Éthique et intelligence artificielle : l'exemple européen

Jérôme De Cooman*
Assistant, Junior Researcher Liege Competition and Innovation Institute (LCII), Université de Liège (ULiège)

« Car ces lois sont ou des lois de la nature ou des lois de la liberté. La science de la première s'appelle physique, celle de la seconde s'appelle éthique »(1).

◆ TABLE DES MATIÈRES ◆

◆ ◆ ◆

* Jerome.decooman@uliege.be.
(1) E. Kant, *Fondement de la métaphysique des mœurs*, 1785.

INTRODUCTION

L'intelligence artificielle (« IA ») est partout. Cette discipline scientifique[2] permet de nombreuses applications anodines de la vie quotidienne comme l'autodestruction du spam[3], le perfectionnement des systèmes d'imagerie médicale[4] et la gestion du trafic aérien[5].

De ce fait, les médias commentent quotidiennement les progrès technologiques enregistrés dans les divers domaines de l'IA, de la reconnaissance faciale aux applications vidéoludiques. Dans l'industrie, une même effervescence s'observe : la digitalisation, la 4e révolution industrielle et l'Internet des objets sont autant de termes porteurs. Après un long hiver, l'IA traverse désormais une nouvelle saison estivale. Ses prouesses techniques impressionnent[6], les *start-up* fleurissent[7], les

---

(2) Nous choisissons de définir l'intelligence artificielle comme « *the science and engineering of making intelligent machines* » (J. McCarthy *et al.*, « A Proposal for the Darmouth Summer Research Project on Artificial Intelligence (31 aout 1955) », reproduit dans *AI Magazine*, vol. 31, no 4, 2006, pp. 12-14 ; voy. aussi J. McCarthy, « What is Artificial Intelligence », 12 novembre 2017, disponible sur http://jmc.stanford.edu/articles/whatisai/whatisai.pdf). L'intelligence artificielle n'est donc pas une machine, mais un ensemble de disciplines scientifiques – regroupant notamment l'apprentissage automatique (supervisé ou non) et l'apprentissage profond – dont l'objectif est de créer des systèmes autonomes. Cette approche distinguant IA des systèmes IA est celle retenue par le groupe d'experts de haut-niveau de la Commission européenne. Voy. The European Commission's High-Level Expert Group on Artificial Intelligence, « A Definition of AI : Main Capabilities and Disciplines », Definition developped for the purpose of the AI HLEG's deliverables, 8 avril 2019, disponible sur https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341.

(3) N. Kumaran, « Spam does not bring us joy – ridding Gmail of 100 million more spam messages with TensorFlow », 6 février 2019, *Google Cloud Blog*, disponible sur https://cloud.google.com/blog/products/g-suite/ridding-gmail-of-100-million-more-spam-messages-with-tensorflow.

(4) B. H. Do, C. Langlotz et C. F. Beaulieu, « Bone Tumor Diagnisis Using a Naïve Bayesian Model of Demographic and Radiographic Features », *Journal of Digital Imaging*, vol. 30, no 5, 2017, pp. 640-647.

(5) T. Fong, « Autonomous Systems : NASA Capability Overview », 24 août 2018, *NASA*, disponible sur https://www.nasa.gov/sites/default/files/atoms/files/nac_tie_aug2018_tfong_tagged.pdf ; R. Demichelis, « L'intelligence artificielle déjà au service de l'aviation », 27 mars 2018, *Les Echos*, disponible sur https://www.lesechos.fr/2018/03/lintelligence-artificielle-deja-au-service-de-laviation-987559 ; K. Sennaar, « How the 4 Largest Airlines Use Artificial Intelligence », 31 janvier 2019, *Emerj*, disponible sur https://emerj.com/ai-sector-overviews/airlines-use-artificial-intelligence/.

(6) Voy. par exemple B. Marr, « Why Every Company Need An Artificial Intelligence (AI) Strategy for 2019 », 21 mars 2019, *Forbes*, disponible sur https://www.forbes.com/sites/bernardmarr/2019/03/21/why-every-company-needs-an-artificial-intelligence-ai-strategy-for-2019/#6fae0c2968ea ; G. Press, « 7 Indicators Of The State-Of-Artificial Intelligence (AI), March 2019 », 3 avril 2019, *Forbes*, disponible sur https://www.forbes.com/sites/gilpress/2019/04/03/7-indicators-of-the-state-of-artificial-intelligence-ai-march-2019/#35de2c10435a ; N. Martin, « Turing Award And $1 Million Given to 3 AI Pioneers », 27 mars 2019, *Forbes*, disponible sur https://www.forbes.com/sites/nicolemartin1/2019/03/27/turing-award-and-1-million-given-to-3-ai-pioneers/?ss=ai-big-data#22feef1a4784.

(7) Venture Scanner, « Artificial Intelligence Sector Overview – Q3 2018 », 10 novembre 2018, disponible sur https://www.venturescanner.com/blog/2018/artificial-intelligence-sector-overview-q3-2018.

entreprises repensent leur *business model*[8]. Aux techno-optimistes qui ne voient que les bénéfices de l'IA, certains répondent par de l'inquiétude, anticipant tantôt un bouleversement de l'emploi[9], tantôt un effilochage des droits fondamentaux[10].

Comme du temps de la guerre froide avec la conquête spatiale, les enjeux liés à l'IA sont économiques, technologiques, et politiques. Parmi les deux grandes puissances qui aujourd'hui s'opposent, l'URSS a été remplacée par la République populaire de Chine. Les technologies liées à l'intelligence artificielle font qu'un contrôle croissant de la population est envisageable. Ce faisant, s'assurer que les entreprises qui développent ces outils n'en font pas un mauvais usage est extrêmement crucial. Or, l'Union européenne est prise entre le marteau chinois et l'enclume étasunienne[11]. Elle accuse un retard dans le développement d'une industrie liée au développement des algorithmes intelligents. Par ailleurs, ces deux grandes puissances tendent chacune à se positionner d'un point de vue éthique[12].

---

(8) Deloitte, « Developing legal talent stepping into the future law firm », février 2016, disponible sur https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/audit/deloitte-uk-developing-legal-talent-2016.pdf ; voy. égal. le récent ouvrage de Me. Van Wassenhove sur le *coworking* des avocats ; S. Van Wassenhove, *Le coworking*, Bruxelles, Anthémis, 2018.

(9) L'étude pessimiste la plus couramment citée est celle réalisée en 2013 par Carl Frey et Michael Osborne, selon laquelle 47 % des emplois américains disparaîtront d'ici 2040 ; C. Frey et M. Osborne, « The future of employment : how susceptible are jobs to computerization? », 17 septembre 2013, disponible sur https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf. Cependant, il convient de noter que cette étude n'est pas la seule et que les résultats présentés ailleurs sont bien moins alarmants. Ainsi, l'OCDE estime à 9 % le nombre d'emplois menacés ; M. Arntz, T. Gregory, U. Zierahn, « The Risk of Automation for Jobs in OECD Countries : A Comparative Analysis », OECD Social, Employement and Migration Working Papers No. 189, 16 juin 2016, disponible sur https://www.oecd-ilibrary.org/docserver/5jlz9h56dvq7-en.pdf?expires=1554463494&id=id&accname=guest&checksum=14A84B7571C52501CF7411985139A818. Pour un recensement des différentes études menées sur le sujet, voy. O. Colin, « Digitalisation et emplois, entre anxiété et nouvelles opportunités », in C. de Salle (éd.), *Accompagner la robotisation de l'économie*, Les études du Centre Jean Gol, 2017, disponible sur http://www.cjg.be/wp-content/uploads/2019/01/CJG_Etude_robots_201712_BD.pdf. Précisons enfin que tout ceci est loin d'être nouveau. Il s'agit très clairement d'un retour à la hantise keynésienne du chômage technologique, défini comme étant « *unemployment due to our discovery of means of economising the use of labour outrunning the pace at which we can find new uses for labour* » (voy. J.M. Keynes, *Economic Possibilities for our Grandchildren*, 1930, disponible en ligne sur http://www.econ.yale.edu/smith/econ116a/keynes1.pdf).

(10) En guise d'exemple, la reconnaissance faciale et la notation citoyenne sont régulièrement sous le feu des critiques. Ce genre de technologie est déjà mise en place en Chine et plus que probablement utilisée contre les minorités ethniques, au rang desquelles on retrouve les Ouïghours ; A. Webb, « Amy Webb on Artificial Intelligence, Humanity and The Big Nine », *EconTalk Podcast*, 11 mars 2019, disponible sur http://www.econtalk.org/amy-webb-on-artificial-intelligence-humanity-and-the-big-nine/.

(11) Le Canada est également très actif dans le développement de l'apprentissage profond ; *ibid.*

(12) Concernant la Chine, voy. The National New Generation Artificial Intelligence Governance Expert Committee, « Governance Principles for a New Generation of Artificial Intelligence : Develop Responsible Artificial Intelligence », 17 juin 2019, disponible sur https://perma.cc/V9FL-H6J7 (texte original en chinois disponible sur https://perma.cc/7USU-5BLX). Nous avons utilisé la traduction reprise par L. Laskai et G. Webster, « Translation : Chinese Expert Group Offers

L'intelligence artificielle soulève en effet des enjeux moraux. Si le dilemme du tramway, tant évoqué au sujet des véhicules autonomes, n'est sans doute plus à présenter[13], il n'est néanmoins pas le seul. Les risques liés aux détournements des technologies d'application automatique de la loi[14], de même que la résolution des biais[15] sont autant de difficultés critiques auxquelles on ne peut se soustraire.

Vu l'importance de ces enjeux, la Commission européenne a nommé 52 experts indépendants afin de rédiger des lignes directrices en matière d'éthique dans le domaine de l'intelligence artificielle[16]. Un premier texte a été adopté le 18 décembre 2018 (ci-après, « le Projet »)[17] et fut ouvert à

'Governance Principles' for 'Responsible AI' », 17 juin 2019, *New America*, disponible sur https://perma.cc/V9FL-H6J7. Concernant les États-Unis, voy. B. L. LAWRENCE *et al.* (renvoyant au House Committe on Science, Space and Technology), « Resolution Supporting the development of guidelines for ethical development of artificial intelligence », House Resolution 153, 116th Congress, 27 février 2019, disponible sur https://www.congress.gov/bill/116th-congress/house-resolution/153/text et Defense Innovation Board, « AI Principles : Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense », 31 octobre 2019, disponible sur https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF et Defense Innovation Board, « AI Principles : Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense – Supporting Document », 31 octobre 2019, disponible sur https://admin.govexec.com/media/dib_ai_principles_-_supporting_document_-_embargoed_copy_(oct_2019).pdf. Ces initiatives ont mené à l'adoption des principes éthiques début 2020, voy. U.S. Department of Defense, « DOD Adopts Ethical Principles for Artificial Intelligence », Immediate Release, 24 février 2020, disponible sur https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

(13) J. J. THOMSON, « The Trolley Problem », *The Yale Law Journal*, 1985, vol. 94, nº 6, pp. 1395-1415.

(14) Voy., pour de plus amples explications à ce sujet, N. PETIT, « Automated Law Enforcement : A Review Paper », *SSRN* (Working Paper), 26 mars 2018, disponible https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3145133.

(15) Les biais algorithmiques peuvent soit être des représentations correctes de la société biaisée, soit être des représentations biaisées d'un fait correct. Dans ce cas, l'explication peut parfois être trouvée dans la non-représentativité de la base de données dans laquelle l'algorithme a puisé ses sources. Voy. M. KRITIKOS, « Artificial Intelligence *ante portas* : Legal & Ethical reflections », *European Parliamentary Research Service*, Mars 2019. Voy. aussi M.-T. WANG et J. L. DEGOL, « Gender Gap in Science, Technology, Engineering and Mathematics (STEM) : Current Knowledge, Implications for Practice, Policy, and Future Directions », *Educational Psychology Review*, 2017, vol. 29, pp. 119-140 et F. ROSSI, « Building Trust in Artificial Intelligence », *Journal of International Affairs*, 2017, vol. 72, nº 1, pp. 127-133.

(16) Peut-être mieux connu sous son appellation anglophone : « *Ethics Guidelines for Trustworthy AI* ».

(17) The European Commission's High Level Expert Group on Artificial Intelligence, « Draft Ethics Guidelines for Trustworthy AI », Working Document for stakeholders' consultation, 18 décembre 2018, disponible sur https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_draft_ethics_guidelines_18_december.pdf.

une consultation publique, ce qui entraîna la publication du texte définitif le 8 avril 2019 (ci-après, « le Rapport »)[18].

Cet article fait le point sur ce document et sa genèse. Pour ce faire, nous proposons un plan tripartite. Dans une première section, nous présenterons le Rapport européen (I). Une deuxième section analysera ensuite les écueils d'une approche éthique (II). Nous exposerons enfin dans une troisième section pourquoi une telle approche reste intéressante malgré ses difficultés intrinsèques (III).

I. ANALYSE DU RAPPORT EUROPÉEN

L'objectif de cette section est de décrire et analyser les éléments établis par le groupe d'experts désigné par la Commission européenne et dont l'objectif était, notamment, de rédiger des lignes directrices éthiques permettant le développement d'une intelligence artificielle digne de confiance. Pour ce faire, il nous faudra tout d'abord rappeler le contexte qui a mené à la constitution du groupe d'experts (A). Une fois cela fait, nous analyserons le Rapport en présentant tout d'abord les trois composantes d'une IA digne de confiance (B). Nous exposerons ensuite les droits fondamentaux à la base de la réflexion (C) et les principes éthiques qu'ils impliquent (D). Nous présenterons subséquemment les exigences, techniques ou non, nécessaires à l'avènement d'une intelligence artificielle digne de confiance, ainsi que les méthodes de mise en œuvre et d'évaluation proposée pour s'assurer du respect desdites exigences (E). Nous analyserons enfin les commentaires reçus à l'issue de la consultation publique et comment les experts y ont répondu (F).

Précisons que nous n'incluons pas dans cette section les éléments du Rapport qui ne sont pas directement liés au raisonnement permettant d'identifier ce que serait une intelligence artificielle digne de confiance. C'est la raison pour laquelle tant le glossaire que la liste d'exemples d'opportunités et de risques soulevés par l'intelligence artificielle ne sont pas analysés ici.

A. Contexte

La réflexion européenne sur l'intelligence artificielle s'inscrit dans la stratégie pour un marché unique numérique[19]. En mai 2017, la Commission européenne publiait son évaluation de mi-parcours de la mise en œuvre de

(18) The European Commission's High Level Expert Group on Artificial Intelligence, « Ethics Guidelines for Trustworthy AI », 8 avril 2019, disponible sur https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477.

(19) Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des Régions, « Stratégie pour un marché unique numérique en Europe », COM(2015) 192, 6 mai 2015, disponible sur https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52015DC0192&from=FR.

ladite stratégie, dans laquelle elle insistait sur l'importance de « construire des capacités d'intelligence artificielle »[20]. Simultanément, tant le Parlement[21] que le Conseil européen[22] recommandaient à la Commission de présenter une approche européenne relative à l'intelligence artificielle. C'est pourquoi la Commission présenta le 25 avril 2018 sa communication relative à l'intelligence artificielle pour l'Europe[23]. Celle-ci poursuivait un triple objectif : renforcer la capacité technologique et industrielle de l'Union européenne ainsi que le recours à l'intelligence artificielle au sein de l'économie (1), préparer aux changements socio-économiques qu'entraine l'intelligence artificielle (2) et garantir un cadre éthique et juridique approprié, fondé sur les valeurs de l'Union européenne et conforme à la Charte des droits fondamentaux de celle-ci (3)[24].

Conscient qu'une analyse profonde de l'impact du développement de l'IA sur les droits fondamentaux ne pouvait être réalisée sans que les États membres unissent leurs forces, une déclaration de coopération fut signée le 10 avril 2018[25]. Les États signataires[26] devaient œuvrer de concert pour que soit publié un plan coordonné sur l'intelligence artificielle, ce qui fut fait le 7 décembre 2018[27]. Outre l'invitation qui était faite aux États membres de

---

(20) Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des Régions sur l'examen à mi-parcours de la mise en œuvre de la stratégie pour le marché unique numérique : un marché unique numérique connecté pour tous, COM(2017)228 final, 10 mai 2017, disponible sur https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52017DC0228&from=EN.

