of the Committee of Ministers to member States on electronic democracy (e-democracy) Recommendation CM/Rec(2017)5

Automatic Processing of Personal Data, regarding supervisory authorities and transborder data flows (ETS No. 181) Charter of Fundamental Rights of the European Union Framework convention for the protection of national minorities and explanatory report	of the Committee of Ministers to member States on standards for e-voting Document of the Copenhagen Meeting of the Conference on the Human Dimension of the OSCE Report on the misuse of administrative resources during electoral processes adopted by the Council for Democratic Elections and by the Venice Commission (CDL-AD(2013)033) Report on electoral rules and affirmative action for national minorities' participation in decision making in European countries adopted by the Council for Democratic Elections and the Venice Commission (CDL-AD(2005)009) Code of good practice on referendum adopted by the Council for Democratic Elections and the Venice Commission (CDL-AD(2007)008rev-cor) Council of Europe Disability Strategy 2017-2023 Resolution 1897 (2012) of the PACE, Ensuring greater democracy in elections Code of Good Practice in the field of Political Parties adopted by the Venice Commission and Explanatory Report adopted by the Venice Commission (CDL-AD(2009)021)
Democracy (excluding issues relating to elections and electoral cycle)	
- Universal Declaration of Human Rights - International Covenant on Civil and Political Rights - International Convention on the Elimination of All Forms of Racial Discrimination - Convention on the Elimination of All Forms of Discrimination against Women - Charter of Fundamental Rights of the European Union - Convention 108+ - Convention for the Protection of Human Rights and Fundamental Freedoms and Protocols - European Charter of Local Self-Government	- Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes See also Compendium Chapter A (separation of powers / good governance) Chapter B (media pluralism & diversity ; protection pf freedom of expression on the Internet) Chapter C (enabling civil society) Chapter E (citizen's participation)

- Framework Convention for the Protection of National Minorities	
Good Governance	
- Universal Declaration of Human Rights - International Covenant on Civil and Political Rights - International Convention on the Elimination of All Forms of Racial Discrimination - Convention on the Elimination of All Forms of Discrimination against Women - Convention for the Protection of Human Rights and Fundamental Freedoms and Protocols - Charter of Fundamental Rights of the European Union - Convention 108+ - European Charter of Local Self-Government and Protocols - Council of Europe Convention on Access to Official Documents	-12 principles of good democratic governance - Recommendation of the Committee of Ministers to member States on supervision of local authorities' activities CM/Rec(2019)3 See also Compendium Chapter A (good governance) Chapter E (Integration policies – standards and mechanisms) And see https://www.coe.int/en/web/good-governance/conventions-recommendations
Gender equality including violence against women	
- Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR) - Council of Europe Convention of Preventing and Combating Violence against Women (article 17§1 on the participation of the ICT sector in the prevention & fight against violence against women, Article 34 cyber stalking) -European Social Charter - UN Convention on the Elimination of All Forms of Discrimination against Women - Universal Declaration of Human Rights - International Covenant on Civil and Political Rights International Covenant on Economic, Social and Cultural Rights - International Convention on the Elimination of All Forms of Racial Discrimination - Charter of Fundamental Rights of the European Union	CM Recommendation (2019)1 on Preventing and Combating Sexism CM Recommendation (2013)1 on gender equality and media ECRI's General Policy Recommendations, no. 15 on hate speech
Culture, Creativity and Heritage	
Universal Declaration of Human Rights International Covenant on Economic, Social and Cultural Rights Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR)	
EFCNM	Numerous CoE/CM and PACE and Congress RECs and Resolutions on issues of cultural identity, diversity

