



HAL
open science

**La Convention-Cadre du Conseil de l'Europe sur l'IA :
premier traité international encadrant les risques de l'IA sur
les droits de l'homme, la démocratie et l'Etat de droit, à la
valeur ajoutée incertaine**

Olivia Tambou

► **To cite this version:**

Olivia Tambou. La Convention-Cadre du Conseil de l'Europe sur l'IA : premier traité international encadrant les risques de l'IA sur les droits de l'homme, la démocratie et l'Etat de droit, à la valeur ajoutée incertaine. Le règlement européen sur l'IA et au delà : Quel encadrement de l'IA ?, Bruylant; Larcier, pp.293-316, 2025, 9782802775461. <hal-05306374>

HAL Id: hal-05306374

<https://hal.science/hal-05306374v1>

Submitted on 9 Oct 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

*CHAPITRE 2 – CONVENTION-CADRE DU CONSEIL DE L'EUROPE
SUR L'IA : PREMIER TRAITÉ INTERNATIONAL
ENCADRANT LES RISQUES DE L'IA SUR LES DROITS
DE L'HOMME, LA DÉMOCRATIE ET L'ÉTAT DE DROIT
À LA VALEUR AJOUTÉE INCERTAINE*

Olivia TAMBOU

MAÎTRE DE CONFÉRENCES HDR À L'UNIVERSITÉ PARIS DAUPHINE-PSL

Le Conseil de l'Europe est la première organisation internationale à avoir adopté un texte contraignant relatif à l'IA ouvert à la ratification d'États non européens. La Convention-cadre du Conseil de l'Europe sur l'IA (ci-après « Convention ») est un texte de trente-six articles, dont l'ambition est de contribuer à l'échelle internationale à l'encadrement du développement de l'usage de l'IA en ce qui concerne les risques sur les droits de l'homme, la démocratie et l'État de droit¹. Cette Convention est donc limitée à ces trois champs qui sont au cœur des compétences et de l'expertise du Conseil de l'Europe. Elle était attendue par de nombreux acteurs notamment de la société civile. Elle avait vocation à compléter utilement l'AI Act très orienté sur le volet économique et commercial, même s'il est teinté d'humanisme au-delà de sa logique de marché intérieur². Ces attentes se sont confrontées à la dure réalité des négociations internationales. Le texte final précise que la Convention n'a pas « pour objet de créer de nouveaux droits de l'homme ou de nouvelles obligations en matière de droits de l'homme »³. Il s'agit bien plus de faire en sorte que le développement de l'IA ne compromette

1. Pour une analyse détaillée du contenu de la Convention, cf. M. HO-DAC, « La Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit : un outil singulier d'interopérabilité normative transnationale au sein de la gouvernance mondiale de l'IA », in H. ASCENSIO, G. CAHIN, A.-Th. NORODOM (dir.), *Annuaire français de droit international*, vol. 2024, Paris, LGDJ (à paraître, 2025). Pour une première analyse de l'ensemble de la Convention, cf. M. GARTNER, « Council of Europe: States Adopt First Binding Framework Treaty on AI », *Journal of AI Law and Regulation*, 2024, vol. 1, Issue 3, pp. 342-349. Pour une approche comparée avec l'AI Act, cf. J. ZILLER, « The Council of Europe Framework Convention on Artificial Intelligence vs. the EU Regulation: Two Quite Different Legal Instruments », *CERIDAP*, avril 2024, disponible sur <https://ceridap.eu/the-council-of-europe-framework-convention-on-artificial-intelligence-vs-the-eu-regulation-two-quite-different-legal-instruments/?lng=en>.

2. Voy. le chapitre de M. HO-DAC dans cet ouvrage, p. 115.

3. Cf. rapport explicatif de la Convention, § 13.

BRUYLANT

pas l'effectivité des protections existantes. La valeur ajoutée d'une telle approche semble d'emblée incertaine. Les plus pessimistes pourraient en déduire que le Conseil de l'Europe a perdu la bataille dans le dialogue compétitif entre les organisations internationales, et les États pour la régulation régionale et mondiale de l'IA. S'il est vrai que le texte de la Convention est décevant, il est possible de s'interroger sur la manière dont le Conseil de l'Europe pourrait maintenir son leadership en garantissant la sauvegarde du patrimoine juridique européen développé depuis plus de soixante-dix ans en son sein, face au développement de l'IA.

C'est à partir de 2018 que progressivement la plupart des comités, des organes intergouvernementaux, organes spécialisés du Conseil de l'Europe, ont entrepris d'examiner l'impact de l'IA dans leur domaine d'activité et ont fait des recommandations. À titre d'illustration⁴, la Commission européenne pour l'efficacité de la justice (CEPEJ) a adopté en 2018 une charte éthique européenne d'utilisation de l'IA dans les systèmes judiciaires et leur environnement. En 2019, le Comité consultatif de la Convention pour la protection des données à l'égard du traitement automatisé des données à caractère personnel (Convention 108) a adopté des lignes directrices sur l'intelligence artificielle et la protection des données. Mais surtout, à la suite d'une conférence de haut niveau organisée à Helsinki en février 2019⁵, le Comité des ministres du Conseil de l'Europe a créé un organe dédié, le Comité *ad hoc* sur l'IA (CAHAI), afin d'avoir une vision plus transversale. Le mandat du CAHAI a été d'examiner la faisabilité et les éléments potentiels d'un cadre juridique pour le développement, la conception et l'application de l'IA, fondés sur les normes du Conseil de l'Europe dans le domaine des droits de l'homme, de la démocratie et de l'État de droit. C'est sur la base des travaux du CAHAI et notamment d'une étude de faisabilité adoptée en 2021⁶ que le Comité des ministres a chargé son successeur, le Comité sur l'IA (CAI), d'élaborer une Convention-cadre sur l'IA, les droits de l'homme, la démocratie et l'État de droit. C'est à la suite d'un processus participatif impliquant de nombreux États y compris non européens, associant d'autres organisations internationales, des représentants de la société civile, et de l'industrie que cette Convention a été élaborée. Le texte final est un texte de compromis

4. Pour une vision complète, cf. la brochure du Conseil de l'Europe, *Le Conseil de l'Europe et l'intelligence artificielle*, Strasbourg, Éditions du Conseil de l'Europe, 2023, <https://rm.coe.int/brochure-artificial-intelligence-fr-march-2023-print/1680aab8e7>.

5. Conseil de l'Europe, « Maîtriser les règles du jeu – l'impact du développement de l'intelligence artificielle sur les droits de l'homme, la démocratie et l'état de droit », Helsinki, 26-27 février 2019, www.coe.int/fr/web/freedom-expression/aiconference2019.

6. Étude de faisabilité sur un cadre juridique pour l'IA base sur les standards du Conseil de l'Europe, adoptée par le CAHAI, le 17 décembre 2020, accessible à <https://rm.coe.int/cahai-2020-23-final-etude-de-faisabilite-fr-2787-2531-2514-v-1/1680a1160f>.

entre les différentes approches actuelles existantes à l'échelle mondiale de l'encadrement de l'IA. Prenant en considération l'évolution rapide de la technologie, susceptible de faire évoluer l'approche de l'encadrement de l'IA par les États, la Convention est avant tout une Convention-cadre. Il s'agit d'un instrument flexible pouvant être complété par d'autres textes contraignants ou non soit pour approfondir certains de ses aspects, soit pour aborder l'utilisation de l'IA dans des secteurs spécifiques.

