Analyse des principaux cadres supranationaux de régulation de l'intelligence artificielle

De l'éthique à la conformité

Projet d'étude par **Yannick Meneceur**, chef de l'unité du développement numérique au Conseil de l'Europe, chercheur associé à l'Institut des Hautes Études sur la Justice (IHEJ) et expert auprès de l'Observatoire Éthique et IA de l'Institut Sapiens – Version de travail du 31 mai 2021 (v.0.4)

Table des matières

Synthèse de l'étude 2

Introduction : L'intervention des régulateurs internationaux pour créer un cadre de confiance au développement des applications d'intelligence artificielle 5

I. Une démarche commune : aller au-delà des déclarations éthiques 7

II. Des travaux à replacer dans le cadre des mandats de chaque organisation 11

III. Un début de convergence pour la définition des systèmes d'IA 15

IV. Des mesures pour sécuriser le marché 19

V. Quels mécanismes de gouvernance de l'IA ? 23

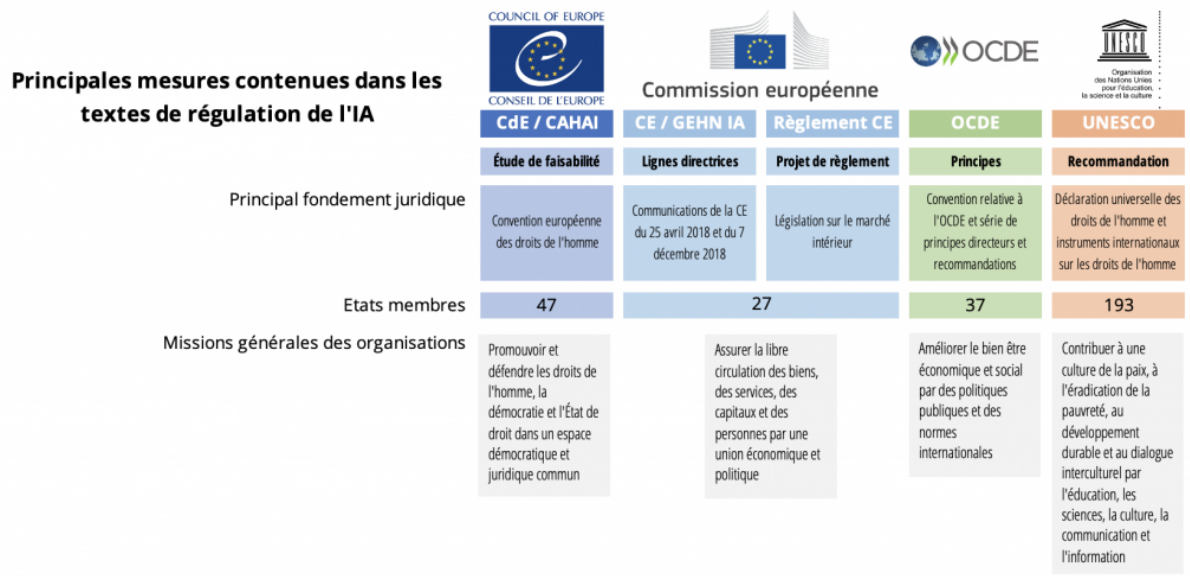
Conclusion : des textes ambitieux, dont l'effectivité reste à garantir 28

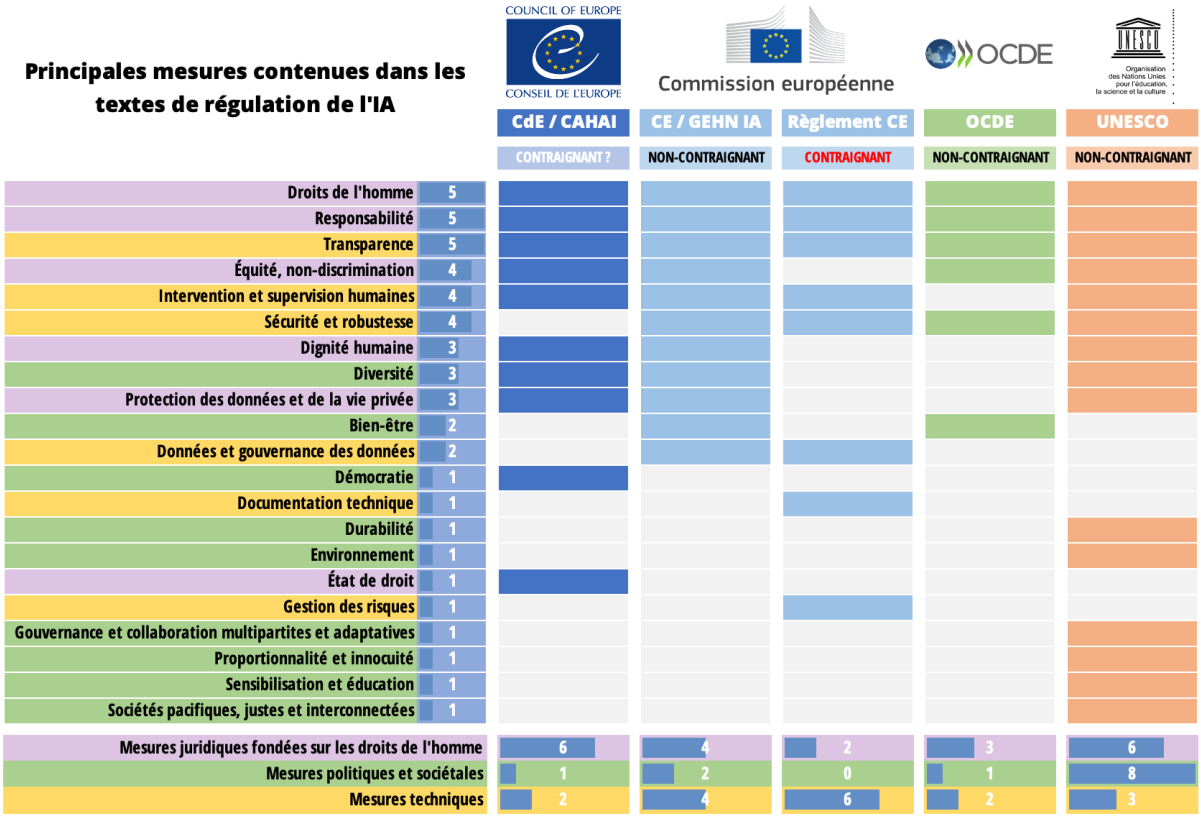
Annexe 31

La présente étude n'engage que son auteur et ne traduit aucune position officielle du Conseil de l'Europe.

L'auteur tient à remercier Bilel Benbouzid et Michèle Dubrocard pour leurs apports.

Synthèse de l'étude





Note méthodologique : les mesures listées correspondent aux principales thématiques développées dans chacun des documents. La simple citation d'une mesure n'a pas systématiquement conduit à son identification dans le tableau.

