Electronic Democracy and Digital Justice: Driving Principles for AI Regulation in the Prism of the Human Rights

Electronic Democracy and Digital Justice: Driving Principles for AI Regulation in the Prism of the Human Rights

ALESSANDRO MANTELERO[1]
Polytechnic University of Turin (Italy).

ABSTRACT: A growing debate in several European fora is paving the way for future rules for Artificial Intelligence (AI). A principles-based approach prevails, with various lists of principles drawn up in recent years. These lists, which are often built on human rights, are only a starting point for a future regulation. It is now necessary to move forward, turning abstract principles into a context-based response to the challenges of AI. This article therefore places the principles and operational rules of the current European and international human rights framework in the context of AI applications in two core, and little explored, areas of digital transformation: electronic democracy and digital justice. Several binding and non-binding legal instruments are available for each of these areas, but they were adopted in a pre-AI era, which affects their effectiveness in providing an adequate and specific response to the challenges of AI. Although the existing guiding principles remain valid, their application should therefore be reconsidered in the light of the social and technical changes induced by AI. To contribute to the ongoing debate on future AI regulation, this article outlines a contextualised application of the principles governing e-democracy and digital justice in view of current and future AI applications.

KEYWORDS: Human Rights; Artificial Intelligence; Electronic Democracy; Digital Justice; Regulation

SUMMARY: 1 Ai challenges and human rights; 2 Ai and electronic democracy; 2.1 Participation 2and good governance; 2.2 Elections; 3 Ai and digital justice; 3.1 Adrs and court decisions; 3.2 Crime prevention; 4 Conclusions; References.

1 AI CHALLENGES AND HUMAN RIGHTS

Artificial Intelligence (AI) is part of our daily life. It is used to moderate public debate, fashion the social environment and support human decision-makers in various fields, including justice. AI is therefore a component of many decision-making processes affecting individuals and groups, actively

1 Orcid: 0000-0001-6020-0571.

shaping our communities and personal lives[2]. This means that AI is no longer a mere technical or marketing trend but a regulatory issue[3], given the social consequences and, in some cases, legal effects.

To correctly frame this debate, it is important to keep in mind the difference between natural and artificial intelligence, where the latter is nothing more than a data-driven and mathematical form of information processing[4]. AI is not able to think, elaborate concepts or develop theories of causality: AI merely takes a path recognition approach to order huge amounts of data and infer new information and correlations.

Data dependence is both the strength and the weakness of these systems. Poor data undermines the quality of their results[5], datafication can only partially represent reality[6] and incredibly large datasets and complex AI solutions often do not allow human decision makers to inspect and check the 'reasoning' of the machine[7]. The upshot of these technical and structural constraints can be summed up under three main headings: bias, obscurity, and ownership.

Regarding bias, the design and development of AI tools can be affected by different biases that, in many cases, differ from human bias[8]. Bias does not only concern the much debated data quality (for example selection bias)[9], but also the methodologies adopted (*e.g.*, pre-processing and data cleaning biases, measurement bias, bias in survey methodologies)[10], the target of investigation (*e.g.*, historical bias in pre-existing data-sets and under- or over-representation of certain groups in new data-sets), and the psychological attitude of the data scientists (*e.g.*, confirmation bias).

This brief listing of potential biases also reveals the human component of AI solutions, often underestimated in a misleading comparison between

2 For an analysis of the different impacts of AI on individuals and society see Council of Europe, 2018c; MANTELERO-ESPOSITO, 2021; ZUIDERVEEN BORGESIUS, 2020.

3 See European Commission, 2021; Council of Europe – Ad hoc Committee on Artificial Intelligence (CAHAI), 2020; Council of Europe, 2020b; Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data (Convention 108), 2019; OECD, 2019; UNESCO, 2021. See also VERONESE-NUNES LOPES ESPIÑEIRA LEMOS, 2021.

4 See HILDEBRANDT, 2021.

5 See European Union Agency for Fundamental Rights, 2019.

6 See AGRE, 1994; HILDEBRANDT, 2019.

7 See KOLKMAN, 2020.

8 See CUMMINGS et al., 2018, 2; CARUANA et al., 2015; EYKHOLT et al., 2018.

9 See AI Now Institute, 2017, 4 and 16-17.

10 See VEALE-BINNS, 2017.

humans and machines. This dichotomy understates the role of human intervention in AI data processing[11] and the intentional or unintentional transposition of developers' views into the AI reference values used for classification[12].

As for obscurity, this concerns both the AI tools used and the way they impact on individuals, whose circumstances are analysed and represented through them. Not only is the way some AI applications actually function and process information unknown[13], even to data scientists, but individuals are often unaware of their being dynamically grouped on the basis of unseen correlations and inferences, without being able to know the identity of the other members of the group. Obscurity therefore entails two different consequences: first, data scientists are unable to clearly justify the specific decisions suggested by AI; and second, people are passively scrutinised by AI without having a meaningful or effective role in AI design or the opportunity to voice their collective interests[14].

This level of obscurity and the limitations to democratic participation in AI development is heightened by a third feature of many AI products: ownership. The proprietary nature of the algorithms used and, in certain cases, of the data silos used to train and implement them mean that intellectual property rights are a further barrier to access to the architecture of these applications and to public oversight[15].

These three inherent constraints – bias, obscurity, and ownership – have a direct impact on the challenges of AI and its social acceptance in monitoring and governing human activities (*e.g.*, smart cities),[16] offering personalised services (*e.g.*, predictive medicine)[17] and, more in general, supporting humans in the decision-making process.