L'ambition de la Convention est donc de dresser un premier cadre juridique commun acceptable par tous et d'ancrer juridiquement certains points sur lesquels un consensus émerge à l'échelle internationale. Le premier est sans aucun doute la définition même de l'objet qu'il faut encadrer⁷. La terminologie retenue est celle de systèmes d'IA (SIA), c'est-à-dire les applicatifs de l'IA et non la technologie en elle-même. L'article 2 de la Convention reprend mot pour mot la définition des SIA issue des recommandations de l'OCDE dans leur version révisée en 2023 et 2024⁸. Un SIA est un « système automatisé, qui pour des objectifs explicites ou implicites, déduit, à partir d'entrées reçues, comment générer des résultats en sortie tels que des prévisions, des contenus, des recommandations ou des décisions qui peuvent influencer sur des environnements physiques ou virtuels. Différents systèmes d'intelligence artificielle présentent des degrés variables d'autonomie et d'adaptabilité après déploiement ». Cette définition large exclurait les logiciels simples qui se contentent d'effectuer des tâches autonomes sur la base de règles définies exclusivement par des humains. En revanche, cette définition fonctionnelle permet de prendre en compte les systèmes automatisés dont les entrées sont fournies par les humains ou des machines. Elle intègre aussi que les objectifs d'un SIA peuvent être définis de manière explicite ou implicite par l'homme. Il peut aussi s'agir d'objectifs implicites qui sont révélés par l'exploitation des données d'entraînement ou par l'entraînement d'un modèle qui grâce à l'apprentissage affine les préférences de ses utilisateurs⁹. Enfin, cette définition inclut les SIA génératives en évoquant clairement que les SIA peuvent produire des contenus. Elle permet également d'inclure certains SIA qui continuent d'évoluer après leur conception et leur déploiement.

7. Contrairement à l'*AI Act*, et au projet zéro de la convention, le texte final de la convention ne comporte que la définition des SIA. Cf. CAI, « Projet zéro révisé de convention [cadre] sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit », Strasbourg, 6 janvier 2023, CAI (2023)01_FR, accessible à <https://rm.coe.int/cai-2023-01-fr-projet-zero-revisé-de-convention-cadre-public/1680aa1942>.

8. Pour une analyse comparée de la définition des SIA, cf. les propos introductifs de cet ouvrage, p. 22.

9. Cf. OCDE, « Explanatory Memorandum on the Updated OECD Definition of an AI System », *OECD Artificial Intelligence Papers*, mars 2024, n° 8, p. 7.

Le deuxième point de consensus international est la nécessité d'avoir une approche globale permettant d'encadrer toutes les activités menées dans le cadre du cycle de vie des SIA¹⁰. Dans un souci de flexibilité et de subsidiarité, le texte ne définit pas ce concept ni ne donne de listes exhaustives des différentes étapes du cycle de vie d'un SIA. Le rapport explicatif renvoie sur ce point aux travaux de l'OCDE qui décrivent ce cycle de vie à travers sept étapes qui, en pratique, ne se succèdent pas toujours de façon linéaire et vont de la conception du SIA à sa mise hors service¹¹.

Le troisième point de consensus international est la nécessité d'aborder l'encadrement des SIA selon une approche par les risques. L'article 1^{er} de la Convention précise que son objectif est d'imposer aux Parties l'adoption de mesures « graduées et différenciées, si nécessaire, en fonction de la gravité et de la probabilité de l'apparition d'impacts négatifs sur les droits de l'homme, la démocratie et l'État de droit ». De même, le chapitre V est entièrement consacré à l'évaluation et l'atténuation des risques et des impacts négatifs, même si à la différence de l'*AI Act*, la Convention ne cherche pas à établir une catégorisation des risques.

Il en résulte qu'à première vue, la Convention-cadre pour l'IA du Conseil de l'Europe semble difficilement qualifiable d'*AI Act light*, mais plutôt d'une sorte de compromis entre les recommandations de l'OCDE et l'*AI Act*. La prise en compte de la diversité des traditions et des systèmes juridiques et politiques aboutit à un texte comportant souvent des dispositions très générales et abstraites. Les pressions exercées par les États tiers au Conseil de l'Europe, notamment les États-Unis, ne sont sans doute pas étrangères à la forme finale de la Convention¹². Pour autant sa vocation à rallier des États non-européens au modèle de l'Union européenne de régulation de l'IA n'est pas évidente, même si cette perspective n'est pas exclue. La Convention est avant tout un texte ouvert offrant une alternative à l'approche de l'*AI Act* tout en étant interopérable avec lui. La Convention laisse beaucoup de flexibilité aux États dans le choix de leur approche par les risques de l'IA. Elle fixe néanmoins des obligations de résultat aux États et des exigences générales

10. Cf. art. 1^{er} de la Convention.

11. Les 7 étapes sont : 1) la planification et la conception, 2) la collecte et le traitement des données, 3) le développement des ISA y compris l'élaboration de modèles et/ou l'adaptation de modèles existants à des tâches spécifiques, 4) les essais, vérification et validation, 5) la fourniture et la mise à disposition de systèmes, 6) le déploiement, 7) l'exploitation et suivi et 8) la mise hors service. Cf. § 15 du rapport explicatif qui renvoie aux travaux de l'OCDE notamment.

12. Cf. E. DE CAPITANI, « The COE Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law. Is the Council of Europe losing its Compass? », *European Area of Freedom Security and Justice*, 4 mars 2024, accessible à <https://free-group.eu/2024/03/04/the-coe-convention-on-artificial-intelligence-human-rights-democracy-and-the-rule-of-law-is-the-council-of-europe-losing-its-compass/>.

qu'ils s'engagent à mettre en œuvre dans leur ordre juridique. Du point de vue matériel, elle peut aboutir à ancrer dans le droit contraignant les recommandations de l'OCDE pour certains États tout en laissant à d'autres le choix d'opter pour un niveau de protection plus élevé. Le choix de cette approche syncrétique des risques est articulé avec une volonté de véritablement structurer la coopération internationale. Ces deux points font à la fois la force et la faiblesse de l'approche du Conseil de l'Europe. À ce stade, l'approche syncrétique des risques qui émanent du texte final de la Convention soulève des inquiétudes (I), mais la Convention-cadre constitue une première pierre rendant possible une utile et future structuration de la coopération internationale sous l'effet de ce que l'on pourrait appeler le *Strasbourg effect* (II).

I. UNE APPROCHE SYNCRÉTIQUE DES RISQUES SOURCES D'INQUIÉTUDES

Le mandat donné par le Comité des ministres au CAI pour l'élaboration de la Convention ne comportait aucune précision sur l'approche par le risque qui devait être choisie. Pourtant, l'étude de faisabilité du CAHAI de 2021 avait recommandé non seulement une méthodologie pour classer les risques des SIA, mais aussi d'intégrer des catégories similaires à celles de l'*AI Act*, de SIA à risque faible, de SIA à haut risque et de SIA inacceptable¹³. Dès le projet zéro de la Convention rendu public en janvier 2023, les négociateurs ont renoncé à toute catégorisation des risques, en laissant cette tâche aux Parties. L'objectif était clairement de proposer une approche syncrétique des risques, acceptable par tous, en se concentrant sur des exigences générales. L'importance des choix laissés aux Parties relativise l'apport de la Convention à la protection des droits fondamentaux, de la démocratie et de l'État de droit.

A. *Le renoncement à la catégorisation des risques au profit d'exigences générales*

La Convention fixe trois grands types d'exigences générales pour les Parties. D'une part, des obligations générales de protection des droits de l'homme et de l'intégrité des processus démocratiques et du respect de l'État de droit. D'autre part, sept principes généraux communs sont listés au chapitre III, qui rappellent ceux que l'on retrouve dans de nombreuses chartes éthiques ou recommandations. Enfin, la Convention pose un cadre général de gestion et de risques des impacts négatifs.

13. Cf. « Possible Elements of a Legal Framework on Artificial Intelligence, Based on The Council of Europe's Standards on Human Rights, Democracy and Rule of Law », CAHAI (2021)09rev, Strasbourg, 3 décembre 2021, p. 5.

1. – L'impact ambivalent des obligations générales

Premièrement, les Parties doivent prendre ou maintenir des mesures pour que les SIA dans leur État soient cohérents avec leurs obligations nationales et internationales en matière de droit de l'homme¹⁴. Cela inclut aussi une obligation pour les Parties de disposer de voies de recours accessibles et effectives contre les éventuelles violations de ces droits de l'homme qui pourraient résulter d'activités menées dans le cycle de vie des SIA¹⁵. Selon le rapport explicatif, cette obligation, quoique générale, constitue une obligation de résultat, qui laisse uniquement aux Parties le choix des moyens pour sa mise en œuvre¹⁶. Cela implique que les Parties doivent lister l'ensemble de leurs obligations aux droits de l'homme et veiller à ce que ces droits soient toujours protégés avec le développement des SIA¹⁷.