L'intervention des régulateurs internationaux pour encadrer le développement et l'application de l'intelligence artificielle vient en réponse à une inquiétude croissante dans l'opinion publique, confortée par la recherche, quant aux effets directs et indirects de cette technologie sur les droits des individus et la société. Les propositions de cadres éthiques n'ayant pas semblé apporter une réponse satisfaisante et convaincante, des organisations intergouvernementales telles que le Conseil de l'Europe, l'Union européenne, l'OCDE et l'UNESCO ont produit, sous l'impulsion de leurs États membres, de nombreux rapports, études, lignes directrices ou recommandations. Si ce qui pourrait être considéré comme du « droit souple » (*soft law*) présente une influence politique, technique et morale bien plus substantielle que de simples déclarations de bonne volonté des acteurs de l'IA, l'année 2021 marque toutefois un nouveau tournant, avec le premier texte juridiquement contraignant proposé par la Commission européenne en avril 2021 pour renforcer la sécurité des produits d'IA. Le Conseil de l'Europe envisage également un mélange d'instruments juridiques contraignants et non contraignants pour prévenir les violations des droits de l'homme et des atteintes à la démocratie et à l'État de droit.

Dans ce contexte de forte activité intergouvernementale, de nombreux observateurs, dont certains des États membres de ces organisations, se sont soucié de la cohérence de ces initiatives. Cette étude tend à démontrer que, même en l'absence de coordination formelle, la cohérence des initiatives les plus significatives sélectionnées (Conseil de l'Europe, Commission européenne – Union européenne, OCDE, UNESCO) apparaît tout d'abord résulter du mandat de chacune des organisations internationales. Des convergences apparaissent également sur le fond, s'agissant par exemple de la définition juridique des systèmes d'intelligence artificielle ou encore d'une approche fondée sur les risques, afin de ne pas imposer de trop fortes contraintes réglementaires à des dispositifs algorithmiques sans enjeux importants. La conformité (compliance) paraît ainsi se substituer progressivement à l'éthique, en annonçant trouver ses fondements dans les droits de l'homme et non dans des principes de philosophie morale ou un transfert de la bioéthique. De manière générale, les mécanismes juridiques semblent chercher à concilier innovation, sécurité et droits des personnes, dans l'intérêt de créer de la confiance pour le développement du marché des systèmes d'IA.

Des points de recouvrement ou de possibles divergences demeurent cependant. Si la réalisation d'études préalables de conformité fait consensus, leurs modalités ne font pas encore l'objet d'une normalisation. Les projets de textes ou les textes étudiés font aussi parfois apparaître des positions de compromis entre les États membres, notamment en Europe, avec des positions n'étant pas encore totalement stabilisées. Il doit être également relevé que des sujets controversés, comme l'emploi de la reconnaissance faciale dans l'espace public à des fins de sécurité publique[1], ne font pas l'objet d'un consensus avec la société civile, même si les préoccupations sécuritaires résultant de la menace terroriste et de la crise sanitaire laisse supposer une généralisation de ces dispositifs à plus ou moins court terme, encadrée par l'autorité judiciaire.

Enfin, les textes étudiés n'envisagent pas à ce stade de modalités procédurales spécifiques (pensons notamment à l'administration de la preuve ou des recours collectifs) contre des décisions prises/assistées par des algorithmes dans certains cas sensibles d'application et laissent encore apparaître la nécessité d'approfondir le débat démocratique au regard de la portée des enjeux de la transition numérique, notamment par des mécanismes inclusifs et participatifs allant plus loin que des consultations multipartites.

[1] Le terme d'identification biométrique à distance « en temps réel » a été préféré au terme de reconnaissance faciale dans la proposition de règlement sur l'IA de la Commission européenne.

Introduction : L'intervention des régulateurs internationaux pour créer un cadre de confiance au développement des applications d'intelligence artificielle

Un contexte de convergence technologique. – Alors que les applications de l'intelligence artificielle (IA), toujours plus diverses, se diffusent et s'installent dans notre quotidien, la question de leur régulation se pose de manière de plus en plus accrue. Ce développement s'insère de plus dans une dynamique globale de convergence avec d'autres technologies et d'accélération de la fourniture de services dématérialisés, au sein de laquelle la production et la valorisation de la donnée est devenue centrale. Avec le déploiement du réseau 5G, toutes les innovations technologiques semblent pouvoir se connecter, s'interconnecter et dialoguer pour faciliter nos vies, proposer de nouvelles fonctionnalités et développer l'industrie numérique.

Une inquiétude croissante dans l'opinion publique, confortée par la recherche. – Des documentaires et des reportages ont toutefois traduit avec une relative acuité les inquiétudes grandissantes de l'opinion publique face à cette transition numérique. La recherche, notamment en droit, en philosophie et en sciences sociales documente également, avec une meilleure précision, les effets concrets de cette transition sur les droits ou l'état des individus, ainsi que sur l'ensemble de la société. Des rapports parlementaires, tels que ceux du Parlement européen ou de l'Assemblée Parlementaire du Conseil de l'Europe (APCE), y font aussi écho. Les preuves s'accumulent donc pour démontrer pêle-mêle le renforcement des discriminations, les capacités de traçage et de ciblage des individus ou encore le renforcement de la fracture numérique. Les pouvoirs publics dans le monde, et particulièrement en Europe, soucieux de ne pas perdre rang dans la compétition économique et technologique, produisent en réponse des discours rassurants mêlant éducation (déploiement en France de conseillers numériques), persuasion douce par des « nudges » et discours volontaristes accompagnés d'un soutien économique substantiel. Des stratégies nationales sur l'IA traduisent déjà dans de très nombreux pays ces politiques publiques.

L'intervention des régulateurs internationaux pour créer de la confiance. – Les régulateurs internationaux ne pouvaient rester à l'écart de ce qui est présenté comme une 4ème révolution industrielle. La proposition de règlement de la Commission européenne, publiée le 21 avril 2021, a retenu toute l'attention des commentateurs, mais il ne faut pas ignorer que cette initiative s'inscrit dans une dynamique beaucoup plus vaste, mobilisant toutes les organisations intergouvernementales, régionales et internationales. Le États membres de ces organisations les incitent en effet, dans le respect de leur mandat, à accompagner cette transformation en contribuant à créer de la confiance dans les systèmes d'IA : confiance pour le développement et la sécurité des marchés (principalement Union européenne et OCDE) et confiance pour les droits des individus et le développement durable des sociétés (principalement Conseil de l'Europe et UNESCO). Depuis 2017, la production a été intense, d'abord sous la forme d'instruments juridiques non-contraignants (recommandations, déclarations ou lignes directrices), de chartes ou de rapports, largement inspirée des initiatives éthiques, puis aujourd'hui en instruments contraignants.

La complémentarité des travaux des organisations internationales. – Ces travaux pourraient laisser craindre une certaine forme de concurrence entre tous ces régulateurs. L'objectif de cette étude sera de clarifier les positions de quatre des principales organisations internationales (Conseil de l'Europe, Union européenne / Commission européenne, OCDE et Nations Unies / UNESCO), à la lumière de leurs mandats respectifs, afin de constater un grand nombre de convergences et de complémentarités. La question de leur effectivité sera posée en conclusion.