(21) Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103(INL)), disponible sur https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_FR.html.

(22) General Secretariat of the Council, « European Council Meeting (19 October 2017) – Conclusions », EUCO 14/17, disponible sur http://data.consilium.europa.eu/doc/document/ST-14-2017-INIT/en/pdf.

(23) Communication de la Commission au Parlement européen, au Conseil européen, au Conseil, au Comité économique et social européen et au Comité des Régions, « l'intelligence artificielle pour l'Europe », COM(2018) 237, 25 avril 2018, disponible sur https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52018DC0237&from=EN.

(24) Voy. égal. le communiqué de presse du 25 avril 2018 de la Commission européenne ; Commission européenne, « Intelligence artificielle : la Commission présente une approche européenne visant à stimuler l'investissement et à fixer des lignes directrices en matière d'éthique », 25 avril 2018, disponible sur http://europa.eu/rapid/press-release_IP-18-3362_fr.htm.

(25) Signed Declaration of Cooperation on AI, 10 avril 2018, disponible sur https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence.

(26) À savoir l'Autriche, la Belgique, la Bulgarie, la République tchèque, le Danemark, l'Estonie, la Finlande, la France, l'Allemagne, la Hongrie, l'Irlande, l'Italie, la Lettonie, la Lituanie, le Luxembourg, Malte, les Pays-Bas, la Pologne, le Portugal, la Slovaquie, la Slovénie, l'Espagne, la Suède, le Royaume-Uni, la Norvège, la Roumanie, la Croatie, la Grèce et Chypre.

(27) Le point F du plan coordonné vient préciser que l'une des clefs de la réussite sera l'intégration des règles éthiques dès le début du processus de conception, c'est-à-dire *by-design*. Voy. Communication de la Commission au Parlement européen, au Conseil européen, au Conseil, au Comité économique et social européen et au Comité des Régions, « un plan coordonné dans le domaine de l'intelligence artificielle », COM(2018) 795, 7 décembre 2018, disponible sur https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52018DC0795&from=EN.

poursuivre leurs travaux relatifs à leurs stratégies nationales[28], un point de ce plan coordonné mérite d'être souligné. Il s'agit du « développement de lignes directrices en matière d'éthique dans une perspective mondiale et mise en place d'un cadre juridique propice à l'innovation ».

Vu sa communication d'avril 2018, vu le plan coordonné de décembre 2018 et afin de répondre aux préoccupations éthiques soulevées par l'intelligence artificielle, la Commission décida d'instaurer un groupe d'experts indépendants issus de filières aussi différentes, mais complémentaires, que les sciences de l'informatique, la philosophie et le droit[29]. La démarche n'est pas sans précédent. La Commission a, par le passé, créé d'autres groupes d'experts[30] sur des thématiques telles que le changement climatique[31], la mobilité[32] et le cancer[33]. De telles démarches démontrent l'importance, mais aussi peut-être la volonté d'associer des individus issus de milieux professionnels différents, représentant tant le monde académique que le secteur privé et la société civile, à la recherche d'une solution à un problème actuel dans le respect des valeurs européennes.

Le 27 juin 2018, le groupe d'expert se réunissait pour la première fois[34]. Le 18 décembre 2018, le Projet était publié. Grâce à la création d'un forum,

---

(28) En octobre 2017, la Finlande a présenté son rapport « Finland's Age of Artificial Intelligence » (Steering Group of the Artificial Intelligence Programme, « Finland's Age of Artificial Intelligence. Turning Finland into a leading country in the application of artificial intelligence : Objective and recommendations for measures », Publications of the Ministry of Economic Affairs and Employment 47/2017, http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y). En mars 2018, la France a publié le rapport Villani, aussi connu sous le nom de « AI for Humanity » (C. VILLANI, « Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne », https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf). En novembre 2018, c'est au tour de l'Allemagne de publier sa stratégie liée à l'IA (Nationale Strategie für Künstliche Intelligenz (AI Made in Germany), « Artificial Intelligence Strategy », novembre 2018, https://www.ki-strategie-deutschland.de/home.html). Enfin, en février 2019, c'est la Suède qui a défini son plan au regard de l'IA (Ministry of Enterprise and Innovation, « National Approach to Artificial Intelligence », février 2018, https://www.government.se/491fa7/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf).

(29) La liste de ces experts est disponible sur https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence.

(30) La liste complète des groupes d'experts récemment créés, modifiés, fermés ou mis en attente est disponible sur http://ec.europa.eu/transparency/regexpert/index.cfm?do=news.news.

(31) Voy. « Commission expert group : Mission Board for adaptation to climate change including societal transformation (E03664) », http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3664&news=1.

(32) Voy. « Group of Experts on the Smart Tachograph (E03663) », http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3663&news=1.

(33) Voy. « Commission expert group : Mission Board for cancer (E03665) », http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3665&news=1.

(34) High-Level Expert Group on Artificial Intelligence (E03591), Group Details, Meetings, disponible sur https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3591&NewSearch=1&NewSearch=1.

l'*European AI Alliance*, les auteurs ont pu recueillir plus de cinq-cents commentaires, qui furent publiés le 12 février 2019[35]. Tenant compte des résultats de cette consultation, une seconde version du Projet fut publiée le 8 avril 2019. Le même jour, la Commission présentait son plan pour bâtir la confiance dans l'intelligence artificielle tournée vers l'humain, dans lequel elle déclara soutenir les exigences éthiques proposées par le groupe d'experts[36]. Le 26 juin 2019, le groupe d'experts publiait ses politiques et recommandations d'investissement pour une intelligence artificielle digne de confiance (ci-après, les Recommandations)[37] et ouvrait une phase d'évaluation des lignes directrices du Rapport[38].

Peu de temps après, le 16 juillet 2019, les lignes politiques de la Commission pour 2019-2024 étaient publiées. Dans celles-ci, la Présidente de la Commission Ursula von der Leyen s'engageait à présenter une proposition législative en vue d'une approche européenne coordonnée relative aux implications humaines et éthiques de l'intelligence artificielle dans les 100 premiers jours de son mandat[39]. C'est finalement le 19 février 2020 que la Commission publia son livre blanc sur l'intelligence artificielle prônant une approche européenne de l'excellence et de la confiance[40].

---

(35) Tous les commentaires reçus par le groupe d'experts de haut niveau sur l'IA peuvent être consultés sur https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf.

(36) « Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions : Building Trust in Human-Centric Artificial Intelligence », COM(2019)168 final, 8 avril 2019, disponible sur https://ec.europa.eu/digital-single-market/en/news/communication-building-trust-human-centric-artificial-intelligence.

(37) The European Commission's High Level Expert Group on Artificial Intelligence, « Policy and Investment Recommendations for Trustworthy AI », 26 juin 2019, disponible sur https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence.

(38) « EU Artificial intelligence ethics checklist ready for testing as new policy recommendations are published », 26 juin 2019, disponible sur https://ec.europa.eu/digital-single-market/en/news/eu-artificial-intelligence-ethics-checklist-ready-testing-new-policy-recommendations-are.

(39) U. VON DER LEYEN, « A Union that strives for more : My agenda for Europe », 16 juillet 2019, disponible sur https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf.

(40) Commission européenne, « White Paper on Artificial Intelligence – a European approach to excellence and trust », COM(2020) 65 final, 19 février 2020, disponible sur https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

Ce long processus, de 2015 à 2020, est schématiquement reproduit à la figure 1 ci-dessous. La figure 2, quant à elle, met en exergue les dates-clefs du travail du groupe d'experts sur les lignes directrices et ses Recommandations(41).

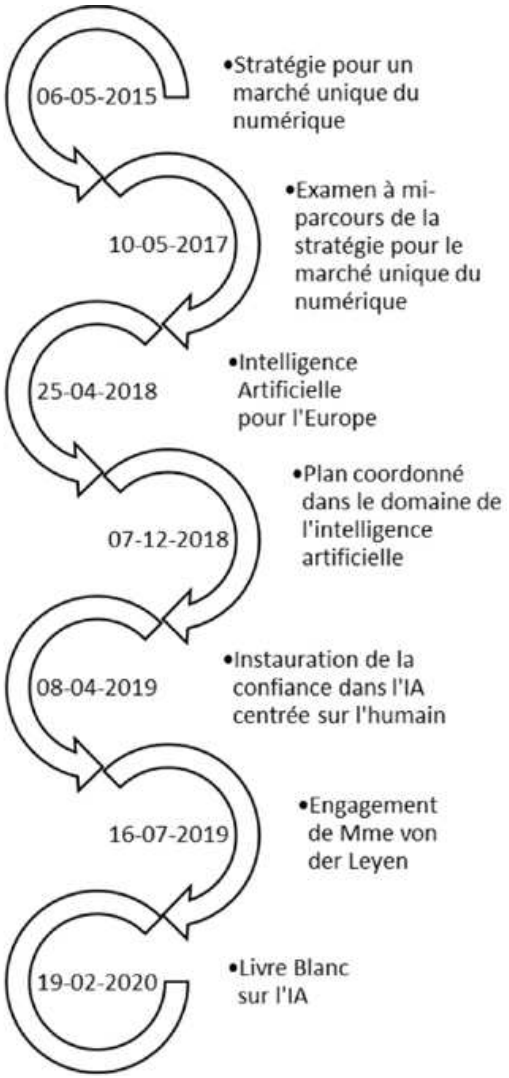

Figure 1 –
Évolution de la stratégie européenne IA

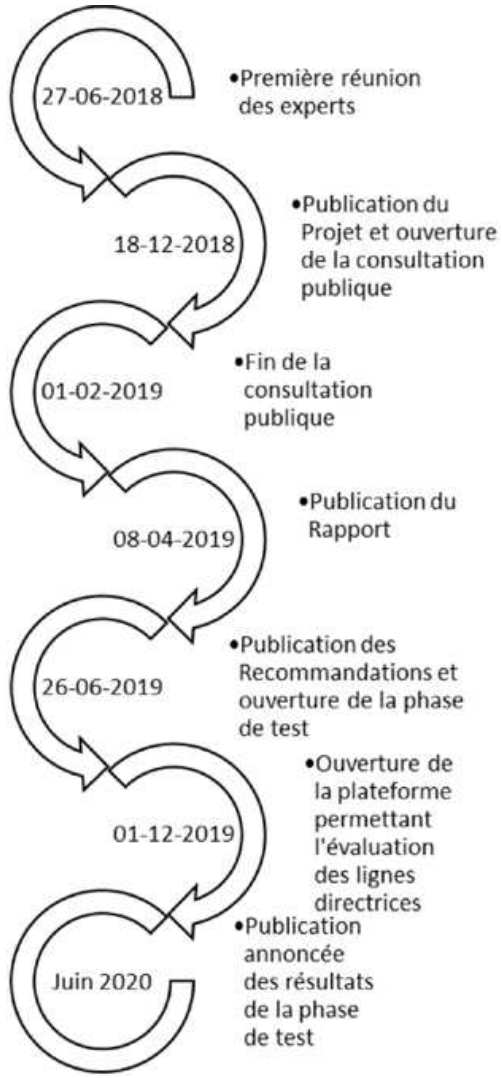

Figure 2 –
Évolution du travail du groupe d'experts

B. Les composantes de l'intelligence artificielle digne de confiance

Le groupe d'experts reconnaît l'IA comme l'une des forces transformatrices de notre temps, force inclusive au plan social et source de croissance,

(41) Notre objectif est d'étudier les efforts éthiques déployés par le groupe d'experts indépendants de la Commission européenne. Seront donc analysés le Projet et le Rapport contenant les lignes directrices éthiques en matière d'intelligence artificielle. Les autres documents mentionnés ci-dessus ne seront utilisés que lorsqu'ils viennent en renfort du travail éthique susmentionné.

d'opportunités économiques et de prospérité. Il en souligne néanmoins les risques, bien que trop brièvement au regard du texte du Projet[42].

L'intelligence artificielle digne de confiance est ainsi érigée en étoile polaire de la réflexion[43]. Une telle IA requiert trois composantes : le respect des dispositions légales applicables (*lawful AI*), l'adhésion aux principes et valeurs éthiques que le Rapport identifie (*ethical AI*) ainsi qu'une robustesse technique et sociale (*robust AI*)[44]. Chacune de ces composantes est nécessaire mais non suffisante. Seule leur combinaison permettra d'atteindre un résultat satisfaisant.

Les systèmes IA ne sont pas destinés à évoluer dans un désert juridique. Ils vont s'insérer dans la structure normative qui organise nos sociétés. C'est la raison pour laquelle la composante de *licéité* est à ce point fondamentale[45]. Paradoxalement, c'est aussi la seule qui n'est pas analysée dans le Rapport[46].

Au-delà du respect de la loi, les systèmes IA doivent également être alignés avec certaines normes morales. L'intérêt de cette deuxième composante, *l'éthique*, permet d'assurer un développement de systèmes IA fiables même si le législateur ne s'est pas encore penché sur la question[47]. Par ailleurs, même si les principes éthiques affirmés dans le Rapport sont largement soutenus par la loi, la portée de l'adhésion à ces principes est bien plus importante que le simple respect des droits et obligations induit par les normes juridiques[48].

La troisième composante assure la *robustesse* des systèmes IA. L'objectif est de graver dans le marbre le vœu pieux d'un système éthique. Plutôt que de se limiter à l'affirmation des principes éthiques, il est impératif de s'assurer que ceux-ci seront respectés, tant au niveau technique que social[49]. Tandis que la robustesse technique assurera la fiabilité et la sécurité du système, la robustesse sociale étudiera le contexte dans lequel ledit système opérera.

L'idée soutenue par le groupe d'experts est donc l'avènement de systèmes IA centrés sur l'humain[50]. Ces systèmes doivent être des outils au service de

---

(42) Voy. *infra*, section I.F.1.

(43) « *Trustworthy AI will be our north star, since human beings will only be able to confidently and fully reap the benefits of AI if they can trust the technology* » ; voy. Draft Ethics Guidelines (executive summary). Le Rapport ne retient pas cette terminologie d'« étoile polaire », mais l'idée reste présente tout au long du document.

(44) Rapport, p. 5.

(45) *Ibid.*, p. 6.

(46) « The Guidelines does not explicitely deal with the first component of Trustworthy AI (lawful AI), but instead aim to offer guidance on fostering and securing the second and third components (ethical and robust AI) ». Rapport, p. 6.

(47) *Ibid.*, p. 6.

(48) L. FLORIDI, « Soft Ethics and the Governance of the Digital », *Philosophy & Technology*, Mars 2018, vol. 31, nº 1.

(49) Rapport, p. 7.

(50) « *AI systems need to be human centric* ». Rapport, p. 4.

l'humanité et non l'inverse[51]. L'intelligence artificielle n'est pas une fin en soi ; c'est un moyen pour améliorer la vie des êtres humains[52]. Une telle formulation n'est pas sans rappeler les impératifs catégoriques kantiens : « agis uniquement d'après une maxime telle que tu puisses vouloir en même temps qu'elle devienne une loi universelle »[53]. Kant cherchait ainsi à identifier des règles à universaliser, comme l'obligation morale de toujours dire la vérité[54]. Il précise qu'il faut agir « de telle sorte que tu traites l'humanité, aussi bien dans ta personne que dans la personne de tout autre, toujours en même temps comme une fin, et jamais simplement comme un moyen »[55]. Autrement dit, nous devons traiter convenablement l'humanité et jamais considérer un individu comme un outil[56].

Par ailleurs, le Rapport précise qu'il est impératif de maximiser les avantages que peuvent susciter les systèmes IA, tout en en réduisant les risques[57]. Cette affirmation illustre un autre courant philosophique, celui du conséquentialisme. Opposé à l'éthique déontologique, la légitimité d'une action est déterminée après une comparaison de ce qu'elle coûtera et du bénéfice que l'on pourrait en retirer[58].

C. Les droits fondamentaux à la base de la réflexion éthique

Le groupe d'experts prend pour point de départ les droits fondamentaux européens[59] dont il dérive les valeurs éthiques à respecter. Ce faisant, il aborde le sujet sous un angle similaire à celui du développement de l'éthique biomé-

(51) Cette idée est clairement transcrite par l'exigence d'action et contrôle humain des systèmes IA (*human agency and oversight*), qui précise que « *AI system does not undermine human autonomy or causes other adverse effects* ». Rapport, p. 16.

