Réglementations européennes de l'intelligence artificielle : la tentation du pragmatisme

Les droits humains à l'épreuve des impératifs économiques, politiques et sécuritaires en Europe

Par Yannick Meneceur[1], magistrat en disponibilité et maître de conférences associé à l'Université de Strasbourg

Introduction : une construction règlementaire en tension avec les paradoxes de notre temps

Les paradoxes et les tensions parcourant nos sociétés contemporaines continuent, comme depuis toujours, à animer et à structurer un monde toujours plus complexe. La haute sophistication de certains développements techniques, comme le numérique, contribuent pleinement à alimenter cette mécanique : tout à la fois outil incontestable d'amélioration de l'efficacité de divers traitements de l'information et instrument à la puissance inédite pour contrôler des masses, l'informatique (au sens large en y incluant les réseaux de communication) est un vecteur radical de transformation des rapports entre individus. Dans cette phase de transition majeure et tout à fait soudaine à l'échelle de l'histoire des civilisations, l'intelligence artificielle (« IA[2] ») est aujourd'hui au cœur des promesses d'optimisation de notre ère moderne, en parvenant à s'infiltrer dans la plupart de ses interstices (loisirs et temps libre). Présentée dans nombre de discours comme plus fiable et plus neutre que les humains, nourrie d'une représentation mathématisée de nos environnements à l'aide de données devenues massives, « l'IA » s'impose comme un moteur d'une prise de décision rationalisée, poursuivant un projet, initié par la statistique, de gouvernance par la quantification du monde[3]. Pourtant, les conséquences de la généralisation de l'usage de ce qu'il conviendrait mieux de qualifier de « systèmes algorithmiques » sont devenues assez claires dans l'opinion publique et ont échappé à la seule sphère des techno-critiques du fait de leur particulière consistance. Risques de discriminations et de biais, impact sur la hiérarchisation et la qualité de l'information[4] ainsi que sur la production de savoirs[5], nouvelle forme de gouvernementalité vidant de tout débat contradictoire la vie publique (et politique) : voici quelques exemples d'une révolution[6] emportant, sur son passage, des équilibres et des acquis supposés pérennes.

Car, dans le même temps que l'informatique commençait à se déployer dans les années 1950, l'horreur de la Seconde Guerre mondiale avait conduit à une prise de conscience inédite sur la place

[1] Les opinions exprimées n'engagent que l'auteur.
[2] Afin de se garder de tout anthropomorphisme et par commodité éditoriale, le terme d'intelligence artificielle sera présenté sous la forme de son acronyme et entre guillemets. Cette présentation a été choisie en remplacement des termes, souvent plus appropriés, de « systèmes d'intelligence artificielle », « d'outils d'intelligence artificielle » ou « d'applications de l'intelligence artificielle ».
[3] V. notamment A. Desrosières, La politique des grands nombres – Histoire de la raison statistique, Éd. La Découverte, 1993 et A. Supiot, La gouvernance par les nombres – Cours au Collège de France (2012-2014), Fayard, coll. Poids et mesures du monde, 2015
[4] J. Christ, De l'intime au public : Habermas à l'épreuve des réseaux sociaux, AOC Media, 25 août 2022, accessible sur : https://aoc.media/analyse/2022/08/24/de-lintime-au-public-habermas-a-lepreuve-des-reseaux-sociaux-2/, consulté le 23 octobre 2022
[5] Ch. Anderson, The End of Theory. The Data deluges makes the scientific method obsolete, Wired, 23 juin 2008
[6] Le terme peut sembler rebattu, mais nous le retiendrons au sens d'un changement transversale, brusque et rapide dans la structure politique, économique et sociale de nos sociétés contemporaines.

CC BY NC ND

de droits universels, inaliénables et indivisibles[7]. Des mécanismes juridiques inédits, comme la Déclaration universelle des droits de l'homme au sein des Nations Unies en 1948 ou la Convention européenne des droits de l'homme (CvEDH) au Conseil de l'Europe en 1950 se sont imposées afin de protéger les individus de la résurgence de projets politiques funestes et de l'oppression de régimes autoritaires. La portée de la CvEDH sera considérable dans cette construction, en ajoutant à l'aspect contraignant du traité international la constituant la force d'une juridiction spécialisée, directement accessible aux citoyens, dont les décisions s'imposent aux États parties. De manière plus large, c'est le projet européen imaginé par Churchill, Adenauer, Schuman, Spaak, de Gasperi et Bevin, hommes politiques édifiés de l'horreur de deux guerres mondiales, qui trouve ses racines dans des valeurs encore portées aujourd'hui par le Conseil de l'Europe : droits de l'Homme, démocratie et État de droit. Mais, assez inévitablement, l'entropie a fait son œuvre en érodant progressivement nos mémoires vives des multiples causes des catastrophes passées. L'énergie politique consacrée à la revivification d'un projet marchand mondial libéral a lentement déplacé ce qui avait été créé pour être une clé de voûte de nos sociétés contemporaines vers un discours d'intérêt, poli et revendiqué comme essentiel, mais concrètement perçu par les élites politiques et économiques comme étant parfois trop extrême ou militant. S'il est vrai que la défense des libertés fondamentales et des droits humains reste vivement saisie par une active société civile, les fondements des droits de l'Homme sont plutôt à trouver dans l'héritage humaniste des Lumières et une construction juridique plaçant au sommet de la hiérarchie de normes des principes et valeurs s'imposant à toute société s'affirmant comme démocratique. Mais, pour le regard de décideurs publics nourris de doctrine économique, la représentation du monde est nécessairement devenue plurielle. Le droit, en général, n'est qu'une externalité parmi d'autres, contribuant surtout à organiser et structurer les échanges entre individus. Sans renier, *prima facie*, les discours plaçant des droits de l'Homme au cœur du projet démocratique, la liberté d'entreprendre, l'innovation technologique et la croissance économique (de plus en plus verdie ou rendue sobre au passage) paraitraient avoir à nouveau acquis dans l'opinion commune une valeur au moins égale pour achever une société de progrès.

Les paradoxes et tensions parcourant les deux univers décrits, technologiques et juridiques, pourraient paraître relativement théoriques et abstraits, en contraste avec notre quotidien immédiat empli de crises (sanitaires, sécuritaires, économiques). La rencontre de deux projets d'instruments juridiques contraignants sur « l'IA » en Europe rend toutefois extrêmement sensible ces tensions : à Bruxelles, le projet de règlement sur l'intelligence artificielle a été présenté par la Commission européenne en avril 2021, en élément d'une vaste politique de réglementation du numérique pour organiser le marché ; à Strasbourg, le projet de traité international du Conseil de l'Europe (sous la dénomination de convention) est entré dans sa phase de négociation depuis janvier 2022 avec un mandat confié à un comité intergouvernemental, dans l'objectif d'établir un standard juridique contraignant fondé sur les droits de l'Homme, la démocratie et l'État de droit. S'il semble évident pour les deux régulateurs régionaux européens, aux missions bien distinctes, que le lourd problème de confiance dans les divers instruments algorithmiques trouve une solution au-delà de l'éthique, les réponses juridiques, présentées un temps comme complémentaires, semblent se trouver aujourd'hui mises en concurrence. Car, dans les politiques publiques, « l'IA » est bien devenue d'un

[7] Ou, selon la Déclaration et Programme d'action de Vienne du Comité des droits de l'homme des Nations unies (1993) : « universels, indissociables, interdépendants et intimement liés », https://treaties.un.org/doc/source/docs/A_CONF.39_11_Add.2-F.pdf

CC BY NC ND

très haut niveau stratégique dans la plupart de ses domaines d'application ; contrôler des applications avancées de « l'IA », c'est assurer un avantage en matière de processus industriels, de commerce, de sécurité et même en termes militaires ; dans ces conditions, ralentir le marché (et l'innovation), même animé de la volonté de protéger les droits humains, c'est prendre le risque d'un retard sur les autres continents en concurrence. D'où la recherche par les régulateurs d'équilibres entre innovation et régulation au travers d'une approche par les risques afin de réserver les contraintes réglementaires à des situations strictement identifiées.

En juxtaposant ces deux projets, nous serions donc supposés révéler assez clairement les tensions entre un « Ancien monde » d'après-guerre où l'idée des fondateurs européens était de faire primer des valeurs universelles sur tout autre projet politique et un « Nouveau monde » post-moderne, cédant à la tentation du pragmatisme pour accommoder les droits fondamentaux avec divers autres impératifs, notamment d'ordre économique, politique et sécuritaire. Cette grille d'analyse, qui pourrait être considérée à juste titre comme simpliste, ignore évidemment les turbulences rencontrées dès le début de la construction européenne, et il serait faux d'idéaliser la primauté politique des droits humains après-guerre : des équilibres ont aussi été recherchés par les rédacteurs de la CvEDH[8]. Mais, nous verrons aussi que cette grille de lecture pour chercher à distinguer les stratégies de deux grands régulateurs européens en matière numérique n'est pas tout à fait pertinente, ceux-ci se rejoignant dans un discours qui pourrait être qualifié de « realpolitik », du fait même qu'ils partagent leurs États membres.

Nous examinerons donc successivement le projet de règlement de l'intelligence artificielle de la Commission européenne (1) et le projet de convention du Conseil de l'Europe, dans ses deux versions (2) : au premier l'ambition de fixer des règles à même de réguler le marché, de donner un avantage aux entrepreneurs européens et de ne pas entraver la lutte contre diverses formes de criminalité. Au second de traduire les principes de la CvEDH dans le contexte de la conception, du développement et de l'application de l'intelligence artificielle, tout en n'ignorant pas les conséquences pratiques de l'édiction de principes dont les effets seraient négatifs sur l'innovation et sur d'autres impératifs imposés par les politiques publiques de ses États membres.

[8] L'article 8 relatif à la protection de la vie privée modère ce droit si une ingérence prévue par la loi est nécessaire « à la sécurité nationale, à la sûreté publique, au bien-être économique du pays, à la défense de l'ordre et à la prévention des infractions pénales, à la protection de la santé ou de la morale, ou à la protection des droits et libertés d'autrui. » La jurisprudence de la Cour a toutefois révélé une interprétation relativement restreinte de ces notions.

CC BY NC ND

1. Proposition de règlement sur l'IA de la Commission européenne : organiser le marché

1.1. Contexte stratégique de l'encadrement du numérique au sein de l'Union européenne

Il doit tout d'abord être rappelé que la proposition de règlement sur l'IA (RIA) n'est qu'un élément d'une stratégie globale de régulation du numérique de l'Union européenne, présentée le 9 mars 2021 sous la forme d'ambitieux objectifs à 10 ans. Le RGPD avait déjà posé une première brique d'un programme ambitieux de réglementation de l'économie numérique. L'Union européenne établit maintenant un nouvel ensemble de réglementations, notamment un règlement sur les marchés numériques (DMA, Digital Market Act – publié au journal officiel de l'Union européenne le 13 octobre 2022), un règlement sur les services numériques (DSA, Digital Service Act – approuvé par le Conseil de l'Union le 4 octobre 2022), un règlement sur la gouvernance des données (DGA, Data Governance Act – adopté en mai 2022 et applicable en 2023), une directive sur le données ouvertes (Open Data Directive – adoptée le 20 juin 2019) et, nous l'avons vu, une nouvelle réglementation sur l'intelligence artificielle (RIA – présenté par la Commission le 21 avril 2022). Ajoutons-y la proposition de directive présentée le 28 septembre 2022 sur la responsabilité non-contractuelle applicable aux systèmes d'intelligence artificielle et une révision conjointe de la directive sur la responsabilité du fait des produits défectueux. Ajoutons également la « Déclaration européenne sur les droits et principes numériques pour la décennie numérique » adoptée en décembre 2022 par le Parlement européen, le Conseil de l'Union européenne et la Commission européenne[9], document juridiquement non contraignant, énumérant une série de principes ayant vocation à irriguer la production juridique de l'Union dans le domaine numérique.

Contrairement à une idée reçue, ces textes ne viennent pas combler un *no man's land* juridique relatif à « l'IA », mais renforcent bien au contraire sur une structure déjà bien établie de réglementations dans l'espace de l'Union. En adoptant de nouveaux textes contraignants, la Commission a conscience qu'elle touche à la chaîne de production de valeur des industries et des entreprises, mais elle en attend des résultats positifs en matière de concurrence, notamment en alignant les opérateurs américains et asiatiques (plus avancés technologiquement) sur les mêmes règles que les entreprises européennes du secteur et en rassurant les consommateurs d'employer des biens et services numériques en invoquant les droits fondamentaux.

1.2. Historique de l'encadrement de l'intelligence artificielle par la Commission européenne

La stratégie sur l'intelligence artificielle de l'Union européenne s'est progressivement précisée depuis 2018, en pleine expansion de l'éthique de « l'IA ». Concrètement, l'Institution a commencé de communiquer sur ses réflexions relatives à la place de l'intelligence artificielle dans le cadre du marché unique numérique en publiant en avril 2018 une communication intitulée « L'intelligence artificielle pour l'Europe ». Elle y expose son ambition de devenir un acteur mondial de premier

[9] (2023/C 23/01), publiée au Journal officiel de l'Union européenne le 23 janvier 2023, accessible sur : https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32023C0123(01)&from=EN, consulté le 10 mars 2023

CC BY NC ND

plan en matière d'innovation dans l'économie fondée sur les données et dans ses applications, au vu de l'importance stratégique de « l'IA » tant sur les plans juridiques, géopolitiques, commerciaux et sécuritaires. Cette approche européenne de « l'IA », qui inclut notamment la nécessité de garantir un cadre éthique et juridique approprié fondé sur les valeurs de l'Union et conforme à la Charte des droits fondamentaux, n'ouvrait pas alors la porte à la création d'une nouvelle législation contraignante spécifique à cette technologie. La communication notait en effet que l'Union disposait d'un cadre réglementaire existant sur lequel elle pouvait s'appuyer, par exemple en ce qui concerne la protection des données personnelles, la non-discrimination, la sécurité des produits et la responsabilité du fait des produits.

C'est dans ce contexte que la Commission a constitué dès juin 2018 un groupe indépendant d'experts de haut niveau (GEHN, AI HLEG en anglais pour *Artificial Intelligence High Level Expert Group*) composé d'une cinquantaine personnalités représentatives des différentes parties prenantes (secteur privé, public et société civile) afin d'établir des lignes directrices pour parvenir à une « IA de confiance » (*Trustworthy AI*). Celles-ci ont été publiées, à l'issue d'intenses discussions et de la mobilisation d'une vaste communauté de « l'Alliance IA», en avril 2019 en proposant 7 « exigences clés » : facteur humain et contrôle humain ; robustesse et sécurité ; respect de la vie privée et gouvernance des données ; transparence ; diversité, non-discrimination et équité ; bien-être sociétal et environnemental ; responsabilisation[10]. La Commission a exprimé son soutien à ce travail indépendant par la communication « *Building Trust in Human-Centric Artificial Intelligence* » d'avril 2019[11]. Le groupe a également publié en juin 2019 des recommandations en matière de politique et d'investissement pour une IA digne de confiance, ainsi qu'un document relatif à l'opérationnalisation de leurs lignes directrices contenant une liste d'évaluation pour une IA digne de confiance (*Assessment List for Trustworthy AI* - ALTAI) en juillet 2020.

Précisions que, dans le même temps, en décembre 2018 et en avril 2021, la Commission a présenté une succession de deux plans coordonnés sur « l'IA » avec les États membres pour aligner les politiques de manière à favoriser le développement et l'utilisation de « l'IA » en Europe. Les aspects éthiques et réglementaires de « l'IA » sont également discutés depuis 2018 au sein du groupe « DEI and AI » (*Diversity, Equity and Inclusion*), composé de représentants des États membres.

Mais la réelle impulsion de la proposition de RIA est à trouver dans les orientations politiques de la présidente von der Leyen en juillet 2019, où elle annonçait une approche européenne coordonnée des implications humaines et éthiques de « l'IA », confiant au passage le pilotage de ces travaux principalement au Commissaire Thierry Breton et à la DG CNECT (responsable de la création d'un marché unique numérique - alors que le RGPD avait été piloté par la DG JUST, en charge des politiques de l'Union en matière de justice, de droits des consommateurs et d'égalité femmes-hommes).