I. Une démarche commune : aller au-delà des déclarations éthiques

L'éthique comme réponse à l'inquiétude de l'opinion publique. – Depuis le milieu des années 2010, le développement des applications de l'IA a conduit à une intense production de cadres éthiques, notamment de la part du secteur privé, afin de répondre aux inquiétudes substantielles de l'opinion publique. L'enquête en 2016 de l'ONG ProPublica sur les discriminations résultant de la prédiction des risques de récidive à l'aide d'algorithmes aux États-Unis[2], l'affaire Cambridge Analytica en 2018 révélée par The Guardian démontrant l'exploitation de données détournées de Facebook pour de la propagande électorale[3], le succès de documentaires télévisuels tels que The Social Dilemma (*Derrière nos écrans de fumée*) ou iHuman (2020), sans parler du succès de livres comme Weapons of Maths Destruction (Cathy O'Neil, Crown Books, 2016) ou The Age of Surveillance Capitalism (Soshana Zuboff, PublicAffairs, 2019), voilà autant d'événements ou de productions qui ont constitué à forger une critique de l'application des algorithmes, dont les nouvelles techniques fondées sur les données depuis les années 2010 (apprentissage automatique - *machine learning*).

Des réponses éthiques critiquées. – En réponse, les grandes entreprises numériques, mais également une partie de la communauté académique, se sont engagées dans la constitution de cadres, non-juridiquement contraignants, inspirés de la bioéthique et de principes de philosophie morale. La biomédecine s'est en effet confrontée très tôt à toute une série de dilemmes moraux face au développement de technologies très complexes, autorisant par exemple la manipulation du génome. La philosophie morale, quant à elle, est venue apporter son cadre de référence et ses propres tensions internes, entre éthiciens de la vertu, déontologistes et conséquentialistes[4]. L'année 2018 a vu ainsi publier toute une série de documents, aux qualités extrêmement hétérogènes, visant à constituer des cadres d'actions pour les développeurs de solutions d'IA,

[2] J.Angwin, J.Larson, S.Mattu and L.Kirchner, Machine Bias – There's software used across the country to predict future criminals. And it's biased against blacks, ProPublica, 23 mai 2016 – Accessible sur : https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, consulté le 25 mai 2021

[3] Voir à ce titre le site internet du journal « The Guardian », The Cambridge Analytica Files, accessible sur : https://www.theguardian.com/news/series/cambridge-analytica-files, consulté le 25 mai 2021

[4] P. Gautrin, Conséquentialistes contre universalistes – une guerre de critères en éthique de l'IA ?, CS Science IA, 6 mars 2021 – Accessible sur : https://www.cscience.ca/2021/03/06/analyse-consequentialistes-contre-universalistes-une-guerre-de-criteres-en-ethique-de-lia/ , consulté le 21 mai 2021

énumérant des listes plus ou moins longues de principes de haut niveau[5]. Rodrigo Ochigame, ancien chercheur du MIT Media Lab, a publié un article en 2019 dans The Intercept explicitant, au moins en partie, le succès de cette dynamique : le souhait, pour les grandes entreprises technologiques, de différer toute prétention de régulation contraignante[6]. Sur le fond, Brent Mittelstadt, chercheur au Oxford Internet Institut, démontrait la même année les raisons pour lesquelles les principes éthiques de l'IA ne pouvaient constituer des fondements sérieux à une régulation effective[7]. Voilà quelques exemples qui ont contribué à documenter les insuffisances de l'éthique produite par les acteurs privés ou académiques et à écarter tout espoir d'autorégulation des acteurs, malgré un intense lobbying.

Un engagement des organisations internationales résultant de multiples impulsions. – Les organisations internationales ne pouvaient rester à l'écart de ce mouvement. Non pour « monter dans le train » de la régulation, stimulées par l'effervescence de l'époque, mais pour répondre aux obligations imposées par leurs mandats et à la demande de leurs États membres afin de contribuer à créer un cadre mondial de confiance à même de développer et tirer profit des technologies numériques. Il doit être relevé que cet engagement trouve également son impulsion dans l'activité parlementaire, puisque le Parlement européen a publié, entre la publication du livre blanc de la Commission européenne[8] et la proposition de règlement toute une série de rapports sur l'IA[9]. De même, le Conseil de l'Europe, qui partage son Assemblée Parlementaire avec l'OCDE, a publié dans la même période sept rapports thématiques sur l'IA, encourageant à l'édiction de règles contraignantes sur les applications de cette technologie[10]. Pour le Conseil de l'Europe, la Commissaire aux droits de l'homme avait également publié en mai 2019 une Recommandation mettant l'accent sur la nécessité d'études d'impact préalable sur les droits de l'homme pour l'emploi de

[5] Voir la base de données accessible en ligne des Temps Électriques, réunissant plus de 400 textes : https://lestempselectriques.net/index.php/2020/05/06/ia-algorithmes-big-data-data-science-inventaire-des-cadres-ethiques-et-politiques/, consulté le 25 mai 2021

[6] R. Ochigame, The invention of 'Ethical AI', The Intercept, 20 décembre 2019 – Accessible en ligne : https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/, consulté le 25 mai 2021

[7] B. Mittelstadt, Principles Alone Cannot Guarantee Ethical AI, Nature Machine Intelligence, novembre 2019 – Accessible en ligne : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3391293, consulté le 25 mai 2021

[8] Ce livre blanc préfigurait la proposition de règlement étudiée ici. Commission européenne, Livre blanc – Intelligence artificielle, une approche par l'excellence et la confiance, COM(2020)65final, 19 février 2020 – Accessible sur : https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_fr.pdf, consulté le 27 mai 2021

[9] Voir notamment la série de rapports publiés en octobre 2020, accessible sur : https://www.europarl.europa.eu/news/fr/press-room/20200925IPR87932/rendre-l-intelligence-artificielle-ethique-sure-et-innovante, consulté le 27 mai 2021

[10] Rapports publiés sur la page de l'APCE « Intelligence artificielle – Respecter démocratie, droits de l'homme et État de droit », accessible sur : https://pace.coe.int/fr/pages/artificial-intelligence, consulté le 27 mai 2021

l'IA dans le secteur public[11] ; le Comité des ministres s'était également positionné dans une Déclaration[12] et une Recommandation sur la nécessité de réguler les algorithmes afin de prévenir des violations des droits de l'homme[13]. Enfin il doit être souligné que tous les documents les plus importants ont fait l'objet de consultations externes dites multipartites, plus ou moins ouvertes, afin d'associer au plus près toutes les parties prenantes au travail de régulation. Ainsi la société civile, notamment des associations comme Access Now ou Algorithm Watch, le secteur privé, des associations professionnelles, comme l'IEEE ou l'ISO, des think-tanks, comme The Future Society, les milieux académiques ont été associés non seulement par le biais d'un statut d'observateur dans les groupes de travail pour certains textes, mais aussi par des questionnaires en ligne. Le grand public a été également associé, notamment à l'UNESCO par le biais d'une consultation mondiale sur l'intelligence artificielle pilotée par l'un des fondateurs de la Déclaration de Montréal, dont le processus de démocratie participative avait été remarqué.