(52) « *AI is not an end in itself, but rather a promising means to increase human florishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation* ». Rapport, p. 4.

(53) E. KANT, *Groundwork of the Metaphysics of Morals* (1785), cité in K. AMERIKS et D.M. CLARKE (éd.), *Cambridge Texts in the History of Philosophy*, traduit par M. GREGOR, Cambridge, Cambridge University Press, 1997, p. 15.

(54) Nous soulignons que l'obligation de toujours dire la vérité peut être délicate en pratique, par exemple lorsque dire la vérité engendrera des conséquences néfastes. H. VARDEN, « Kant and Lying to the Murderer at the Door… One More Time : Kant's Legal Philosophy and Lies to Murderers and Nazis », *Journal of Social Philosophy*, 2010, vol. 41, nº 4.

(55) E. KANT, *Groundwork of the Metaphysics of Morals* (1785), *op. cit.*, p. 38.

(56) S. KERSTEIN, « Treating Others Merely as Means », *Utilitas*, vol. 21, nº 2, 2009, pp. 163-180.

(57) « *To maximise the benefits of AI systems while at the same time preventing and minimizing their risks* ». Rapport, p. 4. Voy. *infra*, section I.F.1.

(58) J. FIESER, « Consequentialist theories », *Internet Encyclopedia of Philosophy*, https://www.iep.utm.edu/ethics/#SH2c.

(59) Ceux-ci sont extraits des droits de l'homme, des traités européens et de la Charte des droits fondamentaux.

dicale[60]. Sont ainsi identifiés les droits fondamentaux suivants : le respect de la dignité humaine (*respect for human dignity*), la liberté des individus (*freedom of the individual*), le respect de la démocratie, de la justice et de l'État de droit (*respect for democracy, justice and the rule of law*)[61], l'égalité, la non-discrimination et la solidarité (*equality, non-discrimination and solidarity*)[62] ainsi que les droits des citoyens (*citizen's rights*)[63].

Dans ce contexte technologique, la *dignité* implique que tous les êtres humains ont de la valeur et que celle-ci ne peut en aucune façon être réduite en nous considérant comme des objets, sources de données dont se nourrissent les systèmes IA[64].

La liberté des individus implique que tous les êtres humains disposent du droit de prendre leurs propres décisions par eux-mêmes[65]. *In casu*, cela signifie que tous doivent être protégés contre les intrusions abusives d'entités, qu'elles soient gouvernementales ou non.

Au regard de la *démocratie*, les systèmes IA doivent ne pas être en mesure d'interférer dans les processus démocratiques, notamment le droit de vote, ce qui conduirait à une diminution de la pluralité des opinions[66].

Lorsque *l'égalité et la non-discrimination* rencontrent l'intelligence artificielle, l'accent est mis sur l'identification des biais et leur élimination[67].

Enfin, les *droits des citoyens*, au rang desquels on retrouve le droit de vote, le principe de bonne administration et le droit d'accéder aux ressources publiques, doivent être améliorés au contact des systèmes IA. Ceux-ci doivent

---

(60) Une telle approche a en effet été utilisée dans la Convention Oviedo ; voy. Conseil de l'Europe, « Convention pour la protection des Droits de l'Homme et de la dignité de l'être humain à l'égard des applications de la biologie et de la médecine : Convention sur les Droits de l'Homme et la biomédecine », série des traités européens n° 164, adoptée le 4 avril 1997, entrée en vigueur le 1er décembre 1999, disponible sur https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168007cf99.

(61) « *Respect for democracy, justice and the rule of law* » : l'IA ne peut en aucune façon interférer dans les processus démocratiques ou diminuer la pluralité d'opinions qui fait la force de nos démocraties.

(62) « *Equality, non-discrimination and solidarity including the rights of persons belonging to minorities* » : l'égalité signifie plus que la non-discrimination, qui reconnaît la possibilité de traiter différemment des situations différentes. En effet, l'égalité implique un traitement identique pour tous les êtres humains, indépendamment de la situation dans laquelle ils se trouvent. Dans un contexte d'IA, tous doivent avoir le même accès à l'information, aux données, au marché et à une distribution équitable de la valeur ajoutée générée par les technologies.

(63) « *Citizen's rights* » : dans leurs interactions avec les autorités publiques, les citoyens ont le droit de savoir quand ils font l'objet d'un traitement automatique de leurs données par les entités gouvernementales. Ils doivent également se voir offrir la possibilité de se retirer d'un tel système (*opt-out*).

(64) Rapport, p. 10.

(65) *Ibid.*

(66) *Ibid.*, p. 11.

(67) *Ibid.*

permettre une amélioration tant de l'efficacité que de l'efficience des organes de l'État[68].

D. Principes éthiques induits par les droits fondamentaux

À partir de ces cinq droits fondamentaux, les experts vont identifier les principes éthiques à nécessairement respecter lors du développement d'un système IA. On retrouve ici encore l'idée kantienne d'impératifs[69]. Ceux-ci sont au nombre de quatre : le respect de l'autonomie humaine (*respect for human autonomy*), la prévention de toute atteinte (*prevention of harm*), l'équité (*fairness*) et l'explicabilité (*explicability*).

Dérivant particulièrement du droit fondamental protégeant la liberté des invididus (*freedom of the individual*), l'*autonomie humaine* assure un contrôle plein et entier des êtres humains sur eux-mêmes. Nous ne pouvons être soumis ou manipulés par des systèmes IA. Au contraire, la technologie doit être un outil permettant d'améliorer nos connaissances et capacités sociales, professionnelles ou culturelles. Par ailleurs, plus proche des droits des citoyens (*citizens' rights*), l'autonomie humaine affirme également un maintien de l'être humain au centre des processus démocratiques[70].

La *prévention de toute atteinte* (*prevention of harm*) est sans aucun doute le principe éthique le plus évident, à ce point même qu'il s'agit de la consécration dans un document officiel de la première loi de la robotique de l'écrivain de science-fiction Isaac Asimov[71]. Ce principe est dérivé de la dignité (*human dignity*) et assure une protection de l'intégrité physique et mentale. L'emphase est également mise sur le principe d'égalité et de non-discrimination (*equality, non-discrimination and solidarity*) puisqu'une attention toute particulière est apportée aux asymétries d'information dont les conséquences néfastes pourraient être exacerbées par l'utilisation d'outils algorithmiques performants. Empêcher ces situations d'empirer fait également partie du principe de prévention de toute atteinte[72].

---

(68) Le Rapport insiste sur le fait que la terminologie employée («*citizens*») ne l'est pas pour tracer une distinction entre les citoyens d'un État et ceux qui ne le sont pas. Sont également visés les ressortissants étrangers, même clandestins, dès lors que ce sont des sujets de droit au regard du droit international et européen. Rapport, p. 11.

(69) Voy. *supra*, notes 52 et 54.

(70) Rapport, p. 12.

(71) Ces lois ont pour la première fois été présentées en 1942 dans la nouvelle intitulée «cercle vicieux» (*runaround*). La première loi s'expose comme suit : «un robot ne peut nuire à un être humain ni laisser sans assistance un être humain en danger» («*A robot may not injure a human being or, through inaction, allow a human being to come to harm*»). Voy. I. ASIMOV, «Runaround», *Astounding Science Fiction*, mars 1942. Pour la traduction française, voy. I. ASIMOV, «Cercle Vicieux», in I. ASIMOV, *Le Grand Livre des robots*, vol. I, *Prélude à Trantor*, Omnibus, 1990, p. 208.

(72) Rapport, p. 12.

L'utilisation de l'intelligence artificielle doit être équitable. De tous les concepts utilisés par le Rapport, celui de *l'équité* (*fairness*) est sans doute le plus abstrait. C'est pourquoi les auteurs de ce document le concrétisent en distinguant sa dimension substantive de sa dimension procédurale. D'un point de vue matériel, l'équité implique un engagement à assurer une répartition juste et équitable des avantages et des coûts et à garantir que personne, groupe ou individu, ne subisse de préjugés. Bien plus, les systèmes IA doivent assurer une égalité des chances relative à l'accès à l'éduction, aux biens, aux services et à la technologie(73). D'un point de vue procédural, la justice implique de prévoir la possibilité de contester les décisions algorithmiques et, le cas échéant, d'obtenir réparation du dommage causé(74).

Enfin, *l'explicabilité* (*explicability*) est envisagée comme une condition *sine qua non* de la confiance des utilisateurs des systèmes IA. En synthèse, il implique qu'il faut s'assurer que les processus décisionnels soient transparents et que les décisions prises par les algorithmes soient explicables(75). Ce principe est particulièrement important car l'absence d'explicabilité de la décision implique l'absence de voie de recours effective(76).

Lorsque le système est à effet « boîte noire » (*black box algorithm*), c'est-à-dire lorsque les opérations appliquées aux *inputs* sont à ce point opaques qu'il est impossible d'expliquer pourquoi tel *output* a été préféré plutôt qu'un autre, d'autres méthodes doivent être mises en place pour assurer l'explicabilité, notamment la traçabilité et l'audit(77).

---

(73) *Ibid.*

(74) *Ibid.*, p. 13.

(75) *Ibid.* Voy également l'article 22 et le considérant nº 71 du RGPD organisant l'ébauche d'un droit à l'explicabilité des prises de décisions automatisées ; Règlement (UE) nº 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données), *J.O.U.E.*, L 119, 4 mai 2016, p. 1. Voy. égal. B. GOODMAND et S. FLAXMAN, « European Union Regulations on Algorithmic Decision Making and a "Right to Explanation" », *AI Magazine*, 2017, pp. 50-57.

(76) Voy. *Contra State v. Loomis*, 881 N.W.2d 749 (Wis. 2016). Dans cette affaire, la décision du juge de maintenir Loomis en détention et non de le libérer sous condition était notamment basée sur Compass, l'algorithme développé par Northpointe, Inc. En l'occurrence, l'absence d'explicabilité de l'algorithme était volontaire et non technique, afin de garantir les droits de propriété intellectuelle de la société ayant développé le produit et pour éviter que les détenus puissent manipuler l'algorithme s'ils en connaissaient les rouages. Voy. A. VAN DEN BRANDEN, *Les robots à l'assaut de la justice*, Bruxelles, Bruylant, 2019. Voy. aussi L. BENNETT MOSES et J. CHAN, « Using big data for legal and law enforcement decisions : Testing the new tools », *University of New South Wales Law Journal*, 2014, vol. 37, nº 2, p. 643.

(77) Rapport, pp. 13, 37 et 38. Une telle exigence de traçabilité est particulièrement importante au regard de l'apprentissage profond (deep learning) et de l'apprentissage par renforcement (reinforcement learning). Voy. V. BUHRMESTER, D. MÜNCH et M. ARENS, « Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey », *arXiv* (Cornell University), 27 novembre 2019, disponible en ligne sur : https://arxiv.org/pdf/1911.12116.pdf et voy. R. GUIDOTTI, A. MONREALE, S. RUGGIERI, F. TURINI, F. GIANNOTTI et D. PEDRESCHI, « A survey of methods for explaining black box models », vol. 51, nº 5, *ACM computing surveys*, 2019, pp. 93 et s.

Dans le Projet, ces principes étaient présentés différemment. Les experts avaient retenu au rang des principes éthiques la bienfaisance(78), la non-malfaisance(79), l'autonomie(80), la justice(81) et l'explicabilité(82).

On constate que le principe de bienfaisance (*principle of beneficience : "do good"*) a été supprimé du Rapport, en raison de sa réelle impraticabilité(83). Sa substance, considérablement développée, peut néanmoins être retrouvée d'une part, dans la nouvelle exigence de bien-être sociétal et environnemental insérée dans le deuxième chapitre(84) et d'autre part, dans la liste des opportunités soulevées en fin de document et servant de justification par l'exemple à la critique selon laquelle rien ne permet de prouver que les avantages liés au développement de l'intelligence artificielle l'emportent sur ses risques(85).

Le principe de non-malfaisance (*principle of non maleficience : "do no harm"*) a, quant à lui, été renommé « prévention de toute atteinte » (*prevention of harm*). Son contenu est cependant identique tant dans le Projet que dans le Rapport. Il s'agit donc simplement d'une modification de libellé.

Il en va de même pour l'autonomie humaine (*principle of autonomy : "preserve human agency"*) et l'explicabilité (*principle of explicability : "operate transparently"*). Malgré une légère reformulation de leurs titres, ces principes restent inchangés dans le Rapport(86).

Il n'en va pas de même pour le principe de justice (*principle of justice : "be fair"*). D'une part, son appellation a changé : à partir du principe de justice dans le Projet, c'est un principe d'équité qui a éclos dans le Rapport(87). D'autre part, ce principe a vu sa substance se préciser grâce à l'ajout de sa double dimension matérielle et procédurale, présentée ci-dessus(88).

---

(78) « *The Principle of Beneficience : 'Do Good'* » : les systèmes IA doivent être développés dans l'optique d'améliorer constamment le bien-être individuel et collectif.

(79) Projet, p. 9.

(80) *Ibid.*

(81) *Ibid.*, p. 10.

(82) *Ibid.*

(83) À cet égard, la Commission note que « *The revised Guidelines no longer contain a reference to the "do good" principle. However, they stipulate clearly that one of the goals of Trustworthy AI is to improve individual and collective wellbeing, and provide guidance concerning how ethics can help to achieve this objective* » ; European AI Alliance, « Ethics Guidelines for Trustworthy AI – Overview of the Main Comments received Through the Open Consultation », 16 avril 2019, disponible sur https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58480.

(84) Rapport, p. 19.

(85) *Ibid.*, pp. 32-33.

(86) Dans le Projet, on parle de « *principle of autonomy : preserve human agency* », tandis que dans le Rapport, on parle de « *principle of respect for human autonomy* ». Parallèlement, dans le Projet, on parle de « *principle of explicability: "operate transparently"* » tandis que dans le Rapport, on parle de « *principle of explicability* », sans insister sur l'objectif de transparence. Voy. Projet, pp. 9-10 et Rapport, pp. 12-13.

(87) Du « *principle of justice: "be fair"* » du Projet, on obtient le « *principle of fairness* » dans le Rapport. Voy. Projet, p. 10 et Rapport, pp. 12-13.

(88) *Ibid.*, pp. 12-13.

Ces principes éthiques n'ont rien d'original. Il résulte de plusieurs méta-analyses récentes[89] que les nombreuses recommandations éthiques et lignes directrices adoptées de par le monde tendent à converger vers les principes de bienfaisance, non-malfaisance, autonomie et justice[90]. Il s'agit, comme on vient de le constater, des appellations retenues au sein du Projet. Ces principes sont directement issus de la bioéthique[91], à l'exception de l'explicabilité[92], comme illustré à la figure 3 ci-après. Cela n'est guère étonnant. L'intelligence artificielle n'a pas le monopole des questions philosophiques. La bioéthique, tout comme l'éthique de l'intelligence artificielle, relève de l'éthique appliquée[93]. Concernant l'éthique médicale, sexuelle et sociale, l'évolution des mœurs a fait que la notion de « bien » a changé, entraînant des modifications au niveau juridique. En Belgique, la dépénalisation partielle de l'avortement en 1990[94] et de l'euthana-

---

(89) J. LEIKAS *et al.*, « Ethical Framework for Designing Autonomous Intelligent Systems », *Journal of Open Innovation, Technology, Market and Complexity*, 2019, vol. 5, nº 1 ; Algorithm Watch, « The AI Ethics Guidelines Global Inventory », 2019, https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/, cité par L. FLORIDI, « Translating Principles into Practices of Digital Ethics : Five Risks of Being Unethical », *Philosophy & Technology*, 2019, vol. 32, p. 186 ; L. FLORIDI, « A Unified Framework of Five Principles for AI in Society », *Harvard Data Science Review*, 2019, vol. 1, nº 1 ; T. HAGENDORFF, « The Ethics of AI Ethics – An Evaluation of Guidelines », *arXiv* (Cornell University), 28 février 2019, https://arxiv.org/abs/1903.03425 ; A. JOBIN, M. IENCA et E. VAYENA, « The global landscape of AI ethics guidelines », *Nature Machine Intelligence*, 2019, vol. 1, pp. 389-399 ; D. GREENE, A. L. HOFFMAN et L. STARK, « Better, nicer, clearer, fairer : a critical assessment of the movement for ethical artificial intelligence and machine learning », *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019, https://scholarspace.manoa.hawaii.edu/bitstream/10125/59651/0211.pdf ; A. BENSOUSSAN et J. BENSOUSSAN, *IA, Robots et Droits*, Bruxelles, Bruylant,2019, spécialement l'annexe 1 : synthèse des textes de droit souple en robotique et IA, pp. 453-458.