Deuxièmement, les Parties doivent garantir que les SIA ne sont pas utilisés afin de compromettre « l'intégrité des processus démocratiques, l'indépendance et l'efficacité des institutions y compris le principe de la séparation des pouvoirs, l'indépendance de la justice et l'accès à la justice »¹⁸.

Il s'agit aussi de permettre un accès équitable et la participation des personnes au débat public afin qu'elles puissent se forger librement une opinion. La formulation choisie reste ambivalente au-delà d'une dimension négative, elle aurait pu comporter une obligation cette fois-ci positive imposant aux Parties de prendre des mesures pour permettre l'utilisation des SIA pour renforcer les institutions et les processus démocratiques. L'affirmation d'une véritable obligation positive donnerait plus de corps aux dispositions relatives à la mise en œuvre de la Convention. Cela permettrait de consacrer une réelle dimension inclusive, promouvant l'utilisation de l'IA pour respecter les droits des personnes handicapées et des enfants¹⁹, garantissant le débat public autour des questions importantes soulevées par l'usage des SIA²⁰, et favorisant la maîtrise du numérique et les compétences numériques pour tous²¹.

14. Cf. art. 4 de la Convention.

15. Cf. art. 14 de la Convention.

16. Cf. rapport explicatif, § 38.

17. Cf. rapport explicatif, §§ 39-41, qui établit une liste indicative des principaux traités internationaux en matière de droit de l'homme auxquels les Parties peuvent être soumises, sachant qu'en outre, chaque Partie doit y ajouter ses dispositions nationales relatives à la protection des droits de l'homme.

18. Cf. l'article 5 de la Convention et les paragraphes 43 et 46 du rapport explicatif qui illustrent les contextes sensibles spécifiques dans lesquels l'utilisation de l'IA peut potentiellement porter atteinte à la démocratie et à l'État de droit.

19. Art. 18 de la Convention.

20. Art. 19 de la Convention.

21. Art. 20 de la Convention.

2. – La consécration de sept principes généraux à défaut de véritables droits

Le chapitre III de la Convention énonce selon une approche générale des principes communs qui doivent être intégrés dans les réglementations nationales de chacune des Parties et qui sont ainsi applicables aux activités des SIA. Il s'agit 1°) de la dignité humaine et de l'autonomie personnelle, 2°) de la transparence et du contrôle, 3°) d'obligation de rendre des comptes et de responsabilité, 4°) de l'égalité et de la non-discrimination, 5°) du respect de la vie privée et de la protection des données à caractères personnels, 6°) du principe de fiabilité, 7°) de principe d'innovation sûre.

Cette liste interroge non seulement quant à la sélection des principes communs qui ont été retenus, mais aussi plus globalement quant à la signification juridique du recours à cette notion de principes communs.

D'une part, certains de ces principes sont des droits fondamentaux garantis notamment par la Cour EDH²². La valeur ajoutée de les lister ici sous la nature de principes communs paraît limitée. D'autre part, pour les cinq premiers principes, la Convention se contente de préciser que chaque partie « adopte ou maintient des mesures » pour les garantir dans les activités au sein du cycle de vie des SIA. La liste de ces principes symbolise le socle des principes sur lesquels États tiers et États parties du Conseil de l'Europe peuvent se retrouver. Ces principes font d'ailleurs écho à la plupart des recommandations d'autres organisations internationales ou institutions.

Le premier principe commun énoncé est celui de dignité humaine et d'autonomie personnelle. Il s'agit de veiller à ce que la réglementation de l'IA soit centrée sur l'humain et respecte la valeur inhérente de chaque individu. Si la dignité humaine n'est pas explicitement consacrée par la Convention européenne des droits de l'homme, elle est au cœur de la jurisprudence de la Cour EDH²³. Elle a aussi été inscrite à l'article 1^{er} de la charte des droits fondamentaux de l'UE. Sa consécration explicite en tant que premier principe commun au sein de la Convention rappelle

22. Il s'agit du principe de non-discrimination (art. 14 CEDH) et le respect de la vie privée et la protection des données à caractère personnel (art. 8 CEDH).

23. Cf. C. GREWE, « La dignité humaine dans la jurisprudence de la Cour européenne des droits de l'homme, Intervention à la 7^e conférence-débat du Centre de droit public comparé, Université Panthéon-Assas Paris II, 30 octobre 2014 », *Revue générale du droit on line*, 2014, n° 18323. Pour une illustration récente dans une affaire de mineur non accompagné demandeur de protection internationale dans laquelle la Cour a condamné la Grèce pour traitements inhumains et dégradants ; cf. Cour EDH, 23 janvier 2024, *OR c/ Grèce*, n° 24650/19. La Cour souligne que le requérant resté sans abri et sans accès aux soins pendant 6 mois se trouvait dans « une situation de précarité, d'insécurité et de dénuement physique et psychologique qui [...] portait atteinte à l'essence même de la dignité humaine », cf. pt 65.

que la dignité humaine est universellement reconnue comme le fondement des droits de l'homme²⁴. Appliquée au contexte de l'IA, c'est affirmer clairement que l'humain doit toujours prévaloir sur la machine et ne pas être ravalé au rang d'objet. Le Conseil d'État insistait sur cette dimension en parlant de primauté humaine²⁵. Si l'affirmation de la dignité humaine était inévitable, la pratique montre que l'invocabilité de ce concept n'est pas toujours aisée²⁶, et ce d'autant plus qu'il peut y avoir des acceptations différentes selon les États²⁷, voire les juges nationaux²⁸.

L'intérêt de l'article 7 de la Convention aurait pu être d'imposer aux Parties des mesures spécifiques pour garantir de véritables droits attachés à cette dignité humaine. Or, l'article 7 ne permet pas de consacrer véritablement un droit à la supervision humaine, voire un contrôle humain, ou un droit garantissant l'intégrité physique et mentale des humains incluant un droit à la déconnexion des SIA. L'article 7 ne donne pas non plus véritablement de corps à la notion d'autonomie personnelle.

L'article 8 relie le contrôle à la transparence ce qui renvoie aux mécanismes de gouvernance des activités des SIA. Le rapport explicatif précise que la transparence vise à rendre compréhensible et accessible aux Parties concernées le fonctionnement des SIA tout au long de leur cycle de vie. Il s'agit pour l'essentiel de disposer d'informations appropriées sur les choix techniques faits notamment le niveau d'automatisation. L'article 8 oblige explicitement les Parties à adopter des mesures permettant l'identification des contenus générés par des SIA. Cette obligation de transparence renforcée applicable aux IA dites « génératives » est aujourd'hui largement admise²⁹. Cela peut se traduire techniquement par l'apposition d'un filigrane permettant aux personnes de savoir que le contenu a été généré par un SIA. En complément, l'article 14 de la Convention impose aussi aux Parties de garantir que les personnes concernées soient bien informées qu'elles interagissent avec un SIA et non avec un humain.

24. Cf. § 1^{er} du préambule de la Déclaration universelle des droits de l'homme ; Pacte international relatif aux droits civils et politiques.

25. Cf. CE, *Intelligence artificielle et action publique : construire la confiance, servir la performance*, étude, mars 2022, Paris, La Documentation française, pp. 98 et s.

26. Cf. M. BREUER, « The Council of Europe as an AI Standard Setter », *Verfassungsblog*, 4 avril 2022.

27. B. LIGER et K. ORFALI (dir.), *La dignité de la personne : quelles réalités ?*, 1^{re} éd., Bruxelles, Bruylant, 2016.

28. L. BURGOGUE-LARSEN (dir.), *La dignité saisie par les juges en Europe*, 1^{re} éd., Bruxelles, Bruylant, 2011.

29. Cf. art 50 de l'*AI Act*, ou encore G-7 Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems, <https://www.mofa.go.jp/files/100573473.pdf>.