C'est ainsi qu'en février 2020, la Commission européenne a publié dans un livre blanc intitulé « Intelligence artificielle - Une approche européenne axée sur l'excellence et la confiance » les premières orientations relatives à des options politiques pour un cadre réglementaire et une

[10] Lignes directrices accessibles sur : https://op.europa.eu/fr/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1, consulté le 23 octobre 2022
[11] Non traduite en français, accessible sur : https://digital-strategy.ec.europa.eu/fr/node/1951, consulté le 15 octobre 2022

approche en matière d'investissement. Ce livre blanc reprend les acquis des lignes directrices pour introduire l'idée d'une approche par les risques et d'une labellisation des systèmes d'IA. Ce document est sous-tendu par deux préoccupations précises : d'une part, prévenir les atteintes aux droits fondamentaux du fait de l'utilisation de ces nouvelles technologies et, d'autres part, les atteintes sur la sécurité et le bon fonctionnement du régime de responsabilité. Il doit être précisé que l'appréhension de la notion de risque dans ce livre blanc a été considérée comme assez rudimentaire et a provoqué de nombreux commentaires dans le cadre de la consultation publique, qui s'est déroulée entre février et mai 2020. La Commission entendait alors restreindre les nouvelles mesures aux seules applications « à haut risque », ce haut risque étant déterminé en fonction de deux critères cumulatifs : d'une part, en fonction du secteur (par ex. « les soins de santé, les transports, l'énergie et certains pans du secteur public ») et, d'autre part, en fonction de la finalité de l'application elle-même (excluant par exemple un système de planification des rendez-vous dans un hôpital). Par ailleurs, la Commission préconise l'adoption d'une définition « suffisamment souple pour tenir compte des progrès techniques tout en étant suffisamment précise pour garantir la sécurité juridique nécessaire » et rappelle que tout nouvel apport réglementaire doit venir adapter et renforcer les règles en matière de protection des consommateurs et de protection des données à caractère personnel.

1.3. La proposition de règlement sur l'intelligence artificielle : un texte encore en construction en début 2023

Héritière des lignes directrices et nourrie des discussions du livre blanc sur « l'IA », la proposition de RIA, publiée le 21 avril 2021, a déjà été abondamment commentée, analysée et critiquée. Il en sera rappelé les traits essentiels, étant précisé que les co-législateurs (Conseil de l'Union européenne et Parlement européen) sont déjà intervenus de manière assez substantielle sur divers éléments de la proposition au moment de l'écriture de ces lignes.

Nous disposons en effet, du côté du Conseil de l'Union, de révisions portées dans un texte de compromis écrit sous les présidences successives slovène, française et tchèque, avec l'ambition de réviser l'ensemble des articles. Sous présidence slovène, le Conseil a ainsi révisé les articles 1 à 7 (incluant les annexes I et III) sur la définition de l'IA et les applications à haut risque. Sous présidence française, les révisions ont concerné l'ensemble des autres dispositions (articles 8 à 85 et annexes IV à IX), conduisant ainsi à la publication du premier texte de compromis le 17 juin 2022. La présidence tchèque a réouvert les débats sur la définition de « l'IA », la classification des systèmes d'IA comme étant à haut risque, la gouvernance et les mesures d'exécution du texte, ainsi que l'exclusion de la sécurité nationale. Une nouvelle version du texte de compromis a été publiée le 23 septembre 2022, avec une nouvelle révision des articles 1 à 29 (et des annexes I à IV). Le texte d'orientation générale définitif, adopté le 6 décembre 2022 par le Conseil[12], supprime l'annexe I définitionnelle et restreint la définition de l'IA dans l'article 3. L'article 3 ajoute également une définition des systèmes d'IA à usage général et un critère supplémentaire pour définir un système

[12] Communiqué de presse du Conseil de l'Union européenne, Législation sur l'intelligence artificielle: le Conseil appelle à promouvoir une IA sûre et respectueuse des droits fondamentaux, accessible sur : https://www.consilium.europa.eu/fr/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/, consulté le 30 décembre 2022

comme étant à haut risque (le système d'IA ne devra pas avoir qu'une fonction purement accessoire dans le processus de prise de décision).

Du côté du Parlement européen, les deux co-rapporteurs de la Commission IMCO (Commission du marché intérieur et de la protection des consommateurs) et LIBE (Commission des libertés civiles, de la justice et des affaires intérieures) ont présenté leur projet de rapport conjoint le 20 avril 2022. La phase de négociation politique et de rédaction du compromis a débuté le 1er juin 2022 à l'expiration du délai pour recevoir les amendements. Pour complexifier le paysage, les avis de 5 autres commissions seront également rendus nécessaires, sur la base des article 56 ou 57 du règlement intérieur du Parlement européen : les commissions JURI (Commission des affaires juridiques), ITRE (Commission de l'industrie, de la recherche et de l'énergie), CULT (Commission de la culture et de l'éducation), ENVI (Commission de l'environnement, de la santé publique et de la sécurité alimentaire) et TRAN (Commission des transports et du tourisme). Un 10e lot d'amendements a été diffusé mi-décembre 2022 avec des modifications assez substantielles, comme l'exclusion des systèmes d'IA à usage général de la liste des systèmes à haut risque[13]. Les rapporteurs ont également fait circuler en mi-février 2023 un programme pour des rencontres avec de nouvelles propositions de compromis, comme celles d'utiliser la définition de « l'IA » du NIST (National Institute of Standards and Technology) américain, de redéfinir le champ d'application, d'enrichir les pratiques prohibées et des catégories d'applications à considérer comme étant à haut risque[14]. Le vote en plénière est envisagé en début 2023, du fait de longues luttes de compétences qui ont eu lieu en 2021 ainsi que la difficile coopération entre les deux co-rapporteurs.

Les efforts de réécriture du Conseil de l'Union et du Parlement européen ne sont toutefois qu'une étape du long processus législatif, puisque la rédaction finale du texte sera assurée, sur la base des textes de compromis et avec l'assistance de la Commission, dans le cadre de négociations interinstitutionnelles (un « trilogue »).

1.4. Contenu de la proposition de règlement sur l'intelligence artificielle de la Commission européenne

La lecture des différentes versions des textes de compromis et des amendements révèle une grande instabilité des dispositions de la proposition de RIA, notamment en ce qui concerne la définition de « l'IA » et les dispositions relatives à l'emploi de « l'IA » à des fins de sécurité intérieure. Il sera donc pris le parti de s'en tenir à la rédaction initiale du texte proposé par la Commission pour en extraire les lignes essentielles[15] : une définition extensive et précise de ce qu'il faut entendre par « IA », une approche par les risques pour ne faire porter les effets juridiques les plus contraignants sur les applications présentant le plus d'enjeux et un dispositif de certification et de labellisation.

[13] L. Bertuzzi, Leading MEPs exclude general-purpose AI from high-risk categories – for now, Leading MEPs exclude general-purpose AI from high-risk categories – for now, Euractiv, 12 décembre 2022, accessible sur : https://www.euractiv.com/section/artificial-intelligence/news/leading-meps-exclude-general-purpose-ai-from-high-risk-categories-for-now/, consulté le 30 décembre 2022

[14] L. Bertuzzi, AI Act: All the open political questions in the European Parliament, Euractiv, 13 février 2023, accessible sur : https://www.euractiv.com/section/artificial-intelligence/news/ai-act-all-the-open-political-questions-in-the-european-parliament/, consulté le 10 mars 2023

[15] Il sera fait référence dans cette étude à la version anglaise de la proposition et de ses annexes, disponibles sur : https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence

CC BY NC ND

En transition avec l'analyse du projet de Convention du Conseil de l'Europe, nous situerons la place donnée aux droits fondamentaux dans la proposition de RIA.

À titre liminaire il doit être relevé qu'après des décennies d'une extrême prudence à réguler les développements numériques, hormis sous l'angle de la protection des données ou de la lutte contre la cybercriminalité, cette réglementation constitue un tournant stratégique majeur. L'Europe compte ainsi pleinement s'imposer en gardien des valeurs dans le nouvel environnement numérique et – surtout – jouer pleinement de son expérience de régulateur pour aligner les opérations de ses concurrents directs (américains et asiatiques) sur son rythme.

Une définition extensive de « l'IA » - Dans son texte, la Commission commence par trancher la question de l'objet à appréhender sous le terme vague « d'IA ». La Commission a décidé d'employer la notion de « système d'IA », qu'elle a entendu définir clairement pour concilier sécurité juridique et flexibilité pour l'avenir (point 6 du préambule, p.18).

En l'état de la rédaction du texte présenté par la Commission le 21 avril 2021, l'article 3.1 définit donc un système d'IA comme « un logiciel développé à l'aide d'une ou plusieurs des techniques et approches énumérées à l'annexe I et capable, pour un ensemble donné d'objectifs définis par l'homme, de générer des résultats tels que du contenu, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels ils interagissent ». Même si la Commission entend rester le plus neutre technologiquement à ce stade du texte (section 3.1 p.8 et 5.2 p.12), l'article 3 renvoie tout de même à une annexe I technique, qui couvre de manière très large diverses techniques algorithmiques, tant « symboliques » que « connexionnistes[16] ». Cette définition se rapproche volontairement de celle posée par l'OCDE, dans sa Recommandation du Conseil sur l'intelligence artificielle publiée en 2019 :

> **Système d'IA** : Un système d'intelligence artificielle (ou système d'IA) est un système automatisé qui, pour un ensemble donné d'objectifs définis par l'homme, est en mesure d'établir des prévisions, de formuler des recommandations, ou de prendre des décisions influant sur des environnements réels ou virtuels. Les systèmes d'IA sont conçus pour fonctionner à des degrés d'autonomie divers.

La Commission entend ainsi couvrir de manière assez large tout type de système autonome, en gardant une maîtrise précise des objets régulés en employant une technique d'énumération[17]. Cette volonté de contrôle contextualisé de la définition permet ainsi de circonscrire avec précision les objets de la réglementation, en donnant le moins d'adhérence possible aux débats extérieurs sur la notion « d'IA ».

[16] Systèmes énumérés dans l'annexe I : « (a) Approches d'apprentissage automatique, y compris l'apprentissage supervisé, non supervisé et par renforcement, utilisant une grande variété de méthodes, y compris l'apprentissage profond ; (b) Approches fondées sur la logique et la connaissance, y compris la représentation des connaissances, la programmation inductive (logique), les bases de connaissances, les moteurs d'inférence et déductifs, le raisonnement (symbolique) et les systèmes experts ; (c) les approches statistiques, l'estimation bayésienne, les méthodes de recherche et d'optimisation.

[17] B. Benbouzid, N. Smuha, Y.Meneceur, Quatre nuances de régulation de l'intelligence artificielle: Une cartographie des conflits de définition. Réseaux, n°232-233, pp. 29-64

CC BY NC ND

Cette définition est toutefois âprement discutée tant au Conseil de l'Union qu'au Parlement européen, avec des stratégies variables, dont certains tendent à réduire le champ aux seuls algorithmes d'apprentissage automatique[18]. Les débats autour de la notion des systèmes d'IA à usage général, ainsi que la propre difficulté à les définir[19] et à les inclure éventuellement dans la catégorie des systèmes à haut risque, ont encore ajouté à la complexité des discussions.

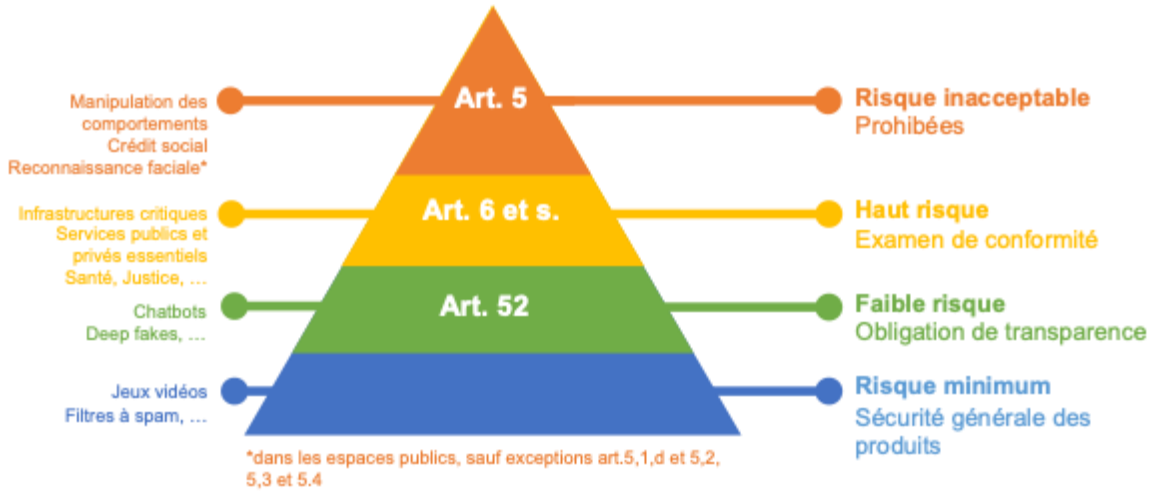
Une approche par les risques – Le cœur du mécanisme du RIA est une approche graduée fondée sur les risques, afin de faire porter aux applications présentant les risques les plus importants des dispositions plus contraignantes, allant du bannissement à un examen de conformité. Les applications à faible risque ou à risque minimum se voient octroyer peu ou pas d'obligations en application de ce texte. Il doit être précisé que le RIA ne se substitue pas aux autres mécanismes juridiques existants et aux obligations spécifiques en découlant (RGPD, responsabilité du fait des produits défectueux et autres mécanismes de responsabilité – dont un projet de directive spécifique aux systèmes d'IA publié en septembre 2022). Si des systèmes d'IA, même à faibles risques, provoquent des dommages, les victimes pourront donc bénéficier de voies de recours sur leur fondement, *ex post*. Les fournisseurs, quant à eux, sont tenus d'appliquer toutes les règlementations, y compris pour les applications à risque minimum.

Le mécanisme du RIA fondé sur les risques a été présenté par la Commission sous la forme d'une pyramide avec, à son sommet, les applications bannies. Ainsi, plus précisément, la proposition distingue 4 types d'applications, avec des contraintes réglementaires d'une intensité décroissante : 1) les **applications prohibées** du fait de leur nature (influence sur les comportements avec des conséquences physiques ou psychologiques, crédit social et reconnaissance faciale dans l'espace public – avec de très larges exceptions pour un usage d'enquête judiciaire), 2) les **applications à haut risque** devant répondre à des exigences clés et à un mécanisme d'évaluation de la conformité préalable à la mise sur le marché, 3) les **applications à risque limité**, auxquels sera imposé une obligation spécifique de transparence (comme les chatbots ou les deepfakes, afin de savoir que l'on a affaire à une machine), et 4) celles enfin présentant un **risque minimum**.

[18] L. Bertuzzi, EU Council presidency pitches significant changes to AI Act proposal, Euractiv, 1er décembre 2021, accessible sur : https://www.euractiv.com/section/digital/news/eu-council-presidency-pitches-significant-changes-to-ai-act-proposal/, consulté le 16 octobre 2022

[19] C.I. Gutierrez, Carlos Ignacio A. Aguirre, R. Uuk, C. Boine, M. Franklin, A Proposal for a Definition of General Purpose Artificial Intelligence Systems, SSRN, 5 octobre 2022, accessible sur : https://ssrn.com/abstract=4238951, consulté le 27 décembre 2022

CC BY NC ND



Source : Y. Meneceur, d'après European Commission, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence

Il doit être relevé que ce mécanisme a été fortement inspiré des travaux de la Daten Ethik kommission allemande qui, dans une opinion d'octobre 2019 avait présenté une proposition énumérant 5 niveaux de risques[20].

L'examen de la conformité des systèmes d'IA à haut risque - Le point le plus substantiel de la proposition est le mécanisme d'examen de la conformité des systèmes d'IA représentant le plus de risques pour les individus.

Nous ne développerons ici l'analyse des articles relatif aux systèmes créant des risques inacceptables (définis dans l'art.5), regroupant notamment des systèmes ayant des risques physiques ou psychologiques sur les individus ou encore les systèmes dits de crédit social. La présidence slovène du Conseil a, par exemple, proposé d'étendre l'interdiction du système de crédit social des autorités publiques aux entités privées. En outre, l'interdiction de l'utilisation de notation a également été étendue pour inclure l'exploitation d'une « situation sociale ou économique », conduisant à de profondes répercussions sur le secteur financier, car, par exemple, les taux d'intérêt des prêts sont actuellement calculés en fonction de la probabilité de remboursement. La reconnaissance faciale dans les espaces publics est aussi prohibée par principe, sauf exception pour des enquêtes judiciaires (cf.*infra*).

