

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

About some international documents relating to the ethics of Artificial Intelligence

Poullet, Yves

Published in:

Time to reshape the digital society

Publication date:

2021

Document Version

le PDF de l'éditeur

[Link to publication](#)

Citation for pulished version (HARVARD):

Poullet, Y 2021, About some international documents relating to the ethics of Artificial Intelligence: some insights. dans *Time to reshape the digital society: 40th anniversary of the CRIDS*. Collection du CRIDS, numéro 52, Larcier , Bruxelles, pp. 523-540.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

CHAPTER 6

About some international documents relating to the ethics of Artificial Intelligence – Some insights

Yves POULLET

Co-chairman of the Nadi Institute, Emeritus professor at the University of Namur (Belgium), Associate Professor at the UCLille (France) – IFAP Vice-President (WG Info-Ethics) – Member of the Royal Academy of Belgium.

1. Digitalization has definitively invaded our daily lives, our business activities and our administration management and decisions. Today, different technologies (Internet of things, Big Data and Artificial Intelligence applied together) are radically modifying the methods and the impact of all our activities and decisions. At the same time, it must be recognized that AI systems often operate, with potential bias and errors, in an opaque manner combining the functioning of multiple neural networks. They rely on random statistical combinations taking into account millions of data items, which now render possible unprecedented efficient and predictive profiling of individual behavior. This reality and the possible lack of human mastership of the technology increase the risks incurred individually by the data subjects but also collectively by groups of individuals and even for our democracy and rule of law, as demonstrated by the Cambridge Analytica case and the increasing phenomenon of disinformation, clearly linked by the use of AI systems. These reasons militate in favor of a re-built and renewed regulatory framework for AI that must take into account ethical issues like discrimination, democracy and groups' privacy, as requested by certain recent and increasingly numerous national and international initiatives.

2. The ethical aspects of the development of AI meet the concerns of many public (OECD, UNESCO, Council of Europe, Arab League...) ¹ and private ² international bodies. This topic is at the heart of the “third way” called for by the European Union concerning an Artificial Intelligence (in short AI) characterized by the words: “Excellence and Trust”. This article does not aim to provide a detailed overview of all the existing or planned instruments by public international organizations. This article will refer to these initiatives in not much detail ³ and will only cover those of the OECD, UNESCO, the Council of Europe, and the European Union. The mere observation of the multiplication of documents on AI ethics deserves to be questioned as to why there is such an efflorescence. Beyond that, what catches our attention are the major trends, common (or not) to these documents.

We will analyze a few selected themes. The first is the very nature of the proposed regulation. Is asserting ethical principles sufficient or do ethics ultimately refer to the claim of a legal framework? Beyond that, does the regulatory model refer to self-regulation, co-regulation, or public regulation? On this point, there are widely differing points of view. The second is an analysis of the ethical principles asserted. The list is often long, reflecting confusion between ethical values and the means of ensuring that these values are respected. What is striking is that, in all cases, the values proclaimed intend to go beyond the protection of individual interests and freedoms to highlight the collective stakes involved in AI development, social justice, democracy, and the environment. Some documents point out that while individual issues seem to be well covered by many texts thereby rendering the protection of individual freedoms effective, the lack of effectiveness in the implementation of collective issues is problematic. The third theme pursues this second reflection: the

¹ See also the ITU World Summit on Artificial Intelligence for the Public Interest (AI for Good). See <https://www.itu.int/fr/mediacentre/backgrounders/Pages/artificial-intelligence-for-good.aspx>.

² Thus, among many texts, those of the Institute of Electrical and Electronic Engineering (IEEE): IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (The “IEEE Global Initiative”), accessible on the IEEE website <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html> and the “Montreal Declaration on Artificial Intelligence”, which is the result of an academic initiative, <https://ia-ethique.be/declaration-montreal-intelligence-artificielle/>. The reader will find an impressive list (more than fifty) of “ethical or non-binding frameworks applicable to artificial intelligence and data science”, in Y. MENECEUR, *L'intelligence artificielle en procès*, Bruylant, 2020, Appendix, pp. 409-417 and will read his very critical analysis of this movement in favor of these ethical frameworks, pp. 199 *et seq.*

³ On these texts and our reflections, see Y. Poullet, *Ethique et Droits de l'homme dans notre société du numérique*, Brussels, Académie royale de Belgique, 2020.

approach centered on the risks incurred either individually by individuals or collectively by our societies leads us to question the need to distinguish according to the seriousness of these risks, to put the emphasis on the procedures for evaluating these risks, on the actors at the root of these risks and their duties to collaborate in taking them into account and controlling them and, finally, on the role of the State in this area. On this last field of reflection, it is worth noting the boldness of the European texts, advocating the creation of national *Data Ethics* commissions.

It will then be time to attempt to arrive certain conclusions and to answer this question: what is the value of this so-called ethical approach? Is this “ethics bashing” or a path to a real solution?