Issues surrounding data-intensive solutions and their use in decision--making processes concern a variety of interests related to human rights and freedoms[18]. To address the growing concern about the potential impact of

11 See TUBARO-CASILLI-COVILLE, 2020; CRAWFORD-JOLER, 2018; LOIDEAIN-ADAMS, 2020.
12 See also WEST-WHITTAKER-CRAWFORD, 2019.
13 See SELBST, 163; BURRELL, 2016; BRAUNEIS-GOODMAN, 131.
14 See also GRABER, 2020; MANTELERO, 2016.
15 See PASQUALE, 2015, 193.
16 See also Privacy International, 2017; GOODMAN-POWLES, 2019. See also COHEN, 2019, 62-3.
17 See FERRYMAN-PITCAN, 2018.
18 See MANTELERO-ESPOSITO, 2021; Council of Europe, 2018c.

AI on human rights and freedoms, several initiatives have been proposed at local, national and international levels, and a variety of guidelines have been drawn up by NGOs, research centres and corporate entities. Several proposals have focused on ethics[19], often blurring the line between law and ethics, describing human rights and freedoms as ethical values with their 'ethicisation' and relativization.

This emphasis on the ethical dimension can entail the risk of extending to the field of data science an ethical imperialism whose effects are already known in biomedicine and the social sciences[20]. In this regard, previous experience in ethical assessment of scientific research suggests that careful consideration should be given to the distinction between ethical and legal values and the differences between ethical approaches[21]. Several documents providing guidelines on AI refer to ethics in a fairly broad and indefinite manner, with no clarification (or justification) of the ethical framework used[22].

Ethical responses to uncertainty in a rapidly changing technological and social environment may paradoxically become a new source of ambiguity. Discretionary and, in some cases, interest-based values risk weakening the legal framework or indirectly redefining it without following an appropriate procedure as required by the regulatory process[23].

Without underestimating the role of ethics in technology development, these considerations suggest a more balanced integration of law and ethics in AI regulation, based on the emphasis on the role of human rights as the universal cornerstone of the future architecture of AI regulation. From a regulatory perspective, the main challenge is to contextualise the legal principles and provisions enshrined in international human rights instruments, drafted in a pre-AI era, within the current scenario where predictive policing tools, automated digital propaganda and other new AI-based applications are reshaping many aspects of our society and human relations.

Regulatory initiatives have been proposed in several countries[24], many of them referring explicitly to all or some human rights. However, these are

19 See JOBIN-IENCA-VAYENA, 2019; HAGENDORFF, 2020.
20 See SCHRAG, 2017.
21 See HILDEBRANDT, 2021.
22 See RAAB, 2020. See also Independent High-Level Group on Artificial Intelligence, 2019.
23 See NEMITZ, 2018.
24 See GESLEY, 2019; Council of Europe, 2020a. See also MANHEIM-KAPLAN, 2019, 160.

often generic statements without a proper contextualisation of the rights and freedoms considered. Although it is relatively easy to agree on a general list of rights and freedoms that should underpin AI development, these lists do little to advance the regulatory process, since general principles, such as transparency or participation, can be interpreted in many different ways.

An effective contribution to the human rights debate in this field can therefore only come from a proper contextualisation of these guiding principles within the AI scenario. This means placing such rules, including the operational ones, in the context of the changes to society produced by AI and providing a more refined and specific formulation of the guiding principles with a view to possible future AI regulation.

This contextualisation of the guiding principles and rules can provide a more refined and elaborate formulation, taking into account the specific nature of AI products and services, and helping to better address the challenges arising from AI.

From a methodological perspective, an analysis of international legally binding instruments is the obligatory starting point in defining the existing legal framework, identifying its guiding values and verifying whether this framework and its principles properly address the issues raised by AI, with a view to preserving the harmonisation of the existing legal framework in the fields of democracy and justice.

The methodology adopted is therefore necessarily deductive, extracting the guiding principles from the variety of regulations concerning the fields in question. 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[25].

These guiding principles should be considered within the scenario of AI-driven transformation, which in many cases requires adaptation. They remain valid, but their implementation must be reconsidered in the light of the social and technical changes brought about by AI. This will deliver a more contextualised and granular application of these principles so that they can make a concrete contribution to the shape of future AI regulation.

25 See Council of Europe, 2020b.

Against this background, the following sections examine two critical areas of AI application: electronic democracy and digital justice. While in other areas, such as data protection and biomedicine, the specific nature of the sectors and recent soft-law regulatory initiatives[26] make it possible to draft some provisions for future AI regulation[27], in these two realms this is much more difficult. In addition, key principles that can be seen as guiding elements of future AI regulation, such as transparency and explainability[28], are open to varying interpretations and implementations, given the higher political significance of both democracy and justice. The analysis therefore focuses on high-level principles and their contextualisation, resulting in a more limited elaboration of key guiding provisions.

2 AI AND ELECTRONIC DEMOCRACY

Democracy covers an extremely wide array of societal and legal issues[29], most of them likely to be implemented with the support of ICT[30]. In this scenario, AI can play an important role in the present and future development of digital democracy in an information society.

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 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 related to electronic democracy: (i) participation[31] 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,

26 See for example Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data, 2019; CEPEJ, 2018.

27 See MANTELERO, 2020.

28 *E.g.*, SELBST-BAROCAS, 2018; EDWARDS-VEALE, 2017; DIAKOPOULPS, 2013. See also KAMINSKI-MALGIERI, 2020.

