

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

À propos de quelques documents internationaux relatifs à l'éthique de l'intelligence artificielle

Poullet, Yves

Published in:

Revue du Droit des Technologies de l'information

Publication date:

2021

Document Version

le PDF de l'éditeur

[Link to publication](#)

Citation for pulished version (HARVARD):

Poullet, Y 2021, 'À propos de quelques documents internationaux relatifs à l'éthique de l'intelligence artificielle: pistes de réflexion', *Revue du Droit des Technologies de l'information*, numéro 81, pp. 5-14.

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.

ÉDITORIAL

À propos de quelques documents internationaux relatifs à l'éthique de l'intelligence artificielle – Pistes de réflexion

Yves Pouillet

1. L'éthique réclamée à propos du développement de l'intelligence artificielle (en abrégé: IA) rencontre les préoccupations de nombre d'instances internationales publiques (OCDE, UNESCO, Conseil de l'Europe, Ligue arabe, ...)¹ et privées². Elle est au cœur de la « troisième voie » réclamée par l'Union européenne à propos d'une IA caractérisée par les mots: « Excellence et Confiance ». Notre propos n'est pas de détailler tous les instruments mis en place ou projetés par les organisations publiques internationales. Nous les citerons sans les détailler outre mesure dans le cadre de cet article³ et ne retiendrons que celles de l'OCDE, de l'UNESCO, du Conseil de l'Europe et de l'Union européenne. La seule constatation de la multiplication des documents en matière d'éthique de l'IA mérite qu'on s'interroge sur le « Pourquoi ? » d'une telle efflorescence. Au-delà, ce qui retient notre attention, ce sont les tendances lourdes, communes ou non à ces documents.

Nous le ferons à propos de quelques thèmes choisis. Le premier est la nature même de la régulation proposée. Affirmer des principes éthiques est-il suffisant ou l'éthique renvoie-t-elle *in fine* à la réclamation d'un cadre juridique ? Au-delà, le modèle régulateur renvoie-t-il à de l'autorégulation, de la corégulation ou de la régulation publique ? Sur ce point, les points de vue sont largement différents. Le second entend analyser les principes éthiques affirmés. Leur liste est souvent longue, témoigne de confusion entre valeurs éthiques et moyens d'assurer le respect de ces valeurs. Ce qui frappe, c'est que dans tous les cas, les valeurs proclamées entendent déborder la protection d'intérêts et de libertés individuels pour mettre en évidence les enjeux collectifs du développement de l'IA, la justice sociale, la démocratie, l'environnement. Certains documents soulignent que si les enjeux individuels semblent bien couverts par de nombreux textes qui rendent effective la protection des libertés individuelles, par contre, l'absence d'effectivité dans la mise en œuvre

¹ Voy. également le Sommet mondial de l'UIT sur l'intelligence artificielle au service de l'intérêt général (AI for Good). Sur ce Sommet, consulter le site: <https://www.itu.int/fr/mediacentre/backgrounders/Pages/artificial-intelligence-for-good.aspx>.

² Ainsi, parmi de nombreux textes, ceux de l'Institut d'ingénierie électrique et électronique (IEEE): « *IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems* (The IEEE Global Initiative), accessible sur le site de l'IEEE: <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html> » et la « Déclaration de Montréal sur l'intelligence artificielle », qui résulte d'une initiative académique: <https://ia-ethique.be/declaration-montreal-intelligence-artificielle/>. Le lecteur trouvera une liste impressionnante des « cadres éthiques ou non contraignants applicables à l'intelligence artificielle et à la science des données », in Y. MENECEUR, *L'intelligence artificielle en procès*, Bruxelles, Bruylant, 2020, Appendice, pp. 409 à 417 et lira son analyse très critique de ce mouvement en faveur de ces cadres éthiques, pp. 199 et s.

³ Sur ces textes et nos réflexions, voy. Y. POULLET, *Ethique et Droits de l'homme dans notre société du numérique*, Bruxelles, Académie royale de Belgique, 2020.



des enjeux collectifs pose problème. Le troisième poursuit cette seconde réflexion : l'approche centrée sur les risques encourus soit individuellement par les individus, soit collectivement par nos sociétés conduit à s'interroger sur la nécessité de distinguer suivant la gravité de ces risques, à mettre l'accent sur les procédures d'évaluation de ces risques, sur les acteurs à la base de ces risques et leurs devoirs de collaborer à la prise en compte et à la maîtrise de ceux-ci et, enfin, sur le rôle de l'État en la matière. Sur ce dernier champ de réflexion, on notera la hardiesse des textes européens, prônant y compris la création de commissions nationales de « Data Ethics ».