3. Artificial intelligence (AI for short) is a “buzzword”, the object of many fantasies evoking a society as much of transhumanism as of total transparency of people and of “purely algorithmic governmentality” of our societies and our individual lives, according to the expression of A. Rouvroy.⁴ Villani⁵ describes it as “one of the keys to tomorrow’s power in a digital world”. No doubt, these prophecies are exaggerated. Nevertheless, recent scandals such as the *Cambridge Analytica* scandal, the profiling of Bristol city council schoolchildren on the basis of past and future school performance for the purposes of school selection but also for commercial purposes, force us to question ourselves. We know that AI helps or even substitutes business managers in the selection of candidates for employment, it allows estate agencies to select people interested in a house according to the preferences expressed by the owners, etc. The control of migrants at European borders will soon use Artificial Intelligence systems combining graphological analysis, facial recognition, and big data that can calculate the risk of terrorism presented by a candidate for immigration.⁶ A well-known multinational company is announcing the possibility, thanks to AI, of detecting Alzheimer’s disease in Internet users at an early stage, based on messages sent and keystrokes on the keyboard. In short, the particularly predictive capabilities offered by AI justify the

⁴ A. ROUVROY, “L’*homo juridicus* est-il soluble dans les données?”, in E. DEGRAVE *et al.* (eds.), *Droits, normes et libertés dans le cybermonde. Liber Amicorum Yves Poulet, Cahiers du Crids*, n° 43, Bruxelles, Larcier, 2018, pp. 417-443.

⁵ C. VILLANI, “Giving meaning to artificial intelligence. For a national and European strategy”, *Report to the Prime Minister of the French Republic*, March 2018, p. 11.

⁶ For example, the “I-Border Ctrl” system currently being developed by Europe for border control, the modules of which are described by the report by Algorithmwatch and the Bertelsman Foundation (“Automating Society Taking Stock of Automated Decision-Making in the EU”, a report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations, January 2019).

questioning of our societies on the limits to be imposed on the public and private designers and operators of Artificial Intelligence systems in the name of essential ethical values.

4. This question justifies the growing number of reports and often recommendations published on AI and ethics. The Organisation for Economic Co-operation and Development (OECD) Recommendation⁷ on Artificial Intelligence, issued by the Council of Ministers in 2019, is the first intergovernmental standard for AI policies and has been the foundation upon which many other documents have been based since then.⁸ As far as UNESCO is concerned, the 40th General Conference, following the example of what had been achieved in the field of bioethics, called for an international standard-setting instrument in its Resolution 40 C/37, a recommendation on the ethics of Artificial Intelligence. A draft recommendation⁹ prepared by a so-called *ad hoc* group of experts is currently being discussed by national delegations and a final text will be presented for adoption at the next General Assembly in Autumn 2021. On the Council of Europe side, the Committee of Ministers appointed, on September 11, 2019, an equally *ad hoc* Committee of Experts: the CAHAI.¹⁰ The Committee is “entrusted with studying the feasibility and potential elements, on the basis of broad multi-stakeholder consultations, of a legal framework for the development, design, and application of Artificial Intelligence, based on Council of Europe standards in the fields of *human rights, democracy and the rule of law*”. The recent adoption on December 17, 2020 by the CAHAI of the Council of Europe of the so-called “Study of feasibility concerning a “feasibility study on a legal framework on AI design, development and application based on CoE standards”¹¹ is noteworthy and a first draft of a Council of Europe Convention is expected for the end of the year. The particular interest of this text will be underlined in the rest of the article.

⁷ Council Recommendation on Artificial Intelligence, adopted by the OECD Ministerial Council on May 22, 2019, <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449>.

⁸ In June 2019, at the Osaka Summit, G20 Leaders welcomed the G20 Principles on AI from the OECD Recommendation (Council Recommendation on Artificial Intelligence, adopted by the OECD Ministerial Council on May 22, 2019, <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449>).

⁹ Preliminary Report on the Draft Recommendation on the Ethics of Artificial Intelligence, https://unesdoc.unesco.org/ark:/48223/pf0000374266_fre.

¹⁰ Either Ad Hoc Artificial Intelligence Committee whose work is accessible on the page www.coe.int/cahai.

¹¹ Available at <https://rm.coe.int/cahai-2020-23-final-etude-de-faisabilite-fr-2787-2531-2514-v-1/1680a1160f>.

The European Union is not to be outdone. If certain documents under the previous Juncker Presidency had already sensed the importance of the topic, it is clear that it is to the current von der Leyen Presidency that we owe this desire for an ethical framework for the development of AI, as the touchstone of this famous “third way” that Europe intends to follow at a distance from both the American and Chinese models. “We want the application of these new technologies to be worthy of the trust of our citizens [...]. We encourage a responsible approach to Artificial Intelligence centered on the human being”.¹² This is a good introduction and summary of the Commission’s White Paper on its strategy in this area, which is perfectly, summed up in the¹³ words “Excellence and Trust”. This White Paper is based in particular on the work of a High Level Group of Experts (*HLGE on AI*) commissioned by the Commission, which in April 2019 produced ethical recommendations for a trusted Artificial Intelligence system.¹⁴ In response to this ethical impetus from the Commission and in full consultation with the latter, the European Parliament adopted a resolution on October 20, 2020 containing recommendations to the Commission on a framework for the ethical aspects of Artificial Intelligence, robotics and related technologies.¹⁵ It is emphasized that the text is more than a resolution, it contains a proposal for a Regulation, a “ready-to-sign” text for the Commission and the Council, which will be difficult to deviate from what was agreed by the Parliament. From its side, the European Commission has issued on April 21, 2021 a “Proposal for a Regulation of the EU Parliament and of the Council laying down harmonized rules on Artificial Intelligence and amending certain Union legislative Acts”.¹⁶

¹² Statement by U. VAN DER LEYEN, President of the Commission at the announcement of the European Union’s strategic plan for Artificial Intelligence, February 19, 2020. This statement is to be read in conjunction with the OECD text, which in its preamble notes: “Recognizing that in view of the rapid evolution and implementation of AI, it is necessary to build a stable policy framework, which favors a trustworthy human-centered AI ...”.