Le rapport explicatif rappelle que la transparence est étroitement liée à d'autres principes tels que celui de l'explicabilité et de l'interprétabilité. Ainsi, la transparence peut impliquer de donner une liste de facteurs qui sont utilisés par le SIA pour prendre une décision, mais aussi d'être informé sur comment un SIA prend ses décisions. Au-delà de la transparence, le contrôle repose sur l'ensemble des mécanismes mis en place pour surveiller, évaluer les activités menées par les SIA.

Ces principes de transparence et de contrôle se complètent avec une obligation de rendre des comptes et de responsabilité. L'objectif est ici de pouvoir clairement identifier les responsabilités des acteurs dans le cycle de vie des SIA lorsqu'il y a des impacts négatifs sur les droits de l'homme, la démocratie et l'État de droit.

Les principes d'égalité, de non-discrimination, de respect de la vie privée et des données à caractère personnel renvoient pour l'essentiel au respect de ces droits fondamentaux comme cela a déjà été évoqué antérieurement. Le principe de fiabilité renvoie à des exigences de qualité notamment des données et de la cybersécurité tout au long du cycle de vie des SIA. Le rapport explicatif souligne que les normes techniques peuvent contribuer de façon concrète à garantir ce principe de fiabilité.

Le dernier principe d'innovation sûre fait écho à la nécessité d'encadrer le besoin d'expérimentation, d'essais pour le développement de SIA qui ne pourraient l'être autrement. Ce principe rappelle simplement que ces méthodologies innovantes types bac à sable doivent avoir pour but d'éviter les impacts négatifs sur les droits de l'homme, la démocratie et l'État de droit. De tels environnements innovants doivent nécessairement être placés sous la surveillance d'autorités compétentes.

3. – La description d'un cadre commun de gestion des risques devant être concrétisé par les Parties

L'article 16 de la Convention impose aux Parties l'existence d'un cadre de gestion des risques prenant en compte les impacts réels et potentiels sur les droits de l'homme, la démocratie et l'État de droit. À défaut de catégorisation théorique des risques, ce cadre propose une approche prenant en compte le contexte et l'utilisation prévue des SIA. Il appartient alors à chaque Partie de prendre des mesures permettant d'identifier, d'analyser et d'évaluer les risques réels et potentiels sur les droits de l'homme, la démocratie et l'État de droit ainsi que d'évaluer l'impact des SIA. Le cadre de risque doit prendre en compte la gravité et la probabilité des impacts potentiels, le point de

vue des parties prenantes, comprendre un suivi des risques et des impacts négatifs, et documenter tout cela. Cela inclut de documenter en cas d'impact négatif les mesures prises pour y remédier.

B. Le risque de relativiser la protection des droits fondamentaux, de la démocratie et de l'État de droit

L'article 3 de la Convention comporte en lui un risque de relativiser la protection des droits fondamentaux, de la démocratie et de l'État de droit en raison des choix faits pour réduire le champ d'application de la Convention.

1. – L'importance des choix laissés aux Parties pour le meilleur et pour le pire

L'article 3 de la Convention laisse deux choix importants aux Parties quant à son champ d'application.

Le premier choix concerne la possibilité d'avoir une approche différenciée permettant de limiter l'application de la Convention aux activités « des pouvoirs publics ou des acteurs privés qui agissent pour leur compte ». L'article 3.1., b), précise que les Parties peuvent simplement répondre aux risques et aux impacts découlant des activités des SIA des acteurs privés en prenant des « mesures appropriées ». Le rapport explicatif précise qu'il peut s'agir de « mesures administratives, ou volontaires »³⁰. Autrement dit, la formulation finale de l'article 3.1 illustre les tensions entre les États partisans d'une approche réglementaire globale couvrant les acteurs publics et privés, et ceux souhaitant limiter l'encadrement des entreprises privées à des formules plus souples et agiles de régulation au nom de la nécessité de ne pas contraindre l'innovation³¹. Cette tension entre ces deux approches a été au cœur des négociations de l'article 3 révélées dans les médias³² y compris par des tribunes de représentants de la société civile³³. Il est indéniable que la

30. Rapport explicatif, pt 29.

31. C'est d'ailleurs la raison pour laquelle, la Convention exclue de son champ d'application les activités de recherche et de développement qui ne sont pas encore rendus disponibles, cf. art. 3.3 de la Convention. Une exclusion similaire se retrouve à l'article 2.6 de l'*AI Act*. Pour autant la Convention pose un principe d'innovation sûre, cf. art. 13. L'article 3.3. rappelle également que des essais ou activités similaires ne peuvent être entrepris d'une manière qui puisse porter atteinte aux droits de l'homme, à la démocratie et à l'État de droit.

32. www.euractiv.com/section/digital/news/us-obtains-exclusion-of-ngos-from-drafting-ai-treaty/.

33. Cf. « Open Letter to COE AI Convention Negotiators: Do Not Water Down Our Rights », mars 2024, accessible à https://algorithmwatch.org/de/wp-content/uploads/2024/03/Open_letter_Council_of_Europe_AI_Convention.pdf.

version finale de l'article 3 reflète largement la position d'États tiers comme les États-Unis, le Canada et le Japon, mais aussi de certains États européens non membres de l'Union européenne comme la Suisse³⁴. Dans ce contexte, il est difficile de comprendre comment le risque d'avoir une application à géométrie variable de la Convention selon l'option choisie par la Partie pourra être évité. Plus globalement, cette approche comme celle de l'Union européenne³⁵ pose avec acuité la question de comment intégrer dans des normes souples des exigences de respect de droits fondamentaux, d'État de droit et de démocratie³⁶. Enfin, l'affirmation selon laquelle la Partie ayant fait le choix d'une approche différenciée « ne peut déroger ou limiter l'application des obligations internationales qui lui incombent en matière de protection des droits de l'homme, de la démocratie et de l'État de droit » laisse perplexe. D'une part, une telle affirmation relève de l'évidence et semble symbolique. D'autre part, sa présence atteste encore une fois de la tension possible entre une approche souple et une approche réglementaire et de la nécessité de subordonner la première à une hiérarchie des normes classiques.

L'article 3.1., b) précise que les Parties devront exprimer leur choix par le biais d'une déclaration soumise au secrétaire général du Conseil de l'Europe³⁷. Ce choix peut évoluer à tout moment. À ce stade, il est difficile de savoir quelle sera l'ampleur des Parties qui opteront pour une approche différenciée et dans quelle mesure ce choix restera définitif ou provisoire.

Le second choix laissé aux Parties leur permet d'exclure du champ de la Convention les activités des SIA « liées à la protection de leurs intérêts de sécurité nationale ». Ce choix formulé à l'article 3.2 fait écho à l'article 2.3 de l'*AI Act*³⁸. Pour autant, la confrontation entre ces deux dispositions

34. www.swissinfo.ch/eng/foreign-affairs/ai-regulation-is-swiss-negotiator-a-us-stooge/73480128.

35. Cf. contribution de L. FROMONT et A. VAN WAEYENBERGE dans cet ouvrage, p. 61.

36. Cf. contribution de J. MARANHÃO, B. FICO et B. SOUSA dans cet ouvrage, p. 335, sur HUDERIA II-A.

37. Actuellement, seule la Norvège a fait une telle déclaration pour préciser qu'elle appliquerait directement la Convention aux entreprises privées. Il est vrai que cette déclaration peut se faire à des stades ultérieurs notamment celle de l'adhésion ou de la ratification de la Convention. Cf. art. 3.1, b).