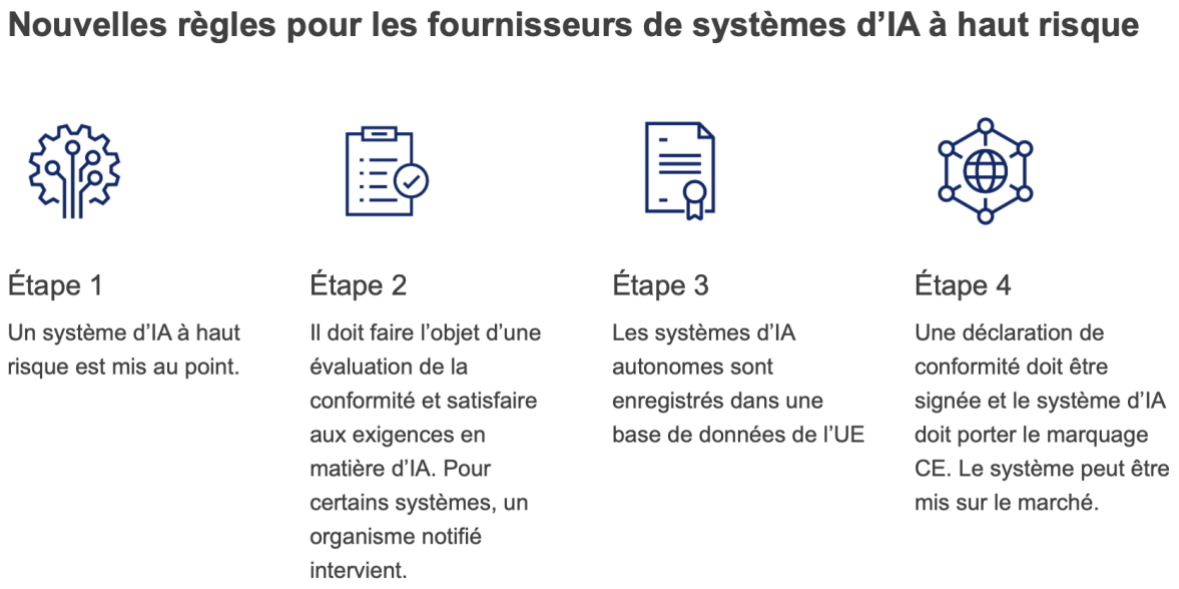
Les systèmes à haut risque, quant à eux, sont définis par plusieurs critères, plus précis que ceux présentés en février 2020 dans le livre blanc de la Commission. La proposition identifie tout d'abord les systèmes d'IA étant une composante de produits faisant déjà l'objet d'une certification tierce-partie (p.13 et art.6.1) des systèmes autonomes d'IA ayant des impacts prévisibles sur les droits fondamentaux (p.13 et art.6.2). Une annexe vient compléter l'article 6.2, en listant 8 domaines sensibles : identification biométrique des individus, gestion des infrastructures critiques, éducation, recrutement et gestion des carrières, divers services publics et privés (dont services d'urgences,

[20] Daten Ethik Kommission, Opinion of the Data Ethics Commission, octobre 2019, p.19, accessible sur : https://www.bmj.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&v=1, consulté le 16 octobre 2022

attribution de crédits, aides sociales), maintien de l'ordre et enquêtes pénales, migration et contrôle des frontières, justice (annexe III).

La Commission s'autorise à mettre à jour l'annexe listant les applications sensibles, dès lors que des systèmes a) peuvent créer des dommages sur la santé ou la sécurité des individus, ou b) sont susceptibles de provoquer des atteintes à leurs droits fondamentaux dans les domaines précédemment spécifiés (art. 7).

Très concrètement, pour être mis sur le marché ces systèmes devront répondre à des exigences clés, principalement issues des lignes directrices du groupe d'experts indépendants de haut niveau mandaté par la Commission, vérifiés lors d'une procédure d'évaluation de conformité ex-ante sous la responsabilité des fournisseurs. Pour ces systèmes à haut risque, la Commission impose par ailleurs un système de gestion des risques à mettre en œuvre tout au long du cycle de vie des applications concernées (art. 9) ainsi qu'une méthodologie de gouvernance des données, afin de garantir un niveau élevé de qualité (art. 10).



En cas de changements substantiels au cours du cycle de vie du système d'IA, retour à l'étape 2.

Source : European Commission, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence

Ces systèmes se voient aussi imposer une documentation spécifique (ce qui est un point faible de la plupart des développements informatiques – art. 11 et annexe IV) et un enregistrement d'événements à même d'assurer la traçabilité des opérations (art. 12). La transparence de ces systèmes et l'information aux usagers est garantie par la fourniture obligatoire d'un certain nombre d'informations (art. 13). La supervision humaine doit être garantie en tenant notamment compte de biais cognitifs connus (biais d'automatisation) conduisant des professionnels à s'appuyer sur des systèmes automatisés sans avoir recours à leur expertise (art. 14, 4, b). Enfin la précision, la robustesse et la sécurité de ces systèmes doivent être particulièrement garanties (art. 15). On retrouve là traduit les 7 exigences des lignes directrices en matière d'éthique pour une IA digne de

confiance du groupe d'experts indépendants de haut niveau, auxquels sont ajoutées un certain nombre d'obligations aux fournisseurs de ces systèmes à haut risque et autres parties (Chapitre 3, art. 16 à 29). Les États membres sont, par ailleurs, invités à désigner une autorité notifiante comme responsable du suivi des procédures relatives aux systèmes à haut risque et un organisme notifié (Chapitre 4, art. 30 à 39), tout à fait classique des mécanismes de certification déjà en œuvre. Un label « CE » sera délivré aux systèmes conformes (art. 49).

Les procédures d'évaluation de la conformité à suivre pour chaque type de système d'IA à haut risque tentent de minimiser la charge pour les opérateurs économiques, ainsi que pour les organismes notifiés (Chapitre 5, art. 40 à 51). Les systèmes d'IA destinés à être utilisés comme composants de sécurité de produits réglementés par la législation relative au nouveau cadre législatif (par exemple, machines, jouets, dispositifs médicaux, etc.) seront soumis aux mêmes mécanismes de conformité et d'application ex ante et ex post que les produits dont ils sont un composant, en y ajoutant les exigences établies par le nouveau règlement. Le fournisseur devra suivre soit la procédure d'évaluation de la conformité sur la base du contrôle interne visée à l'annexe VI, soit la procédure d'évaluation de la conformité fondée sur l'évaluation du système de gestion de la qualité et l'évaluation de la documentation technique, avec l'intervention d'un organisme notifié, visée à l'annexe VII (art.43). L'initiative de la mise en conformité repose, en toute hypothèse, sur le fournisseur. Une base de données enregistrera les systèmes autonomes d'IA à haut risque (art. 60).

Enfin, et cela intéressera particulièrement les micros, petites et moyennes entreprises, des expérimentations pourront être menées dans un cadre juridique adapté (boîtes à sable réglementaires), même dans des domaines sensibles et avec des données à caractère personnel (art. 53 et s.). La mise en œuvre de codes de conduite sectoriels est également encouragée (art. 69).

L'exception sécuritaire conduisant à un déploiement de la reconnaissance faciale dans l'espace public - Les risques de surveillance de masse de la population avec les technologies numériques sont régulièrement soulevés par la société civile et les défenseurs des droits individuels. Avec la reconnaissance faciale et sa mise en œuvre effective dans l'espace public, dans des cadres juridiques relativement flous, ces inquiétudes se sont renforcées substantiellement, et même soutenues par des fournisseurs privés de ces solutions. La Commission fait ici le choix de bannir de l'espace public ces dispositifs d'identification en temps réel des individus, avec une très large série d'exceptions au bénéfice des procédures judiciaires.

Ainsi, « l'identification biométrique à distance en temps réel » est autorisée a) pour la recherche d'enfants disparus, b) en cas de menace imminente pour la sécurité des personnes ou d'attaques terroristes, c) la détection, la localisation, l'identification ou la poursuite d'un suspect recherché pour la commission d'une infraction pénale[21] punie dans l'État membre concerné d'une peine d'emprisonnement d'au moins trois ans (art. 5, 1, d).

La Commission souligne que l'utilisation de ces systèmes à des fins répressives doit également tenir compte des éléments suivants de la nature de la situation donnant lieu à l'utilisation éventuelle, et

[21] La liste des infractions pénales visées par l'article 2, paragraphe 2, de la décision-cadre 2002/584/JAI du Conseil est très large, allant de faits criminels graves à des faits délictuels comme l'escroquerie ou l'aide au séjour irrégulier.

CC BY NC ND

des conséquences de l'utilisation du système sur les droits et libertés de toutes les personnes concernées (art. 5.2), en y ajoutant des limites temporelles et géographiques. Comme pour les contrôles d'identité déjà encadrés par les lois processuelles en matière pénale, une autorisation préalable devra être accordée par une autorité judiciaire ou par une autorité administrative indépendante, sur demande motivée des services de police. Toutefois, dans une situation d'urgence dûment justifiée, l'utilisation du système peut être commencée sans autorisation et l'autorisation ne peut être demandée que pendant ou après l'utilisation (art. 5.3). Ces dispositions devront être transcrites dans une loi spécifique, aménageant recours et dispositifs de supervision (art.5.4).

Bien que la reconnaissance faciale se trouve par principe bannie de l'espace public, les exceptions aménagées conduisent – de fait à une autorisation d'une extension généralisée de cette technologie. Les impératifs de sécurité, notamment le risque terroriste, ont vraisemblablement motivé la Commission à aménager l'exploitation de cette technologie dans un strict cadre judiciaire (et non de simple prévention d'infractions de manière vague). Mais l'équipement sera réalisé dans les espaces publics et l'efficacité du contrôle judiciaire sera intimement liée à la réelle indépendance des autorités en charge de l'enquête, variable selon les États membres, tout comme la capacité de se libérer suffisamment de temps pour exercer un réel contrôle.

Sanctions et suivi – Des sanctions financières allant jusqu'à 6% du chiffre d'affaires annuel pourront être prononcées en cas de manquement d'un opérateur privé (art. 71 de la proposition, qui fixe le montant d'amendes entre 2, 4 ou 6% du chiffre d'affaires annuel selon les hypothèses).

Par ailleurs, le texte constitue un comité européen sur l'intelligence artificielle, composé de représentants des 27 États membres et du contrôleur européen de la protection des données, sera également créé (art. 56 et suivants).

Ce comité, présidé par la Commission, sera chargé de plusieurs missions : a) contribuer à la coopération efficace des autorités nationales de surveillance ; b) coordonner et contribuer aux orientations et analyses de la Commission, des autorités nationales de surveillance et des autres autorités compétentes sur les questions émergentes dans le marché intérieur ; c) aider les autorités nationales de surveillance et la Commission à assurer l'application du règlement. Le comité contribuera de plus à partager les meilleures pratiques entre les États membres, à l'uniformisation des pratiques administratives (dont le fonctionnement des « bacs à sable réglementaires ») et à émettre avis ou recommandations sur les spécifications techniques des normes (art. 57).

La création d'un tel organe de suivi, avec une mission si large, pose encore à ce stade bien des questions. Tout d'abord, son poids sur la régulation du marché pourrait être considérable et, en cas de divergence de vues probables entre ses participants, l'on peut s'interroger sur la paralysie possible du dispositif. Ensuite, la technicité très probable de ce comité ne se trouve pas directement balancée par une représentation démocratique ou de la société civile, même si des observateurs (sans qu'ils ne soient énumérés) pourraient être admis. Enfin la coordination avec les autorités de protection des données, même si elle est prévue *via* le contrôleur européen, reste à parfaire : en première analyse, se pose une difficulté liée à la nature de ces autorités, indépendantes afin d'assurer effectivement leurs missions (quelle sera leur réelle influence dans un comité

intergouvernemental ?), et à une grande dispersion des efforts (n'aurait-il pas été plus simple d'étendre les missions et les moyens des autorités de protection des données ?).

Une proposition de règlement cherchant à concilier protection des individus, sécurité juridique et innovation – Malgré de forts vents contraires[22], la Commission engage donc l'UE sur un terrain redouté par nombre d'entrepreneurs, notamment les micros, petites et moyennes entreprises, pour faire peser sur leur activité de nouvelles contraintes que certains ne manqueront pas de considérer comme bureaucratiques. La solution imaginée par la Commission consiste en une approche « robuste et flexible » fondée sur les risques (p.3 de la proposition), anticipée en Allemagne par la Commission éthique des données, afin de ne pas faire peser les mêmes contraintes sur toutes les applications de l'IA. Approche, dévoilée dès le livre blanc de février 2020, qui n'avait toutefois pas manqué de susciter des réactions extrêmement mitigées, voire hostiles, en se demandant comment définir une application « à haut risque » des autres et – surtout – en évoquant un réel affaiblissement d'une approche de réglementation fondée sur des droits invocables pour tout emploi de cette technologie[23]. Car c'est bien là la difficulté majeure d'un cadre juridique intervenant pour réguler l'application d'une technologie aussi transversale que l'IA : ouvrir des brèches dans la logique et l'enchevêtrement des mécanismes juridiques déjà existants, complexes et « interreliés ».

Cette proposition ouvre également une autre brèche en ce qui concerne des applications sensibles de « l'IA », notamment à des fins d'enquête judiciaire. L'emploi de technologies comme la reconnaissance faciale dans l'espace public, pourtant régulièrement dénoncé même par des opérateurs privés comme Microsoft[24], trouve ainsi un fondement légal. Les exceptions pourraient d'ailleurs apparaître si larges pour certains observateurs que le principe annoncé de sa prohibition (art.5) pourrait venir à en perdre une partie de son sens. Par ailleurs, les risques les plus substantiels, comme les biais pouvant renforcer ou créer des discriminations dans des algorithmes de services publics par exemple, sont traitées par des procédures de conformité préalables à même de garantir un certain nombre d'exigences. Ainsi énoncé, le principe paraît tout à fait satisfaisant, mais les métriques des standards matérialisant ces exigences et l'autorité en charge de les définir[25] demeurent encore trop vagues pour pouvoir en conclure une pleine efficacité.

Même si l'adoption d'un texte définitif pourrait encore prendre au moins 1 à 2 années au moment de l'écriture de cette étude (il aura fallu près de 4 années au RGPD pour voir le jour), l'on peut déjà constater en l'état que la Commission n'a pu échapper à une certaine complexité pour la mise en conformité des applications présentant le plus de risque de dommage sur les individus. La

[22] Voir la note blanche de 14 Etats membres de l'Union européenne « Innovative and Trustworthy AI: Two sides of the same coin », accessible sur : https://em.dk/media/13914/non-paper-innovative-and-trustworthy-ai-two-side-of-the-same-coin.pdf – Consulté le 15 octobre 2022

[23] F. Hidvegi, D. Leufer, E. Massé, The EU should regulate AI on the basis of rights, not risks, Access Now, 17 février 2021 – https://www.accessnow.org/eu-regulation-ai-risk-based-approach/ – Consulté le 15 octobre 2022

[24] Il ne sera pas discuté ici des motivations distinctes de la société civile et d'entreprises comme Microsoft sur la régulation de la reconnaissance faciale – Voir M. Tual, Intelligence artificielle : « Microsoft ne veut pas fournir d'outils qui pourraient violer les droits de l'homme », Le Monde, 3 juillet 2018, accessible sur : https://www.lemonde.fr/pixels/article/2018/07/03/eric-horvitz-microsoft-ne-veut-pas-fournir-d-outils-qui-pourraient-violer-les-droits-de-l-homme_5324975_4408996.html, Consulté le 16 octobre 2022

[25] V. à ce sujet L. Bertuzzi, Commission leaves European standardisation body out of AI standard-setting, Euractiv, 7 décembre 2022, accessible sur : https://www.euractiv.com/section/digital/news/commission-leaves-european-standardisation-body-out-of-ai-standard-setting/, consulté le 3 janvier 2023

classification des systèmes va certainement ouvrir à des interprétations et des débats, des opérateurs pouvant être tentés d'éviter la régulation contraignante pour des applications à la limite du haut risque (est-ce qu'un potentiel système d'optimisation de gestion des enrôlements dans les tribunaux représente un « haut risque » par exemple ?). Le prononcé de sanctions au titre des éventuels manquement devra également prendre en compte la concurrence possible entre différents ordres de juridictions (pénales et administratives), problématique déjà connue entre les autorités de protection des données et les juridictions pénales notamment. Même si ce projet de régulation extrêmement ambitieux cherche à concilier de manière assez efficace protection des individus, sécurité juridique et innovation, il positionne surtout opportunément les institutions de l'UE en véritables métronomes des développements de « l'IA » dans ses États membres (section 5.2.6 p.15 et art.56).

1.5. Quelle est la place pour les droits fondamentaux dans la proposition de règlement sur l'intelligence artificielle ?

Des droits fondamentaux omniprésents dans la proposition de règlement – Les droits fondamentaux (pour employer la terminologie de l'Union européenne en référence à sa Charte des droits fondamentaux) sont loin d'être absents des préoccupations du législateur européen dans la proposition de RIA publiée le 21 avril 2022. Le terme y est employé à 79 reprises, parfois en association avec la nécessité de protéger la santé et la sécurité des personnes ou avec le concept de « valeurs de l'UE », pour parvenir à créer un cadre juridique pour une « IA digne de confiance ».

Plus précisément, les nouvelles dispositions du règlement s'intéressent principalement à la protection des données (le texte vise expressément l'article 16 du TFUE), la non-discrimination et l'égalité entre les femmes et les hommes. La protection de ces droits ne s'avère toutefois pas absolue puisque le texte s'inscrit dans une mécanique d'ensemble cherchant à établir des équilibres entre différents ordres d'impératifs. Le texte rappelle bien que ses objectifs spécifiques sont les suivants (extrait de l'exposé des motifs du RIA, 1.1) :

- Veiller à ce que les systèmes d'IA mis sur le marché de l'Union et utilisés soient sûrs et respectent la législation en vigueur en matière de droits fondamentaux et les valeurs de l'Union ;
- Garantir la sécurité juridique pour faciliter les investissements et l'innovation dans le domaine de l'IA ;
- Renforcer la gouvernance et l'application effective de la législation existante en matière de droits fondamentaux et des exigences de sécurité applicables aux systèmes d'IA ;
- Faciliter le développement d'un marché unique pour des applications d'IA légales, sûres et dignes de confiance, et empêcher la fragmentation du marché.