¹³ Commission White Paper on Artificial Intelligence – *A European approach based on excellence and trust*, Brussels, February 19, 2020, COM(2020)65 final, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_fr.pdf.

¹⁴ HLGE (High Level Group of experts) on AI, *Ethical guidelines for Trustworthy AI*, April 8, 2019, n° 67, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

¹⁵ European Parliament Resolution of October 20, 2020 with recommendations to the Commission on a framework for ethical aspects of Artificial Intelligence, robotics and related technologies (2020/2012(INL)) P9 TA (2020) 0275. This text was supported by the European Commission. It should be noted that the Resolution contains a proposal for a Regulation which is only waiting to be taken up by the other authorities of the European Union. It is expected that a draft Regulation will be proposed in June by the EU Commission.

¹⁶ Brussels, April 21, 2021, COM(2021) 206 final.

5. Many of the texts are called “Ethical Principles” and do not necessarily refer to the need for a regulatory framework. They are recommendations based on ethical values around which AI systems must be designed and developed and must operate, while respecting existing laws, some of which are criticized for their inadequacy. How can we understand this reference to ethics and this refusal of any public regulation in certain texts? *Ethics* refers to acting, to “doing the right thing”; it means that humans and their “artefacts” act or are designed for the Good and the Just. Ethics, according to Spinoza, is not, however, the affirmation of a particular morality. It is a value-based approach and does not receive a single answer. The plural of the English translation of the word, *Ethics*, expresses this singularity and the plurality of ethical responses. Ethics is a questioning, an individual and sometimes collective research about the attitude to adapt in the face of facts, realities, or social changes, such as those brought about by Artificial Intelligence. A distinction is made between ethics and codes of ethics. In the world of economic or professional activities, a code of conduct or ethics constitutes a public declaration by its authors and signatories of the values and practices followed. The code formalizes a certain number of principles of action and “minimum” standards. By publishing their code of conduct, companies or professions undertake to observe these standards and to ensure that they are observed by their subcontractors and suppliers. The code of ethics can be analyzed as a *soft law* with an uncertain binding value, whether or not it is explicitly or implicitly relayed by the law. Ethics could therefore serve as an alibi for rejecting any new binding regulations and justifying recourse to self-regulation alone.¹⁷ Without doubt, the OECD Recommendation, notwithstanding the quality of its content, can be read in this sense. The OECD has always been an advocate of serious self-regulation. The current draft tabled by UNESCO experts, essentially based on non-binding recommendations, was judged to be too “soft” in this respect. The preliminary report of the Director-General of UNESCO (n° 27) states¹⁸: “Last but not least, it was suggested that the draft Recommendation be

¹⁷ On that point among others, see B. WAGNER, “Ethics as an Escape from Regulation: From Ethics-Washing to Ethics-Shopping?”, in M. HILDEBRANDT. (ed.), *Being Profiling. Cogitas ergo sum*, Amsterdam University Press, 2017. “Striving for ethics and ethical decision-making, it is argued, that ethics will make technologies better. While this may be true in many cases, much of the debate about ethics seems to provide an easy alternative to government regulation. Unable or unwilling to properly provide regulatory solutions, ethics is seen as the “easy” or “soft” option which can help structure and give meaning to existing self-regulatory initiatives. In this world, “ethics” is the new industry self-regulation”.

¹⁸ In the text of the project, references to the need for a legal framework to address ethical challenges are rare. See, however, point 9 of the project: “The values and principles

more ambitious. This implies making bold proposals and being more assertive in suggesting that a stronger international legal framework is needed". During the intergovernmental discussion, it was clearly asserted that Ethics might not be an alibi to avoid the full application of Human Rights. UNESCO recommendations consider that ethical guidelines are not a substitute but a useful complement to the Law due to the lack of agility of the latter,¹⁹ more dynamic and potential source of legislation: "This Recommendation addresses ethical issues related to the domain of Artificial Intelligence to the extent that they are within UNESCO's mandate. It approaches AI ethics as a systematic normative reflection, based on a holistic, comprehensive, multicultural, and evolving framework of interdependent values, principles and actions that can guide societies in dealing responsibly with the known and unknown impacts of AI technologies on human beings, societies, and the environment and ecosystems, and offers them a basis to accept or reject AI technologies. Rather than equating ethics to law, human rights, or a normative add-on to technologies, it considers ethics as a dynamic basis for the normative evaluation and guidance of AI technologies, referring to human dignity, well-being and the prevention of harm as a compass and rooted in the ethics of science and technology".