European Charter for Regional and Minority Languages[158]	Numerous CoE/CM and PACE and Congress RECs and Resolutions on issues of cultural identity, diversity and dialogue and minorities
CoE Conventions in the Cultural Heritage Sector (Nicosia Convention =not yet in force; Faro Convention; La Valetta Convention; Granada Convention)	Numerous CoE/CM and PACE RECs on issues of Cultural heritage
UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expressions	CoE Declaration on Cultural diversity CoE CM Rec on the UNESCO Convention
Council of Europe Convention on Cinematographic Co-production (revised) EU's Audiovisual Media Services Directive (AVMSD) / Directive (EU) 2018/1808 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC	Resolution (88)15 amended setting up a European Support Fund for the Co-production and Distribution of Creative Cinematographic and Audiovisual Works ("Eurimages") Recommendation CM/Rec(2017)9 of the Committee of Minister to member States on gender equality in the audiovisual sector
Universal Declaration of Human Rights International Covenant on Economic, Social and Cultural Rights Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR)	
GRECO	
Criminal and Civil Law Conventions on Corruption; GRECO monitoring	CM recommendations on model code of conduct for public officials; lobbying whistleblower protection; transparency of political party funding, etc…
Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR)	Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR)
ESC	
European Social Charter (1961 Charter, 1988 Protocol and 1996 revised Charter)	
European Social Charter rights more specifically	Some examples: - New Strategy and Council of Europe Action Plan for Social Cohesion (approved by CM on 7 July 2010); - CM Rec(2000)3, proposing an individual universal and enforceable right to the satisfaction of basic material needs; - etc.
In addition, there are many social rights (and Charter) aspects related to subjects covered by a wide range of other areas of CoE work:	Some examples: - CM Rec(93)1 on effective access to the law and to justice to the very poor; - social rights aspects of the prison rules (health care, living conditions, employment, education, family rights,…); - etc.

[158] Study forthcoming (spring 2020) on how AI will impact the areas covered by the European Charter for Regional and Minority Languages.

Annex 2- Impacted areas

Impacted areas (applications)
Biomedicine • AI-based surveillance, prevention, diagnosis and intervention in healthcare settings • Prediction-based surveillance, diagnosis, monitoring, financing (insurance) treatments (e.g. user facing apps and online services beyond healthcare settings)
Antidiscrimination • Automated Decision-making covering different areas in both public and private sectors (e.g. job applications, welfare/social benefits, access to goods and services, such as bank loans, insurance) • Predictive policing (which holds high risk of racial profiling) • Predictive justice • Facial recognition • Behavioural prediction technologies such as emotional recognition and AI-based lie detection • Personal assistance tools (e.g. Siri) • Content moderation • Data protection
Cybercrime and electronic evidence • Automated cybercrime and cyberattacks, such as: - Distributed denial of service (DDOS) attacks - Criticial information infrastructure attacks - Man-in-the-middle attacks - Phishing and similar social engineering techniques - Scanning for vulnerabilities - Etc. • Cybercrime investigations and computer forensics: - Collection and analysis of electronic evidence (in relation to any crime). - Attribution - Reverse engineering • Cybersecurity and prevention of cybercrime: - Detection of malware, intrusions, etc. - Automated patching of vulnerabilities
Justice sector • Processing of judicial decisions and data: - to support judicial decision-making or judicial research) - On-line dispute resolution - Provision of legal advice to litigants • Predictive policing
Congress of Local and Regional Authorities • Provision of local public services. • Instruments to promote citizen participation. • Wide variety of digital and electronic applications in cities and local communities. • Application of information and communication technologies (ICT) to improve the quality of life and working environments in cities. • Smart city-governance. The embedding of ICT within government systems.

- Local roll-out of practices that bring people and ICT together in order to foster innovation and enhance the knowledge that they offer.