(90) « *Beneficience, non-maleficience, autonomy, and justice* ». L. FLORIDI et J. COWLS, « A Unified Framework of Five Principles for AI in Society », *Harvard Data Science Review*, 2019, vol. 1, nº 1, p. 5. Voy. aussi. B. MITTELSTADT, « Principles alone cannot guarantee ethical AI », *Nature Machine Intelligence*, 2019, vol. 1, p. 501.

(91) J. WHITTLESTONE, R. NYRUP, A. ALEXANDROVA, K. DIHAL et S. CAVE, « Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence : A Roadmap For Research », *Nuffield Foundation*, 2019, https://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf.

(92) L. FLORIDI et J. COWLS, « A Unified Framework of Five Principles for AI in Society », *Harvard Data Science Review*, 2019, vol. 1, nº 1, p. 5.

(93) L'éthique appliquée est l'une des trois branches de l'éthique, au côté de la méta-éthique et de l'éthique normative. Tandis que la méta-éthique étudie l'origine et la signification des concepts éthiques, l'éthique normative cherche à dégager des normes morales permettant de distinguer les bonnes conduites des mauvaises. Pour le dire autrement, plutôt que de se demander ce qu'il faut faire pour être une bonne personne, ce qui relève de l'éthique normative, la méta-éthique s'intéresse à la question de savoir ce qu'est le bonheur. L'éthique appliquée, quant à elle, est celle qui analyse des controverses morales précises. Voy. J. FIESER, « Ethics », Internet Encyclopedia of Philosophy, https://www.iep.utm.edu/ethics/.

(94) Loi relative à l'interruption de grossesse, modifiant les articles 348, 350, 351 et 352 du Code pénal et abrogeant l'article 353 du même code, *M.B.*, 5 avril 1990.

sie en 2002[95], ainsi que l'autorisation du mariage homosexuel en 2003[96] en sont trois bonnes illustrations.

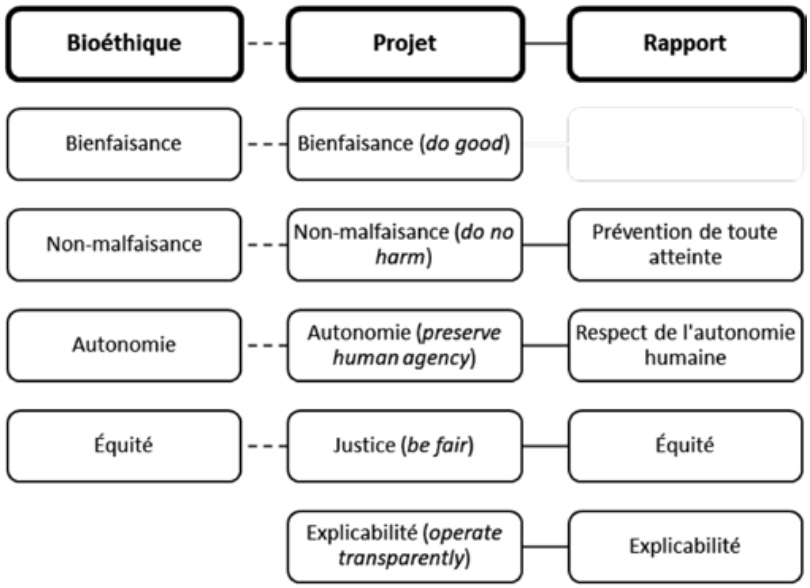

Figure 3 – Comparaison des principes IA et bioéthique

En ce qui concerne l'intelligence artificielle, ce n'est pas un changement de mœurs qui entraîne la réflexion éthique, mais une innovation technologique. L'expérience nous enseigne que l'intelligence artificielle n'est pas la première discipline scientifique à soulever des préoccupations éthiques. Avant elle, d'autres technologies ont entraîné une réflexion philosophique. Un exemple classique est la fission atomique : l'absence de sérieuses considérations éthiques a entraîné l'utilisation de l'arme nucléaire[97]. Plus proche de nous, les nanotechnologies font beaucoup parler d'elles, en raison de leurs très nombreuses applications, notamment dans les puces de radio-identification (RFID)[98].

---

(95) Loi du 28 mai 2002 relative à l'euthanasie, *M.B.*, 22 juin 2002.

(96) Loi du 13 février 2003 ouvrant le mariage à des personnes de même sexe et modifiant certaines dispositions du Code civil, *M.B.*, 28 février 2003.

(97) On se souviendra de la déclaration de Robert Oppenheimer, « *We knew the world would not be the same. A few people laughed, a few people cried, most people were silent. I remembered the line from the Hindu scripture, the Bhagavad-Gita. Vishnu is trying to persuade the Prince that he should do his duty and to impress him takes on his multi-armed form and says, "Now, I am become Death, the destroyer of worlds". I suppose we all thought that one way or another* » ; R. OPPENHEIMER, disponible sur https://www.youtube.com/watch?v=_LmxIptS3cw.

(98) M. GUPTA et N. CHANDEKAR, « Nanotechnology and its applications in RFID », in *International Journal of Advances Computational Engineering and Networking*, vol. 5, Issue 11, Novembre 2017 ; R. E. MCGINN, « What's Different, Ethically, About Nanotechnology ? : Foundational Questions and Answers », *NanoEthics*, 2010, vol. 4, pp. 115-128 ; J. VAN DEN HOVEN, « Nanotechnology and Privacy : The Instructive Case of RFID », *International Journal of Apllied Philosophy*, 2006, vol. 20, pp. 215-228 ; J. R. HERKERT, « Ethical Challenges of Emerging Technologies », in

Certains ont critiqué ce rapprochement de la bioéthique et de l'éthique de l'intelligence artificielle, considérant que cette analogie n'était pas exempte de défauts. Alors que les intérêts du patient et ceux du médecins sont alignés, il n'en va pas de même pour le développeur d'un système IA et son utilisateur, entre lesquels s'insèrent les intérêts financiers des entreprises privées(99). S'il est vrai que le secteur de la santé est lui aussi soumis à des contraintes financières, le cadre régulatoire en place empêche les hôpitaux de faire passer des contraintes budgétaires avant l'intérêt du patient(100). Or, mis à part certaines activités particulières(101), les systèmes IA ne disposent pas d'un cadre régulatoire suffisant. Sans ce soutien aux principes éthiques, il faut compter sur la bonne volonté des développeurs pour respecter les principes éthiques. Or, il a été démontré que les déclarations éthiques avaient peu ou pas d'impact sur la pratique quotidienne de ceux-ci(102). Concernant le Rapport européen relatif à l'intelligence artificielle, il est un fait que les considérations morales développées par le groupe d'experts ne peuvent en aucune façon être source de droits et obligations soumises à la menace de sanction.

E. Les exigences d'une IA digne de confiance : définition, mise en œuvre et évaluation

À partir des quatre principes d'autonomie, de prévention de toute atteinte, d'équité et d'explicabilité, le Rapport offre des indications relatives à la réalisation d'une IA digne de confiance. L'idée est que les principes éthiques doivent être concrétisés par des exigences que les personnes qui développent, conçoivent et déploient les systèmes IA devront respecter(103). Ces exigences sont l'action et le contrôle humain (*human agency and oversight*), la robustesse technique et la sécurité (*technical robustness and safety*), le respect de la vie privée et la gouvernance des données (*privacy and data governance*), la transparence (*transparency*),

---

G. E. ARCHANT, B. R. ALLENBY et J. R. HERKERT (éd.), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight : The Pacing Problem*, Springer, New-York, 2011.

(99) B. MITTELSTADT, « Principles alone cannot guarantee ethical AI », *Nature Machine Intelligence*, 2019, vol. 1, p. 501.

(100) *Ibid.*, p. 502.

(101) On pense principalement à la protection des données personnelles, protégées en Europe par le RGPD.

(102) A. MCNAMARA, J. SMITH et E. MURPHY-HILL, « Does ACM's code of ethics change ethical decision making in software development? », *Proceedings of the 2018 26th ACM Joint Meetin on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, octobre 2018, pp. 729-733. B. MITTELSTADT, « Principles alone cannot guarantee ethical AI », *Nature Machine Intelligence*, 2019, vol. 1, p. 504. T. HAGENDORFF, « The Ethics of AI Ethics – An Evaluation of Guidelines », *arXiv* (Cornell University), 28 février 2019, https://arxiv.org/abs/1903.03425.

(103) Le glossaire du Rapport définit les « *stakeholders* » comme étant « […] *all those that research develop, design, deploy or use AI, as well as those that are (directly or indirectly) affected by AI – including but not limited to companies, organisations, researchers, public services, institutions, civil society organisations, governments, regulators, social partners, individuals, citizens, workers and consumers* ». Voy. Rapport, pp. 37-38.

la diversité, la non-discrimination et l'équité (*diversity, non-discrimination and fairness*), le bien-être sociétal et environnemental (*social and environmental wellbeing*) ainsi que la responsabilité (*accountability*)[104] (1). Ces exigences ne sont pas exhaustives et n'ont pas d'ordre hiérarchique, ce qui ne manquera pas de soulever des difficultés lorsque des problèmes de compatibilité surviendront.

Par ailleurs, la mise en œuvre de ces exigences doit avoir lieu continuellement et à tous les stades du développement du système IA. À cet égard, le rapport propose différentes méthodes, techniques ou non, permettant d'implémenter ces exigences au sein des systèmes IA (2).

Enfin, l'exécution et le respect de ces exigences doivent être évalués (3). C'est la raison pour laquelle une liste de questions permettant ladite évaluation a été insérée dans le Rapport.

1. Les exigences d'une IA digne de confiance

Premièrement, l'exigence d'*action et de contrôle humain* est dérivée du respect de l'autonomie et implique que les systèmes IA doivent soutenir et non pas subordonner les êtres humains dans l'exercice de leurs activités[105]. Ce principe présente une triple dimension : l'amélioration de l'exécution des droits fondamentaux (*fundamental rights*), l'action humaine (*human agency*) et le contrôle humain (*human oversight*).

Laissons de côté les droits fondamentaux, sur lesquels le texte ne dit peu, voire rien. L'action humaine implique que les êtres humains doivent toujours être en mesure de décider, indépendamment de l'utilisation de systèmes IA. Tout l'intérêt de cette technologie est de permettre aux êtres humains de prendre de meilleures décisions, ce qui implique que ceux-ci doivent avoir la capacité nécessaire à l'utilisation des outils technologiques[106]. Pour le dire autrement, « les machines sont faites pour servir l'homme, c'est entendu, mais encore faut-il que l'homme sache s'en servir »[107].

Le contrôle humain, quant à lui, implique que les systèmes IA doivent être monitorés. L'objectif est de s'assurer qu'ils ne subordonnent personne[108].

Deuxièmement, le principe de *robustesse technique et de sécurité* est induit par le principe de prévention de toute atteinte[109]. Il requiert d'agir préventivement en évaluant les éléments qui pourraient affecter le système et en

(104) Rapport, p. 14.
(105) *Ibid.*, p. 15.
(106) *Ibid.*, p. 16.
(107) Robert Silvenberg, dans son introduction au texte de Lewis Padget, nom de plume de Henry Kuttner ; L. PADGET, « Le Twonky », in *Des hommes et des machines*, Bibliothèque Marabout, n° 434, Edition Gérard & C°, 1973, p. 163 (publication originale in *Astounding Science-Fiction*, 1942).
(108) Rapport, p. 16.
(109) *Ibid.*

en définissant le risque. Cela concerne tant la résilience aux cyberattaques(110) que la mise en place d'un plan de secours si les choses devaient mal tourner(111).

Ce principe implique également que les algorithmes doivent être conçus de sorte qu'ils soient particulièrement précis (*accuracy*), fiables (*reliability*) et que leurs raisonnements puissent être reproduits (*reproducibility*)(112). Sans ces trois caractéristiques, il est impossible d'envisager l'utilisation d'algorithmes dans les domaines liés à l'aide à la décision.

Troisièmement, la protection de la *vie privée* découle du principe de prévention de toute atteinte. Ce dernier implique également une certaine forme de *surveillance des données*. Si les données doivent être exemptes de biais ou d'erreurs, leurs accès et traitements doivent être réalisés de sorte que la vie privée soit protégée(113).

Quatrièmement, le principe d'explicabilité justifie l'exigence de *transparence*. Celle-ci implique la traçabilité des données utilisées par un algorithme, l'explication des décisions prises par un système IA et le droit de savoir si une personne communique avec un autre être humain ou avec un algorithme(114).

Cinquièmement, le principe d'équité invite au respect de l'exigence de *diversité*, *non-discrimination* et *équité*. Cela implique qu'il faut éviter de répliquer les biais historiques de nos sociétés dans les systèmes IA, afin de ne pas les aggraver(115). Par ailleurs, les systèmes IA doivent être développés de manière à ce que tous puissent les utiliser, indépendamment des questions relatives à l'âge, au genre ou aux capacités.

Sixièmement, les principes d'équité et de prévention de toute atteinte impliquent que soit développé un *bien-être sociétal et environnemental*. Cela implique qu'une attention toute particulière soit apportée d'une part, à l'impact environnemental des systèmes IA, par exemple en ce qui concerne la consommation d'énergie qu'ils requièrent, et d'autre part, à l'impact social de ces technologies dans nos sociétés, notamment leur impact sur les relations humaines et les compétences sociales(116).

Septièmement, le principe d'équité requiert que soit mis en place l'exigence de *responsabilité*. L'audit, tant interne qu'externe, peut soutenir cette exigence et permettre aux utilisateurs d'avoir confiance dans le système IA qu'ils utilisent(117). En cas de survenance d'un dommage, des mécanismes de réparation devront être prévus(118).

---

(110) *Ibid.*
(111) *Ibid.*, pp. 16 et 17.
(112) *Ibid.*, p. 17.
(113) *Ibid.*
(114) *Ibid.*, p. 18.
(115) *Ibid.*
(116) *Ibid.*, p. 19.
(117) *Ibid.*
(118) *Ibid.*, p. 20.

Les exigences présentées dans le Rapport ne sont que partiellement similaires à celles du Projet. Alors que le Rapport identifie les sept exigences décrites ci-dessus, le Projet en proposait dix, présentées dans l'ordre alphabétique[119] : la responsabilité (*accountability*), la gouvernance des données (*data governance*), le design pour tous (*design for all*), la gouvernance de l'autonomie de l'IA et le contrôle humain (*governance of AI autonomy : human oversight*), la non-discrimination (*non-discrimination*), le respect et le renforcement de l'autonomie humaine (*respect for (& enhancement of) human autonomy*), le respect de la vie privée (*respect for privacy*), la robustesse (*robustness*), la sécurité (*safety*) et la transparence (*transparency*)[120].

Seules deux exigences ont un libellé identique dans le Projet et le Rapport : la transparence et la responsabilité. Les cinq autres exigences du Rapport constituent des fusions de celles du Projet, afin d'éviter les redondances reprochées lors de la consultation publique. Lesdites fusions sont présentées à la figure 4 ci-dessous. Une exception est cependant identifiée, celle du bien-être sociétal et environnemental. Celui-ci est une semi-nouveauté puisqu'il fait écho au principe de bienfaisance, initialement développé dans le Projet mais finalement supprimé du Rapport[121].