6. The European texts are clearer in this respect. The European Parliament resolution strongly calls for a regulatory framework to address the ethical challenges that AI poses to our societies²⁰ and this approach amply justifies the Parliament's proposal for a binding Regulation. The

set out below should be respected by all actors in the life cycle of AI systems, in the first instance, and be promoted through the development of new legislation, regulations and commercial guidelines and the modification of existing ones".

¹⁹ "Taking fully into account that the rapid development of AI technologies challenges their ethical implementation and governance, because of the diversity of ethical orientations and cultures around the world, the lack of agility of the law in relation to technology and knowledge societies, and the risk that local and regional ethical standards and values be disrupted by AI technologies". (Preamble, § 24).

²⁰ Cf. Resolution of the European Parliament, Preamble point L: "Whereas it is necessary not only to adapt existing legislation, but also to address the legal and ethical issues related to AI technologies by means of an effective, comprehensive and sustainable regulatory framework enshrined in Union law, reflecting the values and principles of the Union as enshrined in the Treaties and in the Charter of Fundamental Rights of the European Union (hereinafter referred to as "the Charter"), which would be limited to filling existing legal gaps by avoiding over-regulation and which would enhance legal certainty for businesses and citizens alike, in particular by providing for mandatory measures...". See also point Y: "Whereas common ethical principles are only effective when they are also enshrined in law and when the parties responsible for ensuring, assessing and monitoring conformity are identified...".

European Commission in the same line states: “This Regulation supports the objective of the Union of being a global leader in the development of secure, trustworthy and ethical Artificial Intelligence, as stated by the European Council, and it ensures the protection of ethical principles, as specifically requested by the European Parliament”.

In its above-mentioned feasibility report, the CAHAI stresses the inadequacy of an approach based on ethics alone and condemns self-regulation as a solution to ethical issues²¹: “It was also underlined that soft law approaches cannot substitute mandatory governance. In some instances, due to the fact that the interests of those developing and commercializing the technology and those who might suffer negative consequences thereof are not always fully aligned, there is a particular risk that self-regulation by private actors can bypass or avoid mandatory governance by (inter)governmental authorities. Soft law instruments and self-regulation initiatives can however play an important role in complementing mandatory governance, especially where the interests of the different actors are more aligned and where no substantive risk of negative effects on human rights, democracy and the rule of law is present”. Thus, the Council of Europe, while it sees ethical values as the driving force behind reflection on the risks incurred by AI, clearly advocates for a top-down co-regulation approach in which the public authority sets the regulatory principles, even if this means leaving private actors a certain margin of maneuver in the application of these principles.²² As for the European Parliament’s draft regulation on the ethical aspects of AI, we note that its purpose is “to establish a comprehensive and sustainable regulatory

²¹ Feasibility Study, *op. cit.*, n° 79; see also, n° 93: “Should a regulatory approach that combines a binding instrument with soft law tools be supported by the CAHAI, private actors, civil society organisations, academia and other stakeholders would have an important role not only in assisting states in the development of a binding legal instrument, but also in contributing to the development of sectorial soft law instruments that can complement as well as aid in the implementation of the binding provisions in a context-specific manner (for instance through sectorial guidelines, certifications and technical standards)”.

²² The fight against disinformation, studied in another article in the same issue (Y. POULLET, N. BONTRIDDER, “The European Union and the regulation of disinformation”) is an example of the necessary shift from self-regulation to top-down co-regulation. Top-down co-regulation is what the European authorities had been calling for since 2003 under the Interinstitutional Agreement on Better Lawmaking (OJEC, C 321, December 31, 2003, pp. 1-5). Article 18 of the Agreement defined *co-regulation* as “the mechanism by which a Community legislative act confers the attainment of the objectives defined by the legislative authority on the parties concerned recognized in the field (in particular economic operators, social partners, non-governmental organizations or associations)”. On this notion, the distinction between self-regulation, bottom-up and top-down co-regulation, see our reflections in Y. POULLET, “Information and communication technologies and ‘co-regulation’: a new approach?”, in *Liber amicorum Michel Coipel*, Brussels, Kluwer, 2004, p. 173.

framework of *ethical principles and legal obligations* relating to the development, deployment and use of Artificial Intelligence, robotics and related technologies within the Union”.

7. Finally, it should be recalled²³ that the four traditional universal ethical principles (dignity, autonomy, social justice, and “do good, do not harm”) are enshrined in binding European texts on human rights (European Convention on Human Rights and the Charter of Fundamental Rights of the European Union), which ensure their effectiveness both through the duty of states to give them concrete meaning and through the action of the courts in Strasbourg and Luxembourg. As expressly asserted by the European Commission in the introduction of its proposal, the reference to ethical values is aimed to ensure a better protection of our human rights: “The use of AI with its specific characteristics (*e.g.* opacity, complexity, dependency on data, autonomous behavior) can adversely affect a number of fundamental rights enshrined in the EU Charter of Fundamental Rights (hereinafter “the Charter”). This proposal seeks to ensure a high level of protection for those fundamental rights and aims to address various sources of risks through a clearly defined risk-based approach. With a set of requirements for trustworthy AI and proportionate obligations on all value chain participants, the proposal will enhance and promote the protection of the rights protected by the Charter: the right to human dignity (Article 1), respect for private life and protection of personal data (Articles 7 and 8), non-discrimination (Article 21) and equality between women and men (Article 23). It aims to prevent a chilling effect on the rights to freedom of expression (Article 11) and freedom of assembly (Article 12)...”.