Freedom of expression

- Individual communication (through automated content moderation and restriction – algorithmic sorting, classification, optimisation and recommender systems)
- Media production and distribution (robo-journalism, data-journalism, NLP, micro-targeting of reader-base, automated newsfeeds based on reader profile)
- Societal and political communication/ fragmentation/polarisation of public discourse, political redlining (micro-targeting of voterbase, opinion swaying through bots, proliferation of automated local media sites)

Elections

Pre-electoral period:
- Planification of electoral calendar
- Training of electoral stakeholders
- Delimitation of electoral constituency
- Registration of voters and candidates
- Accreditation of observers (international and domestic)
- Update of the list of voters
- Update of legal framework
- Financing of political parties
- Electoral propaganda by administration and by political parties/candidates

Electoral period:
- Financing of electoral campaigns
- Access to media
- Voting
- Counting of ballots
- Tabulation of results

Post-election period:
- Publication of electoral results
- Electoral dispute resolution

Democracy (excluding issues relating to elections and electoral cycle)

- Separation of power
- Civil society participation
- Citizen's participation
- Privacy
- Citizenship
- Protection of minorities
- Pluralism & diversity
- Legitimacy

Good Governance
- Local governments
- Regional
- Administration
- Service delivery
- Budgetary allocation
- Social security and social benefit systems
- Police and judiciary
- Smart cities
- Public tender and procurement
- Institutional capacities

Gender equality including violence against women (VAW)

General issue of inherited gender bias from the data systems algorithms train on (valid for many areas), which may lead to aggravated gender and social inequalities.

General issues related to AI as an employment sector:

- The lack of participation /under-representation of women exacerbates the potential gender biases and excludes them from a powerful sector
- Exploitation of "click workers" in Europe and worldwide (low salaries, no social protection, no labour rights , long term exposition to damaging content for content moderators etc.)

Specific challenges

- Discriminatory job screening
- Automated decision-making for public and private services
- Facial & speech recognition (performing worse for women, especially some groups)
- Surveillance /stalking facilitated by AI tools ex in the context of domestic violence
- Automated decision-making exacerbating the possibility for multiple discrimination based on sex/gender, race and social origin by combining secondary data like level of education, address, level of income.
- Predictive justice (ex VAW)
- Predictive health based on gender-biased data (ex some diseases characterised as "female" or "male")
- Inherited biases in machine-led content moderation (high tolerance for sexism, sexist hate speech & VAW)
- Gendered virtual assistants / robots perpetuating gender stereotypes
- Gendered marketing perpetuating gender stereotypes
- Differential pricing based on sex/gender

Positive impacts

- Use of GPS tracking devices to ensure respect of protection orders in cases of VAW
- Use of AI by law enforcement agencies to conduct risk assessment in DV cases
- Use of AI to identify and track gender bias and being able to quarantine or eliminate the spreading of (sexist) hate speech on platforms
- Developments of Apps to support and inform victims of VAW
- Use of AI-based tools to analyse content and track gender bias / analyse representation (ex in movies or other media)

Culture, Creativity and Heritage

- Access and participation in public / cultural life;
- FoE (incl. freedom of artistic expression)
- Access to impartial information?
- Automated decision making, targeting, profiling
- Automated decision making, targeting, profiling;
- But also learning of endangered languages to preserve/ protect them
- Automated assistance in administration, health etc. for speakers from minority groups/ languages
- Geolocalisation, Predictive policing, criminal analytics (re destruction, looting, trafficking of cultural property; targeting; learning re endangered heritage can help with its protection
- Automated creation of content, targeting, profiling (re cultural creation, exchange, consumption)
- Audiovisual content development & production:
  - Predictive audience analysis
  - Automated script analysis
  - Assisted or automated script writing
  - Computer Generated Images (SFX, Animation...)
  - Automated location scouting, scheduling and budgeting (impact yet to be assessed)
- Content distribution

- Recommendation algorithms
- Targeted advertising
- Automated control of content (compliance with regulations) / Censorship (ref. Study "Entering the new paradigm of artificial intelligence and series" commissioned by DG2 and Eurimages)
- Access and participation in public / cultural life;
- FoE (incl. freedom of artistic expression)
- Access to impartial information?