38. Art. 2.3. *AI Act* : « Le présent règlement ne s'applique pas aux domaines qui ne relèvent pas du champ d'application du droit de l'Union et n'affecte en aucun cas les compétences des États membres en matière de sécurité nationale, quel que soit le type d'entité chargée par les États membres d'exécuter des tâches en rapport avec ces compétences. Le présent règlement ne s'applique pas aux systèmes d'IA lorsque et dans la mesure où ils sont mis sur le marché, mis en service ou utilisés, avec ou sans modification, exclusivement à des fins militaires, de défense ou de sécurité nationale, quel que soit le type d'entité exerçant ces activités. Le présent règlement ne s'applique pas aux systèmes d'IA qui ne sont pas mis sur le marché ou mis en service dans l'Union, lorsque la production est utilisée dans l'Union exclusivement à des fins militaires, de défense ou de sécurité nationale, quel que soit le type d'entité exerçant ces activités ».

amène à trois types de commentaires. Premièrement, l'exclusion de principe de la sécurité nationale dans l'*AI Act* semblait plus naturelle, étant donné que selon l'article 4 TUE, la sécurité nationale est de la compétence des États membres. En revanche, la sécurité nationale est incluse dans de nombreux traités internationaux en matière de droits de l'homme notamment du Conseil de l'Europe³⁹, mais aussi la Convention américaine relative aux droits de l'homme, le Pacte international relatif aux droits civils et politiques, le Pacte international relatif aux droits économiques, sociaux et culturels (CESCR) des Nations unies. La société civile, consciente des risques de l'utilisation de SIA dans le cadre de la sécurité nationale au regard des droits de l'homme, a rappelé à plusieurs reprises durant les négociations la nécessité que la Convention s'applique au SIA utilisé à des fins de sécurité nationale⁴⁰. Deuxièmement, le texte final opte plutôt pour une non-application de principe. Il précise que les Parties ne sont pas tenues d'appliquer la Convention à ces activités de sécurité nationale. Le fait que l'article 3.2 ne prévoit aucune procédure permettant aux Parties d'inclure la sécurité nationale dans le champ d'application de la Convention contrairement au paragraphe précédent en dit long sur le consensus des parties autour de cette exclusion. Troisièmement, le rapport explicatif tente de minimiser l'ampleur de cette exception en rappelant son champ d'application fonctionnel⁴¹. D'une part, cette exclusion ne concerne que les activités liées à la protection des intérêts de la sécurité nationale des Parties. Ce concept de sécurité nationale est en principe entendu de façon strict⁴² et ne concerne pas les activités courantes de maintien de l'ordre (prévention, détection, enquêtes et poursuite des crimes), ni même les menaces à la sécurité publique. Pour autant et en pratique, la frontière entre ces différents concepts peut paraître un peu floue. D'autre part, le rapport explicatif soutient que la formulation de l'article 3.2 permet d'inclure

39. Cf. la Convention de 2005 pour la prévention du terrorisme et son Protocole additionnel de 2015. Sur l'articulation CEDH, sécurité nationale, cf. Rapport préparé par la division de la recherche de la Cour européenne des droits de l'homme sur la sécurité nationale et la jurisprudence européenne. Le Conseil de l'Europe a également adopté de nombreux textes non contraignants en la matière, cf. l'avis sur la recommandation 2024(2013) sur la sécurité nationale et l'accès à l'information ; T-PD(2018)0 ; *guide pratique sur l'utilisation des données à caractère personnel dans le secteur de la police* ; recommandation de l'Assemblée parlementaire 1983 (2011) sur les recours abusifs au secret d'État et à la sécurité nationale : obstacles au contrôle parlementaire et judiciaire des violations des droits de l'homme sur les questions de sécurité nationale. En revanche, les questions relatives à la défense nationale ne sont pas de la compétence du Conseil de l'Europe (cf. art. 1, d, du statut du Conseil de l'Europe). Cela explique largement l'exclusion prévue à l'article 3.4 de la Convention.

40. « Open Letter to Council Of Europe Ai Convention Negotiators: Do Not Water Down Our Rights », 5 mars 2024, accessible à https://ecnl.org/sites/default/files/2024-03/CSOs_CoE_Calls_2501.docx.pdf.

41. Cf. rapport explicatif, pt 32.

42. Cf. Cour EDH, 8 novembre 2022, *Saure c/ Allemagne*, n° 8819/16, § 52.

dans le champ d'application de la Convention les activités concernant les SIA à « double usage », c'est-à-dire ceux qui peuvent être utilisés pour d'autres buts. Pour autant, la formulation de l'article 3.2 « d'activités liées à la protection des intérêts de sécurité nationale » est ambiguë. À titre de comparaison, celle de l'*AI Act* parle explicitement de fins « exclusivement » de sécurité nationale.

Au-delà du champ d'application, l'article 16.4 laisse aussi aux Parties le choix d'adopter ou non un moratoire, une interdiction ou d'autres mesures appropriées lorsque l'utilisation de certains SIA est incompatible avec le respect des droits de l'homme, la démocratie et l'état de droit. Cette disposition témoigne encore une fois de la volonté de ne pas se caler sur le modèle de l'Union européenne qui impose une catégorie prédéterminée de pratiques interdites.

2. – Des silences qui en disent long

Trois sujets évoqués durant les négociations sont pratiquement absents dans le texte final. Premièrement, la Convention n'évoque presque pas la préservation de l'environnement et de la santé publique. Dans le projet zéro, un article 11 rappelait que les Parties devaient prendre « les mesures nécessaires pour préserver la santé publique et l'environnement dans le contexte de l'application d'un SIA »⁴³. Dans le texte final, seul le préambule précise que la Convention vise « à encourager la prise en compte des risques et des impacts [...] notamment sur la santé humaine et l'environnement ». De son côté l'article 19 de la Convention finale incite les Parties à garantir une consultation publique multipartite sur les incidences environnementales des SIA. Le rapport explicatif donne quelques exemples de la forme que ces consultations pourraient prendre. Il pourrait s'agir d'enquêtes, de questionnaires, d'ateliers publics, de groupes de discussion, de jurys citoyens, de sondages, d'auditions publiques, de comités consultatifs, de groupes d'experts. Pour autant, il y est rappelé que la détermination des sujets, les modalités et la fréquence, ainsi que la prise en compte des résultats de ces consultations restent entièrement dans les mains des Parties. On ne peut que constater l'absence d'obligations précises en la matière, alors même qu'un contentieux émerge y compris devant la CEDH visant à garantir une

43. Cf. CAI, « Projet zéro révisé de convention [cadre] sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit », Strasbourg, 6 janvier 2023, CAI (2023)01_FR. Initialement envisagée comme une obligation, la préservation de la santé et de l'environnement avait même été érigée dans une version ultérieure au rang des principes relatifs aux activités menées dans le cadre de cycle de vie des SIA, cf. version ultérieure du 18 décembre 2023, CAI (2023) 28, avant finalement de ne plus apparaître dans les articles.

protection effective par les États contre les effets néfastes du changement climatique⁴⁴. Or, la conception et l'utilisation de SIA ont nécessairement un coût environnemental qui, même s'il n'est pas encore clairement et objectivement évalué, devrait à terme conduire à un débat public sur les usages entre une IA utile et une IA futile⁴⁵.

Deuxièmement, la Convention ne comporte plus aucune disposition concernant l'éventuelle protection de lanceurs d'alerte⁴⁶ qui dénonceraient les risques que des SIA pourraient poser au regard de droits de l'homme, de la démocratie et de l'État de droit. Il est vrai que dans de nombreux États européens et l'Union européenne se sont dotés d'un statut du lanceur d'alerte⁴⁷ susceptible de s'appliquer pour les personnes qui dénonceraient une violation de la Convention-cadre sur l'IA⁴⁸. Cette possibilité est néanmoins moins évidente dans les pays non membres de l'UE.

Troisièmement, le texte final ne comporte aucune disposition spécifique sur la participation parlementaire pour suivre et examiner les mesures prises pour sa mise en œuvre. L'insertion d'une telle participation avait pourtant été demandée par l'Assemblée parlementaire du Conseil de l'Europe⁴⁹. Il est pensable que la participation des parlements nationaux puisse être couverte par l'incitation des Parties à faire des consultations multipartites. Pour autant, l'absence de mention explicite relative au rôle des parlements nationaux mais aussi de l'APCE laisse clairement entendre que la mise en œuvre de la Convention sera laissée pleinement aux mains des exécutifs des Parties.

44. Cf. Cour EDH, gde ch., 9 avril 2024, *Verein KlimaSeniorinnen Schweiz et a. c/ Suisse*, n° 53600/20 ; Cour EDH, gde ch., 9 avril 2024, *Carème c/ France*, n° 7189/21 ; Cour EDH, gde ch., 9 avr. 2024, *Duarte et a. c/ Portugal*, n° 39371/20, et autres.