Il sera alors temps de risquer certaines conclusions et de répondre à cette interrogation : quelle est la valeur de cette approche dite éthique ? « *Ethics bashing* » ou véritable piste de solution ?

2. L'intelligence artificielle est un « *buzzword* », objet de nombreux fantasmes évoquant une société tant de transhumanisme que de totale transparence des personnes et de « *gouvernementalité purement algorithmique* » de nos sociétés et de nos vies individuelles, selon l'expression d'A. Rouvroy⁴. Villani⁵ la décrit comme « une des clés du pouvoir de demain dans un monde numérique ». Sans doute, ces prophéties sont-elles exagérées. Il n'empêche des scandales récents comme celui du *Cambridge Analytica*, du profilage des enfants des écoles communales de Bristol sur la base du calcul des performances scolaires passées et futures à des fins de sélection scolaire mais également commerciales, nous obligent à nous interroger. On sait que l'IA aide les dirigeants d'entreprise, voire se substitue à eux dans la sélection des candidats employés ; permet à des agences immobilières de sélectionner les personnes intéressées par un logement en fonction des préférences exprimées par les propriétaires. Le contrôle des migrants aux frontières européennes recourra sous peu à des systèmes d'intelligence artificielle alliant l'analyse graphologique, la reconnaissance faciale, l'utilisation de mégadonnées pouvant calculer le risque de terrorisme présenté par un candidat à l'immigration⁶. Une société multinationale bien connue annonce la possibilité grâce à l'IA de détecter chez des internautes, de manière précoce, la maladie d'Alzheimer, et ce à partir des messages envoyés et de la frappe sur le clavier. Bref, les capacités en particulier prédictives offertes dès maintenant par l'IA justifient l'interrogation de nos sociétés sur les limites à imposer aux concepteurs et exploitants publics et privés de systèmes d'intelligence artificielle au nom de valeurs éthiques essentielles.

3. Cette interrogation justifie le nombre croissant de rapports et souvent de recommandations publiés sur l'IA et l'éthique. La Recommandation de l'Organisation de coopération et de développement économiques (OCDE)⁷, sur l'intelligence artificielle émise dès 2019 par le Conseil des ministres représente la première norme intergouvernementale pour les politiques relatives à l'IA

⁴ A. ROUVROY, « L'homo juridicus est-il soluble dans les données ? », in E. DEGRAVE et al. (éds), *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, Rapport au premier ministre de la République française, « Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne », mars 2018, p. 11.

⁶ Ainsi, par exemple, le système « I-Border Ctrl » actuellement développé par l'Europe pour le contrôle des frontières, dont les modules sont décrits comme suit par le rapport d'Algorithmwatch et de la fondation Bertelsman (*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*, Janvier 2019).

⁷ Recommandation du Conseil sur l'intelligence artificielle, adoptée par le Conseil des ministres de l'OCDE, le 22 mai 2019, disponible à l'adresse suivante : <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449>.



et constitue un socle sur lequel s'appuient depuis bien d'autres documents⁸. En ce qui concerne l'UNESCO, la 40^e Conférence générale a souhaité par sa Résolution 40 C/37, à l'instar de ce qui avait été réalisé en matière de bioéthique, un instrument normatif international, soit une recommandation sur l'éthique de l'intelligence artificielle. Un projet de recommandation⁹ préparé par un groupe d'experts dits *ad hoc* est actuellement soumis à l'examen des délégations nationales. Des réunions sont prévues en avril et en juin de cette année pour finaliser le texte qui, le cas échéant, sera présenté à l'adoption de la prochaine assemblée générale. Du côté du Conseil de l'Europe, le Comité des ministres a nommé, le 11 septembre 2019, un Comité d'experts également *ad hoc*: le CAHAI¹⁰. Le Comité, présidé par une belge, bien connue de notre rédaction, la professeure Peggy Valcke, est « chargé d'étudier la faisabilité et les éléments potentiels, sur la base de larges consultations multipartites, d'un cadre juridique pour le développement, la conception et l'application de l'intelligence artificielle, fondé sur les normes du Conseil de l'Europe en matière de *droits de l'homme, de démocratie et d'État de droit* ». On souligne l'adoption récente soit le 17 décembre 2020, par le CAHAI de l'« Étude de faisabilité relative à un cadre juridique pour la création, le développement et l'application de l'IA sur la base des normes du Conseil de l'Europe »¹¹. L'intérêt particulier de ce texte sera souligné dans la suite de l'exposé.