GRECO

- Anti-corruption
- Criminal liability related to the use
- of automated vehicles
- Article 8: Right to respect for private and family life

ESC

All areas of social rights, social security, social cohesion, etc.
Including, but not limited to:

- many aspects of employment (including but not limited to monitoring and surveillance, job screening and work in the platform economy, etc);
- ditto different aspects of health (the right to enjoy the highest standard of health attainable);
- ditto education;
- equally for social protection, integration and participation;
- let alone non-discrimination;
- housing and protection from social exclusion;

For example:

- justice (both as regards the administration of justice, and criminal justice and prisons;
- trafficking in human beings (forced labour and exploitation, …);
- migration and refugees;
- gender equality, plus violence against women;
- children and youth, plus education;
- bioethics;
- non-discrimination, Roma and Travellers, SOGI ;
- drug policy;
- participation and culture;
- sport;

Annex 3- Principles

Guiding principles and legal values	Missing principles
Biomedicine	
Primacy of the human being Privacy and confidentiality Informed consent Autonomy Non-discrimination Non-maleficence/beneficence Accountability Transparency and Equitable Access Public debate	Precautionary principle Human control/oversight Explainability Liability for AI-based decision making Gender equality/equity
Antidiscrimination	
Non-discrimination and equality Diversity and inclusion Intersectionality Right to an effective remedy Right to a fair trial Right to privacy Presumption of innocence and burden of proof Transparency Impartiality Fairness Human control/oversight Access to digital skills	Explainability of AI systems Inclusiveness in design, development and deployment of AI systems
Cybercrime and electronic evidence	
Specific conduct to be criminalised. Specified data in specific criminal investigations to be secured for use as evidence. Effective powers to secure electronic evidence limited by the rule of law conditions and safeguards.	Problem of evidence in the cloud versus territorial enforcement jurisdiction for criminal justice (to be addressed in the 2nd Additional Protocol to the Budapest Convention).
Justice sector	
Non-discrimination Data quality & security Transparency Impartiality Fairness Freedom of choice/ Independence of judges (decision-making process) Human control/oversight Guarantees of the right of access to the judge Guarantees of the right to a fair trial	Precautionary principle for applications missing fundamental transparency requirements
Congress	
Transparency Human control (oversight) Impartiality Right to privacy Data security	Democracy and participation – Deep fakes, Microtargeting and propaganda in the framework of electoral processes

Cyber security Non-discrimination Inclusive cities Financial sustainability Monitoring safety Service efficiency Digital literacy	
Democracy and participation	
Right to free elections Freedom of expression Right of individuals to access the internet Right to private life; Data protection Equality of opportunity for parties and candidates Requirement of a neutral attitude by state authorities with regard to the election campaign, to coverage by the media, and to public funding of parties and campaigns Requirement of a minimum access to privately owned audio-visual media, with regard to the election campaign and to advertising, for all participants in elections Transparency in campaign funding Prevention of improper influence on political decisions through financial donations Responsible, accurate and fair media coverage of electoral campaigns; right of reply, modalities of disseminating opinion polls, transparency requirements on paid advertising content; media pluralism Network neutrality Protection of individuals with regard to the collection and processing of personal data on information highways Non-discrimination Data quality & security Transparency Impartiality Fairness Freedom of choice/ Independence of judges (decision-making process) Human control/oversight Guarantees of the right of access to the judge Guarantees of the right to a fair trial	Balance between sometimes conflicting rights such as e.g. - right to free elections / freedom of expression - right of access to information including on the internet / right to private life, data protection Standards which would be applicable and adequate for digital advertising/campaigns, e.g. with respect to - equality of opportunity for parties and candidates - election campaign and campaign funding, transparency and enforcement - fair media coverage, media pluralism - accountability of internet intermediaries in terms of transparency and access to data enhancing transparency of spending, specifically for political advertising - net neutrality - data protection
Freedom of expression	
Individual autonomy Equality Democratic security Transparency and accountability Independence of the media Diversity and pluralism	Precautionary principle for applications missing fundamental transparency requirements
Elections	