8. What ethical values are enshrined in these documents?²⁴ One could easily be struck by the number of values “discovered” by the documents.

²³ On this point, see our reflections in the book cited in note 3.

²⁴ Note the list proposed by the OECD: it (the Recommendation) sets out five complementary value-based principles, laying the foundation for a responsible approach to supporting trustworthy AI, and calls on AI stakeholders to promote and implement them. These principles are: inclusive growth; sustainable development and well-being; people-centered values and equity; transparency and accountability; robustness, safety and security; and accountability. The UNESCO list distinguishes, without specifying the criteria for this distinction, between values (respect and promotion of human dignity, human rights and fundamental freedoms, ensuring diversity and inclusion, sustainable development, living in harmony and peace) and principles (safety, security, equity and non-discrimination, sustainability, privacy, transparency and accountability, human oversight and decision-making, accountability, awareness and education, multi-stakeholder and adaptive governance and collaboration);. In comparison, the equally numerous values cited by the Montreal

In particular, UNESCO recommendations mention dignity, autonomy, privacy, freedom of expression and disinformation, education, research, monopolization, multiculturalism, wellbeing, non-discrimination, equality of genders, prevention of environmental damage. The same text evokes the need for transparency or explainability, the robustness of the systems used (including their reliability), the data protection requirements, and the responsibility of actors.

9. We are pleading for a simplification of this long list. To do so, inspiration shall be drawn from the UNESCO Convention on Bioethics,²⁵ which was based on the four universal values proposed by Beauchamp and Childress²⁶: dignity, autonomy, social justice, and *beneficent and non-maleficent* technologies. The profusion of values and principles asserted by the text seems to be dangerous and might contribute to regrettable confusion. Thus, it would certainly be useful to gather around the value of social justice, the imperative of non-discrimination, and gender equality. In the same perspective, sustainability as principle and environment and flourishing ecosystems as value could be finally joined together, especially since it seems that the list regrettably mixes values and the means to ensure their respect.²⁷ In our opinion, transparency is not a value in itself but a way of making the values of autonomy or non-discrimination effective. Prudence, or the precautionary principle, as well as respect for the environment, is indicated if one subscribes to the value “Do good and do not harm”, etc. On the one hand, extending the list might be considered as a source of confusion for the public and, above all, of devaluation of true ethical values. On the other hand, it illustrates the variety of problems that linked to the development and use of AI applications.

Declaration, without prioritizing them: well-being, autonomy, privacy and confidentiality, solidarity, democratic participation, equity, inclusion of diversity, prudence, responsibility and sustainable development. The ALLEA Discussion Paper about Ethics and AI drafted jointly by the EU National Academies (2019) mentions more concisely: “Privacy, Autonomy, Rationality, Equality, Dignity and Human flourishing”.

²⁵ Universal Declaration on Bioethics and Human Rights, adopted on October 19, 2005, http://portal.unesco.org/fr/ev.php-URL.ID=31058&URL.DO=DO_TOPIC&URL_SECTION=201.html.

²⁶ T. L. BEAUCHAMP, J. F. CHILDRESS, *Principles of Biomedical Ethics*, 3rd ed., New York, Oxford University Press, 2001.

²⁷ The same remark might be also applied to the EU HLGE Ethical guidelines for Trustworthy AI, which enumerate as ethical principles: “human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental well-being; and accountability”. The assessment of each of these principles is detailed in the so-called ALTAI list (Assessment list for a trustworthy AI), <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>.

On that point, the UNESCO draft recommendations must be pinpointed since they extract so many ethical issues from the traditional ones. I pinpoint the focus on societal questions such as the environment, cultural diversity, living in a peaceful and just society, public participation, and multilingualism. This Recommendation addresses ethical issues related to the domain of AI to the extent that they are within UNESCO's mandate. It approaches AI ethics as a systematic normative reflection, based on a holistic, comprehensive, multicultural, and evolving framework of interdependent values, principles and actions that can guide societies in dealing responsibly with the known and unknown impacts of AI technologies on human beings, societies, and the environment and ecosystems, and offers them a basis to accept or reject AI technologies.

10. But this is not the most important point revealed by the long but instructive list of ethical principles found in the various documents. What is important is the fact that these texts, especially the Unesco, Council of Europe and EU Parliament ones, are clearly envisaging not only traditional risks like threats to our individual liberties but also collective and even societal risks. Indeed, AI systems might cause harms not only to an individual person, for instance a person who would be imprisoned due to the malfunctioning of an AI facial recognition system, but also to a group of persons that are not clearly or *a priori* identifiable. So, elderly persons might be discriminated against as regards their insurance premium; people who are natives of certain local areas will be suspected of certain potential criminal infringements, etc. Collectively, people might be discriminated against by conscious or unconscious bias placed in AI's functioning. That is what van der Sloot²⁸ calls "Group's privacy" threats. Finally, we might also consider societal risks, when the harm to be taken into consideration concerns not only individuals or the collectivity but puts into question societal interest: "Societal harm occurs when one or more interests of society are wrongfully thwarted. In contrast with the above, societal harm is thus not concerned with the interests of a particular individual or the interests shared by a collective of individuals. Instead, it concerns