45. Sur cette question, cf. CONSEIL ÉCONOMIQUE SOCIAL ET ENVIRONNEMENTAL, « Impacts de l'intelligence : risques et opportunités pour l'environnement », avis septembre 2024 ou encore l'AFNOR SPEC IA Frugale 2314 adoptée en juin 2024, accessible à www.boutique.afnor.org/fr-fr/norme/afnor-spec-2314/referentiel-general-pour-lia-frugale-mesurer-et-reduire-limpact-environneme/fa208976/421140.

46. L'article 19 de la version précitée du 18 décembre 2023 précisait : « *Each Party shall take appropriate measures to ensure protection of whistleblowers in relation to the activities within the lifecycle of AI systems which could adversely impact human rights, democracy, and the rule of law* ».

47. Directive (UE) 2019/1937 sur la protection des personnes qui signalent des violations du droit de l'Union européenne, *JOUE*, L 305 du 26 novembre 2019.

48. La directive UE assure une protection en cas de violation du droit de l'UE, ce qui pourrait couvrir la violation de l'*AI Act* en lien avec les textes listés à l'annexe 1 de cette directive notamment ceux relatifs à la sécurité des produits, la protection de la vie privée et des données à caractère personnel, et la sécurité des réseaux et des systèmes d'information. Il serait possible d'envisager une modification de l'annexe 1 de la directive lanceurs d'alertes incluant de manière explicite la protection des lanceurs d'alerte pour violation de l'*AI Act*.

49. Cf. article 9.11 de l'avis 303 (2024) adopté le 18 avril 2024 par l'Assemblée parlementaire du Conseil de l'Europe sur le projet de convention-cadre sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit.

À l'issue de cette première analyse, le texte même de la Convention est décevant. Les importantes concessions faites pour permettre sa future ratification par des États tiers laissent dubitatifs tant du point de vue du fond que de la forme. En effet, n'aurait-il pas été plus pertinent de permettre aux États demandeurs de flexibilité de déroger à la Convention dont le contenu aurait alors pu être plus protecteur pour les autres. Le choix d'une convention reposant sur une approche syncrétique des risques, et interdisant aux États toutes réserves, pourrait être une victoire à la Pyrrhus pour le modèle européen de droits de l'homme que symbolise le Conseil de l'Europe. Il appartiendra au Conseil de l'Europe de démontrer que la structuration de la coopération internationale qu'il propose apporte une réelle plus-value. Pour cela il pourrait s'appuyer sur son expertise et son expérience en matière de monitoring du respect des droits de l'homme, de la démocratie et de l'État de droit. L'effectivité de la protection future accordée par la Convention-cadre est subordonnée à la concrétisation d'un *soft power* du Conseil de l'Europe que certains n'hésitent plus à décrire dans d'autres contextes sous les termes de *Strasbourg effect*⁵⁰.

II. UNE STRUCTURATION UTILE DE LA COOPÉRATION INTERNATIONALE POSSIBLE GRÂCE AU *STRASBOURG EFFECT* ?

La valeur ajoutée de la Convention-cadre du Conseil de l'Europe pourrait être à rechercher dans les outils inhérents à sa mise en application et susceptibles d'être porteurs d'amélioration selon une approche intergouvernementale classique qui dans d'autres domaines a pu faire ses preuves. Pour autant, là aussi les outils proposés semblent limités et le contexte géopolitique actuel peut faire douter que l'adhésion espérée d'États tiers permette au Conseil de l'Europe de peser véritablement dans la structuration de la coopération internationale face à d'autres organisations telles que l'Union européenne et l'OCDE, voire l'ONU.

50. Expression employée jusqu'à présent dans d'autres contextes et qui caractérisent soit l'influence de la Cour européenne des droits de l'homme, cf. B. ÇALI, « Russia and the European Court of Human Rights: The Strasbourg Effect », *IJCL*, janvier 2019, vol. 17, n° 1, pp. 361-365, <https://doi.org/10.1093/icon/moz014> ou encore D. SARMIENTO et S. IGLESIAS SÁNCHEZ, « The Brussels Effect, insight », *EU Law Live*, 14 mai 2024 ; soit l'effet extraterritorial de la CEDH, cf. B. VAN DER SLOOT, « The Strasbourg-Effect: Wieder and Guarnieri and the Extra-Territorial effect of the European Convention on Human Rights », *The European Convention on Human Rights Law Review*, 2024, voire l'influence du Conseil de l'Europe dans l'élaboration d'un droit mondial de la protection des données à caractère personnel, cf. L. A. BYGRAVE, « The "Strasbourg Effect" on Data Protection in Light of the "Brussels Effect": Logic, Mechanics and Prospects », *Forthcoming in Computer Law & Security Review*, 2020, vol. 38, *University of Oslo Faculty of Law, Research Paper No. 2020-14*, disponible sur SSRN: <https://ssrn.com/abstract=3617871> or <http://dx.doi.org/10.2139/ssrn.3617871>.

A. *Des outils d'amélioration à la portée limitée*

La Convention-cadre du Conseil de l'Europe comporte en son sein deux types d'outils classiques. D'une part, elle dispose d'une gouvernance souple qui bien que sommaire pourrait progressivement instiller une amélioration de la coopération internationale en matière de régulation commune des SIA. D'autre part, la Convention permet l'adoption de protocole d'amendement pour des Parties désireuses d'aller plus loin et d'ajouter ou de modifier certains articles. Au-delà de ces mécanismes classiques, la valeur ajoutée du Conseil de l'Europe pourrait être de mettre à profit son expertise en proposant un référentiel méthodologique permettant une approche harmonisée des risques et des études d'impact des SIA sur les droits fondamentaux, la démocratie et l'État de droit.

1. – Des mécanismes classiques

Comme pour d'autres conventions, la Convention-cadre sur l'IA organise sa gouvernance selon une logique intergouvernementale très caractéristique du Conseil de l'Europe. La mise en œuvre effective de la Convention est confiée à la Conférence des Parties qui peut examiner toute question liée à la Convention. Cela passe par l'identification de difficultés, l'adoption de recommandations, l'échange d'informations, la facilitation de règlement amiable de différends entre les Parties, la proposition d'amendements⁵¹ à la Convention. Lorsque l'amendement est proposé par une Partie ou par le Comité des ministres, la Conférence des Parties donne son avis sur cette modification. Au-delà d'une énumération de ce que pourra faire cette Conférence des Parties qui s'apparente à un mécanisme de suivi, l'article 23 de la Convention-cadre reste assez vague. Aucune périodicité de la réunion de cette Conférence des parties n'est indiquée. Tout au plus est-il précisé que la Conférence des Parties est convoquée chaque fois que nécessaire soit par le secrétariat du Conseil de l'Europe, soit par la majorité des Parties, soit par le Comité des ministres. Rien n'est précisé sur les modalités de vote au sein de la Conférence des Parties, ce qui implicitement laisse entendre que le principe de l'égalité des Parties est garanti⁵². Enfin, les Parties s'engagent à faire des rapports périodiques à la Conférence des Parties sur les activités qu'elles ont entreprises pour assurer l'application de la Convention aux pouvoirs publics et aux acteurs privés. Pour autant, ces rapports ne

51. Comme le précise le rapport explicatif, les amendements ne permettent que « des modifications mineures de nature procédurale et technique », *cf.* pt 149.

52. Le point 130 du rapport explicatif fait expressément référence à l'existence d'une partie égale de toutes les Parties au mécanisme de suivi. L'UE devrait donc avoir une seule voix.