29 *E.g.*, Council of Europe, Directorate General of Democracy – European Committee on Democracy and Governance, 2016.

30 *E.g.*, Council of Europe Directorate General of Democracy and Political Affairs and Directorate of Democratic Institutions, 2009; Council of Europe, 2009a, Article 2.2.iii.

31 For a more detailed analysis see Faye Jacobsen, 2013. See also MAISLEY, 2017.

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.

2.1 Participation 2and good governance

The right to participate in public affairs (Article 25 Covenant) is based on a broad concept of public affairs[32], which includes public debate and dialogue between citizens and their representatives, with a close link to freedom of expression, assembly, and association[33]. 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 (for example 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.

AI-based participative platforms (*e.g.*, Consul, Citizenlab, Decidim[34]) can make a significant contribution to the democratic process, facilitating

32 See UN Human Rights Committee, 1996.

33 See also UN Committee on Economic, Social and Cultural Rights, 1981, para 5.

34 Information on these platforms is available at https://decidim.org/; https://consulproject.org/en/; https://www.citizenlab.co/. Accessed: 29 dec. 2019.

citizen interaction, prioritising of objectives, and collaborative approaches in decision-making[35] on topics of general interests at different levels (neighbourhood, municipality, metropolitan area, region, country)[36].

Specific issues arise in relation to AI tools for democratic participation (including those for preventing and fighting corruption[37]), 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 of the Council of Europe 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[38]. This principle is common to other fields, such as healthcare[39], but is a context-based notion. While in healthcare transparency is closely related to self-determination, here it 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[40]. 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 and their justification[41].

Moreover, the implementation of this notion of transparency should also consider the range of different users of these tools, adopting an accessible approach[42] 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.

35 See also Council of Europe, 2017a.

36 See also Council of Europe, 2009c.

37 See United Nations, Convention against Corruption, 2003, Article 13.

38 See Council of Europe, 2009b, para 6.

39 See Council of Europe, 1997.

40 See also Council of Europe, 2017a.

41 See Council of Europe, 2009b, para 6 and Appendix, para P.57. See also Council of Europe, 2016b, Appendix, paras 2.1.3 and 3.2. On the importance of justification see Hildebrandt, 2018b, 271-3.

42 See also Council of Europe, 2018b, Appendix, para B.IV.

Transparency and accessibility are closely related to the nature of the architecture used to build AI systems. Open source and open standards can therefore contribute to democratic oversight of the most critical AI applications[43]. There are cases where openness is affected by limitations, due to the nature of the specific AI application (for example 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[44].

In the context of AI applications to foster democratic participation, an important role can be also played by interoperability[45] 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[46], and should therefore be extended to the design of AI-based systems.

Another key principle in e-democracy is accountability. In this regard, to be accountable, AI service providers and entities using AI-based solutions for e-democracy 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[47].

Finally, given the role of media in the context of democratic participation[48], AI applications must not compromise the confidentiality and security of communications and protection of journalistic sources and whistle-blowers[49].

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, legal values such as the protection of minorities, pluralism and

43 See also Council of Europe, 2009b, para 6 and Appendix, paras G.58 and P.54.

44 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 Council of Europe, 2009b, Appendix, paras P. 55 and G.57.

45 See also Council of Europe, 2009b, Appendix, paras P. 56, G.56, 59 and 60.

46 See also Council of Europe, 2009b, para 6.

47 See also Council of Europe, 2020b.

48 See Council of Europe, 2016a, Appendix, para 2; Council of Europe Parliamentary Assembly, 2019a.

49 See also Council of Europe Parliamentary Assembly, 2019b; Council of Europe, 2014.

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[50], 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[51].

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[52]. 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[53], 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.

50 See also Council of Europe, 2020b.

51 See also Council of Europe, 2019a, para 7.

52 See Council of Europe, 2009b, paras 5 and 6, and Appendix, para G.67. See also Mantelero, 2018.

53 See MIKHAYLOV-ESTEVE-CAMPION, 2018.

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[54].

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[55], 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 they may also be unintentionally embedded into AI solutions, due to the cultural, social and gender composition of AI developer teams. For this reason, inclusiveness has an added value here, in terms of inclusion and diversity[56] in AI development.

The principles discussed for e-democracy can be repeated with regard to good governance[57]. This is the case with smart cities and sensor-based environmental management, where open, transparent and inclusive decision-making processes play a central role[58]. Similarly, the use of AI to supervise the activities of local authorities[59], for auditing and anticorruption purposes[60], 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[61]. However, in these cases, it is important to guarantee the right to know we are interacting with a machine[62] 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.

54 See also Council of Europe, 2009b, Appendix, para P. 58.

55 See also VERBEEK, 2011, 41-65.

56 See also Council of Europe, 2020b, Appendix, para 3.5.

57 See also Council of Europe, 2009b, Appendix, para P. 4; Council of Europe, 2004; Committee of Ministers of the Council of Europe, 2008.

58 See also Privacy International, 2017.

59 See also Council of Europe, 2019b, Appendix, paras 4 and 9.

60 See also SAVAGET-CHIARINI-EVANS, 2019 (discussing the Brazilian case of the 'Operação Serenata de Amor').

61 See MEHR, 2017.

62 See also Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data (Convention 108), 2019, para 2.11.

Special attention should also be paid to the potential use of AI in human-machine interaction to implement nudging strategies[63]. 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[64]. 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[65].