Des travaux restant sous le contrôle politique des organisations internationales. – Alors que le processus de travail intergouvernemental s'appuie traditionnellement sur une expertise venant des États membres, avec l'appui d'experts externes venant le plus souvent des milieux académiques, l'on voit que le sujet de la régulation de l'intelligence artificielle s'est inscrit dans des démarches associant étroitement une variété plus importante d'acteurs, notamment les entrepreneurs et l'industrie. La Commission européenne a aussi réuni au sein d'une « AI Alliance » une communauté de 4000 participants invités à contribuer aux travaux. Cette évolution ne tend toutefois pas vers de la co-rédaction, puisque, qu'il s'agisse de la proposition de règlement de la Commission européenne, du Conseil de l'Europe (et son Comité ad hoc sur l'intelligence artificielle – CAHAI), de l'OCDE (et son Comité de la politique de l'économie numérique – CPEN) ou de l'UNESCO (et son groupe de 24 experts ad hoc – ad hoc experts group, AHEG qui soumettent leurs travaux en réunion intergouvernementale d'un Comité spécial dédié), le processus a été à chaque fois placé sous le contrôle des États membres. Seuls les travaux sur

[11] Commissaire aux droits de l'homme, Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme, mai 2019 – Accessible sur : https://rm.coe.int/decoder-l-intelligence-artificielle-10-mesures-pour-proteger-les-droit/168094b6e2, consulté le 27 mai 2021

[12] Comité des ministres du Conseil de l'Europe, Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques, Decl(13/02/2019)1 – Accessible sur : https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4c, consulté le 30 mai 2021

[13] Comité des ministres du Conseil de l'Europe, Recommandation aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme, CM/Rec(2020)1 – Accessible sur : https://search.coe.int/cm/pages/result_details.aspx?ObjectId=09000016809e1124, consulté le 30 mai 2021

les lignes directrices pour une IA digne de confiance ont été présentés comme indépendants par la Commission européenne.

II. Des travaux à replacer dans le cadre des mandats de chaque organisation

Les organisations internationales retenues pour l'analyse. – Nous retiendrons dans le cadre de cette étude quatre organisations pour la nature de leurs missions et le caractère extrêmement significatif de leurs productions : le Conseil de l'Europe, l'Union européenne et plus spécifiquement la Commission européenne, l'OCDE et une agence des Nations-Unies, l'UNESCO.

Les textes retenus pour l'analyse. – Nous sélectionnerons, par ailleurs, cinq textes spécifiques émanant de ces organisations parmi le nombre élevé de documents produits par elles (documents de politiques publiques, recommandations, principes, lignes directrices ou chartes, études ou rapports). Ainsi, seront étudiés: a) l'étude de faisabilité du Comité ad hoc sur l'intelligence artificielle du Conseil de l'Europe (CAHAI, décembre 2020), b) les lignes directrices éthiques pour une IA digne de confiance (Groupe d'Experts indépendants de Haut Niveau - GEHN mis en place par la Commission européenne, avril 2019), c) la proposition de règlement de l'intelligence artificielle (Commission européenne, avril 2021), d) la Recommandation du Conseil sur l'intelligence artificielle (OCDE, mai 2019) et e) le projet de Recommandation sur l'éthique de l'intelligence artificielle (UNESCO, avril 2021).

Des réponses cohérentes avec les mandats. – Le premier constat à dresser pour l'ensemble de ces textes est qu'ils sont naturellement tout à fait conformes aux mandats des organisations dont ils émanent. Ainsi, même s'ils traitent d'un objet commun, utilisent parfois une terminologie identique et emploient des experts[14] communs, leurs angles et leurs objectifs sont à chaque fois singuliers.

L'étude de faisabilité du CAHAI du Conseil de l'Europe. – Le Conseil de l'Europe s'inscrit ainsi dans sa longue expérience de régulation des technologies numériques, initiée dès 1981 avec la Convention 108, premier et (encore) seul texte international de protection des données (modernisé en 2018 avec un protocole additionnel). La directive du Parlement européen et du Conseil pour la

[14] Dans le langage des organisations internationales, la fonction d'expert fait référence aux représentants désignés par les États ou choisis par l'organisation elle-même, avec une expertise reconnue dans le domaine d'activité. Ces experts sont le plus souvent issus des rangs des gouvernements eux-mêmes ou des milieux académiques et universitaires. Plus rarement, ils peuvent également venir du secteur privé ou de la société civile.

protection des données de 1995 et le RGPD ont directement puisé dans les acquis de ce texte fondateur. Il doit d'ailleurs être précisé que le Conseil de l'Europe est la seule organisation internationale dont le mandat l'autorise à élaborer des traités internationaux (donc contraignants) sur les droits de l'homme, la démocratie et l'État de droit, auxquels d'autres États que ses 47 États membres peuvent adhérer. 55 États ont ainsi signé ou ratifié la Convention 108, 66 ont fait de même s'agissant de la Convention sur la Cybercriminalité. L'étude de faisabilité du CAHAI pose ainsi la première pierre d'un instrument juridique potentiellement contraignant (nous y reviendrons) fondé, notamment, sur la Convention européenne des droits de l'homme.

Les lignes directrices pour une IA digne de confiance et le projet de règlement de la Commission européenne (AI Act). – La Commission européenne, quant à elle, a produit deux documents centraux dont la nature et la portée diffèrent significativement. Les lignes directrices du GEHN sont le fruit d'un travail d'experts indépendants de la Commission, qui n'ont aucun caractère contraignant. Critiquées pour traduire des intérêts de l'industrie numérique[15], elles sont à lire avant tout comme des mesures à même de soutenir le développement d'un marché en Europe fondé sur la libre circulation des biens et des individus, mission principale de l'Union européenne. Même si les droits fondamentaux y sont cités, il doit être rappelé qu'un mémorandum d'accord lie l'Union au Conseil de l'Europe depuis 2007, selon lequel le Conseil de l'Europe reste pour le continent l'Organisation de référence sur les droits de l'homme. Il n'est donc pas surprenant que la proposition de règlement publiée en avril 2021, qui s'intègre dans un paquet comprenant également une Communication favorisant une approche européenne de l'intelligence artificielle (COM(2021) 205 final), un nouveau plan de coordination (révision du plan de 2018), ainsi qu'une proposition de règlement réformant la directive sur les machines (COM(2021) 202 final), soit essentiellement fondé sur l'article 114 du Traité de Fonctionnement de l'UE (TFUE) qui est relatif au marché intérieur[16]. Comme pour le RGPD, ce texte sera juridiquement contraignant à l'issue de la procédure législative.

La Recommandation du Conseil de l'OCDE sur l'intelligence artificielle. – La Recommandation du Conseil de l'OCDE sur l'intelligence artificielle, parue dans le

[15] T. Metzinger, Ethics washing made in Europe, Der Tagesspiegel, 8 avril 2018

[16] La proposition est également fondée sur l'article 16 du TFUE, puisqu'elle contient certaines règles spécifiques relatives à la protection des personnes physiques à l'égard du traitement des données à caractère personnel, notamment des restrictions à l'utilisation de systèmes d'IA pour la reconnaissance faciale dans des espaces accessibles au public à des fins répressive.

même temps que les lignes directrices du GEHN de la Commission européenne, est un instrument juridique intergouvernemental non contraignant qui vise à favoriser l'innovation et la confiance dans l'IA, traduit et adopté par le G20 dans la déclaration ministérielle pour le commerce et l'économie numérique. L'OCDE est spécifiquement mandaté pour améliorer le bien-être économique et social au sein de ses États membres, dans le prolongement - à sa création - du plan Marshall. Il doit être relevé que la Recommandation de l'OCDE aura servi non seulement à la Commission européenne pour définir l'IA (voir ci-dessous) mais également au Partenariat Mondial sur l'Intelligence Artificielle (PMIA ou GPAI - Global Partnership on Artificial Intelligence), équivalent du GIEC pour l'IA, pour fonder ses travaux. Comme pour la Commission européenne, le respect des droits de l'homme et des valeurs démocratiques y sont aussi cités comme objectif pour le développement d'une IA digne de confiance, avec un visa aux objectifs de développement durable des Nations-Unies et à la Déclaration Universelle des Droits de l'Homme de 1948. Même si elle revêt une nature d'instrument juridique, cette Recommandation ne créé aucune obligation juridique directe pour les États qui en font partie. Elle dispose en revanche d'une importante valeur politique et morale, enrichissant les dialogues multilatéraux entre les États concernés.