• Free and fair elections • Freedom of choice/opinion/speech • Universal suffrage • Equal suffrage • Free suffrage • Secret suffrage • Direct suffrage • Frequency of elections • Transparency of electoral process • Inclusiveness of electoral process • Gender balanced participation/representation in public decision-making	• Principle of use of AI systems in electoral processes (especially e-voting systems, etc.) • Opportunities offered by AI to have more inclusive electoral processes (AI as tool for the Electoral Management Bodies and election commissions, AI as an assistant for the voters).
Democracy (excluding issues relating to elections and electoral cycle)	
Transparency Impartiality Fairness Freedom of choice Freedom of expression Freedom of assembly and association Access to information Human control/oversight Diversity Equality Non-discrimination Data quality & security Data protection Independence	- Role of intermediaries - Tech & digital literacy - Question of who owns the data - Democratic oversight - Open data and open government - Risk assessment
Good Governance	
- Non-discrimination - Data quality & security - Impartiality - Fairness - Participation, Representation Fair Conduct of Elections - Responsiveness - Efficiency and Effectiveness - Openness and Transparency - Rule of Law - Ethical Conduct - Competence and Capacity - Innovation and Openness to Change - Sustainability and Long-term Orientation - Sound Financial Management - Human rights, Cultural Diversity and Social Cohesion - Accountability - Redress mechanisms - Access to remedy - Independence	- Democratic oversight - Access to remedy and redress mechanisms in case of automated and algorithmic decisions making by public officials - Role of intermediaries - Tech literacy & competences - Questions of who actually owns the data - Open data and open government - Civil and criminal liability - Risk assessments and risk management
Gender equality including violence against women	
Equality and non-discrimination Integrity / Elimination of violence (against women) Equal access to justice Guarantees of the right to a fair trial and to redress	Un-biased data (Gender) inclusiveness of AI as a sector AI as an employment sector respecting labour and social rights Data quality & security Transparency & explainability Accountability Impartiality Fairness

	Human control/oversight Digital literacy and closing existing digital (gender) gaps, essential with regards to right to redress – if citizens & consumers do not understand AI, they will not be able to claim their rights Precautionary principle for applications missing fundamental transparency requirements Ethical principles such as "do no harm" are not respected because some of the spyware apps are developed and advertised for the sole purpose of "knowing what your wife is up to".
Culture, Creativity and Heritage	
Non-discrimination Access, Freedom of Association, Right to participate in cultural life and create and learn (Covenant) Freedom of Expression Access to impartial information	Precautionary principle for applications missing fundamental transparency requirements Need to develop cultural paradigms and techniques to deal with Autonomization (only exist for Automatisation) "Avoid further centralisation of knowledge and power in the hands of those, who already have it and further dis-empower those who don't" (M. Whitaker) Need to stress rules and rights on access to common goods, and to participate in public life (citizen-centred practices)
Non-discrimination – Impartiality (Protection of National Minorities)	
Non-discrimination – Impartiality Protection of Minorities and their cultural expressions (languages / linguistic diversity, cultural heritage)[159]	Ownership and possible bias of information fed into AI-driven learning applications
Promote/ protect European identity, diversity, co-operation Access to and participation in cultural heritage; protection of cultural heritage	Protection of Human creativity (distinctive nature of human creativity)
Human control/oversight over creative process, transparency Protection and Promotion of cultural diversity Creating conditions for culture to flourish and freely interact Recognise the distinctive nature of cultural activities, goods and services as vehicles of identity, values and meaning	IP and copyright management Protection of Human Creativity (distinctive nature of human creativity)
Cultural diversity Cultural cooperation in Europe and beyond Availability of works Non-Discrimination Data protection Freedom of expression and of creation	Visibility of works Transparency of decision-making (to develop and produce / to censor / to recommend a work) IP ownership, Copyright and moral rights issues

[159] An AI-Language Recreation Machine could fill gaps and help develop a living global language archive, a "Louvre of Languages".

