

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

About some international documents relating to the ethics of artificial intelligence

Poullet, Yves

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (HARVARD):

Poullet, Y 2021, 'About some international documents relating to the ethics of artificial intelligence: Some insights : accelerating Actions and Promoting Digital Wellness (DW) in the context of Artificial Intelligence(AI), 2 days Conference organized by Univ. of Hyderabad together with Information Ethics Network @ Future Africa (Univ. of Pretoria)', pp. 1-10.

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.

About some international documents relating to the ethics of artificial intelligence – Some insights.

- *Accelerating Actions and Promoting Digital Wellness (DW) in the context of Artificial Intelligence(AI)*, 2 days Conference organized by Univ. of Hyderabad together with Information Ethics Network @ Future Africa (Univ. of Pretoria) -

Yves Pouillet, 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. The ethics demanded for the development of AI meets the concerns of many public (OECD, UNESCO, Council of Europe, Arab League, ...) ¹and private² international bodies. It is at the heart of the "third way" called for by the European Union concerning an artificial intelligence (in short AI) characterised 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 organisations. This article will refer to these initiatives in not much detail³ 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 do this on a few selected themes. The first is the very nature of the proposed regulation. Is asserting ethical principles sufficient or does 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 of 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 pursues this second reflection: the approach centred 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

¹ See also the ITU World Summit on Artificial Intelligence for the Public Interest (AI for Good). On this Summit, 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 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 favour of these ethical frameworks, pp. 199 et seq.

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

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 risk 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?

2. 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. ROUVROY⁴. 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. The control of migrants at European borders will soon use artificial intelligence systems combining graphological analysis, facial recognition, the use of megadata that can calculate the risk of terrorism presented by a candidate for immigration⁶. 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 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.
3. 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 on which many other documents have been based since then⁸. As far as UNESCO is concerned, the 40th General Conference in its Resolution 40 C/37, following the example of what had been achieved in the field of bioethics, called for an international standard-setting instrument, a

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

⁵ C.VILLANI, Report to the Prime Minister of the French Republic, "Giving meaning to artificial intelligence. For a national and European strategy", 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 as follows 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).

⁷ Council Recommendation on Artificial Intelligence, adopted by the OECD Ministerial Council on 22 May 2019, available at <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.

recommendation on the ethics of artificial intelligence. A draft recommendation⁹ prepared by a so-called *ad hoc* group of experts is currently being submitted for consideration by national delegations. Meetings are planned for April and June of this year to finalise the text which, if necessary, will be presented for adoption at the next General Assembly. On the Council of Europe side, the Committee of Ministers appointed, on 11 September 2019, an equally *ad hoc* Committee of Experts: the CAHAI¹⁰. 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 17 December 2020 by the CAHAI of the "*Feasibility study on a legal framework for the creation, development and application of AI based on Council of Europe standards*"¹¹ is noteworthy. The particular interest of this text will be underlined in the rest of the article.

The European Union is not to be outdone. If certain documents under the previous Presidency, had already sensed the importance of the topic, it is clear that it is to the current 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 centred 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 system¹⁴. In response to this ethical impetus from the Commission and in full consultation with the latter, the European Parliament adopted a resolution on 20 October 2020 containing recommendations to the Commission on a framework for the ethical aspects of artificial intelligence, robotics and related technologies¹⁵. It is emphasised that the text is more than a resolution, it contains a proposal

⁹ PRELIMINARY REPORT ON THE DRAFT RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE, available at: 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>.

¹² Statement by U. van der LEYEN, President of the Commission at the announcement of the European Union's strategic plan for artificial intelligence, 19 February 2020. This statement is to be read in conjunction with the OECD text, which in its preamble notes: "RECOGNISING 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, 19 February 2020, COM(2020)65 final, available at: 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*, 8 April 2019, n° 67, text available at: [Ethics guidelines for trustworthyAI - Publications Office of the EU \(europa.eu\)](https://publications.ec.europa.eu/en/ethical-guidelines-for-trustworthy-ai)

¹⁵ European Parliament Resolution of 20 October 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.

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.

4. 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, developed and operate, while respecting existing laws, some of which are criticized for their inadequacy. How can we understand this reference in certain texts to ethics and this refusal of any public regulation? 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 adopt 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 formalises a certain number of principles of action and "minimum" standards. By publishing their code of conduct, the company or the professions undertake to observe these standards and to ensure that they are observed by their subcontractors and suppliers. The code of ethics can be analysed 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 alone¹⁶. 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 *in fine*, was judged to be too "soft" in this respect. The preliminary report of the Director-General of UNESCO (n°27) states¹⁷: "Last but not least, it was suggested that the draft Recommendation be more ambitious. This implies making bold proposals and being more assertive in suggesting that a stronger international legal framework is needed. ». 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 societies¹⁸ and this approach amply justifies the Parliament's proposal for a binding Regulation. 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

¹⁶ A ce propos, parmi d'autres, lire B. WAGNER, " *Ethics as an Escape from Regulation: From Ethics-Washing to Ethics-Shopping?* ", in Hildebrandt M. (éd.), *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. »*

¹⁷ 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 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. ».

¹⁸ 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;...* "

to ethical issues¹⁹: "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 commercialising 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 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 manoeuvre in the application of these principles²⁰. 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 framework of **ethical principles and legal obligations** relating to the development, deployment and use of artificial intelligence, robotics and related technologies within the Union". Finally, it should be recalled²¹ that traditional ethical principles, dignity, autonomy and social justice 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.