Le projet de Recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle. – Enfin, le projet de Recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle, rendue publique dans sa dernière version quasiment dans le même temps que le projet de règlement de la Commission, s'inscrit dans le cadre du mandat général de cette agence qui s'efforce de contribuer au maintien de la paix et de la sécurité par l'éducation, les sciences, la culture, la communication et l'information, afin d'assurer le respect universel de la justice, de la légalité, des droits de l'homme et des libertés fondamentales pour tous les peuples du monde. Le texte répond donc à de très larges préoccupations sociétales sans créer d'obligation juridique directe pour les États. La portée de ce texte, discuté par 193 États, est toutefois unique et a vocation à constituer une des références mondiales en matière d'éthique de l'IA. Il doit être précisé que ce texte entend également fournir des orientations éthiques à tous les acteurs de l'IA, y compris les secteurs public et privé, en fournissant une base pour une évaluation de l'impact éthique des systèmes d'IA tout au long de leur cycle de vie.

Les mesures communes à l'ensemble des documents étudiés. – Les droits de l'homme (terminologie du Conseil de l'Europe et des Nations-Unies) / droits

fondamentaux (terminologie de l'Union européenne, en référence à la Charte européenne des droits fondamentaux) sont présents dans tous les textes, et concrétisés dans des mesures comme la responsabilité, la dignité ou la non-discrimination. Il doit être à nouveau souligné ici que le Conseil de l'Europe est la seule des organisations étant susceptible d'élaborer un texte juridiquement contraignant dans le domaine des droits de l'homme, ouvert largement à la communauté internationale. Pour les textes européens, les mécanismes de responsabilité juridique de l'IA se trouvent également très largement influencés par l'harmonisation du droit de l'Union, notamment au travers de la sécurité des produits et de la directive relative à la responsabilité du fait des produits défectueux[17]. La diversité des systèmes de responsabilité dans le monde ne permet pas d'envisager l'émergence d'un modèle unique en la matière s'agissant de l'application de l'IA. Nous verrons enfin que la mesure portant sur la transparence des systèmes d'IA se traduit principalement dans des mécanismes d'évaluation des risques, sans pour autant converger encore vers des modalités communes.

[17] Directive 85/374/CEE du Conseil relative au rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de responsabilité du fait des produits défectueux

III. Un début de convergence pour la définition des systèmes d'IA

La question de la définition de l'IA. – Avant d'aller plus en avant dans l'analyse des mesures proposées, il convient d'étudier la question de la définition des systèmes d'IA, qui anime encore la communauté scientifique en l'absence de définition harmonisée à ce jour.

Un début de convergence. – En retenant la définition de l'OCDE pour sa proposition de règlement, il semble que la Commission européenne a effectué un pas vers un début de convergence pour définir, au moins juridiquement, ce que nous devrions entendre par ce terme entré dans le langage courant et interprété de manière plus ou moins large. Précisons également que l'ensemble des textes ne tombent pas dans l'écueil de l'emploi du terme IA seul, mais y adjoignent le terme de « système » pour éviter tout anthropomorphisme et qualifier l'ensemble de mécanismes et de composants concourant au fonctionnement de ces algorithmes particuliers. Cette approche a aussi l'avantage de distinguer la discipline (et la recherche dans ce domaine) des applications précises.

La définition du projet de règlement de la Commission européenne à rapprocher de celle de l'OCDE. – Ainsi, l'article 3(1) de la proposition de règlement définit un système d'intelligence artificielle 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[18] ». Dans sa Recommandation, l'OCDE définit un système d'IA comme « un système basé sur une machine qui peut, pour un ensemble donné d'objectifs définis par l'homme, faire des prédictions, des recommandations ou des décisions influençant des systèmes réels ou virtuels ».

[18] L'annexe I citée dans l'article 3(1) liste précisément les systèmes suivants : « (a) Les 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. »

La définition de l'UNESCO. – Sans reprendre aussi explicitement la définition de l'OCDE, l'UNESCO souligne également dans son projet de Recommandation les fonctions de prédiction et de prise de décision : « Les systèmes d'IA sont des technologies de traitement de l'information qui englobent des modèles et des algorithmes produisant une capacité d'apprentissage et d'exécution de tâches cognitives aboutissant à des résultats tels que la prédiction et la prise de décision dans des environnements matériels et virtuels ».

L'approche du Conseil de l'Europe, fondée sur les effets. – Le Conseil de l'Europe, pour sa part, ne procède pas encore à une rédaction précise mais appelle dans son étude de faisabilité à adopter une définition simplifiée et technologiquement neutre « couvrant les pratiques ou les cas d'application où le développement et l'utilisation de systèmes d'IA, ou plus généralement de systèmes automatisés de prise de décision, peuvent avoir un impact sur les droits de l'homme, la démocratie et l'État de droit », tout en tenant compte de toutes les implications socio-techniques des systèmes. Il doit être noté que plusieurs États membres ont soutenu dès la première réunion plénière du CAHAI l'adoption de la définition de l'OCDE, ce qui aurait l'avantage d'une certaine cohérence matérielle.

Des définitions allant bien au-delà de la prise de décision sur la base de modèles d'apprentissage automatique. – Parmi les éléments notables résultant de la juxtaposition de ces définitions, l'on voit que les régulateurs sont intéressés par des capacités précises des systèmes d'IA : la prise de décision, tant des environnements réels que virtuels. Le second élément notable tient aux techniques précises pour y parvenir et l'on voit, notamment à la lecture de l'annexe I de la proposition de règlement de la Commission, que les régulateurs ne se sont pas limités à la technologie ayant déclenché ce regain d'intérêt pour l'IA dans les années 2010 : l'apprentissage automatique (*machine learning*) et ses dérivés comme l'apprentissage profond (*deep learning*). Même si l'on retrouve la notion d'apprentissage dans ces documents, il est clair que les quatre organisations ont souhaité traiter de manière large les différentes pratiques, tant symboliques (utilisant un code formel signifiant pour établir les règles de traitement) que connexionnistes (employant des données pour induire les règles de traitement).

Un changement de paradigme pour la régulation du numérique. – Potentiellement, il s'agit donc pour l'ensemble des organisations internationales d'apporter un encadrement assez large pour les prises de décisions réalisées ou assistées des systèmes informatiques, même avec des techniques plus anciennes,

dès lors bien entendu qu'elles entrainent des conséquences significatives sur les individus. Loin d'être tout à fait technologiquement neutres (l'annexe I de la Commission européenne est très précise), ces cadres constituent une évolution très significative dans la doctrine d'intervention des régulateurs face aux technologies numériques. En effet, seules la protection des données et la lutte contre la cybercriminalité semblaient jusqu'à lors avoir mobilisé une partie de la communauté internationale sous la forme d'instruments juridiques contraignants. Le développement des réseaux et d'internet dans les années 1990 et 2000 n'avaient appelé aucune réponse structurée de la plupart des régulateurs et le « laissez-faire » a conduit à l'émergence de Big Techs comme Facebook et Google, qui cristallisent aujourd'hui nombre de critiques. Les inquiétudes dans l'opinion publique avec les développements des systèmes d'IA ont donc provoqué une réponse des organisations internationales et paraissent avoir servi à rouvrir le dossier de la régulation du numérique en général.