CC BY NC ND

Mais les droits fondamentaux restent bien invoqués comme l'une des bases de cette proposition (section 1.1 p.1 et p.3 notamment, et 3.5 p.11), en plus de garantir un fonctionnement harmonieux du marché interne (section 2 p.6 et 7). C'est d'ailleurs sur la base des droits fondamentaux que sont prohibées certaines applications ayant de fortes conséquences sur les individus (art.5).

Des questions de principes restant à traiter – Si l'on peut souligner de la mention de certains problèmes sous-estimés (comme les biais d'automatisation – art. 14, 4, b), l'on voit bien que des questions de principe restent à approfondir, notamment celle de l'opportunité du recours à des algorithmes dans certaines matières comme la police ou la justice. Même si, dans ce domaine à « haut risque », une procédure d'examen de conformité s'appliquera, l'annexe III autorise – de fait – le principe d'un appui de systèmes algorithmiques pour la prise de décision judiciaire, le profilage des individus ou le contrôle aux frontières. A titre d'exemple, l'atténuation concrète des biais ou le degré d'explicabilité attendu au travers des mécanismes de mise en conformité reste à prouver, dans un contexte où la représentation de l'équité est bien subjective[26]. La conséquence de cette position sera de ne satisfaire, en définitive, personne : les défenseurs de droits jugeront comme une régression cette légitimation ; les legaltechs et autres opérateurs privés vont se voir imposer une très lourde mise en conformité sans nécessairement répondre au fond des critiques pouvant leur être adressées (comme le sens des prévisions des logiciels de jurimétrie[27]). Des applications avec de très forts enjeux et aux fondements scientifiques incertains risquent donc d'être « blanchies » par un dispositif de mise en conformité.

La prohibition du principe de la reconnaissance faciale dans l'espace public, avec de très larges exceptions, est également à relever[28]. Cette proposition de règlement autorise en réalité une très large diffusion de cette technologie et, comme pour la vidéosurveillance/vidéoprotection, va finir par équiper tous nos espaces publics « au cas où ». Le contrôle de l'autorité judiciaire pourrait être rassurant si celle-ci avait les moyens effectifs de cette nouvelle tâche, sans parler du caractère variable de l'indépendance de ces autorités selon les États qui la mettront en œuvre. La proposition constitue donc potentiellement un recul des droits individuels dans certains États, les exigences sécuritaires de notre société risquant de déconstruire nombre de mécanismes pourtant chèrement acquis.

Parmi les applications prohibées, l'on peut aussi noter un recul de la Commission entre le *leak* d'une version préliminaire, fuité dans les médias, et le document final : alors que le texte préliminaire mentionnait de manière large les algorithmes pouvant avoir un effet sur les individus et leur manière de se comporter, sur leur manière de former une opinion ou de prendre une décision à leur détriment, le texte final restreint la prohibition aux seuls systèmes pouvant causer des dommages physiques ou psychologiques.

[26] J. Angwin, J. Larson, Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say, Ethics of Data and Analytics, Auerbach editions, 2022

[27] Y. Meneceur, Quel avenir pour la « justice prédictive » ?, La Semaine Juridique édition générale n°7, 12 février 2018

[28] Voir le commentaire au Wall Street Journal de Sarah Chander, conseillère politique principale à European Digital Rights : « La liste des exemptions est incroyablement large », cela "va en quelque sorte à l'encontre du but de prétendre que quelque chose est une interdiction" – S. Schechner, P. Olson, Artificial Intelligence, Facial Recognition Face Curbs in New EU Proposal, Wall Street Journal, 21 avril 2021

CC BY NC ND

Des mécanismes de responsabilité réactualisés dans deux projets distincts de directives – Il doit également être relevé l'absence de mécanisme clair facilitant le recours des individus contre des prises de décision algorithmiques (ou assistées par des algorithmes) au sein de la proposition de RIA. Toutefois, la publication le 28 septembre 2022 d'une directive sur la responsabilité en matière d'intelligence artificielle apporte un certain nombre de réponses en matière civile[29]. Ces nouvelles règles se présentent comme permettant « aux victimes de dommages causés par des technologies de l'IA d'accéder à une réparation de la même manière que si elles avaient été lésées dans d'autres circonstances ». Cette directive introduit deux mesures principales : la « présomption de causalité », qui dispensera les victimes de l'obligation d'expliquer en détail comment le dommage a été causé par une faute ou une omission spécifique et l'accès aux éléments de preuve détenus par les entreprises ou les fournisseurs, lorsque ces derniers utilisent de « l'IA » à haut risque.

Précisons aussi que la directe relative à la responsabilité du fait des produits défectueux a également fait l'objet d'un projet de mise à jour dans le même temps, et s'appliquera aux actions intentées par des particuliers contre le fabricant pour les dommages causés par des produits défectueux, aux pertes matérielles dues à la mort, aux dommages corporels ou matériels et aux pertes de données[30].

Voilà de quoi tenter de compenser l'asymétrie entre les parties, avec, d'une part, un régime de responsabilité civile pour faute allégeant la charge de preuve du consommateur et introduisant des outils supplémentaires (droit à la divulgation d'éléments de preuve, présomptions réfragables) en réparation d'un vaste panel de dommages (y compris les dommages causés par un acte de discrimination ou par une violation de droits fondamentaux tels que le droit à la vie privée) et un régime de responsabilité civile stricte en cas de pertes matérielles dues à la mort, aux dommages corporels ou matériels et aux pertes de données.

Même si l'examen approfondi de ces deux textes excède le champ de cette étude, il doit toutefois être relevé que certains dommages, faisant aujourd'hui l'objet à la fois de sanctions pénales au vu de l'importance du trouble causé à la société et de sanctions civiles (discrimination, violations de la vie privée) devraient pouvoir continuer à faire l'objet d'incriminations. Il restera donc à examiner si la portée du premier projet de directive viendra effectivement renforcer la protection des individus en contaminant le droit pénal en matière d'administration de la preuve ou si, au contraire, le maintien de différents régimes risque de fragmenter la réponse judiciaire.

Un grand absent : la protection de la société des impacts négatifs de la transformation numérique – Peut-être de manière plus substantielle, la dimension collective de la transformation numérique de notre société est peu développée dans ce texte. Il faut dire qu'il s'insère dans une stratégie très offensive de la Commission de développement des technologies numériques et que les notions de frugalité dans le domaine (frugalité des données, de la consommation électrique et énergétique) paraissent pour certains analystes encore trop souvent instrumentalisées dans un certain nombre de discours publics, sans effet concret à court terme. La prohibition de systèmes de crédit social répond bien à une préoccupation collective, mais la régulation de la propagande à

[29] Directive accessible sur : https://ec.europa.eu/info/files/proposal-directive-adapting-non-contractual-civil-liability-rules-artificial-intelligence_fr, consulté le 23 octobre 2022

[30] Directive accessible sur : https://single-market-economy.ec.europa.eu/document/3193da9a-cecb-44ad-9a9c-7b6b23220bcd_en, consulté le 23 octobre 2022

l'ère des réseaux sociaux, le renforcement de la primauté du droit vis-à-vis du développement de nouvelles formes de gouvernementalité (*data-driven law* par exemple[31]) ou encore le développement d'un principe de solidarité auraient pu nourrir cette proposition.

Enfin, si les États membres de diverses organisations internationales ou régionales (Nations Unies et ses agences dont l'UNESCO, OCDE, Conseil de l'Europe, …) ont régulièrement appelé à une coordination entre ces institutions pour bâtir un cadre global de régulation du numérique, la proposition de la Commission aurait pu mieux s'ancrer à ces acquis. Même si le texte s'affirme « largement conformes à d'autres recommandations et principes internationaux, ce qui garantit que le cadre IA proposé est compatible avec ceux adoptés par les partenaires commerciaux internationaux de l'UE » (p.13) et que la définition de la proposition initiale d'avril 2021 est largement inspirée de celle de l'OCDE, rien ne démontre que cette conformité sera atteinte à l'issue du long processus législatif. S'agissant d'un texte de la Commission, l'on comprend que les références soient réalisées dans l'écosystème de l'UE. Mais n'oublions pas que les négociations pour l'adhésion de l'UE à la Convention européenne des droits de l'homme ont repris de manière favorable : cette possible – et souhaitable – adhésion entraînera certainement des conséquences sur les dispositions ici proposées et il aurait été rassurant de les voir anticipées.

De manière générale, et comme le démontre les intenses discussions tant au Conseil de l'Union qu'au Parlement européen, il est à craindre que cette proposition n'attire en définitive que des critiques de tous les bords[32], oscillant parfois entre le trop ou le trop peu de précisions. Le pire des scénarios consisterait à adopter un cadre juridique qui, à la fois, légitimerait des applications tout à fait critiquables d'un point de vue des droits fondamentaux et complexifierait encore l'entrée sur le marché de jeunes entreprises, déboussolés par le nouveau maquis juridique applicable.

[31] M. Hildebrant, Data-Driven Prediction of Judgment. Law's New Mode of Existence?, SSRN, 30 mars 2020 – https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3548504 – Consulté le 16 octobre 2022

[32] W. Knight, Europe's Proposed Limits on AI Would Have Global Consequences, Wired, 21 avril 2021 – https://www.wired.com/story/europes-proposed-limits-ai-global-consequences/ – Consulté le 16 octobre 2022

CC BY NC ND

2. Le projet de convention du Conseil de l'Europe sur l'intelligence artificielle : protéger les droits de l'homme, la démocratie et l'État de droit

2.1. Rappel du rôle et du mandat du Conseil de l'Europe

Pour mémoire, le Conseil de l'Europe est une organisation internationale unique en son genre tout à fait distincte de l'Union européenne, pour avoir été créé bien avant elle dès 1949, avec un mandat portant sur les droits de l'Homme et disposant aujourd'hui d'acquis considérables. Par son statut, l'Organisation se consacre à l'établissement de normes juridiques contraignantes et non contraignantes pour réaliser une plus grande unité entre ses États membres et encourager leur progrès économique et social[33]. En tant que plus ancienne organisation intergouvernementale paneuropéenne, le Conseil de l'Europe rassemble 46 États membres[34] autour d'une idée : les peuples d'Europe ne peuvent construire un avenir commun que sur la base de ce qui les unit plutôt que de ce qui les divise. L'Organisation s'est progressivement construite autour de trois piliers - les droits de l'Homme, la démocratie et l'État de droit - et défend des valeurs universelles à travers de près de 225 conventions, dont la Convention européenne de sauvegarde des droits de l'Homme et des libertés fondamentales (plus connue sous le nom simplifié de Convention européenne des droits de l'Homme).

Parallèlement, le Conseil de l'Europe a également mis en place un ensemble sans précédent de mécanismes de mise en œuvre, de suivi et d'assistance à ses États membres dans des domaines aussi variés que le fonctionnement des institutions démocratiques, l'efficacité et l'indépendance des systèmes judiciaires, les droits sociaux, la culture, l'éducation, la jeunesse et le sport, afin de devenir au fil du temps la véritable conscience démocratique de l'Europe, laissant le développement du marché et l'intégration économique à l'Union européenne. Il convient de souligner que nombre des conventions du Conseil de l'Europe ne sont pas seulement ouvertes à ses 46 États membres, puisque des États non-membres peuvent également les négocier, y adhérer et devenir des parties à part entière.

Son action aujourd'hui peut être résumée autour d'un « triangle dynamique » liant création de standards, suivi des États membres et coopération.

2.2. Expérience du Conseil de l'Europe dans l'encadrement de technologies numériques

Les technologies développées par l'homme ont invariablement un impact sur son environnement et sa condition. Le Conseil de l'Europe est donc intervenu très tôt après sa création pour veiller à ce que le développement des innovations les plus diverses respecte les valeurs fondamentales. Au début des années 1960 par exemple, en pleine guerre froide, le Conseil de l'Europe a hérité de certaines des missions de l'Union de l'Europe occidentale (UEO) liées à la normalisation

[33] Statut du Conseil de l'Europe, art.1.a) : Le but du Conseil de l'Europe est de réaliser une union plus étroite entre ses Membres afin de sauvegarder et de promouvoir les idéaux et les principes qui sont leur patrimoine commun et de favoriser leur progrès économique et social.

[34] Le Comité des Ministres du Conseil de l'Europe a décidé de l'expulsion de la Fédération de Russie de l'Organisation le 16 mars 2022.

CC BY NC ND

pharmaceutique dans une perspective de protection civile. C'est ainsi que pour assurer la sécurité et la qualité des médicaments, le concept de Pharmacopée européenne a émergé en 1964, lorsque huit pays pionniers, tous membres du Conseil de l'Europe, se sont engagés à signer la nouvelle Convention internationale relative à l'élaboration d'une Pharmacopée européenne[35], soulignant leur volonté de doter l'Europe d'un outil de référence unique pour la protection de la santé publique sur leurs territoires respectifs et de faciliter l'échange international de médicaments, essentiels en cas de catastrophe ou de besoin urgent. Grâce à cet effort commun, coordonné avec l'Union européenne, nous disposons aujourd'hui d'une véritable Europe de la santé, coordonnée entre Bruxelles et Strasbourg, qui assure non seulement la protection des populations mais aussi la sécurité juridique d'innovations industrielles hautement techniques par le biais de la certification de médicaments. La Convention d'Oviedo sur la biomédecine de 1997[36] et son protocole de 1998 sur l'interdiction du clonage humain[37], ainsi que la Convention MEDICRIME[38] de 2011 sur la lutte contre la contrefaçon de médicaments, sont des instruments juridiques pionniers élaborés par le Conseil de l'Europe pour répondre à ces mêmes objectifs.

La première contribution du Conseil de l'Europe pour réguler le développement numérique remonte à 1981 avec la « Convention 108[39] », modernisée en 2018 par un protocole d'amendement (« Convention 108+[40] »). Cette Convention a établi les toutes premières normes mondiales garantissant le droit à la vie privée et la protection des données personnelles. Sans être affaiblie par les très importantes évolutions technologiques intervenues depuis son adoption, ce texte encadre toujours avec la même efficacité le traitement informatique des données sensibles (par exemple les opinions politiques ou la vie sexuelle) et crée un droit pour chacun de connaître les données existantes le concernant, avec un droit de rectification. Il instaure également une libre circulation des données entre les parties qui ont ratifié la Convention. Ces mécanismes, largement inspirés des législations nationales de la Suède (1973), de l'Allemagne (1976) et de la France (1978), et qui sont aujourd'hui développés dans le Règlement général sur la protection des données (RGPD) de l'Union européenne, ont été modernisés en 2018 pour s'adapter aux réalités du monde en ligne et à certaines pratiques. De nouveaux principes ont été reconnus, tels que la transparence, la proportionnalité, la responsabilité, la limitation des données et la protection de la vie privée dès la conception. La Convention 108 a été ratifiée par les 46 États membres du Conseil de l'Europe et les 9 États non-membres (Argentine, Cap-Vert, Fédération de Russie, Maurice, Mexique, Maroc, Sénégal, Tunisie et Uruguay). 43 États membres et non membres ont déjà signé le protocole d'amendement, et 20 États ont le statut d'observateur.

La deuxième contribution majeure du Conseil de l'Europe à la réglementation des technologies numériques est la Convention de Budapest sur la cybercriminalité, ouverte à la signature en 2001[41]. Il s'agit de la première et unique norme juridique mondiale traitant des infractions spécifiques commises contre et au moyen de systèmes informatiques. Elle encadre les pouvoirs procéduraux avec des garanties pour enquêter sur la cybercriminalité et collecter des preuves électroniques en

[35] ETS n°50 - https://rm.coe.int/168090293b
[36] ETS n°164 - https://rm.coe.int/168007cf98
[37] Deux ans seulement après le clonage de la brebis Dolly, l'ETS n°168 - https://rm.coe.int/168007f2ca
[38] ETS n°211 - https://rm.coe.int/168008482f
[39] ETS n°108 - https://rm.coe.int/1680078b37
[40] STCE n°223 - https://rm.coe.int/16808ac918
[41] ETS n°185 - https://rm.coe.int/1680081561

relation avec tout crime et édicte une coopération internationale sur la cybercriminalité et l'échange de preuves électroniques. Un protocole spécifique sur la xénophobie et le racisme via les systèmes informatiques a été adopté en 2003[42], ainsi qu'un autre protocole additionnel sur la coopération renforcée en matière de cybercriminalité et de preuves électroniques en 2021[43]. Il y a actuellement 66 Parties à la Convention et 11 États observateurs. En 2020, le Conseil de l'Europe a aidé quelque 130 pays dans leurs efforts de renforcement des capacités dans ce domaine.

2.3. Des travaux sur l'encadrement des algorithmes, du Big Data et de l'intelligence artificielle antérieurs à ceux de la Commission européenne

Pour commencer par le niveau politique, l'Assemblée parlementaire du Conseil de l'Europe (APCE) a publié un rapport et adopté une Recommandation sur l'intelligence artificielle en 2017, soulignant la nécessité de renforcer la réglementation existante. Le rapport intitulé « Les droits de l'Homme à l'ère des robots, les défis découlant de l'utilisation de la robotique, de l'intelligence artificielle et de la réalité virtuelle et augmentée » a été rédigé avec le Rathenau Instituut[44], tandis que la Recommandation sur la convergence technologique, l'intelligence artificielle et les droits de l'Homme[45] est fondée sur un rapport rédigé par Jean-Yves Le Déaut, membre de l'Assemblée nationale française[46].