²⁸ We refer to the very interesting book L. TAYLOR, L. FLORIDI, B. VAN DER SLOOT (eds.), "Group Privacy: new challenges of data technologies", Dordrecht, Springer, 2017. On that point, see also, Y. POULLET, *Le RGPD face à l'intelligence artificielle*, Cahier du CRIDS, n° 49, Bruxelles, Larcier, 2020 et le rapport B. FRENAY and Y. POULLET, *Profiling and Convention 108+: Report on developments after the adoption of Recommendation (2010)13 on profiling*, (Rapport établi pour le Conseil consultatif de la Convention n° 108 en novembre 2019 dans le cadre de la révision de la recommandation de 2010 sur le profilage, publié sur le site du Conseil de l'Europe).

harm to an interest held by society at large”.²⁹ According to the text, we see a progressive enlargement, through ethical values, of the absolute need for legislators to take into account from now on in the context of the development of AI applications, not only the risks incurred by each of us individually (essentially restrictions on our individual freedoms), and those that affect groups of individuals, but also the risks that society at large may face. The discrimination that may be caused by Artificial Intelligence techniques involves, beyond individuals, groups of individuals united by gender and race, but also by the presence of genetic data that makes it possible to predict future illness or by residence in a neighborhood synonymous with the risk of terrorism or difficulties at school. Beyond these individual and collective risks, there are societal risks that we must envisage. *Fake news* is a risk to our democracy or the health of an entire society; private regulation by platforms creates a risk of normalization of behavior, which affects the vitality of our democracies;³⁰ our technological systems are energy-intensive and threaten the survival of the planet; AI weakens the rule of law, where it no longer allows legislators to challenge the truths that come out of computers and judges to exercise their role of controlling respect for the law. This broadening of concerns is vital and all the more so because, as CAHAI incidentally points out,³¹

²⁹ N. SMUHA, “Beyond the individual: Governing societal harms”, *Internet policy review - Special issue: “Governing “European values” inside data flows*, preprint. See also, J. E. COHEN, *Between Truth and Power: The legal Construction of Informational capitalism*, Oxford University Press, 2019.

³⁰ What the EU Commission proposal on the Digital Services Act (DSA) of December 20, 2020 (Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC) has called the three systemic risks linked with the activities of ‘very large platforms’: “Three categories of systemic risks should be assessed in-depth. A first category concerns the risks associated with the misuse of their service through the dissemination of illegal content A second category concerns the impact of the service on the exercise of fundamental rights, as protected by the Charter of Fundamental Rights, including the freedom of expression and information, the right to private life, the right to non-discrimination and the rights of the child. Such risks may arise, for example, in relation to the design of the algorithmic systems used by the very large online platform or the misuse of their service through the submission of abusive notices or other methods for silencing speech or hampering competition. A third category of risks concerns the intentional and, oftentimes, coordinated manipulation of the platform’s service, with a foreseeable impact on health, civic discourse, electoral processes, public security and protection of minors, having regard to the need to safeguard public order, protect privacy and fight fraudulent and deceptive commercial practices. Such risks may arise, for example, through the creation of fake accounts, the use of bots, and other automated or partially automated behaviours...” (Recital 57).

³¹ See n° 86: “Moreover, the societal dimension of AI’s risks that surpasses the impact on individuals, such as the impact on the electoral process and the democratic institutions or the legal system, is not yet sufficiently considered. While a number of national and

our regulatory arsenal, while it is well armed against what we have called individual risks,³² is much less so in relation to these collective or societal risks. Definitively, it is time that policymakers shift their attention from an individualistic one to a societal one, by recognizing the protection of societal interests as such and deeming them worthy of protection to be addressed by adequate legislation. On that precise point, we regret that the European Commission's Proposal on AI does not address more clearly these issues and only evokes traditional threats to Human Rights. In our opinion, our obligation to envisage societal risks justifies a right to public participation in the decision-making process about AI innovations that have a societal impact. The setting up of a commission carrying out a "data ethics" assessment and ensuring a public debate with the different stakeholders is to be welcomed (see *infra*, n° 12). Access to justice must be granted by the law to associations without them having to demonstrate individual damage (therefore being able to directly invoke the collective and societal interest they are promoting).

11. The risk-based approach related to the development of certain AI tools is particularly highlighted by European documents, whether they originate from the Council of Europe or the European Union. This approach³³ is also present, but more incidentally, in the UNESCO project.³⁴ The CAHAI text is clear and recommends: "that a future Council of Europe legal framework on AI should pursue a risk-based approach targeting the specific application context". The first recitals of the European parliamentary project are devoted to the need for an ethical and legal

international mechanisms allow individuals to seek redress before a court when a human right is breached in the context of AI, this mechanism is currently underdeveloped as regards an interference with democracy or the rule of law, which concern broader societal issues. Their protection necessitates public oversight over the responsible design, development and use of AI systems whenever such risks exist, by setting out clear obligations or requirements to this end".

³² The GDPR (the EU General data Protection Regulation (2016) and all the national institutions created throughout the world around the cause of data protection are a good example of this approach.