L'Union européenne n'est pas en reste. Si certains documents, sous l'ancienne Présidence, avaient déjà pressenti l'importance du sujet, il est clair que c'est à la Présidence actuelle que l'on doit cette volonté d'un encadrement éthique du développement de l'IA, comme pierre de touche de cette fameuse troisième voie que l'Europe entend suivre à distance tant du modèle américain que celui chinois. « Nous voulons que l'application de ces nouvelles technologies soit digne de la confiance de nos citoyens [...] Nous encourageons une approche responsable de l'intelligence artificielle centrée sur l'humain »¹². Cette réflexion introduit et résume bien le Livre blanc consacré par la Commission à sa stratégie en la matière que les mots « Excellence and Trust »¹³ résument parfaitement. Ce Livre blanc s'appuie notamment sur les travaux d'un groupe d'experts de haut niveau (*HLGE on AI*) commandité par la Commission, qui avaient abouti en avril 2019 à des recommandations éthiques pour un système d'intelligence artificielle digne de confiance¹⁴. À cette impulsion éthique de la Commission et en pleine concertation avec cette dernière, le Parlement européen a entendu répondre par une résolution du 20 octobre 2020 contenant des recommandations à

⁸ En juin 2019, à l'occasion du Sommet d'Osaka, les dirigeants du G20 ont accueilli les Principes du G20 sur l'IA, issus de la Recommandation de l'OCDE.

⁹ Rapport préliminaire sur l'avant-projet de recommandation sur l'éthique de l'intelligence artificielle, disponible à l'adresse : https://unesdoc.unesco.org/ark:/48223/pf0000374266_fre.

¹⁰ Soit Comité Ad Hoc Artificial Intelligence dont les travaux sont accessibles sur la page : www.coe.int/cahai.

¹¹ Disponible à l'adresse : <https://rm.coe.int/cahai-2020-23-final-etude-de-faisabilite-fr-2787-2531-2514-v-1/1680a1160f>.

¹² Déclaration d'U. van der Leyen, Présidente de la Commission lors de l'annonce du plan stratégique de l'Union européenne en matière d'intelligence artificielle, le 19 février 2020. On rapprochera cette déclaration de celle du texte de l'OCDE qui dans son préambule note : « RECONNAISSANT que compte tenu de l'évolution et de la mise en œuvre rapides de l'IA, il est nécessaire de bâtir un cadre d'action stable, qui privilégie une IA digne de confiance centrée sur l'humain, ... ».

¹³ Livre Blanc de la Commission sur l'intelligence artificielle (*Livre blanc – Intelligence artificielle – Une approche européenne basée sur l'excellence et la confiance*, Bruxelles, 19 février 2020, COM(2020)65 final, disponible sur le site : 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, *Lignes directrices en matière d'éthique pour une IA digne de confiance*, 8 avril 2019, n° 67, texte disponible sur le site : Ethics guidelines for trustworthy AI – Publications Office of the EU (europa.eu).



la Commission concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes¹⁵. On souligne que le texte constitue plus qu'une résolution, il contient une proposition de Règlement, bref un « prêt-à-signer » pour la Commission et le Conseil qui, par ailleurs, pourront difficilement s'écarter de ce qui a constitué l'accord du Parlement.

4. Nombre des textes sont baptisés «Principes éthiques» et ne font pas nécessairement référence à la nécessité d'un encadrement réglementaire. Il s'agit de recommandations fondées sur des valeurs éthiques autour desquelles doivent se concevoir, s'élaborer et fonctionner les systèmes d'IA, certes dans le respect des lois existantes dont l'inadéquation de certaines est par ailleurs dénoncée. Comment comprendre cette référence de certains textes à l'éthique et ce refus de toute réglementation publique? L'éthique renvoie à l'agir, au « bien agir », elle entend que l'homme et ses « artefacts » agissent ou soient conçus pour le Bien et le Juste. L'éthique, selon Spinoza, n'est cependant pas l'affirmation d'une morale particulière. Elle est une démarche appuyée sur des valeurs et ne reçoit pas une réponse unique. Le pluriel de la traduction anglaise du mot, *Ethics*, exprime bien la singularité et la pluralité des réponses éthiques. L'éthique se veut un questionnement, une recherche individuelle et parfois collective à propos de l'attitude à adopter face à des faits, des réalités ou des changements sociaux, comme ceux apportés par l'intelligence artificielle. On distinguera l'éthique des codes d'éthique. Dans le monde des activités économiques ou professionnelles, un code de conduite ou d'éthique constitue une déclaration publique par leurs auteurs et signataires des valeurs et des pratiques suivies. Le code formalise un certain nombre de principes d'action et de normes « minimales ». En publiant leur code de conduite, l'entreprise ou les professions s'engagent à observer ces normes et à les faire observer par leurs sous-traitants et fournisseurs. Le code d'éthique s'analyse comme une *soft law* à la valeur obligatoire incertaine, relayée ou non, explicitement ou implicitement, par le droit. L'éthique pourrait donc signifier l'alibi pour refuser toute réglementation contraignante nouvelle et justifier le recours à la seule autorégulation¹⁶. Sans doute, la recommandation de l'OCDE, nonobstant la qualité de son contenu, peut se lire en ce sens. L'OCDE a toujours été défenseur d'une autorégulation sérieuse. Le projet actuel déposé par les experts de l'UNESCO, essentiellement basé sur des recommandations *in fine* non contraignantes, a été jugé comme trop « mou » sur ce plan. Le rapport préliminaire de la Directrice générale de l'UNESCO (n° 27) indique¹⁷: « Enfin et surtout,