2.2 Elections

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[66].

63 On the use of nudging in the smart city context, see Ranchordás, 2019; Gandy-Nemorin, 2019. See generally Sunstein, 2015a and 2015b; Thaler-Sunstein, 2008; Sunstein-Thaler, 2003.

64 See also CITRON-CALO, 2021.

65 See Section 3.

66 See Council of Europe, 2018a; The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., 2017. See also Council of Europe, 2016b, Appendix, paras 1.5, 2.1 and 3; European Commission for Democracy trough Law (Venice Commission), 2019, para 151.E; Bukovska, 2020. See also Bychawska-Siniarska, 2017.

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, transparency, openness, risk management and accountability (including the adoption of certification and auditing procedures) are fundamental elements of the technological solutions adopted in these stages of the electoral process[67].

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[68].

More specific and new responses are needed in the case of propaganda and disinformation[69]. Here the existing binding and non-binding instruments do not set specific provisions, given the novelty of the disinformation based on new forms of communication, such as social networks, which differ from traditional media[70] 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[71], 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[72] 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[73] should entail obligations for media and social media

67 See Council of Europe, 2017b, Appendix I, paras 1, 2, 32, and 35-40. See also Council of Europe, Directorate General of democracy and Political Affairs – Directorate of Democratic Institutions, 2011.

68 See Council of Europe, 2010; Council of Europe, Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data, 2019.

69 See MANHEIM-KAPLAN, 2019; European Commission, Networks, Content and Technology- Directorate--General for Communication, 2018.

70 See also Council of Europe, 2011.

71 See Council of Europe, 2007, para I.1.

72 See also The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., 'Joint Declaration on "Fake News," Disinformation and Propaganda', para 2.c.

73 See Council of Europe, 2007, para II.1.

operators regarding the transparency of the logic of the algorithms used for content selection,[74] ensuring pluralism and diversity of voices[75], including critical ones[76].

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[77].

Given the important role played by advertising in disinformation and propaganda, the criteria used by AI-based solutions for political advertising should be transparent[78], auditable and provide equal conditions to all the political parties and candidates[79]. In addition, intermediaries should review their advertising models to ensure that they do not adversely affect the diversity of opinions and ideas[80].

3 AI AND DIGITAL JUSTICE

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

This analysis 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.

74 See also The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., Appendix, paras 2.1.3 and 2.3.5.

75 See also EU Code of Practice on Disinformation, 2018.

76 See also Council of Europe, 2016a, Appendix, para 15.

77 See also The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., para 4.e; European Commission for Democracy trough Law (Venice Commission), 2019, para 151.D.

78 See also Council of Europe Parliamentary Assembly, 2019a, paras 9.2 and 11.1; European Commission for Democracy trough Law (Venice Commission), 2019, paras 151.A and 151.B.

79 See also Council of Europe, 2007, para II.5.

80 See also The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., para 4.e.

Guiding principles for the development of AI in the field of digital 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[81].

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[82].

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 rates. This is the case with legal reasoning in problem solving[83]. At the same time, while a misclassification of an image of a cat may have limited adverse effects, an error rate in legal decisions[84] 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 traditional and non-traditional categories[85].

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

82 See Section 2.

83 See also OSOBA-WELSER, 2017, 18. See also Cummings et al., 2018, 13.

84 See Aletras et al., 2016. See also Pasquale-Cashwell, 2018; Hildebrandt, 2018a.

85 See WACHTER, 2021; MITTELSTADT, 2017, 485. See also TAYLOR-FLORIDI-VAN der Sloot, 2017.

Given the textual nature of legal documents, natural language processing (NLP) can play an important role in AI applications for the justice sphere[86]. 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[87]. 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[88].

3.1 ADRs and court decisions

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[89]. 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).

Although AI has an impact in such cases on legal practice and legal knowledge that raises various ethical issues[90], 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.

86 But see OSWALD, 2018; PASQUALE-CASHWELL, 2018.
87 See Council of Bars & Law Societies of Europe, 2020, 29.
88 See the following Section on the distinction between codified justice and equitable justice.
89 See CEPEJ, 2018, Appendix II.
90 See also NUNEZ, 2017.

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[91], can be countered by suitable education and training of legal professionals and the transparency of AI systems (that is the 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[92].

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 (for example 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[93] 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

91 See the notion of e-justice in Council of Europe, 2009b, Appendix, para 38.
92 See also CEPEJ, 2018.
93 See RE-SOLOW-NIEDERMAN, 2019, 252-4.

can be demanded to AI, preserving in any case human overview that also guarantees legal creativity of decision-makers[94].

Regarding equitable justice, as the literature points out[95], 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 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'[96], 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 International Covenant on Civil and Political Rights (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,

94 See also Clay, 2019), 58. 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 reconsider the foundation of the legal reasoning and extend compensation to non-personal and existential damages.

95 See RE-SOLOW-NIEDERMAN, 2019.

96 See CEPEJ, 2018, 9.

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[97]. Data quality and other forms of quality assessment (impact assessment, audits, etc.) can reduce this risk 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[98].

Further concerns affect the principles of fair trial and of equality of arms[99], when court decisions are based on the results of proprietary algorithms whose training data and structure are not publicly available[100]. 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 become 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.

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[101], its effective impact may be undermined

97 See also CEPEJ, 2018.

98 See also Council of Europe, 2020b, Appendix, para 11. See also Pasquale and Cashwell, 'Prediction, Persuasion, and the Jurisprudence of Behaviourism'.