Un Comité d'experts sur les intermédiaires de l'internet (MSI-NET) a été créé avec un mandat de deux ans (2016-2017) dans le cadre du Comité directeur sur les médias et la société de l'information (CDMSI). Il a produit une étude intitulée « Algorithmes et droits de l'Homme »[47] sur « les dimensions des droits de l'Homme des techniques de traitement automatisé des données et les implications réglementaires possibles », qui est l'un des premiers documents du Conseil de l'Europe à traiter de manière exhaustive des conséquences de la généralisation de l'utilisation des algorithmes. Le texte a posé un certain nombre de principes et d'actions (transparence, responsabilité, cadres éthiques et meilleure évaluation des risques) au regard des impacts possibles des algorithmes sur les droits de l'Homme.

Le Comité consultatif de la Convention 108 a publié des Lignes directrices sur la protection des personnes à l'égard du traitement des données dans un monde de Big Data[48] en 2017, suivies d'un rapport[49] et de nouvelles lignes directrices[50] sur l'IA et la protection des données en 2019.

---

[42] ETS n°189 - https://rm.coe.int/168008160f
[43] CM(2021)57-final - https://search.coe.int/cm/pages/result_details.aspx?objectid=0900001680a48e4c
[44] R. Van Est, J.B.A. Gerritsen, avec l'assistance de L. Kool, Human rights in the robot age : Défis découlant de l'utilisation de la robotique, de l'intelligence artificielle et de la réalité virtuelle et augmentée - Rapport d'expert rédigé pour la Commission de la culture, de la science, de l'éducation et des médias de l'Assemblée parlementaire du Conseil de l'Europe (APCE), La Haye : Rathenau Instituut 2017 - https://www.rathenau.nl/sites/default/files/2018-02/Human%20Rights%20in%20the%20Robot%20Age-Rathenau%20Instituut-2017.pdf
[45] Recommandation 2102(2017) - http://semantic-pace.net/tools/pdf.aspx?doc=aHR0cDovL2Fzc2VtYmx5LmNvZS5pbnQvbncveG1sL1hSZWYvWDJILURXLWV4dHIuYXNwP2ZpbGVpZD0yMzcyNiZsYW5nPUVO&xsl=aHR0cDovL2Fzc2VtYmx5LmNvZS5pbnQvbncveG1sL3hzbC1mby9QZGYvWFJlZi1XRC1BVC1YTUwyUERGLnhzbA==&xsltparams=ZmlsZWlkPTIzNzI2
[46] Doc. 14288 - https://pace.coe.int/pdf/3dec81f354f90fad246e485a8d362e0a3a81f2d71e0f61289f8bae65f24109fd/doc.%2014288.pdf
[47] DGI(2017)12 - https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10
[48] T-PD(2017)1 - https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a
[49] T-PD(2018)09Rev - https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6
[50] T-PD(2019)01 - https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8

En 2018, la Commission européenne pour l'efficacité de la justice (CEPEJ) a adopté la Charte éthique européenne sur l'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement, accompagnée d'une étude approfondie sur l'utilisation de l'IA dans les systèmes judiciaires[51]. La Charte énonce cinq principes fondamentaux qui devraient régir le développement d'outils utilisant l'intelligence artificielle dans l'intérêt d'une bonne administration de la justice : le respect des droits fondamentaux, la non-discrimination, la garantie de la qualité et de la sécurité, la garantie de la transparence, de l'impartialité et de l'équité, et enfin la garantie du contrôle de l'utilisateur, afin que celui-ci soit un acteur informé et maître de ses choix.

Tous ces travaux ont progressivement convergé pour illustrer le besoin accru de régulation, tant à l'intérieur qu'à l'extérieur des nouveaux environnements numériques, bien au-delà de la question de la protection des données personnelles ou de la lutte contre la cybercriminalité. De manière constante, ces travaux reconnaissent la potentialité de ce nouveau contexte socio-technique pour améliorer la vie quotidienne, mais aussi pour porter des atteintes sérieuses aux droits de l'Homme, à la démocratie et à l'État de droit.

C'est pourquoi l'ambition du Conseil de l'Europe d'explorer les questions liées à « l'IA » est progressivement devenue une priorité politique, afin de soutenir les États membres dans leur transformation numérique, mais également les aider à prévenir les risques réels ou potentiels pour les individus et l'ensemble de la société.

2.4. Processus politique pour l'élaboration de standards juridiques sur l'intelligence artificielle

C'est dans le cadre du « processus de réforme d'Elseneur » lancé le 18 mai 2018 lors de la 128ème session du Comité des Ministres sous présidence danoise que le Secrétaire Général du Conseil de l'Europe présentait son cinquième rapport annuel sur la « Situation de la démocratie, des droits de l'Homme et de l'État de droit[52] », dans lequel il identifiait pour la première fois parmi les nouveaux défis identifiés « les progrès rapides des technologies, y compris l'intelligence artificielle ».

Cette première impulsion politique allait aboutir dès le 13 février 2019, avec l'adoption par le Comité des Ministres d'une Déclaration portant spécifiquement sur les capacités de manipulation des processus algorithmiques[53]. De manière inédite, l'organe exécutif d'une organisation intergouvernementale de premier plan reconnaissait, sur la base des travaux d'experts académiques, l'exploitation par les concepteurs de processus algorithmiques de biais cognitifs pour influencer les individus, notamment par des techniques de micro-ciblage. Le texte encourageait ainsi les États membres du Conseil de l'Europe à s'engager sur diverses mesures, notamment « sur la nécessité d'évaluer de manière critique le besoin de mesures réglementaires ou autres plus fortes pour assurer une surveillance adéquate et démocratiquement légitimée de la conception, du développement, du

[51] CEPEJ(2018)14 - https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c
[52] Rapport accessible sur : https://edoc.coe.int/en/an-overview/7584-state-of-democracy-human-rights-and-the-rule-of-law-5th-report-role-of-institutions-threats-to-institutions.html
[53] Decl(13/02/2019)1 - https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b

déploiement et de l'utilisation des outils algorithmiques, en vue de garantir une protection efficace contre les pratiques déloyales ou les abus de position de force sur le marché ».

Dans ce contexte, une conférence de haut niveau intitulée « *Governing the Game Changer - Impacts of artificial intelligence development on human rights, democracy and the rule of law* » se tenait une dizaine de jours plus tard à Helsinki et marquait le réel point de départ du processus politique spécifiquement dédié à la question de l'IA au sein du Conseil de l'Europe. Les conclusions de la conférence[54], éclairées par un panel d'experts pluridisciplinaires, plaçaient le développement de cette technologie dans une perspective à la fois critique et constructive, traduisant les longues discussions entre les ministères finlandais de la justice, de l'économie et des affaires étrangères pour parvenir un texte de compromis. L'originalité de ce texte a été de signifier tout à fait clairement que la gouvernance de l'IA ne pouvait raisonnablement pas se concentrer sur la seule éthique, mais aussi explorer la possibilité de développer de nouvelles normes juridiques fondées sur les droits de l'Homme, la démocratie et l'État de droit.

Quelques mois plus tard, dans une Recommandation publiée en mai 2019 intitulée « *Unboxing AI : 10 steps to protect human rights* », la Commissaire aux droits de l'Homme soutenait également la mise en œuvre de mesures visant à assurer et à superviser la conformité des algorithmes aux droits de l'Homme, notamment par l'adoption d'études d'impact spécifiques allant plus loin que celles déjà existantes sur la vie privée, sans toutefois soutenir explicitement la création d'un nouveau cadre juridique contraignant[55].

Le 17 mai 2019 à Helsinki, à l'occasion de la 129ème session du Comité des Ministres, les ministres des affaires étrangères des États membres du Conseil de l'Europe convenait à la quasi-unanimité d'examiner la possibilité de mettre en place un cadre juridique sur le développement, la conception et l'application de l'intelligence artificielle, fondé sur les normes du Conseil de l'Europe en matière de droits de l'Homme, de démocratie et d'État de droit et sur des consultations multipartites.

2.5. Le Comité ad hoc sur l'intelligence artificielle (CAHAI) et le Comité sur l'intelligence artificielle (CAI)

Lancement du CAHAI – C'est ainsi que le mandat d'un Comité ad hoc sur l'IA (CAHAI) était adopté le 11 septembre 2019, sous présidence française du Comité des Ministres, pour une durée de 2 ans. Ce Comité recevait un soutien tout à fait explicite lors d'une réunion des ministres de la justice du Conseil de l'Europe en octobre 2019 d'étudier l'opportunité d'un nouvel instrument juridique[56]. L'ensemble des organisations intergouvernementales travaillant sur la régulation de l'IA ont été conviés dès le début de ces travaux en tant qu'observateurs (Commission européenne et Agence européenne des droits fondamentaux, OCDE et UNESCO notamment).

[54] https://rm.coe.int/conclusions-from-the-conference/168093368c

[55] Recommandation accessible sur : https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64

[56] Voy. la déclaration finale de la présidence française : https://rm.coe.int/final-declaration-of-the-french-presidency-conference-of-mj-coe-15-oct/168098383f.

La première réunion plénière du CAHAI s'est tenue du 18 au 20 novembre 2019, avec un premier échange de vues sur la manière de relever les défis découlant du développement, de la conception et de l'application de l'IA. Malgré la pandémie de COVID-19, le CAHAI réussira à se réunir virtuellement et proposait en septembre 2020 de créer 3 groupes de travail dans son premier rapport d'étape au Comité des Ministres[57].

Il doit être relevé qu'en parallèle de ce Comité ad hoc, le Comité des Ministres adoptait en avril 2020 une nouvelle Recommandation sur les impacts des systèmes algorithmiques sur les droits de l'Homme[58], avec des lignes directrices sur la prise en compte de ces impacts à l'intention des acteurs des secteurs public et privé. Ce document, préparé le Comité d'experts sur les dimensions des droits de l'Homme du traitement automatisé des données et des différentes formes d'intelligence artificielle (MSI-AUT), ainsi que l'étude « Responsabilité et IA » de 2019 sur les implications des technologies numériques avancées (y compris les systèmes d'IA) pour le concept de responsabilité dans un cadre de droits de l'Homme[59], ont bien entendu alimenté les travaux du CAHAI. L'on pourra naturellement se questionner sur la production de ces documents en parallèle par un Comité d'experts sans lien avec le Comité intergouvernemental dédié. L'explication au fait que le MSI-AUT avait été placé sous l'autorité du CDMSI bien avant la création du CAHAI et qu'il a semblé opportun de les laisser conclure leurs travaux, pour alimenter les réflexions du Comité ad hoc.

Travaux du CAHAI – Entre septembre et novembre 2020, le CAHAI a bénéficié d'un soutien politique extrêmement substantiel, avec la publication par l'APCE de sept rapports portant sur différents thèmes, tous convergeant vers la nécessité d'un cadre juridique pour régir le développement de l'IA. Le rapport « Nécessité d'une gouvernance démocratique de l'IA » insiste particulièrement sur la l'adoption d'un cadre juridiquement contraignant de type conventionnel[60]. En novembre 2020, le cadre stratégique quadriennal du Conseil de l'Europe[61], rédigé en pleine concertation avec les États membres et les différents Comités, a établi une priorité n°9 sur « l'IA et l'impact plus large de la transformation numérique et de l'utilisation des nouvelles technologies ».

En décembre 2020, et sans retard malgré la crise sanitaire, le CAHAI rendait son étude de faisabilité[62], soumise au Comité des Ministres. Le comité propose dans ce document d'élaborer « un cadre juridique approprié » qui consisterait en une « combinaison d'instruments juridiques contraignants et non contraignants, qui se complètent ». Plus précisément, l'étude mentionne « qu'un instrument contraignant, une convention ou une convention-cadre, de caractère horizontal, pourrait consolider les principes généraux communs - contextualisés pour s'appliquer à l'environnement de l'IA et utilisant une approche basée sur le risque - et inclure des dispositions plus granulaires en accord avec les droits, principes et obligations identifiés dans cette étude de faisabilité ». Cette étude identifie également des « droits substantiels clés » et des « obligations clés » allant de la nécessité de protéger la dignité humaine et de prévenir les atteintes aux droits de l'Homme, à la démocratie et à l'État de droit, à la nécessité de garantir l'autonomie humaine - par

[57] Le groupe chargé de l'élaboration des politiques (PDG), le groupe chargé de la consultation et de la sensibilisation (COG) et le groupe chargé du cadre juridique (LFG). Voy. le rapport d'avancement : https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=090000016809fa65b
[58] CM/Rec(2020)1 - https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154
[59] DG1(2019)05 - https://rm.coe.int/responsability-and-ai-en/168097d9c5
[60] Doc.15150 - https://pace.coe.int/en/files/28742/html
[61] SG/Inf(2020)34 - https://rm.coe.int/strategic-framework-of-the-council-of-europe/1680a07810 - *op.cit.*
[62] CAHAI(2020)23 - https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da

exemple par la supervision humaine, la possibilité d'intervention et la prise de décision finale. Elle énonce, de plus, des règles de transparence reconnaissant la nécessité pour les individus de savoir quand et dans quelles conditions les systèmes d'IA sont utilisés et affectent leur vie et leurs droits. L'accent était aussi mis sur la nécessité de garantir l'explicabilité des décisions en matière d'IA, l'auditabilité des systèmes d'IA et l'obligation de rendre compte et la responsabilité des personnes impliquées dans le cycle de vie d'un système d'IA. Toujours selon cette étude de faisabilité, le futur cadre juridique devrait également être dynamique, applicable tout au long du cycle de vie d'un projet d'IA et adaptable aux technologies actuelles et futures. De manière concrète, des mécanismes concrets concernant l'opérationnalisation des droits et des principes sont mentionnés. Il s'agit par exemple d'évaluations de l'impact sur les droits de l'Homme afin d'identifier, de prévenir, d'atténuer et de rendre compte des impacts négatifs sur les droits de l'Homme découlant de ces activités[63], de la mise en place de bacs à sable réglementaires et de la surveillance automatisée du fonctionnement des systèmes d'IA.

Le CAHAI publiait également en décembre 2020 une étude intitulée « *Towards regulation of AI systems* »[64], compilant des contributions universitaires de haut niveau employées par les groupes de travail pour produire l'étude de faisabilité.

Résultats des travaux du CAHAI – La présidence allemande du Comité des Ministres du Conseil de l'Europe entre novembre 2020 et mai 2021 soutiendra le CAHAI dans l'élaboration d'un cadre juridique contraignant sur l'IA en vue de prévenir les violations des droits de l'Homme, conformément à ses engagements au sein des institutions de l'Union européenne de créer un cadre juridique – également contraignant – pour assurer la sécurité des produits de l'IA. Cette complémentarité entre le Conseil de l'Europe et l'Union européenne a été réitérée à l'occasion de la conférence « Les droits de l'Homme à l'ère de l'IA - L'Europe comme norme internationale en matière d'intelligence artificielle » qui a eu lieu en janvier 2021. Elle est explicitement mentionnée dans les conclusions[65] de la conférence, en témoignant d'un très fort soutien aux travaux du CAHAI sur la question spécifique de la protection des droits de l'Homme.

Le soutien politique global des États membres dans cette démarche de régulation se traduira aussi dans les décisions de la 131ème session du Comité des Ministres, tenue à Hambourg le 21 mai 2021. Les représentants des États membres donnaient alors instruction rendre prioritaire les travaux sur un éventuel cadre juridique, potentiellement sous la forme d'un instrument juridique transversal contraignant. Les représentants ajoutaient que ce texte pourrait inclure des principes généraux communs ainsi que des instruments additionnels contraignants ou non contraignants, en vue de lancer les négociations lors de la 132ème session ministérielle en mai 2022.

La déclaration finale de la Conférence ministérielle « Artificial intelligence – Intelligent Politics: Challenges and opportunities for media and democracy » qui s'est tenue à Nicosie en juin 2021 soulignait également la nécessité de développer des « approches collaboratives et/ou corégulatrices, y compris une réglementation internationale et nationale juridiquement contraignante le cas échéant », pour une protection efficace de la liberté d'expression en ce qui concerne la modération

[63] Cette proposition découle directement de la Recommandation du Commissaire aux droits de l'Homme
[64] Étude accessible sur : https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a
[65] Conclusions accessibles sur : https://germanycoe.de/storage/documents/Conference_Conclusions.pdf

des contenus, ainsi que pour traiter spécifiquement de la communication électorale en ligne, aux campagnes et à la couverture médiatique afin de créer les conditions d'un déroulement équitable des processus électoraux qui sont désormais influencés par les systèmes algorithmiques, notamment sur les plateformes de médias sociaux.

Création du CAI pour négocier un projet de traité international sur l'intelligence artificielle – Le mandat pour proposer un instrument juridique sur l'intelligence artificielle est confié depuis janvier 2022 à un tout nouveau CAI (Comité sur l'intelligence artificielle).