5. What ethical values are enshrined in these documents²²? One is struck by the number of values "discovered" by the documents. They mention dignity, prevention of environmental damage,

¹⁹ Feasibility Study, *op.cit.*, n°79; voir également, 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). ..."

²⁰ The fight against disinformation, studied in another article in the same issue (Y. POULLET and 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, 31 December 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 recognised in the field (in particular economic operators, social partners, non-governmental organisations or associations)". On this notion, the distinction between self-regulation, bottom-up and top-down co-regulation, our reflections in Y. POULLET, "Information and communication technologies and 'co-regulation': a new approach?" *Liber amicorum Michel Coipel*, Brussels, Kluwer, 2004, p. 173.

²¹ On this point, our reflections in the book cited in note 3.

²² 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-centred 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, humane 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 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*.

individual autonomy, non-discrimination, sometimes including gender equality, transparency or explainability, the robustness of the systems used, including their reliability, privacy, the responsibility of actors, democracy, respect for human rights, etc. The UNESCO Convention on Bioethics²³ was based on the four universal values proposed by Beauchamp and Childress²⁴: dignity, autonomy, social justice and 'beneficent and non-maleficent' technologies. This profusion of values seems to us to be dangerous and the result of regrettable confusion. Thus, it would certainly be useful to group together around the value of social justice, the imperative of non-discrimination, gender equality, that the definition of "privacy" by the case law of the Council of Europe refers to autonomy, ... especially it seems to us that the list mixes values and the means to ensure their respect. Thus, 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. Extending the list is a source of confusion for the public and, above all, of devaluation of true ethical values.

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 distinction that emerges between the risks incurred by each of us individually, essentially restrictions on our individual freedoms, and those that affect groups of individuals or even our society. The discrimination that may be caused by artificial intelligence techniques involves, beyond individuals, groups of individuals united by gender, race, but also by the presence of genetic data that makes it possible to predict future illness or by residence in a neighbourhood synonymous with the risk of terrorism or difficulties at school. *Fake news* is a risk to our democracy or the health of an entire society; private regulation by platforms creates a risk of normalisation of behaviour, which affects the vitality of our democracies; our technological systems are energy-intensive and threatens the survival of the planet; artificial intelligence 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 out²⁵, our regulatory arsenal, while it is well armed against what we have called individual risks²⁶, is much less so in relation to these collective risks.

6. 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 approach²⁷ is also present, but more incidentally, in the UNESCO

²³ Universal Declaration on Bioethics and Human Rights, adopted on 19 October 2005, text available on the UNESCO website (http://portal.unesco.org/fr/ev.php-URL.ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html).).

²⁴ Beauchamp s T. L. et Childress J. F., *Principles of Biomedical Ethics*, 3e éd., New York, Oxford University Press, 2001.

²⁵ Voir le 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 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."

²⁶ 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.

²⁷ 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

project²⁸. 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 framework adapted to the new risks or those amplified by *machine learning*²⁹ systems.

This focus on risks explains the emphasis placed by the texts of UNESCO, CAHAI and the European Parliament on the obligation of 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 analyse 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, labour 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 risky³⁰, in accordance with the precautionary principle. The European Parliament advocates the distinction between "high-risk" systems and others³¹. 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,³² obligations of security, transparency and reliability of the system on the 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 text is that this obligation to evaluate high-risk systems is preventive, continuous and above all entrusted to accredited auditors external to the company or administration involved in setting up³³ an AI system.

factor in the diffusion and adoption of AI; and that a well-informed public debate across society is necessary to realise the full potential of this technology while limiting the associated risks. »

²⁸ 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.

²⁹ 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. »*

³⁰ CAHAI, *op. cit.*, 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.*"

³¹ 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 - 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 7 November 2019, T-PD(2019)07rev, pp. 38 et seq.

³² 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.

³³ 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

7. 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 realise 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 assessment centres, 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 analysing the impacts in general of AI or in particular of this or that type of application (e.g. facial recognition, intelligent cars, etc.) and of advocating measures to supervise experiments. It is also an evaluation body for public sector applications. It is stressed that these bodies must ensure that they include the participation of different interests in discussions and confront 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 on the side of the European Parliament's draft Regulation that the reflections on the institutionalisation of "national supervisory bodies" are the most complete. Recitals 43 et seq. state: "*The Member States should designate an independent administrative authority as a supervisory body*³⁴ ...". 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

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.

³⁴ Some countries did not wait for the Regulation 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, p. 151 and 152.

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 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 confidence 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. ».

8. 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 OCHIGAME³⁵, see it as an invention intended to make the controversial applications of AI acceptable and to restore the confidence 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 Parliament³⁶. Case law will soon draw from recommendations and *best practices* the source of what constitutes the duties of the 'good father', 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 threat³⁷? Second, these documents indicate the solution to such challenges and our collective responsibility to find it. The insistence, based on the principles of both public participation and precaution, on the need to organise 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 efforts of international organisations in this direction. Let us hope that it will be heard.

³⁵ R. OCHIGAME, "The invention of 'Ethical AI'", *The Intercept*, 20 December 2019, quoted by Y. MENECEUR, *op.cit.*, p. 220.

³⁶ 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.

³⁷ On these dangers and the need to go beyond a purely individualistic approach, A. BASDEVANT et J.P. MIGNARD, *L'empire des données, Essai sur la société, les algorithmes et la loi*, Paris, Don Quichotte, 2018