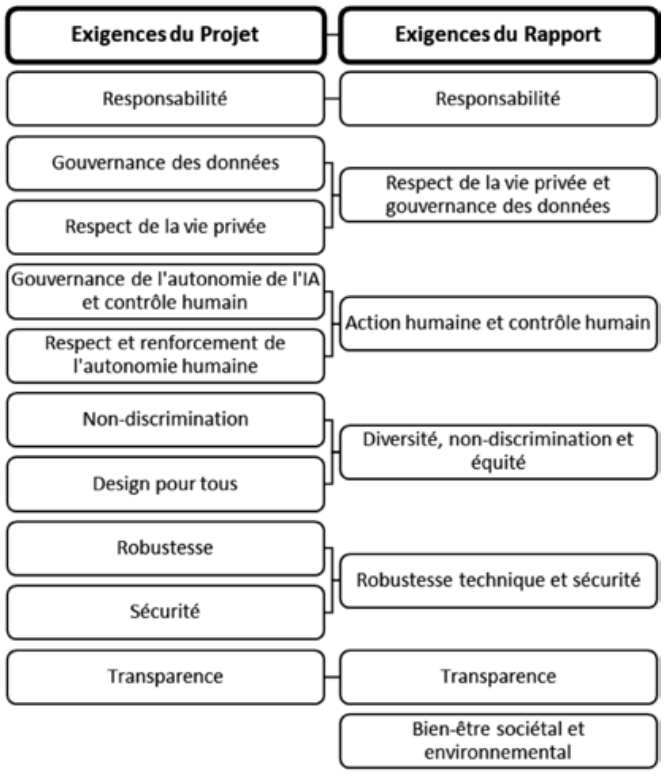

Figure 4 – Corrélations des exigences du Projet et du Rapport

(119) Projet, p. 14.
(120) *Ibid.*, pp. 14-18.
(121) Voy. *supra*, note 83.

Concernant les modifications de contenu, on note que l'exigence de transparence s'est vue adjoindre les méthodes techniques de traçabilité et d'explicabilité(122) tandis que celle de responsabilité s'est vue reconnaitre le besoin d'un audit pour déployer pleinement ses effets(123). Les véritables nouveautés concernent le principe de responsabilité, qui vient requérir la minimisation des impacts négatifs et leur notification(124), la résolution des conflits qui pourraient survenir entre les exigences(125) et la réparation en cas de conséquences négatives(126).

2. *Les méthodes techniques et non techniques de mise en œuvre des exigences*

Parmi les méthodes techniques, on retrouve le développement d'architectures pour IA digne de confiance (*architectures for Trustworthy AI*)(127), l'insertion des règles normatives ou éthiques dès la conception (*Ethics and rule of law by design : X-by-design*)(128), des méthodes d'explication (*explanation methods*)(129), des essais et validations (*testing and validating*)(130) et la mise en place d'indicateurs service de qualité (*quality of service indicators*)(131).

Les méthodes non techniques sont plus variées. On y retrouve l'idée d'un encadrement par la réglementation (*regulation*)(132), les codes de conduite (*codes of conduct*)(133), la normalisation (*standardisation*)(134), la certification (*certification*)(135) et la mise en place d'un cadre gouvernemental garantissant le respect des dispositions éthiques (*accountability via governance*

(122) Rapport, p. 18.
(123) *Ibid.*, pp. 19-20.
(124) « *Minimisation and reporting of negative impact* » : la minimisation des impacts négatifs peut se faire grâce à une vérification continue du développement et du déploiement d'un système IA (pratique du *red teaming*). En outre, afin de garantir une notification effective des conséquences négatives liées à l'application d'un système IA, un mécanisme de protection doit être prévu pour les lanceurs d'alertes. Voy. Rapport, p. 20.
(125) « *Trade-offs* » : pour résoudre les conflits qui surviendraient entre exigences, il faudra se référer à une procédure rationnelle et documentée, respectant les règles de l'art. Cela implique d'identifier les intérêts en jeu et d'évaluer le risque du système IA sur les principes éthiques. Si aucun compromis éthique ne peut être dégagé, le système IA ne pourra ni être développé, ni être déployé. Voy. Rapport, p. 20.
(126) « *Redress* » : les mécanismes de réparation adéquate du dommage doivent être accessibles et connus, afin de garantir la confiance dans le système IA. Voy. Rapport, p. 20.
(127) Rapport, p. 21.
(128) *Ibid.*
(129) *Ibid.*
(130) *Ibid.*, p. 22.
(131) *Ibid.*
(132) *Ibid.*
(133) *Ibid.*
(134) *Ibid.*
(135) *Ibid.*, p. 23.

framework)[136]. Par ailleurs, un autre set de méthodes non-techniques est envisagé et a pour objectif de faire prendre conscience de l'importance des systèmes IA et de leurs impacts. Sont ainsi cités l'éducation et la sensibilisation pour favoriser un état d'esprit éthique (*education and awareness to foster an ethical mind-set*)[137], la participation des parties prenantes et le maintien d'un dialogue social (*stakeholder participation and social dialogue*)[138] ainsi que l'assurance d'une diversité au sein des équipes de développements, tant en termes d'âge, de culture et de genre que de parcours professionnels (*diversity and inclusive design teams*)[139].

La présence d'une méthode de mise en œuvre des principes éthiques, même à ce stade embryonnaire, mérite d'être soulignée. L'une des critiques à l'encontre des réflexions éthiques vis-à-vis des nouvelles technologies est justement sa difficile traduction dans la pratique[140]. Conscient de ce problème, le groupe d'experts a proposé les diverses pistes de solution listées ci-dessus.

3. Méthode d'évaluation des exigences

Le Rapport vient ensuite concrétiser ce qui précède en fournissant une liste non-exhaustive d'éléments à évaluer[141], sous la forme de questions à continuellement se poser, adoptant ainsi le style de l'évaluation cyclique[142]. Chaque question a été liée avec l'une des exigences précédemment exposées.

Cette évaluation continue n'est pas sans rappeler l'éthique aristotélicienne. Comme les citoyens de la Grèce antique, qui apprenaient par l'expérience à atteindre le bonheur[143], les développeurs devront sans aucun doute améliorer leurs systèmes par la méthode essai-erreur, évaluant continuellement ce qu'ils développent[144].

La liste d'évaluation a été testée par 350 entreprises, sur base volontaire, depuis la publication du Rapport en avril 2019[145]. L'idée est de recueil-

(136) *Ibid.*
(137) *Ibid.*
(138) *Ibid.*
(139) *Ibid.*
(140) B. MITTELSTADT, « Principles alone cannot guarantee ethical AI », *Nature Machine Intelligence*, 2019, vol. 1, p. 503.
(141) Rapport, pp. 26-31.
(142) *Ibid.*, p. 15.
(143) ARISTOTLE, *Nicomachean Ethics*, translated by W.D. ROSS, Kitchener, 1999, p. 16.
(144) « *The seven requirements […] should be implemented and evaluated throughout the AI system's lifecycle* ». Rapport, p. 15. « *To verify and validate processing of data, the underlying model must be carefully monitored during both training and deployment […]. Testing and validation of the system should occur as early as possible, ensuring that the system behaves as intended throughout its entire life cycle* ». Rapport, p. 22.
(145) Commission européenne, « White Paper on Artificial Intelligence – a European approach to excellence and trust », COM(2020) 65 final, 19 février 2020, p. 9.

lir un maximum de commentaires pour ensuite rédiger une liste qui puisse constituer un cadre de référence à tous les systèmes applications IA, afin d'assurer la confiance dans tous les domaines. Ces résultats sont attendus pour juin 2020(146). Par la suite, seront rédigés des cadres sectoriels ou ciblant des applications spécifiques(147).

On retrouve ici encore une logique chère à Aristote. Dans l'*Éthique à Nicomaque*, il définit l'éthique comme un art de vie qui doit être pratiqué quotidiennement et qui requiert une analyse au cas par cas(148). Tout comme lui, le Rapport rejette l'idée d'une solution unique à tous les systèmes IA et insiste sur l'importance d'une approche casuistique(149). Une telle approche constitue une autre similitude entre l'éthique de l'intelligence artificielle et la bioéthique(150).

F. Retour sur la consultation publique

Nous avons analysé le retour d'expérience issu de la consultation publique. Parmi les 500 commentaires reçus par la Commission européenne, certains étaient plus longs que d'autres, en ce qu'ils pouvaient contenir des remarques sur plusieurs paragraphes précis du Projet. Après lecture et analyse de ceux-ci, nous estimons à plus de 3.000 le nombre de critiques adressées au groupe d'experts. Nous avons donc procédé à un regroupement de celles-ci lorsqu'elles partageaient la même idée.

Même après élimination des commentaires qui ne constituaient soit que de simples encouragements ou félicitations, soit des remarques orwelliennes et eschatologiques, il eut été impossible de présenter l'intégralité des commentaires pertinents. C'est la raison pour laquelle nous n'en avons sélectionnés que quelques-uns, en fonction de leur récurrence et de l'importance des sujets qu'ils ont mis en exergue.

La principale critique issue de la consultation publique était le trop grand optimiste du Projet. Nous présenterons donc les modifications apportées par les experts afin d'y répondre (1). Nous présenterons ensuite les autres commentaires qui, bien que pertinents, ne méritent pas une analyse aussi approfondie que la précédente, en raison soit de leur nature particulière, soit de la réponse apportée par le Rapport (2).

---

(146) *Ibid.*

(147) Rapport, p. 24.

(148) ARISTOTLE, *Nicomachean Ethics*, translated by W.D. ROSS, Kitchener, 1999, pp. 10 et s. et p. 33.

(149) «*AI systems should not have a one-size-fits-all approach and should consider Universal Design principles addressing the widest possible range of users, following relevant accessibility standards*». Rapport, p. 19.

(150) J. LEIKAS *et al.*, «Ethical Framework for Designing Autonomous Intelligent Systems», *Journal of Open Innovation : Technology, Market, and Complexity*, 2019, vol. 5, nº 1, p. 3.

1. Optimisme du Rapport

La critique la plus récurrente est sans doute que le Projet a adopté un style bien trop optimiste, partant du principe que les avantages de l'IA l'emportent sur ses risques(151). Or, il n'y a pas de preuves que les bénéfices de l'intelligence artificielle l'emportent sur ses risques. Cette assertion doit être développée et prouvée. Cette remarque est d'autant plus importante que c'est sur cette idée qu'est construit tout le Projet. Si elle devait se révéler fausse, tout ce qui suit s'écroulerait comme un château de cartes.

Nous ne pouvons que rejoindre ce commentaire quant au trop grand optimisme du texte du Projet. Une rapide lecture de l'*executive summary* permet de s'en convaincre. Le deuxième paragraphe précise ô combien l'intelligence artificielle est une opportunité pour améliorer la prospérité et la croissance(152). Le paragraphe suivant précise rapidement que cette science présente également des risques qui doivent être convenablement pris en compte(153) mais que, en substance, le jeu en vaut la chandelle.

Néanmoins, le Projet contient une section sur les enjeux critiques soulevés par l'intelligence artificielle. Il s'agit de l'identification sans consentement(154), des risques de confusion entre l'homme et la machine au cours de leurs interactions(155), de l'éva-

(151) «*AI's benefits outweight its risks*»; voy. Draft Ethics Guidelines for Trustworthy AI, 18 décembre 2018, Executive Summary, § 3.

(152) Ce paragraphe compte 127 mots pour 721 caractères, espaces non compris: «*Artificial Intelligence (AI) is one of the most transformative forces of our time, and is bound to alter the fabric of society. It presents a great opportunity to increase prosperity and growth, which Europe must strive to achieve. Over the last decade, major advances were realised due to the availability of vast amounts of digital data, powerful computing architectures, and advances in AI techniques such as machine learning. Major AI-enabled developments in autonomous vehicles, healthcare, home/service robots, education or cybersecurity are improving the quality of our lives every day. Furthermore, AI is key for addressing many of the grand challenges facing the world, such as global health and wellbeing, climate change, reliable legal and democratic systems and others expressed in the United Nations Sustainable Development Goals*».

(153) Cette phrase compte 12 mots pour 56 caractères, espaces non compris: «*AI also gives rise to certain risks that should be properly managed*».

(154) «*Identification without Consent*»: l'intelligence artificielle permet d'identifier de plus en plus efficacement les personnes. À cet égard, il existe une différence entre identifier un individu et le pister, entre la surveillance ciblée et la surveillance de masse. Concernant la question du consentement, il faut noter que celui-ci est actuellement donné sans grande considération de la part des utilisateurs. Un travail éthique doit donc être opéré pour s'assurer que le consentement donné est digne de cette appellation.

(155) «*Covert AI Systems*»: un être humain doit toujours savoir s'il interagit avec un autre être humain ou avec une machine. Il en va de la responsabilité du développeur. Un problème pourrait survenir concernant les androïdes en ce qu'ils sont conçus pour être aussi proche de l'humain que possible. Leur insertion dans la société pourrait conduire à un floutage de la frontière entre l'homme et la machine, entrainant hypothétiquement une réduction de la valeur attribuée à une vie humaine.

luation des citoyens[156], des armes léthales autonomes[157] et de la conscience artificielle[158]. Les experts précisaient néanmoins dans un encadré ne pas être parvenus à un consensus sur ces préoccupations.

Pour répondre à cette critique d'optimisme béat, les experts ont réinséré ces développements dans le Rapport[159]. Les auteurs reconnaissent que les systèmes IA ne sont que des outils et, en tant que tels, leur utilisation peut être détournée à des fins critiquables. Désormais, l'absence de consensus ne porte plus que sur la question de l'avènement d'une super-intelligence[160]. Cela peut probablement s'expliquer par la composition du groupe, opposant ingénieurs et philosophes, les premiers étant plus sceptiques que les second vis-à-vis de la probabilité de développement d'une super-intelligence[161].

Quoi qu'il en soit, dans le Rapport, les risques identifiés ci-dessus sont mis en parallèle avec quelques exemples de systèmes IA dont les effets seraient bénéfiques à l'humanité. Parmi ces illustrations, on retrouve tout d'abord *l'urgence climatique* (*climate change and sustainable infrastructure*)[162]. Les systèmes IA, en ce qu'ils pourront mieux gérer les infrastructures de transport et les dépenses énergétiques, devraient en effet être en mesure de réduire les émissions des gaz à effet de serre dans l'atmosphère. Les experts citent ensuite *les soins de santé* (*health and well-being*)[163]. Selon eux, les systèmes IA peuvent optimiser la médecine, d'abord en constituant des outils de diagnostic plus

---

(156) « *Normative & Mass Citizen Scoring without consent in deviation of Fundamental Rights* » : l'évaluation normative des citoyens met en péril le libre arbitre et l'autonomie de ceux-ci. Si une évaluation devait avoir lieu – après tout, nous sommes d'ores et déjà évalué notamment à l'école, à l'université et avant l'obtention d'un permis de conduire –, une explication claire de la note obtenue doit pouvoir être donnée au sujet de celle-ci.

(157) « *Lethal Autonomous Weapon Systems* (LAWS) » : les armes léthales autonomes peuvent opérer sans aucun contrôle humain, choisissant d'attaquer ou de cibler tel ou tel individu. Il est impératif qu'en fin de compte, un être humain reste responsable de tous les dommages qui pourraient être causés. Le développement de ces armes autonomes pourrait entraîner une course à l'armement sans précédent, avec toutes les conséquences que cela implique. D'un autre côté, interdire purement et simplement le développement de ces technologies n'est pas une solution en ce que ces armes autonomes pourraient réduire le nombre de pertes de vies humaines, en les programmant par exemple pour ne jamais ouvrir le feu sur un enfant.

(158) « *Potential longer-term concerns* » : à plus long terme, certains enjeux voient leur probabilité d'occurrence devenir extrêmement spéculative. Néanmoins, le groupe d'experts, bien que ne parvenant pas à un consensus, a choisi de tout de même intégrer une brève réflexion sur le risque du développement d'une conscience artificielle.

(159) Rapport, pp. 33-35.

(160) Voy. Rapport, p. 35, note de bas de page 76 (« *While some consider that Artificial General Intelligence, Artificial Consciousness, Artificial Moral Agents, Super-Intelligence or Transformative AI ca be examples of such long-term concerns (currently non-existent), many others believe these to be unrealistic* »).

(161) T. HAGENDORFF, « The Ethics of AI Ethics – An Evaluation of Guidelines », arXiv (Cornell University), 28 février 2019, https://arxiv.org/abs/1903.03425, p. 4.

(162) Rapport, p. 32.

(163) *Ibid.*, pp. 32-33.

précis, mais aussi en permettant un travail préventif, assurant un traitement des patients avant aggravation de leur état. Enfin, l'emphase est mise sur *l'enseignement* (*quality education and digital transformation*)(164). Les systèmes IA peuvent aider à identifier dans quels secteurs il y a un besoin de revitalisation des compétences des travailleurs. Par ailleurs, ils peuvent également lutter contre les inégalités au sein de l'enseignement permettant le développement de programmes d'apprentissage personnalisés.