2. – Vers un référentiel méthodologique

Jusqu'à présent, l'une des principales critiques des acteurs de l'écosystème de l'IA est qu'il n'existe aucun document officiel et clair sur lequel ils pourraient s'appuyer pour évaluer l'impact de leur SIA notamment au regard des droits fondamentaux, l'État de droit et la démocratie⁵⁷. C'est la raison pour laquelle le Comité des ministres du Conseil de l'Europe a demandé au Comité sur l'IA d'élaborer une méthodologie appropriée pour permettre une mise en œuvre effective de la Convention-cadre. Le projet est ambitieux car il s'agit de faire la synthèse du cadre actuellement très fragmenté des approches en matière d'impact d'évaluation des risques susceptibles d'être appliqués aux SIA, des normes techniques⁵⁸, des cadres de gouvernance concernant les SIA, et les guides à l'échelle mondiale. Le lancement de ce projet a fait l'objet d'une collaboration avec le centre de recherche Alan Turing⁵⁹ qui a inspiré les travaux du Comité sur l'IA. Un premier document général relatif à la méthodologie HUDERIA a été adopté par le Comité sur l'IA, le 28 novembre 2024⁶⁰. Cette méthodologie devrait être complétée d'ici fin 2025 par un modèle HUDERIA, document plus spécifique fournissant des

57. Sont néanmoins souvent cités en exemple l'outil néerlandais d'étude d'impact des algorithmes au regard des droits fondamentaux dit « IAMA », accessible à www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms, ainsi que l'outil d'impact des algorithmes développé par le gouvernement canadien, accessible à www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html. Pour autant ces deux outils au champ d'application limité aux algorithmes sont généralement jugés trop vagues et insuffisants. Pour une analyse doctrinale faisant un état des lieux des initiatives et des bonnes pratiques en matière d'études d'impact en vue de faire des propositions pour une application aux SIA, cf. les travaux précurseurs d'A. MANTELERO notamment *Beyond Data. Human Rights, Ethical and Social Impact Assessment in AI*, La Haye, Asser Press, 2022, ou encore A. MANTELERO et M. ESPOSITO, « An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems », *Comput. Law Secur. Rev.*, 2021, vol. 41, 105561, ou encore les travaux menés au sein de l'Informatics Institute de l'Université d'Amsterdam par V. SKORIC, G. SILENO et S. GHEBREAB, notamment « Roles of Standardised Criteria in Assessing Societal Impact of AI », in *2024 IEEE Conference on Artificial Intelligence (CAI)*, Singapour, 2024, pp. 1240-1245 ; *id.*, « Human Rights in AI Impact Assessment: Insights from Pilot Implementations, Normative Frameworks, and a Proposal », *International Public Policy Association*, 2023, accessible à www.ippapublicpolicy.org/file/paper/6476ec2fe5c7f.pdf ou encore V. SKORIC, « Critical Criteria for AI Impact Assessment: An Aggregated View », 2 juin 2023, disponible sur SSRN : <https://ssrn.com/abstract=4467502>.

58. Pour un inventaire des normes techniques susceptibles d'être concernées, cf. par ex. J. SOLER GARRIDO *e.a.*, *Analysis of the Preliminary AI Standardisation Work Plan in Support of the AI Act*, EUR 31518 EN, Luxembourg, Office des publications de l'Union européenne, 2023 ; J. SOLER GARRIDO *e.a.*, *AI Watch: Artificial Intelligence Standardisation Landscape Update*, EUR 31343 EN, Luxembourg, Office des publications de l'Union européenne, 2023.

59. D. LESLIE *e.a.*, *Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems: A Proposal*, The Alan Turing Institute, 2021, ci-après « étude de l'Alan Turing Institute ».

60. CAI, « Méthodologie pour l'évaluation des risques et des impacts des systèmes d'intelligence artificielle du point de vue des droits humains, de la démocratie et de l'État de droit » (ci-après « méthodologie HUDERIA »), CAI(2024)16rev2_FR, Strasbourg, 28 novembre 2024.

supports et des ressources permettant la mise en œuvre de cette méthodologie. La structure de la méthodologie adoptée en novembre 2024 s'articule autour de quatre blocs :

- **une analyse des risques contextuelle dite « COBRA »** (*Context-based Risk Analysis*). Cette COBRA repose sur quatre étapes : (1) un cadrage préliminaire, (2) une analyse de risque, (3) une cartographie des impacts potentiels, et (4) un triage des SIA permettant de se concentrer sur ceux qui posent des risques importants. Elle est aussi censée permettre de déterminer « si le SIA est une solution appropriée au problème envisagé »⁶¹ ;
- **un processus d'engagement des parties prenantes dit « SEP »** (*Stakeholder Engagement Process*), qui a vocation à permettre d'intégrer le point de vue des personnes susceptibles d'être affectées par le SIA ;
- **une évaluation des risques et des impacts dite « RIA »** (*Risk and Impact Assessment*) concernant les SIA considérés comme présentant des risques importants pour les droits humains, la démocratie et l'État de droit. Cette évaluation vise à établir l'ampleur d'un impact négatif potentiel et réel (gravité du préjudice potentiel), sa portée (estimation du nombre de personnes affectées, durée des effets), sa réversibilité (degré possible de réparabilité ou de restauration des droits des personnes concernées) et sa probabilité (vraisemblance que le risque se produise) ;
- **un plan d'atténuation dit « MP »** (*Mitigation Plan*) qui propose notamment une hiérarchisation des mesures d'atténuation en quatre niveaux : éviter, atténuer, restaurer, compenser⁶². La hiérarchisation implique qu'éviter et atténuer sont des mesures à prendre au début du cycle de vie du SIA, tandis que les autres niveaux correspondent plus à la phase de déploiement. Ces deux niveaux impliquent notamment la création ou à la mise à disposition de voies de recours pour les personnes affectées.

Sans entrer dans les détails de ces différents blocs, cette méthodologie appelle d'ores et déjà plusieurs commentaires. La lecture de ce texte d'une vingtaine de pages laisse le lecteur sur sa faim. Premièrement, s'il était clair que l'HUDERIA ne devait pas avoir d'effet contraignant, sa portée est d'emblée envisagée de manière ambivalente. L'HUDERIA se définit comme un « guide autonome » et « n'a pas vocation à être une aide à l'interprétation

61. Méthodologie HUDERIA, p. 6.

62. Méthodologie HUDERIA, p. 21.

de la Convention-Cadre [...] »⁶³. Affichée une absence de relation entre la Convention et l'HUDERIA peut paraître étrange, sachant que cet outil a vocation à être utilisée par les « acteurs publics et privés pour aider à l'identification et à la réponse à apporter aux risques et aux impacts sur les droits humains, la démocratie et l'État de droit tout au long du cycle de vie des systèmes d'IA »⁶⁴. L'HUDERIA se veut à la fois être utile à des acteurs non liés par la Convention qui y trouveraient un intérêt et non obligatoire pour les Parties. Ces dernières pourront déterminer si elles souhaitent appliquer la méthodologie HUDERIA ou un autre, voire faire une application combinée de plusieurs méthodologies. Autrement dit, suivre l'HUDERIA n'est pas une étape nécessaire pour mettre en œuvre l'article 16 de la Convention qui fixe aux Parties une obligation générale d'évaluer, de prévenir, et d'atténuer les risques posés par les SIA. Deuxièmement, le document fourni reste très général. Il s'apparente à un inventaire des critères et points d'attention à prendre en compte par les acteurs désireux de construire leur propre méthodologie. Ce texte ne propose donc pas une méthodologie clé en main. Il décrit un certain nombre d'étapes, à travers les quatre blocs précités. Les séquences proposées peuvent être mises en œuvre dans un ordre différent, voire réalisées de façon concomitante, ou partiellement. En outre, l'approche choisie est celle d'un processus de révision itératif⁶⁵ impliquant d'identifier à intervalles réguliers de nouveaux impacts potentiels demandant de reprendre et d'adapter les évaluations précédentes. Une certaine souplesse est inhérente au choix d'une analyse des risques fondée sur le contexte prenant en compte la complexité des activités du cycle de vie d'un SIA⁶⁶. Pour autant, à ce stade, l'HUDERIA ne va pas jusqu'à proposer clairement une méthodologie quantitative ou qualitative pour trier les SIA et se concentrer sur ceux qui posent des risques importants au regard des droits de l'homme, la démocratie et l'État de droit⁶⁷. Le document ne va pas aussi loin que les propositions faites par le centre Alan Turing. Ce dernier préconisait une sorte de *scoring*⁶⁸ reposant sur une grille de trois indicateurs de risques (risque interdit, risque majeur, risque modéré)⁶⁹ et de cinq niveaux de probabilité de préjudices (improbable, possible, probable, très

63. Méthodologie HUDERIA, p. 3.

64. Méthodologie HUDERIA, p. 3.

65. Méthodologie HUDERIA, p. 23.

66. Ainsi trois grandes catégories de facteurs de risques sont envisagées : les risques liés au contexte d'application du SIA, ceux liés au contexte de la conception et du développement, et enfin ceux liés au contexte du déploiement.