L'exception militaire. – Précisons que les textes étudiés ne concernent pas la régulation des applications militaires de l'intelligence artificielle, qui échappent notamment au mandat du Conseil de l'Europe[19]. C'est en effet le mécanisme de la Convention sur certaines armes classiques (CCAC), dans le cadre des Nations-Unies et des Conventions dites de Genève, auquel échoit la responsabilité d'établir des mesures, notamment pour les systèmes d'armes létaux autonomes (SALA). Toutefois, les discussions formelles paraissent actuellement peu progresser depuis l'appel du Secrétaire Général des Nations-Unies à les prohiber en 2019[20]. Le Parlement européen avait également appelé dans une Résolution du 12 septembre 2018 à mettre en place une stratégie de l'UE pour à les prohiber[21] et a adopté une résolution le 20 janvier 2021 traitant notamment des applications militaires de l'IA[22].

Une convergence de définition uniquement juridique. – Il conviendra de voir si les textes définitifs du Conseil de l'Europe, de la Commission européenne (AI act) et de l'UNESCO retiendront les approches que nous venons d'esquisser. Il paraît en toute hypothèse satisfaisant de constater un début de convergence des

[19] Article 1, d) du statut du Conseil de l'Europe

[20] Autonomous weapons that kill must be banned, insists UN chief, UN News, 25 March 2019 – Accessible sur : https://news.un.org/en/story/2019/03/1035381, consulté le 25 mai 2021

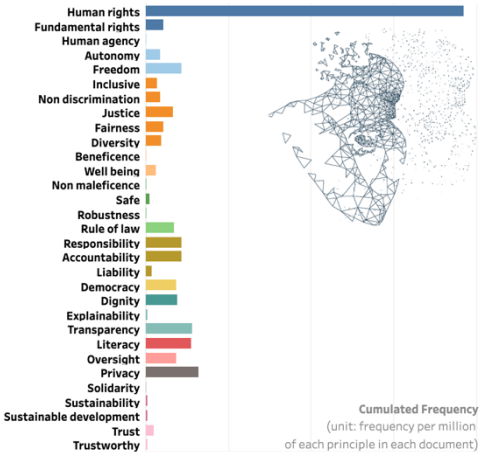
[21] Résolution du Parlement européen du 12 septembre 2018 sur les systèmes d'armes autonomes (2018/2752(RSP)) – Accessible sur : https://www.europarl.europa.eu/doceo/document/TA-8-2018-0341_EN.pdf, consulté le 25 mai 2021

[22] Résolution du Parlement européen du 20 janvier 2021 sur l'intelligence artificielle: questions relatives à l'interprétation et à l'application du droit international (2020/2013(INI)) – Accessible sur : https://www.europarl.europa.eu/doceo/document/TA-9-2021-0009_FR.html,consulté le 25 mai 2021

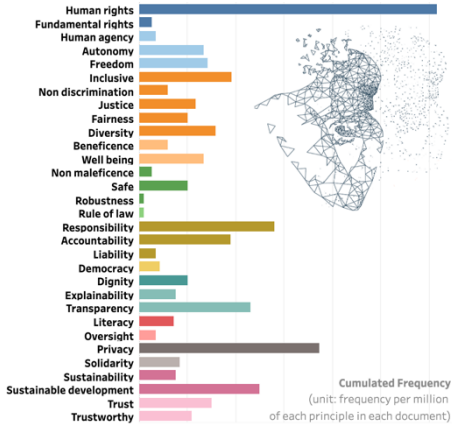
organisations internationales pour définir le domaine qu'elles entendent réguler. Cela pourra vraisemblablement également inspirer les législateurs nationaux et fonder un début de cohérence juridique entre les États membres dans le traitement des contentieux issus de l'application de l'IA. Il sera aussi intéressant de voir si cet effort des juristes aura des effets dans le champ technologique et scientifique afin de contribuer à mieux délimiter le champ de la discipline, aux frontières de l'informatique et des sciences cognitives.

IV. Des mesures pour sécuriser le marché

Créer de la confiance dans les systèmes d'IA. – La question de la confiance apparaît comme centrale dans l'ensemble des documents étudiés. Ainsi le projet de Recommandation de l'UNESCO rappelle que « Les gens devraient avoir de bonnes raisons de croire que les systèmes d'IA peuvent apporter des avantages individuels et partagés, tout en prenant des mesures adéquates pour atténuer les risques » (§12). Le Conseil de l'Europe fait aussi apparaître cette question, notamment dans les mesures relatives à la démocratie (« […] les États membres devraient garantir une approche participative significative et l'implication des différentes parties prenantes (de la société civile, du secteur privé, des universités et des médias) dans les processus décisionnels concernant le déploiement des systèmes d'IA dans le secteur public, en accordant une attention particulière à l'inclusion des individus et des groupes sous-représentés et vulnérables, ce qui est essentiel pour garantir la confiance dans la technologie et son acceptation par toutes les parties prenantes. » – §115).



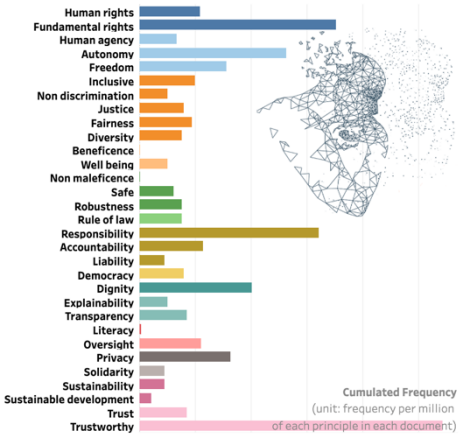
Conseil de l'Europe
Synthèse de 48 documents



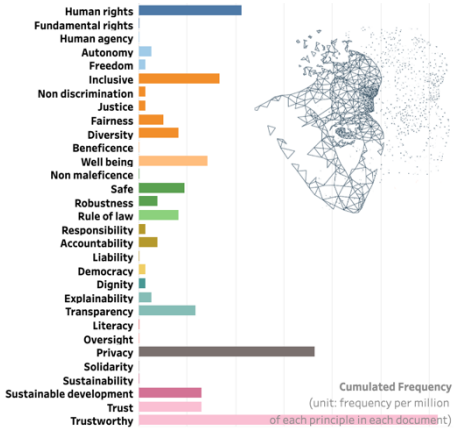
UNESCO
Synthèse de 11 documents

Créer de la confiance pour le développement du marché des systèmes d'IA. – Mais c'est dans les productions de la Commission européenne et de l'OCDE que l'on retrouve le plus fréquemment le concept d'IA digne de confiance (*Trustworthy AI*). Cette tendance est tout à fait confirmée par l'examen d'un corpus plus grand de

documents décomptant la fréquence d'apparition d'une série de concepts communs relatifs à l'IA[23].