³³ On the other hand, the word "risk" is only used once in the OECD recommendations in its explanatory memorandum and in relation to the necessary public confidence: "... confidence is a key determinant of the digital transformation; that although it is difficult to predict the nature of future AI applications and their impacts, confidence in the reliability of AI systems is a key factor in the diffusion and adoption of AI; and that a well-informed public debate across society is necessary to realize the full potential of this technology while limiting the associated risks".

³⁴ It is indeed in the framework of the strategic measures (n° 50 *et seq.*) that the notion of "risk" is evoked and that measures to evaluate or even reduce these risks are proposed.

framework adapted to the new risks or those amplified by *machine learning*³⁵ systems.

This focus on risks explains the emphasis placed by the texts of UNESCO, CAHAI, and the European Parliament and Commission on the obligation to carry out a risk assessment. The UNESCO text remains vague on this subject and seems to envisage this obligation at a general level: “Member States should put in place impact assessments to identify and analyze the benefits and risks of AI systems and the issues they raise, as well as measures to prevent, mitigate and monitor risks. The ethical impact assessment should highlight the impact on human rights, including the rights of vulnerable groups, labor law, the environment and ecosystems, as well as ethical and social impacts in accordance with the principles set out in this document”. The CAHAI is more specific and attaches the obligation of assessment to each AI system and thus to all actors involved in the establishment and operation of an AI system. The document suggests that this evaluation may lead to corrective measures to reduce risks or even to prohibit certain applications that are too risky,³⁶ in accordance with the precautionary principle. The European institutions advocate for the distinction between “high-risk” systems and others.³⁷ In accordance with the principle of proportionality, which requires that any regulation should not impose a burden that is not necessary in view of its objectives, the draft reserves for “high-risk” systems, whether individual or collective,³⁸ obligations of security, transparency and reliability of the system on the

³⁵ Recital 1: “These technologies may involve opportunities and risks, which must be addressed and regulated by a general regulatory framework at Union level, which reflects the ethical principles to be respected from the time of their development and deployment to their use”.

³⁶ CAHAI, *Feasibility study on a legal framework on AI design, development and application based on CoE standards*, n° 42: “This means not only that the risks posed by AI systems should be assessed and reviewed on a systematic and regular basis, but also that any mitigating measures, that are further elaborated ..., should be specifically tailored to these risks. In addition to the risk-based approach, where relevant, a precautionary approach, including potential prohibitions, should be considered”.

³⁷ This is a distinction already present in the GDPR, from which the obligation of risk assessment is also taken up, an assessment which is certainly extended to all the risks described above and which is also incumbent on each of the actors making up the chain leading to a specific application (*i.e.* developers, deployers, users) and not only on the data controllers. Based on this distinction and its validity, the B. FRENAY and Y. POULLET, *Profiling and Convention 108+: Report on developments following the adoption of Recommendation (2010)13 on profiling*, Report to the Advisory Committee on Convention No. 108, Council of Europe, Strasbourg, November 7, 2019, T-PD(2019)07rev, pp. 38 *et seq.*

³⁸ The annex even provides for an exhaustive and cumulative list of high-risk applications using both the sector (*e.g.* education, health, banking and insurance, etc.) and purpose of the application (recruitment of staff, health care, school selection, etc.) criteria.

one hand and, on the other hand, quality controls of external data sources and protocols relating to the purposes of processing and the recipients of the data, which define and guarantee traceable and verifiable access to the data. The latest clarification provided by the European Union texts, both the EU Parliament Resolution and the EU Commission Proposal, is that this obligation to evaluate high-risk systems is preventive and continuous and requires a data governance mechanism. However, on a particular point, the two texts diverge: the first one refers to an external audit, entrusted to accredited auditors or administration involved in setting up³⁹ an AI system, whereas the EU Commission proposal only requires an internal evaluation as regards the conformity to the standards, even if the promotion of the use of standards developed by accredited bodies is mentioned.

12. A final reflection completes this overview of the selected documents. It concerns the role of the State as a driving force for public participation in the discussion on the challenges of AI and the choices to be made regarding its development. The OECD seems to reduce this role to simply informing the public, which is necessary to gain its trust: “trust is a key determinant of the digital transformation; that, although it is difficult to predict the nature of future AI applications and their impacts, confidence in the reliability of AI systems is a key factor in the diffusion and adoption of AI; and that a well-informed public debate across society is necessary to realize the full potential of this technology while limiting the risks associated with it”. The UNESCO Draft Recommendation is bolder (Recommendation No. 53): “Governments should adopt a regulatory framework that sets out a procedure for, in particular, public authorities to carry out impact assessments of AI systems in order to anticipate impacts, mitigate risks, avoid adverse consequences, facilitate citizen participation and address societal challenges. The study should also establish appropriate oversight mechanisms, including the principles of verifiability, traceability, and explicability, to evaluate algorithms, data and design processes, as well as include an external review of AI systems. Ethical impact assessments conducted by public authorities should be transparent and open to the public. They should also be multidisciplinary, multi-stakeholder, multicultural, pluralistic, and inclusive. Member States are encouraged to put in place mechanisms and tools, such as regulatory sandboxes or

³⁹ It is added that this external audit system differs from the internal evaluation system by the controller (with the exception of the consultation of the data protection authority, set up by the GDPR). To this end, certificates of conformity are issued by these bodies. The certificate is compulsory for high-risk technologies, it is produced at the request of the potential beneficiary, for other technologies. The UNESCO text (n° 53) also seems to defend the idea of an external evaluation at least for AI applications of public authorities.

assessment centers, which should enable impacts to be monitored and assessed in a multidisciplinary and multi-stakeholder manner. Public authorities should be required to monitor the AI systems they implement and/or deploy, establishing appropriate mechanisms and tools". The proposal therefore aims at the creation of *Technology Assessment-type* bodies. On the one hand, it is a question of analyzing the impacts of AI in general or in particular of this or that type of application (e.g. facial recognition, intelligent cars, etc.) and of advocating measures to supervise experiments. These bodies would also serve as evaluation bodies for public sector applications. It is stressed that these bodies must ensure that they include the participation of different interests in discussions and confront different disciplinary points of view.