99 See also CEPEJ, 2018, Appendix I, para 138.

100 See also CEPEJ, 2018, Appendix I, para 131.

101 See also ZALNIERIUTE-BENNETT MOSES-WILLIAMS, 2019. In the case of administrative decisions, this propensity may be reinforced by the threat of potential sanctions for taking a decision that ignores results

by the psychological or utilitarian (cost-efficient) propensity of the human decision-maker to take advantage of the solution provided by AI[102].

3.2 Crime prevention

The complexity of crime detection and prevention has stimulated research in AI applications to facilitate human activities. In recent years, several solutions[103] 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[104].

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[105], etc.) discussed in the previous sections and partially addressed by transparency, data quality, data protection, auditing and the other measures[106]. It is worth noting that the role of transparency in the judicial context could be limited so as not to frustrate the deterrent effect of these tools[107]. 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[108]. Self-fulfilling bias, community bias[109] and historical bias[110]

produced by analytics; Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data, 2019, para 3.4.

102 See also CITRON-CALO, 2021; MANTELERO, 2019; BRAUNEIS-GOODMAN, 2018, 127.

103 See ZAVRŠNIK, 2019; European Union Agency for Fundamental Rights, 2018, 98-100; OSOBA-WELSER, 2017.

104 For a taxonomy of predictive methods, see PERRY et al., 2013.

105 See also Council of Europe, 2001, Appendix, para 42.

106 See also RICHARDSON-SCHULTZ-CRAWFORD, 2019.

107 See also OSWALD, 2018; Barrett, 2017, 361-2.

108 See European Union Agency for Fundamental Rights, 2018, 10.

109 See also BARRETT, 358-9.

110 See BENNETT MOSES-Chan, 2018.

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[111], these solutions can also adversely affect the principle of presumption of innocence, procedural fairness, and the right to non-discrimination[112].

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)[113], 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[114]. 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 issues that are often crime-related and require long-term social strategies involving the effective enhancement of individual and social rights and freedoms[115].

111 See LYNSKEY, 2019; MANTELERO-VACIAGO, 2015; Hildebrandt-Gutwirth, 2008.
112 See also Council of Europe, 2001, Appendix, paras 47 and 49.
113 See van BRAKEL-De Hert, 2011, 183.
114 See MEIJER-WESSELS, 2019.
115 See also ROSENBAUM, 2006, 245-66.

4 CONCLUSIONS

The latest wave of AI development is having a growing transformative impact on society and raises new questions in several fields, from predictive medicine and media content moderation to the quantified self and judicial systems.

With a view to preserving the harmonisation of the existing legal framework in the field of human rights, this article sets out to contribute to the debate on future AI regulation by building on existing binding instruments, contextualising their principles and providing key regulatory guidance in the fields of electronic democracy and digital justice.

This approach is based on the assumption that all human activities, including innovation through AI, should be underpinned by the general international principles on human rights. Moreover, only the human rights framework can provide a universal reference for the regulation of AI, while other yardsticks (for example ethics) do not have the same global dimension, are more context-dependent and characterised by a variety of theoretical approaches.

The findings of this analysis show that a limited number of cases do share common principles (for example 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, with the independence of judges or the principles of fair trial and equality of arms, which concern justice alone[116].

Second, some guiding principles are shared by 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.

Transparency, as a means to control the power over data in the hands of public and private entities, is crucial with regard to AI applications for democratic participation and good governance. 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

116 See also the principles of equitable access and of beneficence in health sector, or the principles of non-interference by public authorities in the media to influence elections and the obligation to treat all political parties and candidates equally in electoral advertising.

requiring limitation to avoid prejudicing competing interests (*e.g.*, crime detection and prevention in predictive policing).

We can therefore conclude that transparency is a guiding principle, but we must go beyond a mere claim for 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 needed, with 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. While, for instance, in the field of data protection there are several provisions implementing these principles with a significant degree of detail[117], in the case of democracy and justice these principles are less developed in data--intensive applications such as AI.

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.

Such considerations suggest only partial harmonisation is achievable. The framework of future international AI regulation 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 relevant only in a given field or cases where the same principle is contextualised differently in the different fields.

The analysis carried out in the previous sections has also confirmed that the existing framework based on human rights can provide an appropriate and common context for the development of more specific binding instruments to regulate AI, in line with the principles enshrined in the international legal instruments and capable of effectively addressing the issues raised by AI.

117 See Council of Europe, 2018a, and Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data, 2019.

With a view to future regulation of AI, this study does not rule out a number of gaps, largely due to the fact that in broad areas, such as democracy and justice, differing 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.

REFERENCES

AGRE, P.E. Surveillance and Capture: Two Models of Privacy. *The Information Society* 10, 101,1994.

AI NOW INSTITUTE. AI Now 2017 Report. New York, 2017. Available at https://assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf. Accessed: 26 oct. 2017.

ALETRAS, N. et al. Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective. *PeerJ Computer Science* 2, 93, 2016, doi:10.7717/peerj-cs.93.

BARRETT, L. Reasoably Suspicious Algorithms: Predictive Policing at the United States Border. *New York University Review of Law & Social Change* 41, 327, 2017.

BENNETT MOSES L.; CHAN, J. Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability. *Policing and Society* 28, 806, 2018.

BRAUNEIS R.; GOODMAN, E.P. Algorithmic Transparency for the Smart City. *Yale J.L. & Tech.* 20, 103, 2018.