Commission européenne
Synthèse de 25 documents



OCDE
Synthèse de 9 documents

Cette surreprésentation est intéressante à analyser au regard des mandats de l'Union européenne et de l'OCDE, qui sont des mandats économiques. L'on comprend dès lors pourquoi l'étude de faisabilité du CAHAI et le projet de Recommandation de l'UNESCO investissent la question de la confiance sous l'angle des droits des individus et de la minimisation des risques sur la société dans son ensemble (dont les risques démocratiques), alors que la Commission et l'OCDE cherche plutôt à établir des règles pour le développement du marché.

De l'éthique de l'IA à la conformité ? – Relevons par ailleurs que dans la proposition de la Commission le terme « compliance » (conformité en français) apparait à plus de 80 reprises, contre une quinzaine d'occurrences pour l'éthique. Cette tendance se retrouve également dans de moindres proportions dans l'étude de faisabilité du CAHAI, alors que l'éthique est centrale du projet de l'UNESCO. La Recommandation du Conseil de l'OCDE cite très peu l'éthique et la conformité, mais insiste sur la création d'un écosystème favorable à l'innovation et à la transition numérique. Lors de sa sortie en 2019, la question de mesures *ex-ante* de labellisation ou de certification n'avaient pas encore fait leur chemin et explique vraisemblablement cette prudence à énumérer des mesures précises. Ce constat permet de revenir sur une classification intuitive des documents étudiés, plaçant

[23] Voir la base de données accessible en ligne des Temps Électriques, réunissant plus de 400 textes : https://lestempselectriques.net/index.php/2020/05/06/ia-algorithmes-big-data-data-science-inventaire-des-cadres-ethiques-et-politiques/, consulté le 25 mai 2021

la Commission européenne et l'OCDE au centre de mesures assurant le développement économique, et le Conseil de l'Europe et l'UNESCO de mesures assurant la protection des individus et de la société. Il semble que le Conseil de l'Europe se soit aussi engagé, sur la base de son expérience de certificateurs de médicaments[24], sur l'idée de promouvoir des mécanismes ex-ante de conformité pour les algorithmes[25]. L'étude de faisabilité du CAHAI y fait d'ailleurs référence dès la première note de bas de page, afin de légitimer l'expérience du Conseil de l'Europe à réguler des technologies industrielles complexes. Il en résulte une position originale, où la prévention de violations des droits de l'homme (et d'affaiblissement de la démocratie et de l'État de droit) semblerait pouvoir être rendue effective par la définition de standards ne dénaturant pas la portée et la solennité des obligations en la matière. La convergence entre la proposition de règlement de la Commission et l'étude de faisabilité du Conseil de l'Europe, à laquelle s'ajoute une approche fondée sur le risque dans leurs deux textes, laisse à penser que la nature de l'intervention des deux organisations répond finalement à la même préoccupation : ne pas entraver le développement et l'innovation en proposant des outils et des mesures ayant déjà fait leur preuve dans des environnements industriels à forte compétitivité.

Prépondérance des mesures techniques et juridiques. – De manière générale, il doit être relevé une prépondérance des mesures à caractère technique ou juridique dans les documents étudiés, au détriment de mesures qui pourraient être qualifiées de sociétales ou politiques. Ainsi, l'on ne retrouve aucune mesure relative à la démocratie, l'environnement ou la durabilité parmi les six mesures les plus fréquentes dans l'ensemble des documents. Les trois mesures communes aux cinq documents (droits de l'homme, responsabilité et transparence) et les trois communes à quatre des cinq documents (équité/non-discrimination, intervention humaine/supervision et sécurité/robustesse) témoignent des principales préoccupations des régulateurs cherchant, encore une fois, à créer un cadre de confiance pour accélérer l'adoption de ces outils par le plus grand nombre (et par conséquent le développement du marché). Il faut dire que des

[24] La direction européenne de la qualité du médicament et des soins de santé (EDQM) dépend du Conseil de l'Europe et a pour mission d'œuvrer pour le droit humain fondamental que constitue l'accès à des médicaments et soins de santé de qualité, et de contribuer à la promotion et la protection de la santé humaine et animale. La Convention sur la pharmacopée, adoptée en 1964, est un instrument juridique contraignant dans les 39 États partis au Traité mettant en place le mécanisme d'élaboration de la pharmacopée européenne, composée de plus de 3000 textes. Ces textes décrivent notamment des normes qualité juridiquement contraignantes. Cette direction délivre à ce titre des certificats de conformité aux fabricants dont les substances pharmaceutiques satisfont aux normes.

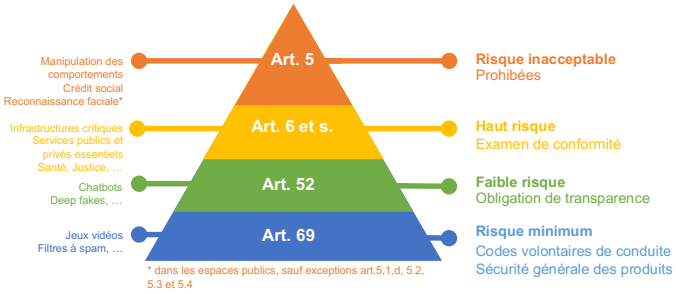
[25] L'idée se retrouve dès 2018 dans la Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, principe n°4 – Accessible sur : https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b,consulté le 25 mai 2021

notions comme la démocratie, par exemple, ne font pas l'objet d'un clair consensus entre les États de ces différentes organisations et il peut sembler délicat de chercher à établir des normes pour l'application de l'IA dans le domaine alors même que les notions fondamentales sont encore en discussion.

Soutenir la croissance économique et l'innovation numérique. – L'on comprend dès lors les raisons pour lesquelles les organisations internationales ont pu être mobilisées, dans un contexte de très forte compétition économique, et les motivations des acteurs. L'Europe (au sens large) entend définir des conditions de développement de l'IA conformes à ses valeurs et ainsi imposer aux autres acteurs internationaux (asiatiques et nord-américains) des conditions strictes d'accès à son marché par des instruments juridiques contraignants. Pour le reformuler, contraindre la Chine par exemple à ne pouvoir vendre ses instruments qu'à des régimes politiques moins stricts s'agissant des libertés individuelles. Sur le terrain économique, l'OCDE nourrit un riche dialogue transatlantique, dont la qualité s'est encore améliorée avec l'administration Biden. Les mesures de sa Recommandation du Conseil et le PMIA, dont l'axe franco-canadien est moteur, vont servir de passerelle naturelle entre les acteurs occidentaux. L'UNESCO, enfin, ancre sa démarche par une approche bien plus holistique que ses homologues. Le premier objectif du projet de Recommandation est d'ailleurs bien de fournir un cadre universel de valeurs, principes et actions aptes à inspirer les législateurs. La richesse des mesures proposées ne permet toutefois pas de lire avec clarté la hiérarchisation opérée, la protection des droits de l'homme semblant par exemple mise au même niveau que la protection de la biodiversité ou la diversité culture. Or les droits de l'homme ne sont bien entendu pas de l'éthique. En définitive, la préoccupation de l'ensemble de ces textes, partagée par tous les États membres, est surtout de ne pas ralentir la croissance économique avec le levier numérique (toujours aussi prometteur malgré les conséquences de la crise sanitaire) et de soutenir l'innovation numérique.