En synthèse, les experts ont partiellement répondu dans le Rapport à la critique formulée à l'encontre de l'optimisme du Projet. Il est vrai qu'aux côtés des risques soulevés par l'IA, ils ont introduit plusieurs exemples des bienfaits que ces technologies devraient nous apporter. Toutefois, les experts n'ont pas répondu à la question de savoir si le jeu en valait vraiment la chandelle en pondérant avantages et inconvénients.

2. Autres critiques du Projet et réponses du Rapport

Les autres critiques qui ont été soulevées au travers de la consultation publique sont plus hétérogènes et ne requièrent pas une analyse aussi approfondie que la critique précédente. Il nous semble néanmoins pertinent de les mentionner en ce que ces commentaires, et surtout les réponses que les experts y ont apportées, tendent à démontrer une vraie prise en compte de la consultation publique lors de la rédaction du Rapport.

On pense tout d'abord aux commentaires selon lesquels il convient de plus *insister sur les minorités et autres groupes vulnérables*. Alors que le Projet mentionnait déjà ce problème(165), le Rapport réinsiste sur ce point d'une part, au moyen du droit fondamental d'égalité et de non-discrimination(166), et d'autre part, en insérant dans le glossaire une définition de ce qu'il y a lieu d'entendre sous le vocable « personne vulnérable »(167).

---

(164) *Ibid.*, p. 33.

(165) Projet, pp. ii, 9, 10 et 13.

(166) Tant le Projet que le Rapport précisent que le droit fondamental d'égalité, de non-discrimination et de solidarité concerne également les minorités et groupes vulnérables (voy. Projet, p. 7 et Rapport, p. 11). On peut aussi lire, dans la synthèse du Chapitre I du Rapport : « *Pay particular attention to situations involving more vulnerable groups such as children, persons with disabilities and others that have historically been disadvantaged or are at risk of exclusion, and to situations which are characterised by asymmetries of power or information, such as between employers and workers, or between businesses and consumers* » (Rapport, p. 13).

(167) « *No commonly accepted or widely agreed legal definition of vulnerable persons exists, due to their heterogeneity. What constitutes a vulnerable person or group is often context-specific. Temporary life events (such as childhood or illness), market factors (such as information asymmetry or market power), economic factors (such as poverty), factors linked to one's identity (such as gender, religion or culture) or other factors can play a role. The Charter of Fundamental Rights of the EU encompasses under Article 21 on non-discrimination the following grounds, which can be a reference point amongst others: namely sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth,*

Par ailleurs, il semblerait que le Projet n'insiste pas suffisamment sur le fait que *le contexte dans lequel opère un système IA joue un rôle prépondérant dans son développement* et que, ce faisant, les exigences – ou plus précisément leur degré – devraient dépendre du cas d'espèce. Le Rapport le précise désormais à de très nombreuses reprises(168).

De plus, selon les remarques de la consultation publique, *l'impact de l'intelligence artificielle sur le marché de l'emploi* aurait dû être davantage analysé. De telles réflexions ont été insérées au sein des principes éthiques du Rapport, sous le principe du respect de l'autonomie humaine(169). Elles ne constituent néanmoins que les prémisses d'une discussion plus large sur le futur de l'emploi et l'avenir des métiers humains. S'il est vrai que ce commentaire est relativement bref, il faut tout de même reconnaître que l'objectif des experts n'était pas de faire de ce rapport une étude quantitative de l'impact de l'intelligence artificielle sur le marché de l'emploi(170).

En outre, on a reproché au Projet son absence de *référencement de normes légales existantes applicables*, à tout le moins partiellement, aux systèmes IA. Si le Projet précise bien qu'il n'y a pas de vide juridique à cet égard, et c'est un heureux rappel que les discussions juridiques sur l'IA tendent à oublier, aucune citation n'est faite à un autre document que le Règlement Général sur la Protection des Données(171). Bien qu'aucune liste n'ait été ajoutée au Rapport

---

disability, age and sexual orientation. Other articles of law address the rights of specific groups, in addition to those listed above. Any such list is not exhaustive, and may change over time. A vulnerable group is a group of persons who share one or several characteristics of vulnerability ». Rapport, p. 38.

(168) « *While these Guidelines aim to offer guidance for AI applications in general by building a horizontal foundation to achieve Trustworthy AI, different situations raise different challenges. It should therefore be explored whether, in addition to this horizontal framework, a sectorial approach is needed, given the context-specificity of AI systems* » (Rapport, p. 3) ; « *Likewise, different opportunities and challenges arise from AI systems used in the context of business-to-consumer, business-to-business, employer-to-employee and public-to-citizen relationships, or more generally, in different sectors or use cases. Given the context-specificity of AI systems, the implementation of these Guidelines needs to be adapted to the particular AI-application* » (Rapport, p. 6) ; « *While all requirements are of equal importance, context and potential tensions between them will need to be taken into account when applying them across different domains and industries […]. While most requirements apply to all AI systems, special attention is given to those directly or indirectly affecting individuals. Therefore, for some applications (for instance in industrial settings), they may be of lesser relevance* » (Rapport, p. 15) ; « *Given the application-specificity of AI systems, the assessment list will need to be tailored to the specific use case and context in which the system operates* » (Rapport, p. 24).

(169) « *AI systems may also fundamentally change the work sphere. It should support humans in the working environment, and aim for the creation of meaningful work* ».

(170) « *The Commission established the High-Level Expert Group on Artificial Intelligence (AI HLEG), an independant group mandated with the drafting of two deliverables: (1) AI Ethics Guidelines and (2) Policy and Investment Recommendations* » ; *voy*. The European Commission's High Level Expert Group on Artificial Intelligence, « Ethics Guidelines for Trustworthy AI », 8 avril 2019, disponible sur https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477.

(171) Projet, pp. 5 et 11.

sous forme d'annexe, la légalité de l'IA (*lawful AI*) a été ajoutée aux composantes d'une IA digne de confiance(172). Le Rapport présente ainsi une liste non-exhaustive des normes dont les systèmes IA devront tenir compte(173) et insiste sur le fait que certaines règles devront ou non être prises en compte en fonction du contexte dans lequel le système opérera(174).

Dans le même ordre d'idées, la consultation publique a indiqué que la *terminologie du Projet manquait de clarté*. Ainsi, selon certains, l'intelligence artificielle n'était pas convenablement définie dans le Projet, qui confondait IA et systèmes IA(175). À la lecture du Rapport, cette critique n'a plus lieu d'être : la définition retenue distingue bien l'intelligence artificielle en tant que science et les systèmes d'intelligence artificielle en tant que logiciel(176).

---

(172) Rapport, pp. 5 et s.

(173) Sont cités le Traité sur l'Union européenne (TUE), le Traité sur le fonctionnement de l'Union européenne (TFUE), la Charte européenne des droits fondamentaux, le Règlement Général sur la Protection des données (RGPD), la Directive relative à la responsabilité du fait des produits défectueux, le Règlement établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne, les directives anti-discrimination, le droit consumériste, les directives liées à la santé et la sécurité au travail, la Déclaration universelle des droits de l'homme, la Convention européenne des droit de l'homme (CEDH) et les différents droits des Etats membres. Voy. Rapport, p. 6.

(174) Dans le secteur de la santé, il faut par exemple également tenir compte du règlement relatif aux dispositifs médicaux. Voy. par exemple Règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) nº 178/2002 et le règlement (CE) nº 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE, *J.O.U.E.*, L117/1, 5 mai 2017.

(175) La définition de l'intelligence artificielle au sein du Projet est la suivante : «*Artificial intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)*». Voy. Projet, p. iv.

(176) La définition de l'intelligence artificielle au sein du Rapport est la suivante : «*Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)*». Voy. The European Commission's High-Level Expert Group on Artificial Intelligence, «A Definition of AI: Main

Outre la définition de l'IA, de nombreux termes clef sont utilisés sans être clairement expliqués[177]. Ces termes étant cruciaux à la bonne compréhension du Projet, l'absence de définitions les concernant est d'autant plus regrettable. Cela étant dit, il faut tout de même reconnaître que le Projet sort du néant : c'est une première au sein de l'Union européenne. Ce faisant, puisque les experts devaient travailler à partir d'une feuille blanche – degré extrême d'abstraction s'il en est –, on doit reconnaître que le Projet participe à la levée du brouillard entourant l'intelligence artificielle. Reconnaissant néanmoins devoir encore préciser la terminologie employée, les experts ont procédé à de larges ajouts dans le glossaire du Rapport publié le 8 avril 2019[178].

Beaucoup ont également reproché l'*absence de priorisation des principes éthiques et des exigences qui y sont liées*. L'intérêt d'une telle priorisation est de régler les éventuels conflits qui pourraient survenir entre plusieurs d'entre eux. Le Rapport publié le 8 avril 2019 mentionne cette question mais laisse la réponse en suspens[179]. Les principes éthiques sont ainsi listés dans le même ordre que les droits de la Charte des droits fondamentaux sur lesquels ils se basent[180]. Cette absence de hiérarchisation ne manquera pas de soulever des difficultés pratiques lorsque des dilemmes moraux seront rencontrés et que l'incompatibilité des principes éthiques et de leurs exigences sera rencontrée[181].

Enfin, nombreux sont ceux qui ont critiqué *le caractère évasif de la liste d'évaluation* en indiquant qu'aucune réponse n'était apportée aux questions soulevées en son sein. À cet égard, le Rapport du 8 avril 2019 ne répond que partiellement aux attentes. Si sa liste d'évaluation a gagné en précision depuis la publication du Projet le 18 décembre 2018, il ne contient toujours aucune réponse aux questions posées. Cela n'est toutefois guère étonnant en ce que l'objectif poursuivi consiste à fournir une liste de questions à systématiquement se poser lors du développement d'un système IA[182]. Les réponses ne doivent

---

Capabilities and Disciplines », *op. cit.*, disponible sur https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341.

(177) Il s'agit notamment de l'explicabilité, de la transparence et de l'audit. Voy. T. METZINGER, « Dialogue seminar on Artificial Intelligence : Ethical Concerns », 19 mars 2019, disponible sur http ://www.europarl.europa.eu/streaming/?event=20190319-1500-SPECIAL-SEMINAR1&start=2019-03-19T15:44:53Z&end=2019-03-19T15:56:00Z&language=en.

(178) Rapport, pp. 36-38.

(179) *Ibid.*, pp. 13 et 20.

(180) « *Without imposing a hierarchy, we list the principles here below in manner that mirrors the order of appearance of the fundamental rights upon which they are based in the EU Charter* ». Rapport, p. 14, note de bas de page 35.

(181) Voy. Rapport, pp. 13 et 20. Voy. aussi J. LEIKAS *et al.*, « Ethical Framework for Designing Autonomous Intelligent Systems », *Journal of Open Innovation : Technology, Market, and Complexity*, 2019, vol. 5, n° 1.

(182) Rapport, p. 24.

dès lors pas être apportées par le groupe d'experts, mais bien par les parties prenantes qui mettront en œuvre ces principes dans la pratique lors d'une phase de test(183).

II. LES ÉCUEILS D'UNE APPROCHE ÉTHIQUE

Naturellement, un exercice éthique tel que celui réalisé par les experts européens présente plusieurs écueils. Nous présenterons tout d'abord la peur que les documents publiés ne soient qu'un cheval de Troie aux intérêts de l'industrie. Un tel phénomène est connu sous le vocable anglophone d'*ethics lobbying* (A). Une fois cela fait, nous analyserons les difficultés qui dérivent du fait qu'il n'existe pas de principes éthiques universels. Cela mène à la prolifération des propositions de principes éthiques. Ce faisant, on peut craindre de voir les entreprises choisir le cadre éthique le plus faible (*ethics shopping*), prétendre être plus éthiques qu'elles ne le sont vraiment auprès des consommateurs (*ethics bluewashing*) voire, plus grave, exporter leurs activités de recherche dans les ordres juridiques présentant les standards éthiques les plus faibles afin de faire importer leurs systèmes IA par l'Union européenne (*ethics dumping*) (B).

A. *Ethics Lobbying*

L'*ethics lobbying*, également parfois désigné sous le terme d'*ethics washing*(184), peut se définir comme la pratique consistant à utiliser les débats éthiques afin de retarder, modifier, voire même remplacer les législations exécutoires(185). S'il était soumis à une telle stratégie, le Rapport ne serait alors qu'un

(183) *Ibid.*, pp. 24-25. L'enregistrement pour cette phase de test (*piloting phase*) avait lieu sur https://ec.europa.eu/futurium/en/ethics-guidelines-trustworthy-ai/register-piloting-process-0.

(184) Le Professeur Metzinger définit l'*ethics washing* comme le comportement consistant à encourager les débats éthiques afin de repousser l'adoption de législations exécutoires. Voy. T. METZINGER, « Dialogue seminar on Artificial Intelligence : Ethical Concerns », 19 mars 2019, disponible sur http://www.europarl.europa.eu/streaming/?event=20190319-1500-SPECIAL-SEMINAR1&start=2019-03-19T15:44:53Z&end=2019-03-19T15:56:00Z&language=en. Nathalie Smuha et Ben Wagner utilisent également l'appelation *ethics washing*. Voy. N. SMUHA, « The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence », *Computer Law Review International*, 2019, n° 4, p. 101. Voy. égal. B. WAGNER, « Ethics as an Escape from Regulation : From ethics-washing to ethics-shopping? », in M. HILDEBRANDT et S. GUTWIRTH (éd.), *Being Profiling. Cogitas ergo sum*, Amsterdam, Amsterdam University Press, 2018.

(185) P. NEMITZ, « Constitutional democracy and technology in the age of artificial intelligence », *The Royal Society Publishing*, 2018, https://royalsocietypublishing.org/doi/full/10.1098/rsta.2018.0089 ; L. FLORIDI, « Translating Principles into Practices of Digital Ethics : Five Risks of Being Unethical », *Philosophy & Technology*, 2019, vol. 32, p. 188 ; T. HAGENDORFF, « The Ethics of AI Ethics – An Evaluation of Guidelines », *arXiv* (Cornell University), 28 février 2019, https://arxiv.org/abs/1903.03425 ; B. MITTELSTADT, « Principles alone cannot guarantee ethical AI », *Nature*

artifice au service d'intérêts catégoriels, notamment industriels, pour s'assurer que, si de futurs développements régulatoires européens devaient avoir lieu, ils iraient dans son sens(186). En synthèse, l'*ethics lobbying* permettrait d'ignorer la régulation, sous couvert de proposer un cadre éthique prétendument mieux adapté aux innovations technologiques(187).

Certains ont considéré ce risque réalisé au vu de la disparition dans le Rapport de l'idée de « *red lines* », c'est-à-dire des principes éthiques non négociables, qu'il s'agisse d'éléments ou de technologies qui ne doivent en aucun cas être développés en Europe(188). Le choix des termes et des concepts employés – ou en l'occurrence abandonnés – reflèterait ainsi une substance relativement pro-industrie, qui aurait été rendue par une surreprésentation de l'industrie au sein du groupe d'experts.

La peur d'une capture par l'industrie de l'effort éthique fut en effet l'objet de nombreux commentaires reçus lors de la consultation publique. La crainte était qu'une surreprésentation d'experts affiliés à l'industrie – et corollairement la sous-représentation de philosophes et éthiciens – puissent imposer des règles éthiques favorables aux entreprises et non aux consommateurs européens. Ce faisant, la Commission européenne se verrait reléguée à la simple fonction d'estampillage d'une politique qu'elle ne contrôlerait pas *de facto*.

Pour vérifier la réalisation de ce risque, nous avons procédé à une répartition desdits experts(189) en fonction d'une part, du type de leur nomination et d'autre part, de leur milieu professionnel. Les experts peuvent d'abord être

---

Machine Intelligence, 2019, vol. 1, p. 501. R. CALO, « Arificial intelligence policy : a primer and a roadmap », *UC Davis Law Review*, 2017, vol. 51, pp. 399-436.

(186) *Ibid.*

(187) « *Ethical frameworks that provide a way to go beyond existing legal frameworks can also provide an opportunity to ignore them* » ; B. WAGNER, « Ethics as an Escape from Regulation: From ethics-washing to ethics shopping ? », in M. HILDEBRANDT et S. GUTWIRTH (éd.), *Being Profiling. Cogitas ergo sum*, Amsterdam, Amsterdam University Press, 2018 p. 84.