¹⁵ Résolution du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes (2020/2012(INL)) P9 TA (2020) 0275. Ce texte a reçu l'appui de la Commission européenne. Il est à noter que la Résolution contient une proposition de règlement qui n'attend que d'être reprise par les autres autorités de l'Union européenne. Sur ce texte, voy. notre Brève.

¹⁶ À ce propos, parmi d'autres, voy. B. WAGNER, « Ethics as an Escape from Regulation: From Ethics-Washing to Ethics-Shopping? », in M. HILDEBRANDT (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 Dans le texte du projet, les références au besoin d'un cadre juridique pour relever les défis éthiques sont rares. Voy. toutefois, le point 9 du projet: « Les valeurs et principes énoncés ci-dessous doivent être respectés par tous les acteurs du cycle de vie des systèmes d'IA, en premier lieu, et être promus par l'élaboration de nouvelles législations, réglementations et directives commerciales et la modification de celles qui existent déjà ».

il a été suggéré que le projet de Recommandation soit plus ambitieux. Cela implique de faire des propositions audacieuses et d'être plus affirmatif en suggérant qu'un cadre juridique international plus fort est nécessaire». Les textes européens sont plus clairs à cet égard. La résolution du Parlement européen réclame fermement un cadre réglementaire pour répondre aux défis éthiques que l'IA pose à nos sociétés¹⁸ et cette approche justifie amplement la proposition par le Parlement d'un Règlement contraignant. Dans son rapport sur la faisabilité déjà cité, le CAHAI souligne l'insuffisance d'une approche fondée sur la seule éthique et condamne l'autorégulation comme solution aux enjeux éthiques¹⁹: «*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*». Ainsi, le Conseil de l'Europe, s'il conçoit les valeurs éthiques comme le moteur d'une réflexion sur les risques encourus par l'IA, plaide clairement pour une approche de corégulation descendante où l'autorité publique fixe les principes réglementaires quitte à laisser aux acteurs privés une certaine marge de manœuvre dans l'application de ces principes²⁰. Quant au projet de règlement sur les aspects éthiques de l'IA du Parlement européen, on note que son objet est «*d'établir un cadre réglementaire complet et pérenne de principes éthiques et d'obligations juridiques* relatifs au développement, au déploiement et à l'utilisation de l'intelligence artificielle, de la robotique et des technologies connexes

¹⁸ Cfr résolution du Parlement européen, Préambule point L.: «*considérant qu'il convient non seulement d'adapter la législation existante, mais également de traiter les questions juridiques et éthiques liées aux technologies de l'IA au moyen d'un cadre réglementaire efficace, complet et durable inscrit dans le droit de l'Union, reflétant les valeurs et les principes de l'Union tels que consacrés par les traités et la Charte des droits fondamentaux de l'Union européenne (ci-après dénommée "Charte"), qui se limiterait à combler les lacunes juridiques actuelles en évitant toute réglementation excessive et qui renforcerait la sécurité juridique pour les entreprises comme pour les citoyens, notamment en prévoyant des mesures obligatoires, ...*». Cfr également le point Y.: «*considérant que les principes éthiques communs ne sont efficaces que lorsqu'ils sont également inscrits dans le droit et que les parties responsables de la garantie, de l'évaluation et du contrôle de la conformité sont identifiées*;».

¹⁹ Feasibility Study, *op. cit.*, n° 79; voy. 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 La lutte contre la désinformation étudiée dans un autre article de ce même numéro (Y. Poullet et N. Bontridder, «*L'Union européenne et la régulation de la désinformation*») est un exemple du passage nécessaire de l'autorégulation à une corégulation descendante. La corégulation descendante est celle que, dès 2003 dans le cadre de l'accord interinstitutionnel «*Mieux légiférer*» (*J.O.C.E.*, C 321, 31 décembre 2003, pp. 1-5.), les autorités européennes avaient souhaité. L'article 18 de l'Accord définissait la *corégulation*, comme «*le mécanisme par lequel un acte législatif communautaire confère la réalisation des objectifs définis par l'autorité législative aux parties concernées reconnues dans le domaine (notamment les opérateurs économiques, les partenaires sociaux, les organisations non gouvernementales ou les associations)*». Sur cette notion, la distinction entre autorégulation, corégulations ascendante et descendante, nos réflexions in Y. Poullet, «*Technologies de l'information et de la communication et 'corégulation': une nouvelle approche?*», *Liber amicorum Michel Coipel*, Bruxelles, Kluwer, 2004, p. 173.