BUKOVSKA, B. Spotlight on Artificial Intelligence and Freedom of Expression #SAIFE'. Organization for Security and Co-operation in Europe, 2020. Available at https://www.osce.org/files/f/documents/9/f/456319_0.pdf. Accessed: 11 aug. 2020.

BURRELL, J. How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms. *Big Data & Society* 3(1), 2016, doi: 10.1177/2053951715622512.

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

CARUANA, R. et al. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.

CEPEJ – European Commission for the Efficiency of Justice. European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment, 2018.

CITRON D.K.; CALO R. The Automated Administrative State: A Crisis of Legitimacy. *Emory Law Journal* 70(4), 2021.

CLAY T. (ed), *L'arbitrage en ligne. Rapport du Club des Juristes*, 2019. Available at https://www.leclubdesjuristes.com/les-commissions/larbitrage-en-ligne/. Accessed: 30 may 2020.

COHEN, J.E. *Between Truth and Power. The Legal Construction of Informational Capitalism*. New York, Oxford University Press, 2019.

Committee of Ministers of the Council of Europe. The 12 Principles of Good Governance enshrined in the Strategy on Innovation and Good Governance at local level, 2008.

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

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

Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data (Convention 108). Guidelines on artificial intelligence and data protection, 2019. Available at https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8. Accessed: 20 feb. 2020.

Council of Europe Directorate General of Democracy and Political Affairs and Directorate of Democratic Institutions, Project «Good Governance in the Information Society», CM(2009)9 Addendum 3, 2009

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

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

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

Council of Europe, Directorate General of Democracy – European Committee on Democracy and Governance. The Compendium of the most relevant Council of Europe texts in the area of democracy, 2016.

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

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

_____. Algorithms and Human Rights. Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications, 2018. Available at https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html. Accessed: 5 may 2018.

_____. Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, 1997.

_____. Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes, 2019a.

_____. Guidelines for civil participation in political decision making, CM(2017)83-final, 2017a.

_____. Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data (Convention 108+), 2018.Council of Europe.

_____. Recommendation CM/Rec(2001)10 on the European Code of Police Ethics, 2001.

_____. Recommendation CM/Rec(2004)15 on electronic governance ("e-governance"), 2004.

_____. Recommendation CM/Rec(2007)15 on measures concerning media coverage of election campaigns, 2007.

_____. Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), 2009b.

_____. Recommendation CM/Rec(2009)2 on the evaluation, auditing and monitoring of participation and participation policies at local and regional level, 2009c.

_____. Recommendation CM/Rec(2010)13 on the protection of individuals with regard to automatic processing of personal data in the context of profiling, 2010.

_____. Recommendation CM/Rec(2011)7 on a new notion of media, 2011.

_____. Recommendation CM/Rec(2014)7 on the protection of whistleblowers, 2014.

_____. Recommendation CM/Rec(2016)4 on the protection of journalism and safety of journalists and other media actors, 2016a.

_____. Recommendation CM/Rec(2016)5 on Internet freedom, 2016b.

_____. Recommendation CM/Rec(2017)5 on standards for e-voting, 2017b.

_____. Recommendation CM/Rec(2018)1 on media pluralism and transparency of media ownership, 2018a.

_____. Recommendation CM/Rec(2018)4 on the participation of citizens in local public life, 2018b.

_____. Recommendation CM/Rec(2019)3 on supervision of local authorities' activities, 2019b.

_____. Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems, 2020b.

_____. Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications, 2018c. Available at https://rm.coe.int/algorithms-and-humanrights-en-rev/16807956b5 Accessed: 15 jan. 2019.

Council of Europe. Graphical visualisation of the distribution of strategic and ethical frameworks relating to artificial intelligence, 2020a.

CRAWFORD K.; JOLER, V. Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources, 2018. Available at http://www.anatomyof.ai. Accessed: 27 dec. 2019.

CUMMINGS, M.L. et al. Chatham House Report. Artificial Intelligence and International Affairs Disruption Anticipated, 2018. Available at https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial--intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf. Accessed: 21 mar. 2020.

DIAKOPOULPS. N. *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*, 2013. Available at https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2. Accessed: 18 mar. 2018.

EDWARDS L.; VEALE, M. Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review* 16, 18, 2017.

EU Code of Practice on Disinformation, 2018. Available at https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation. Accessed: 24 mar. 2021.

European Commission for Democracy trough Law (Venice Commission). 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, 2019.

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

European Commission. Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial

Intelligence Act) and amending certain Union legislative acts, COM(2021) 206 final, 2021.

European Union Agency for Fundamental Rights. #BigData: Discrimination in Data-Supported Decision Making, 2018.

European Union Agency for Fundamental Rights. Data Quality and Artificial Intelligence – Mitigating Bias and Error to Protect Fundamental Rights, 2019.

European Union Agency for Fundamental Rights. Preventing Unlawful Profiling Today and in the Future: A Guide, 2018.

EYKHOLT, K. et al. Robust Physical-World Attacks on Deep Learning Visual Classification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. Available at https://openaccess.thecvf.com/content_cvpr_2018/papers/Eykholt_Robust_Physical-World_Attacks_CVPR_2018_paper.pdf. Accessed: 23 apr. 2021.

FAYE JACOBSEN, A. The Right to Public Participation. A Human Rights Law Update. Issue Paper, 2013. Available at https://www.humanrights.dk/publications/right-public-participation-human-rights-law-update. Accessed: 14 jan. 2021.