(188) T. METZINGER, « Dialogue seminar on Artificial Intelligence : Ethical Concerns », 19 mars 2019, disponible sur http://www.europarl.europa.eu/streaming/?event=20190319-1500-SPECIAL-SEMINAR1&start=2019-03-19T15:44:53Z&end=2019-03-19T15:56:00Z&language=en.

(189) Si la page d'information du groupe d'experts de haut-niveau, hébergée sur le site Internet de la Commission européenne, précise bien que ledit groupe compte 52 membres, seuls 51 d'entre eux sont cités à la fin du Projet et du Rapport. Par ailleurs, sans que cela soit étonnant, la liste annexée au Projet du 18 décembre ne correspond ni à celle du Rapport, ni à celle présentée en ligne, les membres du groupe évoluant constamment. Afin d'analyser la répartition par type de nomination et par domaine d'expertise, nous nous référons aux informations fournies le registre des groupes d'experts de la Commission et d'autres entités similaires. Voy. High-Level Expert Group on Artificial Intelligence (E03591), Group Details, Members, disponible sur https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3591&NewSearch=1&NewSearch=1.

répartis en fonction de quatre catégories : ceux qui ont été nommés en raison de leur qualification personnelle (type A), ceux qui ont été nommés parce que représentants d'un intérêt commun (type B), les organisations (type C) et les autres entités publiques (type E). Les experts du type A ou B peuvent ensuite être répartis selon leur domaine d'expertise, c'est-à-dire, en l'occurrence, technologique, philosophique ou juridique. Les experts de type C ou E sont, quant à eux, distingués selon qu'il s'agit d'une organisation non gouvernementale, d'une entreprise, d'une association professionnelle ou commerciale, d'une institution académique, institut de recherche ou *think tank*, d'une autre association professionnelle, d'un cabinet d'avocats ou d'une autre organisation(190). Le tableau 1 ci-dessous présente la répartition des experts selon cette double classification, tandis que la figure 5 vient graphiquement l'illustrer.

Type	Expertise	Nombre de membres
A	Engineering (IT)	13
	Philosophy	2
	Law	2
B	Engineering (IT)	1
	Philosophy	1
C	NGOs	3
	Companies/Groups	13
	Trade and business associations	5
	Academia, Research Institute and Think Tanks	4
	Professionals' Associations	1
	Law Firms	1
	Other Organisations	2
E	EU Institutions/Bodies	2

Tableau 1 – Répartition des experts de la Commission

(190) La répartition des experts selon leur type et expertise est annexée au présent article.

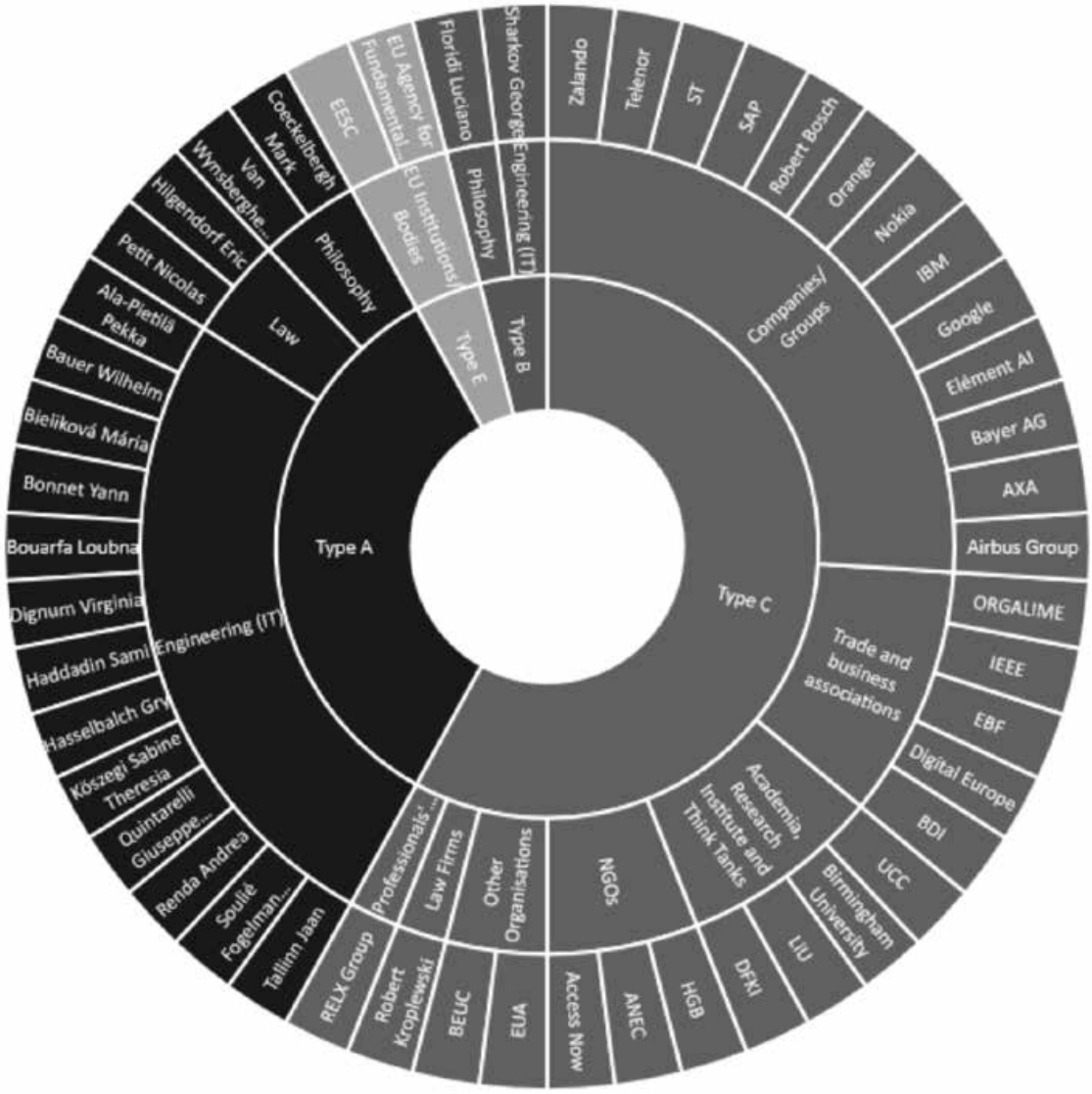

Figure 5 – Répartition des experts de la Commission

On constate aisément que la majorité des membres du groupe sont effectivement issus d'entreprises et ne présentent pas, *a priori*, de formation en philosophie. Si une telle représentation de l'industrie peut surprendre, il faut cependant se souvenir, d'une part, que le groupe d'experts, outre une réflexion sur l'éthique, devait également fournir des recommandations liées à la politique d'investissement de l'Union européenne vis-à-vis de l'intelligence artificielle[191], et d'autre part, que les principes et exigences éthiques avaient pour objectif ini-

(191) « *The Commission established the High-Level Expert Group on Artificial Intelligence (AI HLEG), an independant group mandated with the drafting of two deliverables: (1) AI Ethics Guidelines and (2) Policy and Investment Recommendations* » ; voy. The European Commission's High Level Expert Group on Artificial Intelligence, « Ethics Guidelines for Trustworthy AI », 8 avril 2019, disponible sur https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477.

tial d'être approuvés et appliqués sur base volontaire par les entreprises(192), ce qui impliquait de les intégrer aux discussions.

Quoi qu'il en soit, la question reste de savoir si, *in fine*, il y a eu détournement de l'effort éthique européen par les entreprises, qui se serviraient alors des lignes directrices comme d'un bouclier contre une future réglementation. Deux éléments tendent à indiquer qu'une telle capture n'a pas eu lieu.

Premièrement, l'adoption du document devait avoir lieu par consensus, ce qui empêche *de facto* qu'un groupe puisse imposer ses opinions aux autres(193). Lorsqu'un consensus ne pouvait pas être atteint, le Rapport le précisait dans un encadré ou note de bas de page(194). Par ailleurs, chaque expert avait la possibilité de ne pas signer le document et de rédiger une opinion dissidente, option qu'aucun d'entre eux n'a choisi(195).

Deuxièmement, le Rapport lui-même précise que les lignes directrices ne visent ni à se substituer à toute forme actuelle ou future de politiques ou réglementations, ni à en décourager l'introduction(196). Ce faisant, les experts précisent que les lignes directrices éthiques n'ont aucunement vocation à empêcher l'éventuelle adoption d'une réglementation européenne relative aux systèmes IA.

B. *Ethics Shopping*, *Ethics Bluewashing* et *Ethics Dumping*

Il n'existe pas d'éthique universelle(197). Par nature, le concept de ce que nous estimons être bien ou mal est profondément personnel. Certains préféreront les vertus aristotéliciennes, d'autres le déontologisme kantien, d'autres encore l'utilitarisme benthamien. Face à la diversité de normes éthiques, comment le législateur est-il supposé choisir la théorie qu'il va utiliser ?

L'exemple de la voiture autonome, en ce qu'il est le plus discuté, est sans doute le plus révélateur. Le Massachussets Institute of Technology (MIT) a actualisé le dilemme du tramway à la voiture autonome et en a proposé plu-

(192) Projet, pp. i et 2.

(193) N. SMUHA, « The EU Approach to Ethics Guidelines for Trustworthy Artificial ntelligence », *Computer Law Review International*, 2019, nº 4, p. 103.

(194) C'est le cas par exemple de la super-intelligence. Concernant le Projet, voy. pp. 11-13. Concernant le Rapport, voy. p. 35, note 76.

(195) N. SMUHA, « The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence », *Computer Law Review International*, 2019, nº 4, p. 103.

(196) « *These Guidelines do not intend to substitute any form of current or future policymaking or regulation, nor do they aim to deter the introduction thereof* » ; Rapport, p. 3. Le Projet contenait une déclaration similaire (« *Importantly, these Guidelines are not intended as a substitute to any form of policymaking or regulation […], nor do they aim to deter the introduction thereof* »). Voy. Projet, p. ii.

(197) En guise d'exception, il semblerait que l'interdit de l'inceste soit bien universel. A. RASCOVSKY et M. RASCOVSKY, « The Prohibition of Incest, Filicide and the Sociocultural Process », *International Journal of Psycho-Analysis*, 1991, vol. 53, pp. 271-276.

sieurs variantes dans une large étude ouverte au public[198]. Il en est ressorti une division du monde selon d'une part, des facteurs culturels et d'autre part, des facteurs économiques. D'un point de vue culturel, les cultures individualistes tendent à adopter une logique conséquentialiste, préférant épargner le plus grand nombre en cas d'accident. D'un autre côté, les sociétés collectivistes sont plus enclines à épargner les membres les plus âgés de la population[199]. D'un point de vue économique, les pays les plus prospères ont une préférence pour sacrifier ceux qui sont mis en danger suite à une infraction, par exemple les piétons qui traversent illégalement la route[200].

Parce qu'il n'y a pas d'éthique universelle, le risque est de voir différents pays – ou groupes de pays – adopter des solutions différentes. En l'occurrence, de nombreux auteurs ont déjà démontré que l'éthique appliquée à l'intelligence artificielle générait un grand nombre de recommandations[201]. Or, plus les propositions de cadres éthiques seront nombreuses et variées, plus la confusion entre elles sera grande. La conséquence sera que les entreprises, plutôt que de modifier leurs comportements pour devenir éthiques, s'implanteront au sein des ordres juridiques aux valeurs éthiques les plus faibles, justifiant le mieux leurs comportements actuels[202]. Cette pratique est connue sous le nom d'*ethics shopping* et est similaire au *forum shopping*, ou élection de juridiction, qui consiste à choisir la juridiction qui soit la plus favorable à ses intérêts[203].

La gravité de l'*ethics shopping* est cependant à relativiser. D'une part, la plupart des différences entre les recommandations éthiques ne sont pas des différences sémantiques mais plutôt de vocabulaire[204]. D'autre part, les développements éthiques cristallisés dans le Rapport vont rendre plus difficile cette pratique des entreprises en ce qu'il propose un cadre commun à plusieurs pays[205]. Ce phénomène met en effet en évidence le besoin de concevoir une réglementa-

---

(198) Cette expérience est accessible en ligne à l'adresse http://moralmachine.mit.edu/. Des millions de personnes issues de 233 pays différents ont répondu à l'enquête. E. AWAD, S. DSOUZA, R. KIM, J. SCHULZ, J. ENRICH, A. SHARIFF, J.-F. BONNEFON et L. RAHWAN, « The Moral Machine experiment », *Nature*, 2018, vol. 563, p. 59.

(199) *Ibid.*, p. 62.

(200) *Ibid.*

(201) Voy. *supra*, note 88.

(202) L. FLORIDI, « Translating Principles into Practices of Digital Ethics : Five Risks of Being Unethical », 2019, *Philosophy & Technology*, vol. 32, p. 186.

(203) P. DE VAREILLES-SOMMIÈRES, *Forum Shopping in the European Judicial Area*, Hart Publishing, Oxford, 2007.

(204) L. FLORIDI, « Soft Ethics, the governance of the digital and the General Data Protection Regulation », *The Royal Society Publishing*, 15 octobre 2018, https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0081.

(205) Voy. Rapport, p. 3. Voy. aussi L. FLORIDI, « Translating Principles into Practices of Digital Ethics : Five Risks of Being Unethical », *Philosophy & Technology*, 2019, vol. 32, p. 187.

tion éthique à un niveau supranational, tel que l'Union européenne(206). Le cadre éthico-juridique européen devra donc s'appliquer aux systèmes IA opérant en Europe, indépendamment de leurs origines(207), appliquant ainsi une logique similaire à celle choisie vis-à-vis de la protection des données personnelles(208).

Toutefois, l'*ethics shopping* mène tout droit à deux autres écueils, l'*ethics bluewashing* et l'*ethics dumping*. Dérivé du *greenwashing* environnemental, l'*ethics bluewashing* désigne la stratégie consistant tout simplement pour une entreprise, après avoir sélectionné le cadre éthique qui lui convient le mieux (*shopping*), de chercher à apparaître plus éthique qu'elle ne l'est vraiment au moyen de vastes campagnes marketing (*washing*)(209). À nouveau, le Rapport lutte contre ce phénomène au moyen des exigences de transparence et d'éducation(210).

Les risques induits par l'*ethics dumping* sont, quant à eux, beaucoup plus importants. Se servant des nombreux cadres éthiques proposés par les États, les entreprises peuvent choisir de relocaliser leurs activités de recherches dans l'ordre juridique qui présente les standards éthiques les plus faibles(211). En tant qu'importateur de systèmes IA, l'Union européenne devra veiller à ce que les produits importés n'aient pas été développés au sein d'États aux standards éthiques plus faibles. Une solution peut être trouvée dans le mécanisme de la certification, piste de réflexion proposée par les auteurs du Rapport(212).

(206) European Group on Ethics in Science and New Technologies, « Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems » (European Commission, 9 March 2018), http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf.

(207) N. SMUHA, « The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence », *Computer Law Review International*, 2019, nº 4, p. 101.

(208) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119.

(209) L. FLORIDI, « Translating Principles into Practices of Digital Ethics : Five Risks of Being Unethical », *Philosophy & Technology*, 2019, vol. 32, p. 187.

(210) Rapport, pp. 18 et 23.

(211) Le Professeur Metzinger utilise la terminologie d'*ethics shopping* pour désigner ce phénomène. Voy. T. METZINGER, « Dialogue seminar on Artificial Intelligence : Ethical Concerns », 19 mars 2019, disponible sur http://www.europarl.europa.eu/streaming/?event=20190319-1500-SPECIAL-SEMINAR1&start=2019-03-19T15:44:53Z&end=2019-03-19T15:56:00Z&language=en. Voy. aussi C. WALKER-OSBORN et C. HAYES, « Ethics and AI a moral conundrum », *The British Computer Society*, 11 juin 2018, disponible sur https://www.bcs.org/content-hub/ethics-and-ai-a-moral-conundrum/.

(212) Rapport, p. 23. Voy. égal. L. FLORIDI, « Translating Principles into Practices of Digital Ethics : Five Risks of Being Unethical », *Philosophy & Technology*, 2019, vol. 32, p. 190.