Le Comité a hérité du CAHAI un document adopté lors de sa dernière plénière de décembre 2021, intitulé « Éléments possibles d'un cadre juridique sur l'intelligence artificielle, fondé sur les normes du Conseil de l'Europe en matière de droits de l'Homme, de démocratie et d'État de droit[66] » qui constituera la base d'un tout premier projet de Convention. Ce document s'est appuyé sur une large consultation des milieux universitaires, de la société civile et du secteur privé, afin de garantir que les instruments, contraignants ou non, soient adaptés aux attentes de toutes les parties prenantes.

Ces éléments de cadre juridique reprennent l'idée d'établir des évaluations d'impact, des classifications de risques et des principes généraux pour le développement de l'IA. En l'état, le futur instrument juridique aurait vocation à s'appliquer à toutes les systèmes d'IA en se concentrant sur « les risques potentiels émanant du développement, de la conception et de l'application de systèmes d'IA à des fins de police, d'administration de la justice et d'administration publique ». L'utilisation de l'IA dans les soins de santé, l'éducation et l'octroi d'avantages sociaux, entre autres, devrait faire l'objet de règles sectorielles.

2.6. Le projet de Convention du Conseil de l'Europe sur l'intelligence artificielle : vers un traité international fondé sur les droits de l'Homme, la démocratie et l'État de droit – étude la première « version 0 »

Articulation du calendrier du projet avec la proposition de règlement de la Commission – Les 46 États membres du Conseil de l'Europe semblent déterminés pour proposer des solutions pour un cadre juridique adapté et proportionné en matière d'IA. Malgré l'orientation politique donnée à Hambourg par le Comité des Ministres en 2021 pour un instrument juridique contraignant, la forme de l'instrument juridique n'est pas encore totalement certaine au moment de l'écriture de cette étude, au vu des positions variables et encore incertaines pour des États comme la France et pour la Commission européenne. L'objectif partagé bien est d'établir des règles claires conciliant les droits de l'Homme et l'innovation, pour contribuer à établir la confiance, l'acceptation et la sécurité juridique des systèmes d'IA : mais la Commission européenne a sollicité en août 2022 un mandat exclusif de négociation dans la crainte d'une fragmentation des approches réglementaires et de risques de recouvrement[67]. Il semble toutefois que, comme pour le protocole à la Convention 108 rédigé au moment de l'adoption du RGPD, la Commission cherche

[66] CAHAI(2021)07 - https://rm.coe.int/cahai-2021-07-analysis-msc-23-06-21-2749-8656-4611-v-1/1680a2f228

[67] COM(2022)414 final, accessible sur : https://www.senat.fr/europe/textes_europeens/e17013.pdf, consulté le 23 octobre 2022 ; V. aussi L. Bertuzzi, EU Commission postponed AI treaty negotiations with further delays in sight, Euractiv, 5 octobre 2022, accessible sur : https://www.euractiv.com/section/digital/news/eu-commission-postponed-ai-treaty-negotiations-with-further-delays-in-sight/, consulté le 5 octobre 2022

CC BY NC ND

probablement à conclure d'abord ses propres travaux sur le projet de RIA. Même si ces textes s'avèrent en réalité, tant dans leurs fondements que dans leurs finalités, tout à fait distincts, ils paraissent partager bien plus de compatibilité de vues qu'une rapide première analyse laisserait supposer.

Présentation des deux projets de Convention – Nous commenterons successivement deux « version 0 » du projet de Convention présentées par le président du CAI. La première version n'a fait l'objet que d'une diffusion restreinte en juin 2022 auprès des États membres et de la Commission européenne alors que la seconde a été mise à disposition du public en février 2023[68].

Pour mémoire, le recueil des observations des États membres de la première « version 0 » avait été perturbé par le temps d'instruction de la demande de mandat exclusif de négociation de la Commission, qui avait conduit les 27 États membres de l'Union à réserver leurs réponses. Le calendrier initial du CAI, dont le mandat de 2 années devait conduire à conclure les négociations et produire le texte définitif, n'est pas encore totalement compromis mais semble devoir rester lié aux avancées du processus législatif de la proposition de règlement avec la place prépondérante désormais laissée à la Commission dans les négociations à Strasbourg.

La seconde « version 0 » présente des modifications extrêmement substantielles de structure, de contenu et de hiérarchisation des dispositions. Il semble toutefois opportun de lire successivement les deux versions de texte afin d'illustrer, par leurs différences, les évolutions et les arbitrages intervenus et, partant, l'évolution des priorités politiques relatives à l'encadrement de « l'IA ».

Présentation de la première « version 0 » – À titre liminaire, le texte diffusé en fin juin 2022 est un travail présenté, avec l'appui du Secrétariat, par le président du CAI, Thomas Schneider, ambassadeur et directeur des affaires internationales à l'Office fédéral de la communisation suisse (OFCOM), désigné comme représentant suisse au CAI. Élu l'assemblée plénière du Comité, il détient une longue expérience avec le Conseil de l'Europe, notamment pour avoir présidé le CDMSI (Comité Directeur sur les médias et la société de l'information) qui a supervisé la production de projets de recommandation sur les algorithmes.

Cette première version de 28 articles et une annexe se présente comme un texte de haut niveau, transversal, se concentrant sur la sphère de la protection et de la promotion des droits de l'homme, de la démocratie et de l'État de droit. Conforme au mandat du Conseil de l'Europe, le projet ne vise pas à encadrer les aspects économiques, commerciaux ou militaires des systèmes d'intelligence artificielle. Il ambitionne de fixer des principes fondamentaux et règles régissant la conception, le développement et l'application des systèmes d'intelligence artificielle. Le projet de rapport explicatif, joint au projet de Convention, précise que des normes plus détaillées et spécifiques relatives à la conception, au développement et à l'application des systèmes d'intelligence artificielle pourront « faire l'objet d'instruments sectoriels adéquats et, le cas échéant, de protocoles additionnels à la présente Convention ». Le texte se positionne donc potentiellement comme une convention cadre, dont l'objectif n'est pas de fixer à proprement parler des règles de fond, avec des effets juridiques directs, mais de constituer un cadre dans lequel celles-ci peuvent se développer.

[68] CAHAI(2023)01 - https://rm.coe.int/cai-2023-01-fr-projet-zero-revise-de-convention-cadre-public/1680aa1942

Principes établis par la « première version 0 » – Cette première version du projet de Convention établit des « principes et règles fondamentaux visant à garantir que la conception, le développement et l'application des systèmes d'intelligence artificielle sont pleinement compatibles avec le respect des droits de l'homme, le fonctionnement de la démocratie et le respect de l'État de droit » (art.1.1).

Comme pour la proposition de RIA, la définition vise « les systèmes d'intelligence artificielle », qualifiant un « système basé sur une machine qui est capable d'informer ou de **générer** de manière autonome **des résultats tels que des prédictions, des recommandations ou des décisions** en utilisant des données et des entrées basées sur la machine et/ou l'homme et en **influençant l'environnement avec lequel il interagit** » (art.2).

Des dénominateurs communs se retrouvent avec la définition de la proposition de RIA publiée le 21 avril 2022 :

> On entend par « système d'intelligence artificielle » (système d'IA), un logiciel qui est développé au moyen d'une ou plusieurs des techniques et approches énumérées à l'annexe I et qui peut, pour un ensemble donné d'objectifs définis par l'homme, **générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit**.

Ainsi qu'avec la Recommandation du Conseil sur l'intelligence artificielle de l'OCDE :

> Système d'IA : Un système d'intelligence artificielle (ou système d'IA) est un système automatisé qui, pour un ensemble donné d'objectifs définis par l'homme, est **en mesure d'établir des prévisions, de formuler des recommandations, ou de prendre des décisions influant sur des environnements réels ou virtuels**. Les systèmes d'IA sont conçus pour fonctionner à des degrés d'autonomie divers.

Le principal enjeu du déploiement des systèmes d'IA, aujourd'hui abondamment documenté, est relatif aux biais et aux discriminations produits par certaines applications de prise de décision, parfois de manière indirecte (c'est-à-dire en n'employant pas directement une donnée sensible[69]). Le projet de texte établit dès ses dispositions générales un principe de non-discrimination, en excluant pour la mise en œuvre de ses dispositions toute distinction fondée « sur le sexe, le genre, l'orientation sexuelle, la race, la couleur, la langue, l'âge, la religion, les opinions politiques ou toutes autres opinions, l'origine nationale ou sociale, l'appartenance à une minorité nationale, la fortune, la naissance, l'état de santé, le handicap ou tout autre situation, ou basée sur une combinaison d'une ou plusieurs de ces motifs » (art.3).

Les autres principes fondamentaux établis dans le chapitre II concernent le large spectre de la recherche, la conception, le développement et l'application des systèmes d'IA, en plaçant comme

[69] Il peut être fait référence ici à la notion de « variable proxy », connue des sciences sociales et statistiques, où une donnée, non significative en soi, est fortement corrélée à une autre variable. C'est ainsi qu'en en employant le code postal du lieu de résidence un individu, l'on peut indirectement traiter l'origine ethnique de l'individu dans des zones fortement ségrégées.

CC BY NC ND

socle « la nécessité de sauvegarder et de défendre la dignité humaine, les droits de l'homme et les libertés fondamentales » en plus « de préserver et d'encourager des institutions démocratiques solides et responsables et de sauvegarder l'état de droit comme base institutionnelle permettant d'assurer à la fois la participation démocratique et la protection effective des libertés et droits fondamentaux » (art.5).

Plus précisément, le texte vise des principes : a) de durabilité, solidarité et protection de l'environnement ; b) d'égalité ; c) de protection de la vie privée ; d) de responsabilité et d'obligation de rendre des comptes ; e) de transparence et d'auditabilité ; f) d'environnement réglementaire contrôlé en vue d'encourager la recherche et favoriser l'innovation ; g) de sûreté, sécurité, intégrité des données, cybersécurité et robustesse ; h) de consultation multipartite à la lumière des implications sociales, économiques, éthiques et juridiques ; i) de culture et compétences numériques (art 5.2 à 10).

Des exigences supplémentaires sont imposées aux applications dans le secteur public au vu des risques accrus sur les droits de l'homme, la démocratie et l'État de droit (art.6). Ainsi il est imposé à ces systèmes : a) une base légale adéquate et un « examen préliminaire minutieux de la nécessité et de la proportionnalité de l'utilisation d'un tel système soit effectué compte tenu du contexte du déploiement » ; b) au respect des normes énoncées dans la mesure où toute entité privée agissant pour leur compte est concernée : c) une formation appropriée des utilisateurs sur l'utilisation de tout système d'intelligence artificielle employé ainsi que sur les limites respectives de ces systèmes (art. 6.2 à 4).

Des garanties procédurales sont établies, en créant, le cas échéant, des registres (art.7.1), une obligation d'information de l'usage d'un système d'IA s'il intervient dans un processus de prise de décisions affectant les droits de l'homme et les libertés fondamentales (art.7.2), des obligations d'explication et de justification, dans un langage clair, compréhensible et cohérent, avec des informations relatives à la possibilité de contester une décision affectant les droits de l'homme, les droits et les intérêts juridiques (art.7.3) et un droit à être informé de l'interaction avec un système d'IA (art.7.4).

Des restrictions aux garanties procédurales de l'article 7, reprenant quasiment mot pour mot celles de l'article 8.2 de la Convention européenne des droits de l'Homme, sont énumérées dans l'article 8. Ainsi les États ayant ratifié la Convention seront autorisés à en prévoir par la voie de dispositions législatives lorsqu'elles semblent « nécessaires et proportionnées dans une société démocratique, dans l'intérêt de la sécurité nationale, de la sûreté publique ou du bien-être économique du pays, pour la défense de l'ordre et la prévention du crime, pour la protection de la santé ou de la morale, ou pour la protection des droits et libertés d'autrui ».

Les deux derniers articles précisent que la Convention ne saurait retreindre les droits de l'Homme et les libertés fondamentales déjà garantis (art.9) ni restreindre une protection plus étendue (art.10).

Mesures ex ante imposées aux systèmes d'intelligence artificielle – Le chapitre III du projet de Convention énumère les principes d'une évaluation des risques et de l'impact d'un système d'IA, qu'il soit mis en œuvre par des acteurs privés ou publics, « afin d'identifier, d'analyser et d'évaluer

CC BY NC ND

les risques et d'évaluer l'impact de l'application des systèmes d'intelligence artificielle en ce qui concerne la jouissance des droits de l'homme, le fonctionnement de la démocratie et le respect de l'État de droit » (art.11.1). Le projet de Convention diffusé en juin 2022 ne présentait pas encore de modèle précis de méthodologie dans son annexe, celui-ci ayant fait l'objet de discussions lors de la session plénière du CAI de septembre 2022.

Le texte reprend ensuite une approche par les risques en imposant aux systèmes évalués comme présentant des « risques significatifs » par la méthodologie décrite dans l'article 11 de prendre « toutes les mesures nécessaires de prévention et d'atténuation » (art.12.1). Les systèmes classés comme posant des niveaux de risque inacceptables pourront, conformément au droit national de l'État concerné, imposer un moratoire ou une interdiction totale ou partielle s'il n'existe pas d'autres mesures pour atténuer le risque (art.13.1).

Pour les systèmes posant des risques significatifs, les États ayant ratifié la Convention sont invités à : a) prévenir tout dommage « aux droits de l'hommes, libertés fondamentales, droits et intérêts juridiques des personnes physiques ou morales, aux institutions et processus démocratiques, à l'administration de la justice, à la santé publique et à l'environnement » ; b) préserver « la liberté individuelle, la dignité humaine et l'autonomie, et notamment la capacité de prendre des décisions informées, à l'abri de toute influence indue, de toute manipulation ou de tout effet préjudiciable portant atteinte au droit à la liberté d'expression et de réunion, à la participation démocratique et à l'exercice d'autres droits de l'homme et libertés fondamentales pertinents » ; c) garantir un accès égal et équitable au débat public et à des processus démocratiques inclusifs ; d) préserver le principe d'égalité devant la loi, y compris l'égalité des genres et les droits relatifs aux groupes discriminés et aux personnes en situation de vulnérabilité ; e) veiller à ce que les droits à la vie privée et à la protection des données à caractère personnel soient dûment respectés, notamment par des garanties supplémentaires pour les catégories particulières de données telles que les données médicales (art.15, a à e).

L'article 14 du projet de Convention prohibe les applications suivantes : a) utilisation de la biométrie par des autorités publiques pour « identifier, catégoriser ou déduire les émotions des individus » ; b) notation sociale afin d'accéder à des services essentiels conduisant soit à un traitement préjudiciable dans des contextes sociaux sans rapport avec les contextes dans lesquels les données ont été initialement collectées, soit à une traitement préjudiciable injustifié ou disproportionné par rapport à leur comportement social ; c) toute autre utilisation par les autorités publiques de systèmes d'IA à des fins n'étant pas compatibles avec les valeurs fondamentales des sociétés démocratiques, ne poursuivant pas un objectif légitime et n'étant pas nécessaires dans une société démocratique.

Les dispositions suivantes établissent un mécanisme de suivi et de coopération (art.16 à 18) et des clauses finales habituelles quant aux effets de la Convention, aux amendements, aux révisions, au règlement des différends, aux modalités de signature et d'entrée en vigueur, aux adhésions, à l'application territoriale, aux réserves, aux dénonciations et aux notifications (art.19 à 28).

CC BY NC ND

2.7. Analyse de la première « version 0 » du projet de Convention du Conseil de l'Europe sur l'intelligence artificielle : un texte pragmatique de haut niveau

Le texte présenté par le Conseil de l'Europe aux représentants de ses États membres en juin 2022 est loin d'être définitif, de nombreuses interrogations ayant été soulevées lors des débats de la deuxième réunion plénière de septembre 2022, notamment sur la question de la définition de « l'IA » et sur l'articulation de l'étude d'impact dite HUDERIA (*Human Rights, Democracy and Rule of Law Impact Assessment*) ou AIDHE en français (pour *Analyse d'impact sur les droits de l'homme, la démocratie et l'État de droit*) avec les dispositions de la proposition de règlement de l'Union européenne sur l'intelligence artificielle.

Un projet d'instrument juridique cohérent dans le paysage international et européenne de la régulation de l'intelligence artificielle – En première analyse, le texte répond clairement au mandat du Conseil de l'Europe[70] et complète les initiatives internationales et régionales[71]. Il s'appuie notamment sur une définition assez comparable à celle de l'OCDE et ses dispositions cherchent à compléter les autres initiatives comme celles de l'UNESCO en éclairant précisément les questions des droits de l'homme, de la démocratie et de l'État de droit[72]. Autre élément substantiel notable : le Conseil de l'Europe rejoint la Commission européenne, l'OCDE et l'UNESCO dans une approche de réglementation par les risques en définissant des mesures plus contraignantes applicables aux systèmes présentant des « risques significatifs » (art.15).