FERRYMAN K.; Pitcan, M. *Fairness in Precision Medicine*, 2018. Available at https://datasociety.net/wp-content/uploads/2018/02/Data.Society.Fairness.In_.Precision.Medicine.Feb2018.FINAL-2.26.18.pdf. Accessed: 8 apr. 2018.

GANDY Jr., O.H.; Nemorin, S. Toward a Political Economy of Nudge: Smart City Variations. *Information, Communication & Society* 22, 2112, 2019.

GESLEY, J. Regulation of Artificial Intelligence in Selected Jurisdictions, 2019. Available at https://www.loc.gov/law/help/artificial-intelligence/index.php. Accessed: 30 dec. 2019.

GOODMAN E.; Powles, J. Urbanism Under Google: Lessons from Sidewalk Toronto. *Fordham Law Review* 88(2), 457, 2019.

GRABER, C.B. Artificial Intelligence, Affordances and Fundamental Rights. In Hildebrandt M.; O'Hara K. (eds) *Life and the Law in the Era of Data-Driven Agency*. Edward Elgar, 2020.

HAGENDORFF, T. The Ethics of AI Ethics: An Evaluation of Guidelines', *Minds and Machines* 30, 99, 2020.

HILDEBRANDT M.; Gutwirth, S. (eds) *Profiling the European Citizen: Cross-Disciplinary Perspectives*, Dordrecht, 2008.

HILDEBRANDT, M. Algorithmic Regulation and the Rule of Law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376, 2018a.

_____. Primitives of Legal Protection in the Era of Data-Driven Platforms. *Georgetown Law Technology Review* 2, 252, 2018b.

_____ Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law* 20, 83, 2019.

_____. The Issue of Bias. The Framing Powers of Machine Learning. In Pelillo, M.; Scantamburlo, T. (eds) *Machines We Trust. Perspectives on Dependable AI* (MIT Press: Cambridge, MA) 2021.

Independent High-Level Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI, 2019. Available at https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthyai. Accessed: 2 mar. 2020.

JOBIN, A.; IENCA, M.; VAYENA, E. The Global Landscape of AI Ethics Guidelines', *Nature Machine Intelligence* 1, 389, 2019.

KAMINSKI M. E.; MALGIERI, G. Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, 2020, doi: 10.1145/3351095.3372875.

KOLKMAN, D. The (in)Credibility of Algorithmic Models to Non-Experts. *Information, Communication & Society* 1, 2020, doi: 10.1080/1369118X.2020.1761860.

LOIDEAIN, N. N.; Adams, R. From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of Data Protection Impact Assessments. *Computer Law & Security Review* 36, 2020, doi:10.1016/j.clsr.2019.105366.

LYNSKEY, O. Criminal Justice Profiling and EU Data Protection Law: Precarious Protection from Predictive Policing. *Int J Law Context*, 15, 162, 2019.

MAISLEY, N. The International Right of Rights? Article 25(a) of the ICCPR as a Human Right to Take Part in International Law-Making. *Eur. J. Int. Law* 28, 89, 2017.

MANHEIM, K.; KAPLAN, L. Artificial Intelligence: Risks to Privacy and Democracy. *Yale Journal of Law & Technology* 21, 106, 2019.

MANTELERO, A. AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment. *Computer Law & Security Review* 34(4), 754, 2018.

_____. 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, 2019. Available at https://rm.coe.int/2018-lignes-directrices-sur--l-intelligence-artificielle-et-la-protecti/168098e1b7. Accessed: 20 feb. 2020.

_____. Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection. *Computer Law & Security Review* 32(2), 238, 2016.

_____. Regulating AI Within the Human Rights Framework: A Roadmapping Methodology. In Czech et al. (eds) *European Yearbook on Human Rights 2020*, 477-502, 2020.

_____; Esposito. M.S. An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems. *Computer Law & Sec. Rev.*, 41, 2021, DOI: 10.1016/j.clsr.2021.105561.

MANTELERO, A.; VACIAGO, G. Data Protection in a Big Data Society. Ideas for a Future Regulation. *Digital Investigation* 15, 104, 2015.

MEHR, H. Artificial Intelligence for Citizen Services and Government, 2017. Available at https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf. Accessed: 15 mar. 2021.

MEIJER A.; WESSELS M. Predictive Policing: Review of Benefits and Drawbacks. *Int J Publ Admin* 42, 1031, 2019.

MIKHAYLOV, J.; Esteve, M.; Campion, A. Artificial Intelligence for the Public Sector: Opportunities and Challenges of Cross-Sector Collaboration. *Phil. Trans. R. Soc. A*, 376, 2018, doi: 10.1098/rsta.2017.0357.

MITTELSTADT, B. From Individual to Group Privacy in Big Data Analytics. *Philosophy & Technology* 30, 475, 2017.

NEMITZ, P. Constitutional Democracy and Technology in the Age of Artificial Intelligence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378, 2018, doi: 10.1098/rsta.2018.0089.

NUNEZ, C. Artificial Intelligence and Legal Ethics: Whether AI Lawyers Can Make Ethical Decisions. *Tulane Journal of Technology and Intellectual Property* 20, 189, 2017.

OECD, Recommendation of the Council on Artificial Intelligence, 2019.

OSOBA, O.A.; Welser, W. An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence, 2017. Available at https://www.rand.org/pubs/research_reports/RR1744.html. Accessed: 20 may 2020.

OSWALD, M. Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences'* 376, 2018, doi: 10.1098/rsta.2017.0359.