III. LA RELATION SYMBIOTIQUE DE L'ÉTHIQUE ET DE LA SCIENCE JURIDIQUE

Nous avons vu qu'une approche éthique est sujette à de nombreux écueils. Le Rapport européen permet d'y répondre. On ne peut considérer ses auteurs soumis à une forme d'*ethics lobbying*. Ces derniers précisent en effet que le Rapport n'est pas un substitut de réglementation et qu'ils proposent, dans leurs recommandations liées à l'investissement, de faire concrétiser les exigences éthiques au moyen de normes *ex ante*. Le Rapport, en ce qu'il propose un cadre cohérent au sein de l'Union européenne, réduit également le risque d'*ethics shopping*. Par ailleurs, en insistant sur la transparence et l'éducation, l'*ethics bluewashing* est mise en difficulté, tandis que la certification permet de lutter efficacement contre l'*ethics dumping*.

Il n'en reste pas moins que ces difficultés auraient facilement pu être évitées si, plutôt qu'une approche éthique, une réglementation exécutoire avait été envisagée. Il est donc légitime de se demander pourquoi l'éthique a été envisagée comme prérequis nécessaire à l'adoption d'une législation européenne permettant d'encadrer efficacement l'intelligence artificielle.

Cette question met en évidence un besoin de clarification quant aux relations entre l'éthique et le droit. La distinction n'est pas évidente de prime à bord. Une des premières définitions du droit ne précise-t-elle pas qu'il est l'art du bon et de l'équitable[213] ? Malgré des liens évidents, éthique et droit peuvent être opposés sur plusieurs plans. Tandis que le droit s'intéresse à l'acte qui a été posé, l'éthique insiste plus sur l'intention de l'auteur. Par ailleurs, le droit est créateur de droits et d'obligations, l'éthique, bien que prescrivant certains comportements, ne peut faire naître de droits ni objectifs, ni subjectifs. Enfin, l'éthique échappe aux sanctions étatiques, pas le droit qui repose sur une forme de coercition légitime[214].

Malgré leurs différences, se concentrer exclusivement sur les efforts éthiques de l'Union européenne, c'est ne voir que l'arbre qui cache la forêt. Le groupe d'experts a lui-même appelé à l'adoption d'une régulation en proposant dans ses Recommandations que soit établi un cadre régulatoire approprié[215], notamment au regard de l'organisation d'un audit des développeurs et des exi-

---

(213) Selon Ulpien, citant Celse, *ius est ars boni et aequi* ; Ulpien, (au livre 1er de ses Institutes), Digeste, Livre 1, Titre 1, lex 1, principium.

(214) C. PERELMAN, « Droit et morale » (Rapport présenté au XIVe Congrès international de Philosophie, Vienne, septembre 1968), in C. PERELMAN, *Ethique et Droit*, 2e éd., Bruxelles, Édition de l'Université de Bruxelles, 2012, p. 364.

(215) The European Commission's High Level Expert Group on Artificial Intelligence, « Policy and Investment Recommendations for Trustworthy AI », 26 juin 2019, disponible sur https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence, p. 37.

gences de sécurité à établir *ex ante*(216). Par ailleurs, ce groupe ne constitue pas la seule force vive européenne relative à l'intelligence artificielle, et d'autres initiatives, régulatoires cette fois-ci, sont d'ores et déjà à l'œuvre(217).

Les développements éthiques ont pour objectif d'informer le législateur européen, lui permettant par la suite d'adopter, ou non, les règles qui lui sembleront pertinentes et qui, elles, seront exécutoires(218). C'est en symbiose que l'éthique et le droit déploient optimalement leurs effets. Ces deux disciplines sont chacune une condition nécessaire mais non suffisante à l'encadrement de l'intelligence artificielle(219).

L'éthique et la science juridique s'inspirent mutuellement. Si le législateur s'inspire – ou devrait s'inspirer – de la morale avant de créer du droit, l'éthicien devrait également analyser les règles juridiques. Car si « les principes fondamentaux de la morale […] ne peuvent être contesté *in abstracto* […] dès qu'il s'agit de les appliquer dans des circonstances concrètes, ils donneront lieu à des controverses à l'infini »(220). Si les règles sibyllines de droit sont décryptées et analysées par la doctrine et la jurisprudence, à qui revient la charge de l'explication des règles éthiques ? Malheur à celui qui, voulant faire respecter une règle, en appliquera la lettre et non l'esprit(221).

L'éthique a besoin de la science juridique, et réciproquement. C'est d'autant plus vrai que les recherches éthiques ont pour objectif d'informer le législateur(222). Le Professeur Floridi, filant la métaphore, explique que si les normes adoptées par le législateur sont les règles du jeu, l'éthique consiste en les consignes permettant d'être un bon joueur et de gagner la partie(223). Les règles éthiques peuvent en effet suggérer de ne pas adopter un comportement

(216) N. Smuha, « The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence », *Computer Law Review International*, 2019, n° 4, p. 102.

(217) Commission européenne, « White Paper on Artificial Intelligence – a European approach to excellence and trust », COM(2020) 65 final, 19 février 2020.

(218) L. Floridi, « Establishing the rules for building trustworthy AI », *Nature Machine Intelligence*, 2019, vol. 1, pp. 261-262.

(219) N. Smuha, « The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence », *Computer Law Review International*, 2019, No. 4, p. 101.

(220) C. Perelman, « Droit et morale » (Rapport présenté au XIVe Congrès international de Philosophie, Vienne, septembre 1968), *in* C. Perelman, *Ethique et Droit*, 2e éd., Édition de l'Université de Bruxelles, Bruxelles, 2012, p. 368.

(221) N. Wiener, *The Human Use of Human Beings, Cybernetics and Society*, 2e ed., 1954, Michigan, Houghton Mifflin Harcourt Publishing Company. Pour la traduction française de Pierre-Yves Mistoulon, *voy*. N. Wiener, *Cybernétique et société : l'usage humain des êtres humains*, Édition du Seuil, Paris, 2014, p. 210.

(222) L. Floridi, « Establishing the rules for building trustworthy AI », *Nature Machine Intelligence*, 2019, Vol. 1, p. 261.

(223) *Ibid.*, p. 262. Voy. aussi L. Floridi, « Soft Ethics and the Governance of the Digital », *Philosophy & Technology*, 2018, vol. 31, n° 1, p. 4 et L. Floridi *et al.*, « AI4People – An Ethical Framework for a Good AI Society : Opportunities, Risks, Principles, and Recommendations », *Minds and Machines*, 2018, vol. 28, p. 694.

qui n'est pourtant pas interdit par la loi, ou au contraire d'agir d'une certaine manière sans pourtant y être contraint par le législateur(224). Un développement éthique des systèmes IA implique donc d'identifier *ex ante* les valeurs à respecter(225), dans une optique de transcription des réflexions éthiques *by-design*(226). C'est en fonction de l'identification des bonnes pratiques – et corollairement de leurs dérives – que le législateur pourra par la suite adopter une règle qui soit pertinente. L'approche éthique de l'intelligence artificielle permet ainsi une prise de conscience des parties prenantes, tant au niveau des développeurs que des utilisateurs et des pouvoirs publics(227).

À cet égard, si l'éthique est sous le feu des critiques au regard de ses faiblesses exécutoires(228), les règles législatives le sont également, et ce pour l'exact inverse : l'adoption d'un cadre régulatoire relatif à l'intelligence artificielle pourrait, s'il est prématuré ou trop strict, étouffer l'innovation(229).

Qu'une réglementation excessive ait pour effet de tuer l'innovation dans l'œuf n'est pas un phénomène récent. En 1865, pour encadrer l'utilisation des automobiles, le Parlement britannique adopta le *Locomotives on Highways Act*, qui d'une part, limitait la vitesse des engins en ville à 2 mille par heure (Mph) et en milieu rural à 4 Mph, soit respectivement environ 3,21 et 6,43 kilomètre par heure (Kmh), et d'autre part, imposait qu'un piéton portant un drapeau rouge précède le véhicule de 60 yards, soit environ 54,86 mètres, raison pour laquelle cette législation était connue sous le nom du *Red Flag Act*(230). On accuse aujourd'hui cette norme d'avoir empêché le développement de l'industrie automobile anglaise jusqu'au début du XX^{e} siècle(231).

Par ailleurs, réguler est un processus le plus souvent réactif et donc, par voie de conséquence, inadapté aux progrès technologiques car incapable d'en suivre le rythme d'évolution. Ce risque « d'obsolescence dès l'adoption »(232) s'explique par le fait que le législateur ne peut convenablement évaluer les opportunités et les risques soulevés par la technologie tant qu'elle ne se déve-

(224) L. Floridi, « Establishing the rules for bulding trustworthy AI », *Nature Machine Intelligence*, 2019, vol. 1, p. 262.
(225) J. Leikas *et al.*, « Ethical Framework for Designing Autonomous Intelligent Systems », *Journal of Open Innovation : Technology, Market, and Complexity*, 2019, vol. 5, n° 1.
(226) Rapport, p. 21.
(227) B. Mittelstadt, « Principles alone cannot guarantee ehical AI », *Nature Machine Intelligence*, 2019, vol. 1, p. 501.
(228) L. Floridi, « Establishing the rules for bulding trustworthy AI », *Nature Machine Intelligence*, 2019, vol. 1, p. 261.
(229) *Ibid.*, p. 262. Voy. aussi J. Pelkmans et A. Renda, « Does EU regulation hinder or stimulate innovation ? », *CEPS Special Report*, 2014, p. 96.
(230) Voy. « Automotives on Highways Act », *Encyclopaedia Britannica*, https://www.britannica.com/topic/Locomotives-on-Highways-Act.
(231) Sur cette anecdote, voy. N. Petit, « Law and Regulation of Artificial Intelligence and Robots – Conceptual Framework and Normative Implications », *SSRN*, 14 mars 2017, disponible sur https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931339, p. 12.
(232) *Ibid.*

loppe pas. Cette attente accroît le risque de voir les conséquences négatives se réaliser(233). Ce problème est connu sous le nom de dilemme de Collingridge(234).

On est donc face à un véritable paradoxe : d'une part, l'éthique est critiquée parce que trop souple et, d'autre part, le droit est critiqué parce que trop rigide et inadapté aux rapides évolutions technologiques. Sans doute est-ce dans la combinaison de ces deux approches – éthique et juridique – que se trouve la meilleure option. L'éthique jouerait ainsi le rôle de pionnière en défrichant les terres sauvages de la gouvernance des nouvelles technologies, permettant l'établissement prospère d'une colonie juridique.

C'est cette combinaison éthico-juridique que prône le groupe d'experts : partant d'une page blanche, le Rapport éthique est vu comme un « document vivant »(235), évoluant au gré de sa réception auprès du public concerné. Son texte a déjà été modifié suite à la consultation publique début 2018, il le sera encore en 2020 après la phase de test des exigences qu'il propose(236). Ce n'est qu'ensuite que certains mécanismes – comme la certification – pourront être mis en œuvre par la législation, ce qu'appellent de leurs vœux les experts dans leurs recommandations relatives à une politique d'investissement(237).

CONCLUSION

Dans cet article, nous avons présenté les lignes directrices en matière d'éthique dans le domaine de l'intelligence artificielle rédigées par le groupe d'experts indépendants de haut niveau de la Commission européenne.

Nous avons eu l'occasion d'exposer les principes éthiques mis en avant dans ce Rapport, ainsi que les droits fondamentaux auxquels ils se rapportent. À cet égard, nous avons eu l'occasion de souligner des influences issues tant de l'éthique aristotélicienne de la vertu, de l'éthique déontologique kantienne et, dans une moindre mesure, du conséquentialisme. Nous avons particulièrement insisté sur le rapprochement de l'éthique de l'intelligence artificielle et de la bioéthique, deux illustrations de l'éthique appliquée.

Nous avons ensuite analysé les difficultés qu'un tel exercice éthique ne manque pas de soulever. Après analyse du Rapport, nous avons conclu que les

(233) G.N. MANDEL, « Regulating emerging technologies », *Law, Innovation and Technology*, 2009, vol. 1, nº 1, pp. 75-92.

(234) D. COLLINGRIDGE, *The Social Control of Technology*, London, Francis Pinter Ltd., 1980.

(235) Rapport, p. 3.

(236) Commission européenne, « White Paper on Artificial Intelligence – a European approach to excellence and trust », COM(2020) 65 final, 19 février 2020, p. 9.

(237) The European Commission's High Level Expert Group on Artificial Intelligence, « Policy and Investment Recommendations for Trustworthy AI », 26 juin 2019, disponible sur https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence, p. 37.

experts avaient pu naviguer sur les eaux troubles de la réflexion éthique, tout en en évitant les écueils.

Même s'il est parfois trop flou quant à ses aspects pratiques, le Rapport constitue la première étape d'une réflexion plus vaste sur l'intelligence artificielle. L'absence de mécanisme de sanction en cas de non-respect de son prescrit enlève toute possibilité d'exécution des principes éthiques autres que sur base volontaire. Les experts, conscients de ce problème, ont appelé la mise en place d'une réglementation qui cristalliserait les exigences du Rapport.

Le Rapport constitue donc la pierre angulaire d'une réflexion européenne. Il a le mérite de poser les prémisses de la discussion, d'identifier les risques et de proposer des solutions permettant d'en éviter la réalisation.

En adoptant ces principes éthiques, l'Union européenne vient ainsi se positionner face aux États-Unis et à la Chine. Le message envoyé est fort : les systèmes IA qu'utiliseront les citoyens européens devront être dignes de confiance.

ANNEXE I

Type A – Individual expert apointed in his-her personal capacity	
Ala-Pietilä Pekka	Engineering (IT)
Bauer Wilhelm	Engineering (IT)
Bieliková Mária	Engineering (IT)
Bonnet Yann	Engineering (IT)
Bouarfa Loubna	Engineering (IT)
Coeckelbergh Mark	Philosophy
Dignum Virginia	Engineering (IT)
Haddadin Sami	Engineering (IT)
Hasselbalch Gry	Engineering (IT)
Hilgendorf Eric	Law
Köszegi Sabine Theresia	Engineering (IT)
Petit Nicolas	Law
Quintarelli Giuseppe Stefano	Engineering (IT)
Renda Andrea	Engineering (IT)
Soulié Fogelman Françoise	Engineering (IT)
Tallinn Jaan	Engineering (IT)
Van Wynsberghe Aimee	Philosophy
Type B – Individual expert appointed as representative of a common interest	
Sharkov George	Engineering (IT)
Floridi Luciano	Philosophy
Type C – Organisation	
Access Now	NGOs
Airbus Group SE	Companies/Groups
AXA	Companies/Groups
Bayer AG	Companies/Groups

Bundesverband der Deutschen Industrie e.V. (BDI)	Trade and business associations
Bureau Européen des Unions de Consommateurs (BEUC)	Other Organisations
Deutsches Forschungszentrum für Künstliche Intelligenz DFKI GmbH (DFKI)	Academia, Research Institute and Think Tanks
DIGITALEUROPE (DE)	Trade and business associations
Élément AI inc.	Companies/Groups
European Association for the Co-ordination of Consumer Representation in Standardisation (ANEC)	NGOs
European Banking Federation (EBF)	Trade and business associations
European University Association (EUA)	Other Organisations
Google	Companies/Groups
Hilfsgemeinschaft der Blinden und Sehschwachen Österreichs (HGB)	NGOs
IBM Corporation (IBM)	Companies/Groups
Institute of Electrical and Electronics Engineers, Incorporated (IEEE)	Trade and business associations
Linköpings universitet (LiU)	Academia, Research Institute and Think Tanks
Nokia	Companies/Groups
Orange	Companies/Groups
ORGALIME – The European Technology Industries (ORGALIME)	Trade and business associations
RELX Group	Professionals' Associations
Robert Bosch Gmbh	Companies/Groups
Robert Kroplewski Kancelaria Radcy Prawnego (Robert Kroplewski)	Law Firms
SAP	Companies/Groups
STMicroelectronics (ST)	Companies/Groups
Telenor	Companies/Groups

The University of Birmingham	Academia, Research Institute and Think Tanks
University College Cork, National University of Ireland, Cork (UCC)	Academia, Research Institute and Think Tanks
Zalando SE	Companies/Groups
Type E – Other Public Entity	
European Union Agency for Fundamental Rights	EU Institutions/Bodies
The European Economic and Social Committee (EESC)	EU Institutions/Bodies