au sein de l'Union ». Enfin, on rappellera²¹ que les principes éthiques traditionnels, la dignité, l'autonomie, la justice sociale reçoivent dans les textes européens contraignants relatifs aux droits humains (Convention européenne des droits de l'homme et Charte des droits fondamentaux de l'Union européenne) une consécration qui assure leur effectivité tant par le devoir des États de leur donner une signification concrète, que par l'action des tribunaux de Strasbourg et de Luxembourg.

5. Quelles valeurs éthiques consacrent ces documents²²? On est frappé par le nombre de valeurs « découvertes » par les documents. On cite pêle-mêle la dignité, la prévention des dommages environnementaux, l'autonomie des individus, la non-discrimination dont parfois on distingue l'égalité des genres, la transparence ou l'explicabilité, la robustesse des systèmes utilisés y compris leur fiabilité, la vie privée, la responsabilité des acteurs, la démocratie, le respect des droits de l'homme, ... La Convention de l'UNESCO sur la bioéthique²³ se basait sur les quatre valeurs universelles proposées par Beauchamp et Childress²⁴ : dignité, autonomie, justice sociale et technologies « *beneficent and non maleficent* ». Cette profusion de valeurs nous semble dangereuse et résulter de confusions regrettables. Ainsi, il est certain qu'il serait utile de regrouper autour de la valeur justice sociale, l'impératif de non-discrimination, d'égalité des genres, que la définition de la « *privacy* » par la jurisprudence du Conseil de l'Europe renvoie à l'autonomie, ... surtout il nous apparaît que la liste mélange des valeurs et les moyens d'assurer leur respect. Ainsi, la transparence n'est pas une valeur en soi mais une manière de rendre effectives les valeurs d'autonomie ou de non-discrimination. La prudence ou principe de précaution, de même que le respect de l'environnement s'indiquent si on souscrit à la valeur « *Do good and do not harm* », etc. Allonger la liste est source de confusions pour le public et surtout de dévalorisation des vraies valeurs éthiques.

Mais là n'est pas le point le plus important que révèle la liste de principes éthiques certes longue mais instructive relevée dans les différents documents. Ce qui importe, c'est la distinction qui apparaît entre les risques encourus par chacun de nous individuellement, essentiellement les restrictions à nos libertés individuelles et ceux qui affectent des groupes d'individus voire notre société. Les discriminations que peuvent engendrer les techniques d'intelligence artificielle

²¹ Sur ce point, nos réflexions dans l'ouvrage cité note 3.

²² À noter la liste proposée par l'OCDE : « elle (la Recommandation) énonce cinq principes complémentaires fondés sur des valeurs, jetant les bases d'une approche responsable à l'appui d'une IA digne de confiance, et appelle les acteurs de l'IA à les promouvoir et les mettre en œuvre. Ces principes sont les suivants : croissance inclusive, développement durable et bien-être ; valeurs centrées sur l'humain et équité ; transparence et explicabilité ; robustesse, sûreté et sécurité ; et responsabilité ». La liste de l'Unesco distingue, sans préciser le critère de cette distinction, des valeurs (respect et promotion de la dignité humaine, des droits de l'homme et des libertés fondamentales, assurer la diversité et l'inclusion, le développement durable, vivre en harmonie et en paix) et des principes (innocuité, sûreté et sécurité, équité et non-discrimination, durabilité, protection de la vie privée, transparence et explicabilité, surveillance et décision humaine, responsabilité, sensibilisation et éducation, gouvernance et collaboration multipartites et adaptatives). Par comparaison, les valeurs tout aussi nombreuses citées par la Déclaration de Montréal, sans les hiérarchiser : « le bien-être, l'autonomie, la protection de l'intimité et de la vie privée, la solidarité, la participation démocratique, l'équité, l'inclusion de la diversité, la prudence, la responsabilité et le développement soutenable. Le *discussion paper* de l'ALLEA (2019) mentionne de manière plus concise : *Privacy, Autonomy, Rationality, Equality, Dignity and Human flourishing*.

²³ Déclaration universelle sur la bioéthique et les droits de l'homme adoptée le 19 octobre 2005, texte disponible sur le site de l'UNESCO (http://portal.unesco.org/fr/ev.php-URL.ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html).