_____. Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2018, doi: 10.1098/rsta.2017.0359.

PASQUALE, F. *The Black Box Society. The Secret Algorithms That Control Money and Information*.Harvard University Press, 2015.

PASQUALE F.; CASHWELL, G. Prediction, Persuasion, and the Jurisprudence of Behaviourism. *University of Toronto Law Journal*, 68, 2018.

PERRY, W.L. et al. Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations, 2013. Available at https://www.rand.org/pubs/research_reports/RR233.html. Accessed: 30 mar. 2020.

Privacy International, Smart Cities: Utopian Vision, Dystopian Reality, 2017. Available at https://privacyinternational.org/sites/default/files/2017-12/Smart%20Cities-Utopian%20Vision%2C%20Dystopian%20Reality.pdf. Accessed: 12 may 2020.

RAAB, C.D. Information Privacy, Impact Assessment, and the Place of Ethics. *Computer Law & Security Review* 37, 2020, doi:10.1016/j.clsr.2020.105404.

RANCHORDÁS, S. Nudging Citizens through Technology in Smart Cities. *International Review of Law, Computers & Technology* 34(2), 254, 2020.

RE, R.M.; Solow-Niederman, A. Developing Artificially Intelligent Justice. *Stanford Technology Law Review* 22, 242, 2019.

RICHARDSON, R.; SCHULTZ, J.M.; CRAWFORD, K. Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review* 94, 42, 2019.

ROSENBAUM, D. P. The limits of hot spots policing. In Weisburd D.; Braga A.A. (eds) *Police innovation: contrasting perspectives*, Cambridge University Press, 2006.

SAVAGET, P.; CHIARINI, T.; Evans, S. Empowering Political Participation through Artificial Intelligence. *Science and Public Policy* 46, 369, 2019.

SCHRAG, Z. M. *Ethical Imperialism. Institutional Review Boards and the Social Sciences 1965-2009*. Baltimore, Johns Hopkins University Press, 2017.

SELBST, A. D. Disparate Impact in Big Data Policing. *Georgia Law Review* 52(1), 109, 2017.

SELBST, A. D.; BAROCAS, S. The Intuitive Appeal of Explainable Machines. *Fordham L. Rev.* 87, 1085, 2018.

SUNSTEIN, C. R. The Ethics of Nudging, *Yale Journal on Regulation* 32, 412, 2015.

_____. *Why Nudge? The Politics of Libertarian Paternalism*. Yale University Press, 2015.)

SUNSTEIN C.R.; THALER, R.H. Libertarian Paternalism in Not an Oxymoron', *University of Chicago Law Review* 70, 1159, 2003.

TAYLOR, L.; FLORIDI, L.; van der Sloot, B. (eds). *Group Privacy : New Challenges of Data Technologies*. Cham, Springer, 2017.

THALER, R. H.; SUNSTEIN, C. R. *Nudge. Improving Decisions about Health, Wealth, and Happiness.* New Haven, Yale University Press, 2008.

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. Joint Declaration on "Fake News," Disinformation and Propaganda, 2017.

TUBARO, P.; CASILLI, A. A.; COVILLE, M. The Trainer, the Verifier, the Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence. *Big Data & Society* 7(1), 2020, doi: 10.1177/2053951720919776.

UN Committee on Economic, Social and Cultural Rights (CESCR). General Comment No. 1: Reporting by States Parties, 27 July 1981.

UN Human Rights Committee (HRC). CCPR 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, 12 July 1996.

UNESCO. Draft Text of the Recommendation on the Ethics of Artificial Intelligence, 2021. Available at https://unesdoc.unesco.org/ark:/48223/pf0000377897. Accessed: 3 sep. 2021.

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

VEALE, M.; BINNS, R. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society* 4(2), 2017, doi:10.1177/2053951717743530.

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

VERONESE, A.; NUNES LOPES ESPIÑEIRA LEMOS, A. Trayectoria normativa de la inteligencia artificial en los países de Latinoamérica con un marco jurídico para la protección de datos: límites y posibilidades de las políticas integradoras. *Revista Latinoamericana de Economía y Sociedad Digital* 2, 2021. available at https://revistalatam.digital/article/210207/. Accessed: 27 aug. 2021.

WACHTER, S. Affinity Profiling and Discrimination by Association in Online Behavioural Advertising. *Berkeley Tech. L.J.*, 35(2), 367, 2021.

WEST, S.M.; WHITTAER, M.; Crawford, K. Discriminating Systems. Gender, Race, and Power in AI, 2019. Available at https://ainowinstitute.org/discriminatingsystems.pdf. Accessed: 15 may 2019.

ZALNIERIUTE, M.; Bennett Moses, L.; Williams, G. The Rule of Law and Automation of Government Decision-Making. *The Modern Law Review* 82(3), 425, 2019.

ZAVRŠNIK, A. Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings. *European Journal of Criminology* 1, 2019, doi:10.1177/1477370819876762.

ZUIDERVEEN BORGESIUS, F. Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence. *The International Journal of Human Rights* 24(10), 1572, 2020.

Sobre o autor:

Alessandro Mantelero | *E-mail:* alessandro.mantelero@polito.it

Associate Professor of Private Law and Law & Technology at the Polytechnic University of Turin, and Council of Europe Scientific Expert on AI, data protection and human rights. He is Associate Editor of Computer Law & Security Review and member of the Editorial Board of European Data Protection Law Review.

Artigo convidado.