We find the same emphasis in the CAHAI document (n° 115): "Where relevant and reasonably possible, member States should ensure a meaningful participatory approach and the involvement of different stakeholders (from civil society, the private sector, academia and the media) in the decision-making processes concerning the deployment of AI systems in the public sector, with special attention to the inclusion of under-represented and vulnerable individuals and groups, which is key to ensuring trust in the technology and its acceptance by all stakeholders". The paper further stresses the general supervisory role that public and judicial authorities must have in relation to all AI systems (No 113): "Based on a risk-based approach, effective public oversight and control mechanisms must be guaranteed, to ensure that AI developers and deployers act in compliance with relevant legal requirements, while allowing for intervention by state authorities when it does not happen". It is undoubtedly in the European Parliament's Draft Regulation that we can find the most complete reflections on the institutionalization of "national supervisory bodies". Recitals 43 *et seq.* state: "The Member States should designate an independent administrative authority as a supervisory body"⁴⁰. In particular, each national supervisory body should be responsible for identifying Artificial Intelligence, robotics and related technologies considered to be of high risk in the light of the risk assessment criteria set out in this Regulation, and for assessing and monitoring the compliance of these technologies with the obligations set out in this Regulation. Each national supervisory body should also be responsible for regulating the good governance of those technologies under the coordination of the Commission

⁴⁰ Some countries did not wait for the Regulation in order to set up such bodies. Thus, *Data Ethics Commissions* exist in Denmark, Germany, and the United Kingdom. On these bodies, see our book, *Le RGPD face à l'intelligence artificielle*, Cahier du CRIDS, n° 49, Larcier, Bruxelles, 2020, pp. 151 and 152.

and/or any other competent Union institution, body, office, or agency designated for that purpose. They therefore have an important role to play in fostering the trust and security of the citizens of the Union, as well as in enabling the building of a democratic, pluralist, and equitable society. The text insists on multi-stakeholder membership and these bodies should be the forum for debate between the various interested parties, be they from the world of research, business or civil associations. It must be the *first point of contact in the event of a presumed breach of the obligations arising from the text* and be the relay for citizens' rights, particularly in the event of discriminatory treatment. It is added that these bodies "should provide administrative and professional guidance and support to developers, deployers and users, in particular small and medium-sized enterprises or start-ups which have difficulties in complying with the ethical principles and legal obligations set out in this Regulation".

13. At the end of this rapid analysis, what can we conclude from this recent call for ethics to control the risks linked to the development of AI? Should we, like OCHIGAME,⁴¹ see it as an invention intended to make the controversial applications of AI acceptable and to restore the trust of the citizen? We do not think so. Ethics, as we have said, finds its extensions in human rights texts and its principles will thus support future legislation or even contain such legal prescriptions intended to put them into effect, as is the case with the Regulation proposed by the European Parliament.⁴² Case law will soon draw from recommendations and *best practices* the source of what constitutes the duties of the "reasonable man", in this case those who participate in the establishment of an AI system. Beyond this, the contribution of the texts has a twofold merit: the first is that they singularly broaden the field of risks linked to the development of AI applications and underline that these are "high risks". Are our contemporary concepts of human rights centered on the individual still relevant insofar as it is our environment, our democratic structures, and the rule of law itself that are under substantial threat?⁴³ Second, these documents indicate the solution to such challenges and our collective responsibility to find it. The insistence, based on the principles

⁴¹ R. OCHIGAME, "The invention of 'Ethical AI'", *The Intercept*, December 20, 2019, quoted by Y. MENECEUR, *L'intelligence artificielle en procès, op. cit.*, p. 220.

⁴² Cf. also the transparency obligations imposed by the European Commission's draft regulation called the *Digital Service Act (DSA)* and relating to the AI systems set up by the platforms.

⁴³ On these dangers and the need to go beyond a purely individualistic approach, see A. BASDEVANT, J.P. MIGNARD, *L'empire des données. Essai sur la société, les algorithmes et la loi*, Paris, Don Quichotte, 2018.

TIME TO RESHAPE THE DIGITAL SOCIETY. 40TH ANNIVERSARY OF THE CRIDS

of both public participation and precaution, on the need to organize multidisciplinary, open debates between all interested parties with a view to finding both ethical and legal responses seems to me to be common to all the texts cited. We can only be delighted to hear, at the last Council of Ministers of the Council of Europe, in February 2021, CAHAI's call for a conjunction of all the international organizations' efforts in this direction. Let us hope that it will be heard.