²⁴ T. L. BEAUCHAMPS et J. F. CHILDRESS, *Principles of Biomedical Ethics*, 3^e éd., New York, Oxford University Press, 2001.



mettent en jeu au-delà des personnes des groupes d'individus réunis par le genre, la race mais également par la présence d'une donnée génétique qui permet de prédire la maladie future ou par la résidence dans un quartier synonyme de risque de terrorisme ou de difficultés scolaires. Les *fake news* constituent un risque pour notre démocratie ou la santé de toute une société; la régulation privée par les plateformes engendre un risque de normalisation des comportements, qui affecte la vitalité de nos démocraties; nos systèmes technologiques sont énergivores et menacent la survie de la planète; l'intelligence artificielle affaiblit l'État de droit, là où elle ne permet plus au législatif de contester les vérités sorties des ordinateurs et aux juges d'exercer leur rôle de contrôle du respect du droit. Cet élargissement des préoccupations est vital et l'est d'autant plus que comme le relève incidemment le CAHAI²⁵, notre arsenal réglementaire, s'il est bien armé contre les risques que nous avons appelé d'individuels²⁶, l'est bien moins par rapport à ces risques collectifs

6. L'approche par les risques liés aux développements de certains outils d'IA est particulièrement mise en évidence par les documents européens, qu'ils émanent du Conseil de l'Europe ou de l'Union européenne. Cette approche²⁷ est également présente mais de manière plus incidente dans le projet de l'UNESCO²⁸. Le texte du CAHAI est clair et recommande «*that a future Council of Europe legal framework on AI should pursue a risk-based approach targeting the specific application context*». Les premiers considérants du projet parlementaire européen sont consacrés à la nécessité d'un cadre éthico-juridique adapté aux risques nouveaux ou amplifiés par les systèmes de *machine learning*²⁹.

Cette focalisation sur les risques explique l'emphase mise par les textes de l'UNESCO, du CAHAI et du Parlement européen sur l'obligation d'une évaluation des risques. Le texte de l'UNESCO reste vague à ce sujet et semble envisager cette obligation au niveau général: «Les États membres devraient mettre en place des études d'impact pour identifier et analyser les avantages et les risques des systèmes d'IA et les questions qu'ils soulèvent, ainsi que des mesures de prévention, d'atténuation et de suivi des risques. L'étude d'impact en matière d'éthique devrait mettre en évidence les répercussions sur les droits de l'homme, notamment les droits des groupes vulnérables, le droit du travail, l'environnement et les écosystèmes, ainsi que les incidences éthiques

²⁵ Voy. 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 Le RGPD et l'ensemble des institutions créées autour de la cause de la protection des données en est un bel exemple.

²⁷ Par contre le mot «risque» n'est utilisé qu'une fois par les recommandations de l'OCDE dans son exposé des motifs et en lien avec la confiance nécessaire du public: «... la confiance est un déterminant essentiel de la transformation numérique; que bien qu'il soit difficile de prévoir la nature des futures applications de l'IA et leurs incidences, la confiance dans la fiabilité des systèmes d'IA est un facteur clé de la diffusion et de l'adoption de l'IA; et qu'un débat public bien informé de l'ensemble de la société est nécessaire afin de concrétiser le potentiel entier de cette technologie tout en limitant les *risques* qui y sont associés».

²⁸ C'est en effet dans le cadre des mesures stratégiques (n°s 50 et s.) que la notion de «risque» est évoquée et que des mesures pour évaluer voire réduire ces risques sont proposées.

²⁹ Considérant n° 1: «Ces technologies peuvent comporter des possibilités et des risques, qui doivent être traités et réglementés par un cadre réglementaire général à l'échelle de l'Union, qui traduit les principes éthiques à respecter depuis le moment du développement et du déploiement de ces technologies jusqu'à leur utilisation».



et sociales conformément aux principes énoncés dans la présente». Le CAHAI est plus précis et attache l'obligation d'évaluation à chaque système IA et donc à l'ensemble des acteurs qui participent à la mise sur pied et à l'exploitation d'un système d'IA. Le document suggère que cette évaluation puisse déboucher sur des mesures correctrices aptes à diminuer les risques voire à interdire suivant le principe de précaution, certaines applications trop risquées³⁰. Le Parlement européen prône en la matière la distinction entre les systèmes à « haut risque » et les autres³¹. Suivant le principe de proportionnalité qui exige que toute réglementation n'impose pas de charge non nécessaire au regard de ses objectifs, le projet réserve aux seuls systèmes à « haut risque », qu'ils soient individuels ou collectifs³², d'une part, des obligations de sécurité, transparence et de fiabilité du système et, d'autre part, des contrôles de qualité des sources externes de données et des protocoles relatifs aux finalités des traitements, aux destinataires de ceux-ci et qui définissent et garantissent des accès traçables et vérifiables aux données. Dernière précision apportée par le texte de l'Union européenne, cette obligation d'évaluation des systèmes à « haut risque » est préventive, continue et surtout confiée à des auditeurs accrédités externes à l'entreprise ou à l'administration qui participe à la mise sur pied d'un système d'IA³³.