Human control/oversight, transparency	
Non-discrimination Access, Freedom of Association, Right to participate in cultural life and create and learn (Covenant) Freedom of Expression Access to impartial information	Precautionary principle for applications missing fundamental transparency requirements Need to develop cultural paradigms and techniques to deal with Autonomization (only exist for Automatisation) “Avoid further centralisation of knowledge and power in the hands of those, who already have it and further dis-empower those who don’t” (M. Whitaker) Need to stress rules and rights on access to common goods, and to participate in public life (citizen-centred practices)
GRECO	
Guiding Principles for the Fight against Corruption	Nothing specific on: AI applications to prevent corruption; need to make sure algorithm are not corrupted
	Ongoing work by the CDPC on Criminal liability related to the use of automated vehicles
Article 8: Right to respect for private and family life	Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR)
ESC	
Various broad principles emerge from the Charter and the monitoring activities under the Charter about transparency and participation in decision-making	Automated or computer-assisted or AI-enabled decisions-making would require: - mandatory human oversight in order to mitigate and/or avoid errors in the management, attribution or revocation of entitlements, assistance and related benefits which could amplify disadvantage and/or disenfranchisement; - effective arrangements to protect vulnerable persons from destitution, extreme want or homelessness, and from serious injury or irreparable harm, as a result of the implementation of computer-assisted or AI-enabled decisions in the area of social services; - a proactive approach with a view to ensure that those affected by computer-assisted or AI-enabled decisions in the area of social services, in particular persons in a situation of extreme deprivation or vulnerability, can effectively assert their rights and seek remedies.

Annex 4 – Data Protection
Binding and non-binding instruments in the field of data protection

Convention 108+ Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of personal data (Convention 108). 2019. Guidelines on the data protection implications of artificial intelligence[160]	**Human control** I.6 AI applications should allow meaningful control by data subjects over the data processing and related effects on individuals and on society **Value-oriented design** II.1. AI developers, manufacturers and service providers should adopt a values-oriented approach in the design of their products and services, consistent with Convention 108+, in particular with article 10.2, and other relevant instruments of the Council of Europe. **Precautionary approach** II.2 AI developers, manufacturers and service providers should assess the possible adverse consequences of AI applications on human rights and fundamental freedoms, and, considering these consequences, adopt a precautionary approach based on appropriate risk prevention and mitigation measures. **Human rights by-design approach and bias detection** II.3 In all phases of the processing, including data collection, AI developers, manufacturers and service providers should adopt a human rights by-design approach and avoid any potential biases, including unintentional or hidden, and the risk of discrimination or other adverse impacts on the human rights and fundamental freedoms of data subjects. **Data quality and minimisation** II.4 AI developers should critically assess the quality, nature, origin and amount of personal data used, reducing unnecessary, redundant or marginal data during the development, and training phases and then monitoring the model's accuracy as it is fed with new data. The use of synthetic data may be considered as one possible solution to minimise the amount of personal data processed by AI applications. **Rick of decontextualization** II.5 The risk of adverse impacts on individuals and society due to de-contextualised data and de-contextualised algorithmic models should be adequately considered in developing and using AI applications. **Independent committees of experts** II.6 AI developers, manufacturers and service providers are encouraged to set up and consult independent committees of experts from a range of fields, as well as engage with independent academic institutions, which can contribute to designing human rights-based and ethically and socially-oriented AI applications, and to detecting potential bias. Such committees may play an especially important role in areas where transparency and stakeholder engagement can be more difficult due to competing interests and rights, such as in the fields of predictive justice, crime prevention and detection.

[160] See also T-PD(2019)01, Guidelines on Artificial Intelligence and Data Protection [GAI]; T-PD(2017)1, Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data.