67. Il est uniquement précisé que « la formulation de méthodes quantitatives ou semi-quantitatives de calcul de risques, ou de procédures plus qualitatives ou fondées sur des règles » est envisageable. Cf. p. 11 du méthodologie HUDERIA, CAI(2024)16rev2_FR.

68. Cf. étude de l'Alan Turing Institute, p. 67.

69. Étude de l'Alan Turing Institute, pp. 64 et 208.

probable, non applicable)⁷⁰. De son côté, Alessandro Mantelero propose une méthodologie plus qualitative que quantitative où les indicateurs seraient étudiés selon une échelle allant de faible, moyen, élevé, très élevé assortie d'un code couleur pour chaque niveau⁷¹.

Pour autant, deux points méritent d'être soulignés. Premièrement, la méthodologie HUDERIA invite à intégrer dans cette cartographie les cas où l'utilisation d'un SIA peut avoir un impact positif en faisant « progresser les droits humains, y compris la promotion et la garantie de la non-discrimination »⁷². Autrement dit, la notion d'impact potentiel transcende celle des effets négatifs. Le document propose également différentes voies afin de mettre en balance les avantages sociaux, face aux risques de la conception ou du développement d'un SIA⁷³. Deuxièmement, HUDERIA comporte une dimension participative au sein du processus d'engagement des parties prenantes. Le SEP vise pour l'essentiel à éviter les biais, les discriminations, l'exploitation de vulnérabilités dans la conception et l'utilisation des SIA. Une fois les parties prenantes identifiées, il est préconisé aux utilisateurs d'HUDERIA une sorte d'examen introspectif sur leur positionnement face à ces parties prenantes et leur capacité à les prendre en considération⁷⁴. Des recommandations plus claires concernant la composition des équipes de développement, l'association de personnes extérieures susceptibles d'être des personnes affectées, voire des experts en droit humain à l'évaluation tout long du cycle de vie du SIA⁷⁵ auraient été les bienvenues.

Il résulte de ces premières considérations qu'appliquer HUDERIA ne devrait pas se réduire à une *check-list*, ou à un vague questionnaire à remplir. L'objectif affiché est de conduire les acteurs à modifier, voire à abandonner, certains éléments de leur projet en raison de leur impact potentiel sur les droits de l'homme, la démocratie et l'État de droit. La valeur ajoutée d'HUDERIA pourrait être de compléter l'approche par les normes techniques qui est proposée dans l'*AI Act*, et qui pourrait peiner à véritablement permettre une standardisation des droits fondamentaux⁷⁶. Il reste encore difficile à déterminer dans quelle mesure l'HUDERIA sera

70. Étude de l'Alan Turing Institute, p. 66.

71. A. MANTELERO, « The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template », *Computer Law and Security Review*, septembre 2024, vol. 54, p. 16.

72. Cf. p. 10 du méthodologie HUDERIA, CAI(2024)16rev2_FR.

73. Cf. question zéro p. 12 du méthodologie HUDERIA, CAI(2024)16rev2_FR.

74. Méthodologie HUDERIA, p. 14.

75. Cf. par ex. l'étude de l'Alan Turing Institute, pp. 215 et s.

76. Voy. contribution de L. FROMONT et A. VAN WAEYENBERGE dans cet ouvrage, p. 61.

utilisée dans le cadre de l'UE, même si la Commission européenne a participé à la négociation de ce document⁷⁷. Il n'est pas acquis que l'HUDERIA soit jugée pertinente par les acteurs de l'écosystème de l'IA, tant elle semble nécessiter l'accompagnement d'experts en matière de droits de l'homme, démocratie et État de droit pour la concrétiser. Pourtant, il est de leur intérêt d'intégrer le plus en amont possible selon une logique *by design* les risques au regard des droits de l'homme, de la démocratie et de l'État de droit. L'HUDERIA comporte intrinsèquement une dimension de sensibilisation aux risques pour les droits de l'homme, la démocratie et l'État de droit dont la valeur ajoutée pour les acteurs pourrait être de réduire les possibilités de contentieux *a posteriori*.

3. – Une adhésion espérée de nombreux États

L'un des objectifs de la Convention-cadre est de proposer un cadre susceptible d'être accepté par des États non européens. La réalisation de cet objectif a pesé sur les négociations et explique en grande partie le contenu final du texte. L'absence de réserve possible⁷⁸ constitue l'une des garanties obtenues malgré les compromis déjà évoqués. Une distinction est opérée entre les États tiers qui ont participé à la négociation⁷⁹ et les autres. Les premiers peuvent concourir à l'entrée en vigueur de la Convention. Les autres États tiers sont soumis à une procédure d'adhésion prévue à l'article 31 qui ne pourra être envisagée qu'une fois que la Convention sera en vigueur. L'entrée en vigueur de la Convention est subordonnée à sa ratification par cinq Parties dont au moins trois États membres du Conseil de l'Europe. En novembre 2024, la Convention-cadre avait été signée par dix États⁸⁰ et par l'Union européenne. L'élection de Donald Trump semble rendre plus improbable la ratification future de ce texte par les États-Unis. À ce stade, il est difficile de prédire dans quelle mesure la Convention sera en vigueur au moment de la pleine application de l'*AI Act*, soit en août 2026. De son côté, l'adhésion suppose une invitation prononcée par le Comité des ministres après avoir reçu l'assentiment unanime des Parties à la Convention-cadre

77. La Commission européenne a demandé l'autorisation de participer à la négociation de la méthodologie HUDERIA, cf. CONSEIL DE L'UE, « Note sur l'intention de la Commission européenne d'entamer les discussions sur une méthodologie pour l'évaluation des risques et de l'impact des SIA sur les droits de l'homme, la démocratie et l'État de droit », HUDERIA, Bruxelles, 21 juin 2024, 11364/24.

78. Cf. Art. 34, § 2, de la Convention-cadre.

79. Il s'agit de l'Argentine, l'Australie, le Canada, Costa Rica, Saint-Siège, Israël, Japon, Mexique, Pérou, États-Unis et Uruguay.

80. Il s'agit de l'Andorre, la Géorgie, l'Islande, le Monténégro, la Norvège, la République de Moldavie, le Royaume-Uni, Saint-Marin, les États-Unis d'Amérique, l'Israël.

et une décision prise à la majorité des deux tiers des voix exprimées des membres du Comité des ministres et l'unanimité des représentants des Parties ayant droit de siéger au Comité des ministres.

La future application territoriale de la Convention peut être adaptée par les États de deux façons. D'une manière traditionnelle, l'article 32 de la Convention permet aux États de désigner les parties de leur territoire ou reliées à son territoire dans lesquelles elle s'appliquera. En outre, l'article 33 prévoit une clause fédérale qui permet de prendre en compte la répartition des compétences entre la Fédération et les entités fédérées. Dans ce cas, lorsqu'en vertu de cette répartition l'État fédéral n'est pas compétent pour la mise en œuvre de certaines dispositions de la Convention, il s'engage à informer les états fédérés et à les encourager à prendre des mesures appropriées. Une telle clause fédérale s'inspire de celle prévue à l'article 41 de la Convention sur la cybercriminalité du Conseil de l'Europe. Elle relativise encore l'impact possible d'une éventuelle ratification par les États-Unis.

Pour conclure, l'adoption de la Convention-cadre du Conseil de l'Europe sur l'IA constitue en soi une étape importante puisqu'il s'agit du premier instrument international contraignant encadrant les risques de l'IA sur les droits de l'homme, la démocratie et l'État de droit. Pour autant, l'importance des choix laissés aux Parties rend difficilement appréciable la valeur ajoutée de ce texte. Cela questionne également la capacité du Conseil de l'Europe à s'affirmer comme une enceinte incontournable pour concevoir des outils permettant d'accompagner les acteurs publics et privés vers une meilleure prise en compte de l'impact des SIA sur les droits de l'homme, la démocratie et l'État de droit.