7. Une dernière réflexion achève ce tour d'horizon des documents sélectionnés. Il concerne le rôle de l'État comme moteur de la participation du public à la discussion sur les enjeux de l'IA et les choix à opérer en ce qui concerne son développement. L'OCDE semble réduire ce rôle à une simple information du public, nécessaire à gagner sa confiance : « ..., la confiance est un déterminant essentiel de la transformation numérique ; que, bien qu'il soit difficile de prévoir la nature des futures applications de l'IA et leurs incidences, la confiance dans la fiabilité des systèmes d'IA est un facteur clé de la diffusion et de l'adoption de l'IA ; et qu'un débat public bien informé de l'ensemble de la société est nécessaire afin de concrétiser le potentiel entier de cette technologie tout en limitant les risques qui y sont associés ». Le projet de recommandation de l'UNESCO est plus hardi (recommandation n° 53) : « Les gouvernements devraient adopter un cadre réglementaire qui définisse une procédure permettant en particulier aux autorités publiques de mener à bien des études sur l'impact des systèmes d'IA afin d'anticiper les répercussions, d'atténuer les risques,

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 On retrouve là une distinction déjà présente dans le RGPD, d'où est reprise également l'obligation d'évaluation des risques, évaluation certes élargie à l'ensemble des risques décrits ci-dessus et incombant également à chacun des acteurs qui constituent la chaîne qui mène à une application précise (nommément les développeurs, déployeurs, utilisateurs) et non aux seuls responsables du traitement. Sur cette distinction et son bien-fondé, le rapport B. FRENAY et Y. POULLET, *Profilage et la Convention 108+ : Rapport sur l'évolution de la situation après l'adoption de la Recommandation (2010)13 sur le profilage*, Rapport au Comité consultatif de la Convention n° 108, Conseil de l'Europe, Strasbourg, 7 novembre 2019, T-PD(2019)07rev., pp. 38 et s.

³² L'annexe prévoit même une liste exhaustive et cumulative des applications à hauts risques utilisant à la fois le critère du secteur (ex : éducation, santé, banques et assurances, ...) et celui de la finalité de l'application (engagement de personnel, soins de santé, sélection scolaire, ...).

³³ On ajoute que ce système d'audit externe diffère de celui d'évaluation interne par le responsable du traitement (avec exceptionnellement la consultation de l'autorité de protection des données, mis en place par le RGPD. À cette fin, des certificats de conformité sont émis par ces organismes. Le certificat est obligatoire pour les technologies à hauts risques, il est produit, à la demande du bénéficiaire éventuel, pour les autres technologies. Le texte de l'UNESCO (n° 53) semble également défendre l'idée d'une évaluation externe du moins pour les applications IA des autorités publiques.



d'éviter les conséquences préjudiciables, de faciliter la participation des citoyens et de faire face aux défis sociétaux. L'étude devrait également établir des mécanismes de supervision adaptés, notamment les principes de vérifiabilité, de traçabilité et d'explicabilité, permettant d'évaluer les algorithmes, les données et les processus de conception, ainsi qu'inclure un examen externe des systèmes d'IA. Les études d'impact en matière d'éthique réalisées par les autorités publiques doivent être transparentes et ouvertes au public. Elles devraient par ailleurs être multidisciplinaires, multipartites, multiculturelles, pluralistes et inclusives. Les États membres sont encouragés à mettre en place des mécanismes et des outils, par exemple des bacs à sable réglementaires ou des centres d'évaluation, qui devraient permettre de suivre et d'évaluer les impacts de manière multidisciplinaire et multipartite. Les autorités publiques devraient être tenues de surveiller les systèmes d'IA qu'elles mettent en œuvre et/ou déploient, en établissant des mécanismes et des outils appropriés». Le propos vise donc à la création d'organismes de type « *Technology Assessment* ». D'une part, il s'agit d'analyser les impacts en général de l'IA ou en particulier de tel ou tel type d'application (par exemple : la reconnaissance faciale, les voitures intelligentes, ...) et de prôner des mesures permettant d'encadrer des expérimentations. Il s'agit, d'autre part, d'un organisme d'évaluation pour les applications du secteur public. On souligne que ces organismes doivent veiller à inclure la participation aux discussions des différents intérêts et confronter les points de vue disciplinaires.