	III.7 Appropriate mechanisms should be put in place to ensure the independence of the committees of experts mentioned in Section II.6. **Participation and democratic oversight on AI development** II.7 Participatory forms of risk assessment, based on the active engagement of the individuals and groups potentially affected by AI applications, should be encouraged. III. 8. Individuals, groups, and other stakeholders should be informed and actively involved in the debate on what role AI should play in shaping social dynamics, and in decision-making processes affecting them. **Human oversight** II.8 All products and services should be designed in a manner that ensures the right of individuals not to be subject to a decision significantly affecting them based solely on automated processing, without having their views taken into consideration. **Freedom of choice** II.9 In order to enhance users' trust, AI developers, manufacturers and service providers are encouraged to design their products and services in a manner that safeguards users' freedom of choice over the use of AI, by providing feasible alternatives to AI applications. **Algorithm vigilance** II.10 AI developers, manufacturers, and service providers should adopt forms of algorithm vigilance that promote the accountability of all relevant stakeholders throughout the entire life cycle of these applications, to ensure compliance with data protection and human rights law and principles. **Transparency and expandability** II.11 Data subjects should be informed if they interact with an AI application and have a right to obtain information on the reasoning underlying AI data processing operations applied to them. This should include the consequences of such reasoning. **Right to object** II.12 The right to object should be ensured in relation to processing based on technologies that influence the opinions and personal development of individuals. **Accountability and vigilance** III,2 Without prejudice to confidentiality safeguarded by law, public procurement procedures should impose on AI developers, manufacturers, and service providers specific duties of transparency, prior assessment of the impact of data processing on human rights and fundamental freedoms, and vigilance on the potential adverse effects and consequences of AI applications (hereinafter referred to as algorithm vigilance). **Freedom of human decision makers** III. 4. Overreliance on the solutions provided by AI applications and fears of challenging decisions suggested by AI applications risk altering the autonomy of human intervention in decision-making processes. The role of human intervention in decision-making processes and the freedom of human decision makers not to rely

	on the result of the recommendations provided using AI should therefore be preserved. **Prior assessment** III.5. AI developers, manufacturers, and service providers should consult supervisory authorities when AI applications have the potential to significantly impact the human rights and fundamental freedoms of data subjects. **Cooperation** III.6. Cooperation should be encouraged between data protection supervisory authorities and other bodies having competence related to AI, such as: consumer protection; competition; anti-discrimination; sector regulators and media regulatory authorities. **Digital literacy, education and professional training** III.9. Policy makers should invest resources in digital literacy and education to increase data subjects' awareness and understanding of AI applications and their effects. They should also encourage professional training for AI developers to raise awareness and understanding of the potential effects of AI on individuals and society. They should support research in human rights-oriented AI.
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	Processing of health-related data should always aim to serve the data subject or to enhance the quality and efficiency of care, and to enhance health systems where possible, while respecting individuals' fundamental rights **Interoperability** 1. […] It therefore highlights the importance of developing secure, **interoperable** information systems **Professional standards** 4.4 Data controllers and their processors who are not health professionals should only process health-related data in accordance with rules of confidentiality and security measures that ensure a level of protection equivalent to the one imposed on health **professionals**. **Consent withdrawal** 5.b Health-related data may be processed if the data subject has given their **consent**, except in cases where law provides that a ban on health-related data processing cannot be lifted solely by the data subject's consent. Where consent of the data subject to the processing of health-related data is required, in accordance with law, it should be free, specific, informed and explicit. The data subject shall be informed of their right to withdraw consent at any time and be notified that such withdrawal shall not affect the lawfulness of the processing carried out on the basis of their consent before withdrawal. It shall be **as easy to withdraw consent as it is to give it**. **Right not to know** 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. **Transparency** 11.3. Where necessary and with a view to ensuring fair and transparent processing, the information must also include: […] - the **existence of automated decisions,** including profiling, which is only permissible where prescribed by law and subject to appropriate safeguards. **Interoperability** 14.1. Interoperability may help address important needs in the health sector and may provide technical means to facilitate the updating of information or to avoid storage of identical data in multiple databases, and contribute to data portability. 14.2. It is, however, necessary for interoperability to be implemented in full compliance with the principles provided for in this Recommendation, in particular the principles of lawfulness, necessity and proportionality, and for data protection safeguards to be put in place when interoperable systems are used. 14.3. Reference frameworks based on international norms offering a technical structure which facilitates interoperability should guarantee a high level of security while providing for such interoperability. The monitoring of the implementation of such reference frameworks can be carried out through certification schemes. **Scientific research integrity** 15.10. Where a data subject withdraws from a scientific research project, their health-related data processed in the context of that research should be destroyed or anonymised in a manner which does **not compromise the scientific validity of the research** and the data subject should be informed accordingly.
Recommendation CM/Rec(2016)8 of the Committee of Ministers to the member States on the processing of personal health-related data for insurance purposes, including data resulting from genetic tests	8. The processing for insurance purposes of health-related personal data obtained in a research context involving the insured person should not be permitted.
Recommendation CM/Rec(2010)13 of the Committee of Ministers of the Council of Europe to member States	**Risk of re-identification** 8.5. Suitable measures should be introduced to guard against any possibility that the anonymous and aggregated statistical results used in profiling may result in the re-identification of the data subjects.[161]