Une tension entre général et particulier – Même s'il porte pour l'instant le titre de « convention », ce texte s'apparente toutefois en réalité bien plus à un traité ou convention cadre, au vu de la portée des principes posés (même si cette notion reste, en droit international public, assez vague[73]). Il semble dès lors surprenant d'y retrouver, en projet, un modèle d'étude d'impact d'effet direct, par nature opérationnelle, technique et détaillée pour revêtir un minimum de sens. L'articulation, pour les 27 États membres de l'UE qui vont devoir mettre en œuvre pour les systèmes de haut risque des mécanismes de mise en conformité, est à trouver : en l'état, un « fournisseur d'intelligence artificielle » (pour reprendre les termes du projet de convention) devrait donc se soumettre à la fois à une AIDHE/HUDERIA pour déterminer les dispositions de la Convention s'imposant à lui et, si son système est à haut risque selon les dispositions de la proposition de RIA, réaliser toutes les diligences de mise en conformité. Il aurait semblé plus simple que le Conseil de l'Europe s'en limite à créer une obligation générale d'évaluation des risques, en laissant le soin aux États signataires, membres ou non de l'UE, de mettre en œuvre les dispositions les plus appropriées[74]. Autrement

[70] S'agissant de l'articulation du mandat du Conseil de l'Europe avec celui de l'Union européenne s'agissant des droits de l'Homme, v. P. du Bois, Pierre, L'Union européenne et les droits de l'homme, Relations internationales, vol. 132, no. 4, 2007, pp. 33-39, accessible sur : https://www.cairn.info/revue-relations-internationales-2007-4-page-33.htm, consulté le 23 octobre 2022

[71] Y. Meneceur, Les cadres juridiques des organisations intergouvernementales pour une régulation de l'intelligence artificielle, Revue pratique de la prospective et de l'innovation n°1, Juillet 2021, pages 31-36

[72] Pour avoir accès à une vue globale des initiatives des différents organisations intergouvernementales, v. le site : https://globalpolicy.ai, consulté le 23 octobre 2022

[73] Parmi les éléments pouvant esquisser une définition, l'on relèvera notamment que ces textes énoncent des dispositions dont la traduction législative serait souhaitable, en plus de lignes directrices ne relevant pas à proprement parlé d'obligations juridiques, ainsi que le fait de ne pas fixer des règles de fond avec effet direct, mais de créer un cadre favorable à l'émergence de telles règles. V. notamment A. Kiss, Les traités-cadres : une technique juridique caractéristique du droit international de l'environnement, Annuaire français de droit international XXXIX, Éd. du CNRS, 1993, pp. 792-793.

[74] La raison pour laquelle le principe d'une AIDHE/HUDERIA a été ainsi posé par le Conseil de l'Europe est à trouver dans le fait que ce besoin a été identifié bien avant la publication de la proposition de RIA. Sa pertinence est maintenant à réinterroger au vu des évolutions du paysage réglementaire de cette technologie.

CC BY NC ND

dit, le texte mériterait d'aller jusqu'au bout d'un positionnement en convention cadre, de haut niveau, sans effets juridiques directs, appelant d'autres textes particuliers ultérieurs pour la mise en œuvre sectorielle. L'AIDHE/HUDERIA pourrait très bien être un instrument optionnel, établi de manière autonome et ne ralentissant pas les actuelles discussions de fond pour les dispositions juridiques de l'instrument juridique contraignant[75].

Modérer la protection des droits de l'Homme en fonction des risques encourus ? – Dans la même logique, l'emploi de la notion de « risques significatifs » (résultant de l'AIDHE/HUDERIA) n'est pas clairement articulable avec la pyramide des risques de la proposition de RIA, ouvrant à une double échelle d'évaluation des systèmes d'IA pour les États membres de l'UE. La proposition de la Commission constitue maintenant le standard autour duquel les parties prenantes vont devoir composer. L'espace laissé au Conseil de l'Europe pour imbriquer une méthodologie *ex ante* est sans nul doute extrêmement étroit, sauf à présenter sa propre démarche comme optionnelle.

Sur le fond, la société civile, tout comme des universitaires[76], ont questionné l'approche par les risques pour des textes présentés comme protecteurs des droits fondamentaux[77]. Il est vrai que le RGPD ne hiérarchise pas ses propres effets en fonction de la catégorie technique du traitement, mais prohibe ou limite des emplois en fonction de la nature et de la finalité des données traitées. Dès lors, une violation de la vie privée d'un individu est sanctionnable, quelle que soit la modalité de traitement de l'information.

La préoccupation de ne pas imposer une charge réglementaire pour des applications de « l'IA » considérées comme banales ou triviales a donc conduit les régulateurs internationaux à proposer une approche par les risques et le développement de standards techniques. Il est toutefois surprenant de retrouver également le Conseil de l'Europe sur le terrain de la conformité (*compliance*) en imposant des mesures supplémentaires aux seuls systèmes considérés comme présentant un « risque significatif » à la suite de la conduite d'une AIDHE/HUDERIA (art.15). En fondant ses dispositions sur la prévention spécifique d'une série d'effets ou de dommages, avec une totale neutralité technologique, le texte se serait économisé une série d'ornières tenant à la définition de risques à considérer comme acceptables pour les droits de l'Homme, la démocratie et l'État de droit.

Si d'autres industries technologiques, comme le nucléaire ou la pharmaceutique, ont l'habitude également de définir des niveaux de risques pour pouvoir opérer, la définition de tels risques se fonde sur des faits scientifiques généralement quantifiables et mesurables. Ici, les standards vont devoir s'appuyer sur des données qualitatives, potentiellement interprétables et souvent

[75] Pour une revue critique du AIDHE/HUDERIA, V. A. Mantelero, F. Francesca, Great Ambitions. The International Debate on AI Regulation and Human Rights in the Prism of the Council of Europe's Cahai, Philip Czech et al. (dir.), European Yearbook on Human Rights 2022, Intersentia, 4 avril 2022, accessible sur : https://ssrn.com/abstract=4131299, consulté le 25 octobre 2022

[76] V. notamment N.A. Smuha, E. Ahmed-Rengers, A. Harkens, W. Li, J. MacLaren, R. Piselli, K. Yeung, How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act, 5 août 2021, accessible sur https://ssrn.com/abstract=3899991, consulté le 22 octobre 2022 ou M. Gornet, W. Maxwell, Intelligence artificielle: normes techniques et droits fondamentaux, un mélange risqué, The Conversation, 28 septembre 2022, accessible sur : https://theconversation.com/intelligence-artificielle-normes-techniques-et-droits-fondamentaux-un-melange-risque-189587, consulté le 23 octobre 2022

[77] V. par exemple F. Hidvegi, D. Leufer, E. Massé, The EU should regulate AI on the basis of rights, not risks, Access Now website, 17 février 2021, mis à jour en janvier 2022, accessible sur : https://www.accessnow.org/eu-regulation-ai-risk-based-approach/, consulté le 22 octobre 2022

CC BY NC ND

contextuelles : comment définir un seuil tolérable de biais, si un système d'IA contribue à refuser une prestation sociale ou à minimiser des indemnités du fait de la localisation géographique et, potentiellement par corrélation, de l'origine ethnique (les zones socialement défavorisées regroupant le plus souvent les individus d'immigration récente). Nul doute qu'un gouvernement politiquement marqué très à droite pourrait avoir une interprétation différente d'un gouvernement marqué plutôt à gauche.

Sans remettre en cause le principe de la mise en place d'études préalables identifiant et mesurant l'état d'un risque (avec les limites liées aux outils issus des sciences humaines et sociales s'agissant d'impacts de la mesure d'impacts d'ordre sociaux), il semble toutefois discutable de faire dépendre de seuils pseudo-techniques des mesures de protection juridique. Autrement dit, l'information préalable d'un état des risques produit par une étude peut bien entendu conduire un responsable humain à assumer les risques encourus et à énumérer les mesures pour les atténuer ; mais il est bien plus dans l'esprit des textes relatifs aux droits humains de protéger les individus de tous les types de violations, quelles que soient leurs sources et avec des mécanismes d'exception[78], plutôt que de restreindre, *ab initio*, les sources potentielles de dommages afin de chercher un équilibre avec d'autres intérêts (notamment économiques ou sécuritaires).

Des garanties procédurales à parfaire – Les droits établis par le projet de convention garantissent notamment l'information des usagers quant à une éventuelle interaction avec une machine (art.7.4) ou à l'intervention d'un système d'IA dans une prise de décision (art.7.2), avec la tenue éventuellement de registres (art.7.1), ainsi qu'une information claire explicitant la prise de décision prise par un système d'IA et la possibilité de contester cette décision. Ces garanties apparaissent toutefois relativement peu ambitieuses pour laisser opérer, à titre d'exemple, divers systèmes « prédictifs » aux fonctionnements scientifiques bien incertains. Les restrictions énumérées à l'art.8 pourraient être reformulées, mêmes si les concepts vagues de « bien-être économique du pays » ou de « protection de la morale » (copiés *in extenso* de la Convention européenne des droits de l'Homme) ont déjà été interprétés de manière assez restrictive pour ce qui relevait de la protection de la vie privée par la Cour européenne des droits de l'Homme. Ici, rien ne garantit que ces concepts, élargis à la lutte contre la discrimination, à l'atteinte de la dignité humaine ou à toute autre violation, ne viennent pas justifier des emplois contestables de systèmes d'IA.

Les prohibitions ne semblent également pas encore assez ambitieuses (art.14). Même si elles limitent clairement certains des emplois comme l'identification, la catégorisation et la déduction des émotions, elle laisse la porte ouverte à de nombreuses autres applications aux fondements scientifiques contestés héritées de la physiognomonie. Les « utilisation par les autorités publiques de systèmes d'intelligence artificielle à des fins qui ne sont pas compatibles avec les valeurs fondamentales des sociétés démocratiques, ne poursuivent pas un objectif légitime et ne sont pas nécessaires dans une société démocratique » (art.14.3) paraissent, de même, extrêmement vagues,

[78] Le mécanisme de l'article 10 de la Convention européenne des droits de l'Homme sur la liberté d'expression, pose un principe du droit à la liberté d'expression et énumère trois critères d'ingérence (légalité, légitimité et nécessité dans une société démocratique). V. Guide sur l'article 10 de la Convention européenne des droits de l'Homme, préparé par le greffe de la Cour, accessible sur : https://www.echr.coe.int/Documents/Guide_Art_10_FRA.pdf, consulté le 23 octobre 2022

CC BY NC ND

même si la notion, encore ici, est en réalité dérivée de la Convention européenne des droits de l'Homme sur la liberté d'expression.

Les mécanismes de protection de la démocratie – Des mesures concrètes et originales sont bien à relever, notamment en vue de prévenir les manipulations dans le cadre des élections ou de nature à porter atteinte à la participation démocratique pour les systèmes d'IA présentant des risques significatifs (art.15.b). La protection des valeurs démocratiques et de la démocratie est présente de manière assez constante tout au long du texte (art.1.1, art.5 et 5.1, art.6, art.8, art.11, art.12, art.13.1, art.14.3, art.15.a, b et c, art.17).

La difficulté à laquelle ce texte fait face ne lui est pas propre mais tient au fait que des notions de démocratie ou de société démocratique ne font guère consensus au sein de l'Organisation et peine à émerger de la jurisprudence de la Cour[79]. Même après l'expulsion de la Fédération de Russie en mars 2022, les discussions demeurent parfois âpres et le Conseil de l'Europe n'est pas parvenu en ses plus de 70 ans d'existence à faire émerger un texte standard faisant pleinement consensus. Dès lors, même s'il paraît tout à fait pertinent pour le Conseil de l'Europe d'apporter cette brique « démocratique » dans la construction de règles contraignantes portant sur le développement et l'emploi de systèmes d'IA, l'on ne se sait exactement comment les États membres interprèteront cette notion, au risque de faciliter le « blanchissage » d'usages en réalité très éloignés des attentes des rédacteurs du projet de convention.

2.8. Le projet de Convention du Conseil de l'Europe sur l'intelligence artificielle : vers un traité international fondé sur les droits de l'Homme, la démocratie et l'État de droit – étude la seconde « version 0 »

Une seconde « version 0 » du projet de Convention a été présentée lors de la 4ème réunion plénière du CAI, s'étant tenue entre le 1er et le 3 février 2023[80]. Ce document tente de répondre aux nombreuses critiques développées précédemment, recueillies tant officiellement qu'officieusement auprès de plusieurs parties prenantes.

Comme pour la première version, ce travail est également attribué au président du Comité et se décompose en 8 chapitres et 38 articles, sans annexe. Son préambule reprend une narration très volontariste sur les bienfaits de « l'IA », en prenant garde de minimiser les risques à quelques systèmes :

> « Conscients des développements accélérés de la science et de la technologie et des profonds changements induits par l'application des systèmes d'intelligence artificielle fournissant de nouveaux outils, qui ont le potentiel de promouvoir la prospérité humaine ainsi que le bien-être individuel et social en renforçant le progrès et l'innovation ;

[79] V. par exemple J. Tajadura Tejada, La doctrine de la Cour européenne des droits de l'homme sur l'interdiction des partis politiques, Revue française de droit constitutionnel, vol. 90, no. 2, 2012, pp. 339-371

[80] CAI(2023)01, accessible sur : https://rm.coe.int/cai-2023-01-fr-projet-zero-revise-de-convention-cadre-public/1680aa1942, consulté le 8 mars 2023

CC BY NC ND

> Préoccupés par le risque que **certaines** utilisations de **quelques** systèmes d'intelligence artificielle puissent également avoir le potentiel d'interférer indûment avec l'exercice des droits de l'homme et des libertés fondamentales, de saper la démocratie et de violer l'État de droit ; »

Le texte demeure tout à fait conforme au mandat du Conseil de l'Europe, le projet ne vise pas à encadrer les aspects économiques, commerciaux ou militaires des systèmes d'intelligence artificielle. L'article 4 prend le parti de répéter de manière surabondante l'article 1er du statut du Conseil de l'Europe, qui écarte de son champ de compétence les questions de défense nationale. Le texte se positionne de manière explicite comme une convention cadre, dont l'objectif n'est pas de fixer à proprement parler des règles de fond, avec des effets juridiques directs, mais de constituer un cadre dans lequel celles-ci peuvent se développer.

Définition et principes établis par la seconde version du projet de Convention – L'article 2 recourt à la notion de système algorithmique pour définir la nature d'un « système d'IA » :

> a. « système d'intelligence artificielle » signifie tout **système algorithmique** ou toute combinaison de tels systèmes qui, tels que définis ici et dans le droit interne de chaque Partie, utilise des **méthodes de calcul dérivées de statistiques ou d'autres techniques mathématiques** pour exécuter des fonctions qui sont **communément associées à l'intelligence humaine** ou qui exigeraient autrement cette dernière, et qui **assiste ou remplace le jugement des décideurs humains** dans l'exécution de ces fonctions. Ces fonctions comprennent, sans s'y limiter, la **prédiction**, la **planification**, la **classification**, la **reconnaissance des formes**, l'**organisation**, la **perception**, la **reconnaissance de la parole**, des **sons**, des **images**, la **génération** de textes, de sons, d'images, la **traduction de langues**, la **communication**, l'**apprentissage**, la **représentation** et la **résolution de problèmes** ;

Si la référence à la notion de système algorithmique paraît en résonnance avec les conclusions dressées à la fin de la première partie de cette étude, le reste de la définition délimite un périmètre original par rapport aux autres définitions des organisations internationales intergouvernementales. La longue énumération dans la description des fonctions paraît superflue, alors même que le texte prend la précaution de demeurer non exclusif. La restriction à des algorithmes assistant ou remplaçant des décideurs humains, si elle est bien volontaire, conduit à faire échapper des prévisions de ce texte nombre d'applications et paraît en contradiction avec l'énumération à la suite (une génération automatique d'image ne vise pas nécessairement à assister ou remplacer un décideur humain).

Un alignement avec d'autres définitions d'organisations intergouvernementales, comme celle de l'OCDE, aurait probablement apporté plus de clarté juridique à ce nouveau projet. Cette réécriture apparaît donc comme étant en régression par rapport à la première version présentée précédemment.

Ce nouveau projet établit à nouveau, dès ses dispositions générales et comme dans la première version, un principe de non-discrimination (art.3). Le chapitre II (art.5 à 7) se consacre à développer

CC BY NC ND

des obligations s'imposant spécifiquement aux autorités publiques. Le chapitre III (art.8 à 11) développe des obligations s'imposant tant au secteur public qu'au secteur privé, en ce qui concerne la fourniture de biens, d'équipements et de services. Le chapitre IV (art.12 à 18) énumère une liste de principes fondamentaux applicables à la conception, au développement et à l'application de systèmes d'IA. Le chapitre V (art.19 à 23) présente des mesures et garanties assurant la responsabilité et des voies de recours. Cette version a abandonné l'idée d'intégrer un cadre d'évaluation des risques propre au Conseil de l'Europe pour se limiter à en énoncer le principe dans un chapitre VI (art.24 à 26). Enfin, le projet présente un mécanisme de suivi (chapitre VII, art.27 à 29) et des clauses finales (chapitre VIII, art.30 à 38).

Des dispositions spécifiques au secteur public – Les obligations du chapitre II visent le secteur public, en imposant aux signataires de déléguer à « une autorité nationale compétence » le soin de procéder à « une évaluation approfondie de la nécessité, de la proportionnalité et des risques potentiels, avant et pendant l'application d'un système d'intelligence artificielle à l'élaboration ou à la révision des lois et politiques régissant les rôles respectifs du pouvoir exécutif, des autres institutions démocratiques et du pouvoir judiciaire » (art.7).