On retrouve les mêmes accents dans le document du CAHAI (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* » ». Le document insiste en outre sur le rôle de supervision générale dont doivent disposer les autorités publiques et judiciaires vis-à-vis de tous les systèmes IA (n° 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* ». C'est sans doute du côté du projet de règlement du Parlement européen que les réflexions relatives à l'institutionnalisation d'« organismes nationaux de surveillance » sont les plus achevées. Ainsi, les considérants n^{os} 43 et s. disposent : « Les États membres devraient désigner une autorité administrative indépendante en tant qu'organisme de surveillance³⁴. En particulier, chaque organisme national de surveillance devrait être chargé de recenser l'intelligence artificielle, la robotique et les technologies connexes considérées comme étant à haut risque à la lumière des critères d'évaluation des risques fixés dans le présent règlement, ainsi que d'évaluer et de contrôler la conformité de ces technologies avec les obligations énoncées dans le présent règlement. Chaque organisme national de surveillance devrait également être chargé de réglementer la bonne gouvernance de ces technologies sous la coordination de la Commission et/ou de toute autre institution, organe ou organisme compétent de l'Union désigné à cette fin. Ils ont donc un rôle important à jouer pour favoriser la confiance et la sécurité des

³⁴ Certains pays n'ont pas attendu le règlement pour se doter de ce type d'organismes. Ainsi, des *Data Ethics Commissions* existent au Danemark, en Allemagne et au Royaume-Uni, Sur ces organismes, voy. ainsi, notre ouvrage, *Le RGPD face à l'intelligence artificielle*, Cahier du CRIDS, n° 49, Bruxelles, Larcier, 2020, pp. 151 et 152.



citoyens de l'Union, ainsi que pour permettre l'édification d'une société démocratique, pluraliste et équitable». Le texte insiste sur la composition multipartite et ces organismes doivent être le lieu de débats entre les différentes parties intéressées, qu'il s'agisse du monde de la recherche, des entreprises et des associations civiles. Il doit constituer le «premier point de contact en cas d'atteinte présumée aux obligations nées du texte» et être le relais du droit des citoyens en particulier en cas de traitement discriminatoire. On ajoute que ces organismes «devraient fournir des orientations et un soutien administratif et professionnel aux développeurs, aux déployeurs et aux utilisateurs, en particulier aux petites et moyennes entreprises ou aux start-ups qui ont des difficultés à se conformer aux principes éthiques et aux obligations juridiques énoncés dans le présent règlement».

8. Au terme de cette rapide analyse, que conclure de cet appel récent à l'éthique, pour maîtriser les risques liés au développement de l'IA? Faut-il, comme Ochigame³⁵, y voir une invention destinée à rendre acceptables les applications pourtant controversées de l'IA et endormir la confiance du citoyen? Nous ne le pensons pas. L'éthique, nous l'avons dit, trouve des prolongements dans les textes des droits humains et ses principes viendront ainsi à l'appui de futures législations voire contiennent de tels prescrits juridiques destinés à les mettre en vigueur, comme c'est le cas du règlement proposé par le Parlement européen³⁶. La jurisprudence aura tôt fait de puiser dans les recommandations, les «*best practices*» la source de ce qui constitue les devoirs du «bon père de famille», en l'occurrence de ceux qui participent à la constitution d'un système d'IA. Au-delà, l'apport des textes présente un double mérite: le premier est qu'ils élargissent singulièrement le champ des risques liés aux développements des applications de l'IA et soulignent qu'il s'agit de «hauts risques». Nos conceptions contemporaines des droits humains centrés sur l'individu sont-elles encore de mise dans la mesure où c'est notre environnement, nos structures démocratiques et la règle de droit elle-même qui se trouvent menacés substantiellement³⁷? En second lieu, ces textes indiquent la solution à de tels défis et notre responsabilité collective à la trouver. L'insistance, fondée sur les principes tant de la participation du public que de précaution, autour de la nécessité d'organiser des débats multidisciplinaires, ouverts entre toutes les parties intéressées en vue de trouver des réponses tant éthiques que juridiques m'apparaît comme commune à tous les textes cités. On ne peut que se réjouir d'entendre, lors du dernier Conseil des ministres du Conseil de l'Europe, soit en février 2021, l'appel du CAHAI à une conjonction de tous les efforts des organisations internationales en ce sens. Espérons qu'il soit entendu.

³⁵ R. OCHIGAME, «The invention of 'Ethical AI'», *The Intercept*, 20 décembre 2019, cité par Y. MENECEUR, *op. cit.*, p. 220.

³⁶ Cfr également les obligations de transparence imposées par le projet de la Commission européenne de règlement appelé le *Digital Service Act* (DSA) et relatives aux systèmes d'IA mis en place par les plateformes.

³⁷ Sur ces dangers et la nécessité de sortir d'une approche purement individualiste, A. BASDEVANT et J.P. MIGNARD, *L'empire des données, Essai sur la société, les algorithmes et la loi*, Paris, Don Quichotte, 2018.