[161] See also Convention 108+. Explanatory Report, 19 and 20 ("Data is to be considered as anonymous only as long as it is impossible to re-identify the data subject or if such re-identification would require unreasonable time, effort or resources, taking into consideration the available technology at the time of the processing and technological developments. Data that appears […] When data is made anonymous, appropriate means should be put in place to avoid re-identification of data subjects, in particular, all technical means should be implemented in order to guarantee that the individual is not, or is no longer, identifiable. They should be regularly re-evaluated in light of the fast pace of technological development").

on the protection of individuals with regard to automatic processing of personal data in the context of profiling	
UNESCO. 2019. Preliminary Study on a Possible Standard-Setting Instrument on the Ethics of Artificial Intelligence	[Principles-based approach] • Diversity, inclusion and pluralism (including a multilingual approach should be promoted) • Autonomy • Explainability • Transparency • Awareness and literacy • Responsibility • Accountability • Democracy ("AI should be developed, implemented and used in line with democratic principles") • Good governance ("Governments should provide regular reports about their use of AI in policing, intelligence, and security") • Sustainability • Human oversight • Freedom of expression (including universal access to information, the quality of journalism, and free, independent and pluralistic media, avoiding the spreading of disinformation)
OECD. 2019. Recommendation of the Council on Artificial Intelligence	[Principles-based approach] • Human-centred values and fairness • Transparency and explainability (**awareness** of the interactions with AI systems; understanding of AI outcome; enabling those adversely affected by an AI system to **challenge** its outcome based on plain and **easy-to-understand information** on the factors, and the **logic** that served as the basis for the prediction, recommendation or decision) • Robustness, security and safety (not pose **unreasonable safety risk**; **traceability**, including in relation to datasets, processes and decisions made during the AI system lifecycle; **risk management approach** to each phase of the AI system lifecycle on a continuous) • Accountability
40th International Conference of Data Protection and Privacy Commissioners. 2018. Declaration on Ethics and Data Protection in Artificial Intelligence [ICDPPC]	[Principles-based approach] • Continued attention and vigilance ("establishing demonstrable governance processes for all relevant actors, such as relying on trusted third parties or the setting up of independent **ethics committees**") • Transparency and intelligibility (**explainable AI**, algorithmic **transparency** and the **auditability** of systems, **awareness** of the interactions with AI systems; adequate information on the purpose and effects of AI systems, **overall human control**) • Risk assessment and privacy by default and privacy by design approach ("assessing and documenting the expected **impacts on individuals and society** at the beginning of an artificial intelligence project and for relevant developments during its **entire life cycle**") • **Public engagement** • **Mitigation of unlawful bias and discrimination**