Les obligations de l'article 5 visent à prévenir des dommages résultant d'outils d'aide à la décision (art.5.a), tout en autorisant des exceptions « compatible avec les valeurs fondamentales des sociétés démocratiques, conforme à la loi et nécessaire dans une société démocratique à la poursuite d'un but légitime d'intérêt général » (art.5.b). L'article 6 impose aux États signataires une obligation de diligence pour minimiser (et « prévenir dans la mesure du possible ») tout « préjudice ou dommage illégal ».

Les entités privées agissant pour le compte de l'État sont également concernées par ces dispositions.

Des dispositions communes au secteur public et au secteur privé relatives à la fourniture de biens, d'équipements et de services – Le chapitre III développe des dispositions applicables à la fourniture de biens, d'équipements et de services « d'IA », afin que les acteurs concernés, relevant tant du secteur public que privé, respectent dans le cadre de ces opérations les droits de l'homme, la démocratie et l'État de droit. Plus précisément l'art.8 décompose deux types d'obligations aux États signataires : celle d'assurer la compatibilité des applications de « l'IA » avec les obligations juridiques nationales et internationales (art.8.a) et de donner des orientations effectives sur la manière de prévenir des effets négatifs sur les droits de l'homme, la démocratie et l'État de droit.

L'article 9 impose aux États signataires de prendre les mesures nécessaires pour préserver la liberté individuelle, la dignité humaine et l'autonomie des applications inappropriées de « l'IA » et l'article 10, de manière large, de prendre également des mesures pour « veiller à ce que toutes les parties, tous les groupes et tous les individus bénéficient d'un accès égal et équitable au débat public et à des processus démocratiques inclusifs » sans limiter à l'application de « l'IA » (même si cela pourrait être interprété de manière implicite). La seconde partie de l'article 10 ne parait qu'ajouter au principe posé précédemment, en enjoignant les États signataires de prendre en compte les

CC BY NC ND

développements technologiques et les entités, tant publiques que privées, utilisant des systèmes « d'IA » contribuant à « façonner le débat public ».

Enfin, l'article 11 regroupe les questions de santé publique et d'environnement en une phrase.

Les principes fondamentaux et les garanties de la seconde « version 0 » du projet de Convention – Pour mémoire, alors que la première version du projet de Convention définissait 9 principes dans ses articles 5.2 à 5.10, cette version en comporte 7, avec de nouvelles associations (art.12 à 18) et la disparition des principes de durabilité et de solidarité.

Première « version 0 »	Seconde « version 0 »
Art.5.2 – Durabilité, solidarité et protection de l'environnement	L'environnement est cité dans l'article 11. Les principes du durabilité et solidarité ne sont plus cités.
Art.5.3 – Égalité	Art.12 – Égalité et anti-discrimination
Art.5.4 – Protection de la vie privée	Art.13 – Vie privée et protection des données à caractère personnel
Art.5.5 – Responsabilité et obligation de rendre des comptes	Art.14 - Obligation de rendre des comptes, de responsabilité et de responsabilité juridique
Art.5.6 – Transparence et auditabilité	Art.15 – Transparence et contrôle
Art.5.7 - Environnement réglementaire contrôlé en vue d'encourager la recherche et favoriser l'innovation	Art.17 – Innovation sûre
Art.5.8 – Sûreté, sécurité, intégrité des données, cybersécurité et robustesse	Art.16 - Sûreté
Art.5.9 - Consultation multipartite à la lumière des implications sociales, économiques, éthiques et juridiques	Art.18.1 – Consultation publique
Art.5.10 - Culture et compétences numériques	Art.18.2 – Culture et compétences numériques

Source : Auteur

La garantie de voies de recours effectives en cas de préjudice ou de dommage illégal aux droits de l'homme et aux libertés fondamentales est imposée aux États signataires de la Convention dans l'article 19. Il doit être rappelé que la première « version 0 » avait été critiquée pour ne pas avoir prévu de tels mécanismes.

L'article 20 édicte un droit à un contrôle humain des décisions pour les systèmes « d'IA » « informe ou prend substantiellement une ou plusieurs décisions affectant les droits de l'homme et les libertés fondamentales » (art.20.1), ainsi qu'un droit à l'information si l'on interagit avec un tel système (art.20.2 – et la possibilité alternative d'interagir avec un humain). Des mesures appropriées sont à prendre par les États signataires dans l'intérêt des personnes handicapées pour rendre effectif les articles 20.1 et 20.2 (art.20.3).

Ces trois séries de dispositions (voies de recours, droit à un contrôle humain et droit à l'information) peuvent recevoir exception par la voie législative « lorsqu'elles sont nécessaires et

proportionnées dans une société démocratique à la poursuite d'un but légitime d'intérêt général » (art.21).

Les articles 22 et 23 sont des clauses habituelles visant à s'assurer que cette Convention ne limite pas des protections d'un caractère supérieur.

Simplification de l'approche par les risques – Le précédent projet du Conseil de l'Europe avait été critiqué pour introduire un mécanisme propre d'évaluation des risques, imaginé avec la proposition de règlement de la Commission européenne mais devenu obsolète avec la publication de cette proposition et le début de la procédure législative. Cette nouvelle version se satisfait donc de poser le principe d'une approche par les risques, afin de ne pas entrer en conflit avec le futur mécanisme de mise en conformité de l'Union (art.24 et 25). De manière notable, l'application de certaines dispositions ne sont plus conditionnées à une classification particulière du système, recentrant ainsi cette version du projet en texte fondé sur des droits et non les risques.

Plus spécifiquement, il est demandé aux États signataires de fournir « une orientation effective aux fournisseurs et utilisateurs d'intelligence artificielle pertinents sur la manière d'identifier, d'évaluer, de prévenir et d'atténuer les risques et les impacts négatifs résultant de l'application d'un système d'intelligence artificielle en ce qui concerne la jouissance des droits de l'homme, le fonctionnement de la démocratie et le respect de l'État de droit » (art.24.1). Les détails de cette orientation sont énumérés dans l'article 24.2, en précisant qu'elle devrait :

- Contenir des exigences proportionnées ciblés à des contextes spécifiques pouvant présenter des risques accrus sur les droits de l'homme, la démocratie et l'État de droit ;
- Intégrer la perspective des individus sujets des traitements « d'IA » ;
- Exiger un « enregistrement et la prise en compte particulière des impacts négatifs résultant de l'application de systèmes d'intelligence artificielle améliorés » (la notion de « systèmes améliorés » appelant d'ailleurs une certaine clarification) ;
- Imposer des vérifications régulières, tout au long du cycle de vie des systèmes.

L'article 24.3 laisse toute latitude à un État signataire d'interdire ou de suspendre l'emploi de systèmes « d'IA » lorsque « lorsqu'une telle mesure est considérée appropriée et nécessaire par l'autorité nationale compétente », sans pour autant désigner de types d'application devant faire l'objet de telles dispositions. Il doit être relevé ici la suppression de l'ancien article 14 qui interdisait une série d'applications (systèmes d'intelligence artificielle utilisant la biométrie pour identifier, catégoriser ou déduire les émotions des individus ; systèmes d'intelligence artificielle pour la notation sociale ; systèmes d'intelligence artificielle à des fins qui ne sont pas compatibles avec les valeurs fondamentales des sociétés démocratiques), cette seconde version ne procédant à aucune interdiction de principe.

L'article 25 impose aux États signataires de veiller à ce que les fournisseurs et utilisateurs « d'IA » « appliquent des mesures de prévention et d'atténuation suffisantes résultant de l'application du Cadre de gestion des risques » et par ailleurs documentent et conservent les enregistrements nécessaires.

CC BY NC ND

Une seconde « version 0 » tout aussi pragmatique que la précédente – Ce texte, tout à fait provisoire et encore appelé à évoluer substantiellement à l'occasion des négociations tenues en 2023, apparaît en net progrès quant à son positionnement mais demeure très inégal, tant en ce qui concerne sa clarté que la logique de ses mécanismes. Il se présente, de manière assez originale pour un texte émanant d'une organisation intergouvernementale gardienne de valeurs fondamentales et habitué à des positionnements plus stricts, comme extrêmement pragmatique, avec un préambule prenant soin de ne pas porter trop d'emphase sur les applications contestables de « l'IA » et des dispositions assurant aux États signataires une très large marge d'interprétation.

Même si le texte répond à ses principales critiques, avec des dispositions relatives à des voies de recours et une simplification appréciable d'un mécanisme d'approche par les risques, il ouvre de nouvelles brèches en renonçant à édicter des prohibitions ou des moratoires stricts. Ainsi, alors que la première version interdisait (même de manière timide) des applications aux fondements scientifiques contestables (comme la reconnaissance des émotions, ancien art.14), ce texte se satisfait de renvoyer les États signataires face à leurs responsabilités, sans leur offrir de réelle ligne de conduite (art.24.3). De même, les principes de durabilité et de solidarité (ancien art.5.2) ont disparu du texte alors même que d'autres textes d'organisations intergouvernementales ont intégré ces notions.

Ainsi, la « Déclaration européenne sur les droits et principes numériques pour la décennie numérique » adoptée en décembre 2022 par le Parlement européen, le Conseil de l'Union européenne et la Commission européenne[81], énumère une série de principes apparaissant comme bien plus ambitieux que le projet de Convention examiné ici. Même si cette Déclaration demeure juridiquement non contraignante, la portée politique de ses principes apparaît comme significative pour le développement de nouveaux textes relatifs au numérique. Il pourrait par exemple être cité le chapitre II de cette Déclaration relative à la solidarité et à l'inclusion, dont les articles rappellent que « La transformation numérique devrait contribuer à l'équité et à l'inclusivité sociales et économiques dans l'UE ».

Le pragmatisme et la prudence de cette seconde « version 0 » présentent donc le projet de Convention du Conseil de l'Europe comme peu ambitieux, peut-être à même de fédérer des États sceptiques quant à des politiques publiques numériques protectrices des individus et de la société, mais risquant d'édulcorer des principes pourtant consensuels dans d'autres foras.

[81] (2023/C 23/01), *op.cit.*

CC BY NC ND

Conclusion : des instruments juridiques européens relatifs à l'intelligence artificielle cédant à la tentation du pragmatisme

La juxtaposition des deux initiatives produit bien, par effet de contraste, une image relativement nette de l'évolution de la place des droits humains dans notre société, lorsqu'ils sont mis à l'épreuve de l'innovation technologique et d'impératifs tant économiques que sécuritaires.

S'il y a bien des recouvrements entre la proposition de RIA de la Commission européenne et le projet de Convention du Conseil de l'Europe, ceux-ci sont bien plus d'ordre politique que dans le détail des dispositions. Les deux approches paraissent céder à une forme de « tentation pragmatique » pour modérer les effets des instruments juridiques protégeant les droits humains, la sécurité des individus ou la santé en fonction de risques encourus. Des applications fortement contestées, comme les systèmes de reconnaissance biométriques opérant en temps réel dans les espaces publics (reconnaissance faciale par exemple), ne sont pas pleinement prohibées et bénéficient, au titre d'exigences sécuritaires, d'une tolérance de mise en œuvre à des fins considérées comme « strictement nécessaires ». Sans revenir sur des conséquences concrètes comme le « chilling effect[82] », les décideurs publics amoindrissent la protection des droits fondamentaux et des libertés individuelles en déployant des outils dont la réelle performance est en réalité extrêmement contestable[83]. Il pourrait être répondu de manière tout à fait appropriée que, même si les outils étaient performants, l'amoindrissement serait déjà intolérable : nous insistons toutefois sur le fait que nous nous acculturons par le biais de tels textes à des outils de surveillance[84] en payant un prix qui n'est pas nécessairement à la hauteur des espoirs de meilleure sécurité.

Ce pragmatisme révèle donc non seulement de ce qui pourrait être qualifié de « realpolitik » du numérique, mais également des technologies dans leur ensemble, dès lors que celles-ci conjuguent enjeux économiques, stratégiques et potentiellement sécuritaires. L'on comprend ainsi que pour certains analystes de la société civile, l'invocation des droits humains, dans ce contexte, relèverait d'une sorte adjuvant en capacité de rassurer le public, mais dont les principes actifs seraient en partie désactivés.

S'il n'est pas si surprenant que la Commission européenne, dont l'histoire et l'ADN n'est pas à trouver dans la protection des droits fondamentaux[85], fasse primer sous la direction du Commissaire européen au marché intérieur les questions économiques, le projet de convention du Conseil de l'Europe pourrait encore mieux protéger les droits de l'homme, la démocratie et l'État de droit en assumant de poser des principes de haut niveau complémentaires à ceux de la CvEDH

[82] Découragement d'un exercice légitime de ses droits civils de peur d'une sanction : dans le contexte de la reconnaissance faciale dans les espaces publics, peur d'être fiché et désavantagé ensuite pour l'accès à un emploi public par exemple.
[83] V. notamment Y. Meneceur, Pour un débat public informé, ouvert et démocratique sur la reconnaissance faciale, Acteurs Publics, 20 mai 2022 et Y. Meneceur, Quelle est la probabilité qu'une IA précise à 99% reconnaisse un individu ? Spoiler : ce n'est pas 9 chances sur 10, Blog Les Temps Électriques, 20 février 2022, accessible sur : https://lestempselectriques.net/index.php/2022/02/20/quelle-est-la-probabilite-quune-ia-precise-a-99-reconnaisse-un-individu-spoiler-ce-nest-pas-9-chances-sur-10/, consulté le 23 octobre 2022
[84] V. les développements sur les outils de suivi de contact lors de la crise sanitaire en 2020. Y Meneceur, COVID-19 & Traçage numérique : comment concilier santé et libertés individuelles ?, Texte de l'intervention à la Conférence organisée par Cap Digital, l'UNESCO et la Chaire ITEN de l'Unesco, 27 novembre 2020, accessible sur : https://lestempselectriques.net/index.php/2020/11/29/covid-19-tracage-numerique-comment-concilier-sante-et-libertes-individuelles/, consulté le 23 octobre 2022
[85] Même si la situation a substantiellement évolué ces dernières années, V. P. du Bois, L'Union européenne et les droits de l'homme, *op.cit.*

et en ne cherchant pas nécessairement à « garantir un équilibre adéquat entre le respect des droits de l'homme […] et les divers intérêts économiques, sécuritaires et autres dans le développement et l'utilisation de l'intelligence artificielle » (Préambule du projet, 9e paragraphe). Les droits humains sont un fondement, non une externalité. Et il peut sembler regrettable que ce soit un texte juridiquement non contraignant, par ailleurs présenté dans le cadre de l'Union européenne, qui vienne aujourd'hui constituer la première brique significative de droits et principes applicables à notre ère numérique. Il s'agit de la « Déclaration européenne sur les droits et principes numériques pour la décennie numérique » adoptée en décembre 2022 par le Parlement européen, le Conseil de l'Union européenne et la Commission européenne[86], qui présente une série de principes affirmant « les valeurs et les droits fondamentaux » à garantir en ligne, comme ils sont déjà garantis hors ligne.

Céder à la tentation du pragmatisme, c'est donc aujourd'hui concrètement contribuer à éroder des digues qui ont été érigées en réaction et en pleine conscience de la puissance de l'horreur de la Seconde Guerre mondiale. Même s'il ne s'agit pas de remettre en cause la motivation des organisations concernées à protéger les droits humains, les réglementations sur l'intelligence artificielle ne devraient pas alimenter les critiques de la société civile. Ces droits sont déjà bien mis à mal par la perte de confiance dans la démocratie représentative et la résurgence des idéologies d'extrême-droite, dont les idées prospèrent, également sur l'autel du pragmatisme, et au prix d'une grande confusion[87]. Bien au contraire, il y a là pour les législateurs européens une opportunité d'ajouter aux protections existantes, en instaurant par exemple un principe de minimisation du recours aux divers systèmes algorithmiques afin de limiter non seulement leur emploi inapproprié mais, également, contribuer à une sobriété énergétique plus que jamais nécessaire[88].

Une autre solide confusion résiste aujourd'hui : celle liant innovation et progrès. Les systèmes d'IA sont incontestablement novateurs et apportent incontestablement leurs lots d'amélioration et d'optimisation. Ils ne représentent toutefois pas nécessairement un progrès si l'on examine l'impact de la société dans son ensemble et l'*affordance* de l'écosystème numérique. C'est pourquoi l'impératif débat démocratique qui devrait aujourd'hui avoir lieu est aussi celui de la place que nous souhaitons réserver à ces outils numériques et, ainsi, composer des réglementations traduisant des aspirations collectives qui ne sont pas nécessairement techniciennes.

[86] (2023/C 23/01), *op.cit.*
[87] Ph. Corcuff, La Grande Confusion : Comment l'extrême droite gagne la bataille des idées, éditions Textuel, 2021
[88] Sur la minimisation (ou la proportionnalité) du recours aux algorithmes, v. Y. Meneceur, L'intelligence artificielle en procès – Plaidoyer pour une réglementation internationale et européenne, Bruylant, 2020, p.387