V. Quels mécanismes de gouvernance de l'IA ?

Une approche juridique fondée sur les risques pour l'Europe. – Nous avons déjà abordé les mécanismes de conformité envisagés tant par la Commission européenne que le Conseil de l'Europe, fondés sur l'appréciation des risques encourus. Il doit être précisé que l'approche fondée sur les risques de la Commission européenne est organisée en quatre niveaux, allant de risques inacceptables (et prohibant donc l'emploi de systèmes d'IA), à un risque minimum. Les applications à haut risque sont celles soumises à des procédures de conformité et celles à faible risque ne se voient imposer une obligation particulière de transparence (pour savoir par exemple que l'on interagit avec une machine ou qu'un contenu vidéo a été généré artificiellement). Il doit être précisé que ces dispositions ne sont en aucun cas exclusives les unes des autres[26]. En toute hypothèse, ce nouveau cadre juridique pour les systèmes d'IA entend être complémentaire des autres cadres juridiques existants et n'a donc pas vocation à entrer en concurrence avec le RGPD ou la sécurité générale des produits. En ce qui concerne la protection des données, seule les dispositions relatives à la création d'un « bac à sable règlementaire » ajoutent un fondement juridique supplémentaire[27]. Cette approche fondée sur les risques a été toutefois critiquée, notamment par la société civile, s'alarmant que des droits tels que les droits de l'homme ne pouvaient faire l'objet d'une modulation en fonction des probabilités de survenance d'un danger[28]. L'étude de faisabilité du Conseil de l'Europe n'est pas entrée dans ce niveau de détail en ce qui concerne la catégorisation des risques, mais elle expose aussi un certain nombre de mesures *ex-ante*, telles que les études d'impact préalables sur les droits de l'homme ou la certification, comme étant à même d'assurer l'effectivité des mesures proposées.



Dispositions non mutuellement exclusives

[26] Il doit être précisé, au vu du considérant (70) de la proposition de règlement et de l'article 52, que l'obligation de transparence s'applique aussi aux systèmes à haut risque, dès lors qu'ils interagissent avec des personnes physiques ou de génèrent du contenu pouvant présenter des risques spécifiques d'usurpation d'identité ou de tromperie.

[27] Voir considérant 72, page 34 de la proposition de règlement : « Le présent règlement fournit la base juridique de l'utilisation des données à caractère personnel collectées à d'autres fins pour le développement de certains systèmes d'IA dans l'intérêt public au sein du bac à sable réglementaire de l'IA, conformément à l'article 6, paragraphe 4, du règlement (UE) 2016/679, et à l'article 6 du règlement (UE) 2018/1725, et sans préjudice de l'article 4, paragraphe 2, de la directive (UE) 2016/680 ».

[28] The EU should regulate AI on the basis of rights, not risks, Access Now, 17 février 2021 – Accessible sur : https://www.accessnow.org/eu-regulation-ai-risk-based-approach/, consulté le 26 mai 2021

Cette étude ne tranche pas non plus sur la nature de l'instrument juridique à adopter par le Conseil de l'Europe : un certain nombre d'États membres paraissent en effet encore incertains ou méfiants face à la perspective de la création d'un nouvel instrument contraignant. Il doit être souligné que les Conventions du Conseil de l'Europe sont des traités internationaux, au sens de la Convention de Vienne des Nations-Unies de 1969, et elles doivent être signées et ratifiées avant de créer des effets juridiques : cela laisse donc la possibilité à certains États de rester en retrait, sans bloquer le processus pour d'autres.

Une portée contraignante encore incertaine pour l'instrument juridique du Conseil de l'Europe. – L'étude de faisabilité du CAHAI, analysée ici, vise à apporter au Comité des ministres du Conseil de l'Europe un éclairage sur l'opportunité et le contenu d'une possible Convention sur l'IA. Ce document a été adopté par le CAHAI en décembre 2020 et a été examiné par le Comité des ministres en mai 2021. Il résulte de cet examen l'annonce de l'ouverture de négociations d'ici à mai 2022 sur la base d'un premier projet de rédaction juridique produit par un groupe de travail spécifique (CAHAI-LFG, Legal Framework Group), après adoption par le comité plénier du CAHAI. Ce projet, qui visera à prévenir les violations des droits de l'homme, de la démocratie et de l'État de droit par des systèmes d'IA, ménagera toutefois plusieurs options car les 47 États membres du Conseil de l'Europe n'ont pas encore exprimé leur volonté définitive de s'engager dans un traité juridiquement contraignant. Très précisément, le Comité des ministres a annoncé en mai 2021 que le mécanisme de régulation de l'IA du Conseil de l'Europe pourrait « être composé d'un instrument juridique contraignant [...] ainsi que d'instruments supplémentaires contraignants ou non contraignants destinés à relever les défis liés à l'application de l'intelligence artificielle dans des secteurs spécifiques ».

Une forte influence politique et technique venant de l'OCDE. – La Recommandation du Conseil de l'OCDE, publiée en 2019, paraît bien plus élémentaire en énonçant un nombre restreint de principes à promouvoir et à implémenter pour garantir le développement d'une IA digne de confiance. Cet instrument juridique est donc simplement déclaratoire et paraît encourager le développement de politiques nationales et internationales en la matière en identifiant une série de « principes pour une gestion responsable de l'IA digne de confiance » tels que a) la croissance inclusive, le développement durable et le bien-être, b) les valeurs centrées sur l'humain et équité, c) la transparence et

l'explicabilité, d) la robustesse, la sûreté et la sécurité et enfin e) la responsabilité. L'influence de ce texte n'est toutefois pas à négliger puisque la Commission européenne (et peut-être aussi d'autres organisations) s'est appuyée sur la définition de l'IA de l'OCDE. Les travaux de l'OCDE, qui se prolongent dans un réseau d'experts (ONE.AI – OECD Network of Experts), vont aussi aboutir à un certain nombre de résultats très techniques, comme la classification de systèmes d'IA. L'Observatoire des politiques d'IA (OECD AI Policy Observatory[29]) mis en ligne en 2020 regroupe par ailleurs un nombre remarquable de documents et d'actualités, avec un moteur… d'IA en sélectionnant certains extraits en temps réel. Les travaux du PMIA (GPAI), dont l'OCDE assure le secrétariat, sont à aussi à examiner avec attention, ceux-ci ambitionnant de « combler le fossé entre la théorie et la pratique sur l'IA » en favorisant la coopération internationale[30]. 18 États sont représentés en son sein, en plus de l'Union européenne.

La significative portée morale du projet de l'UNESCO. – Le projet de Recommandation de l'UNESCO a, comme pour l'OCDE, une portée juridique limitée du fait de son caractère simplement déclaratoire. Il n'intégrera donc pas l'ordre juridique de ses 193 États membres comme un instrument contraignant mais constituera – ce qui est déjà substantiel – la plus large référence intergouvernementale mondiale en matière d'éthique de l'IA. Il est également important de souligner que ce texte place l'éthique au cœur de sa démarche, en définissant une série de valeurs (a) Respect, protection et promotion de la dignité humaine, des droits de l'homme et des libertés fondamentales, b) un environnement et des écosystèmes qui prospèrent, c) assurer la diversité et l'inclusion, d) vivre dans des sociétés pacifiques, justes et interdépendantes) et principes (a) proportionnalité et innocuité, b) sûreté et sécurité, c) équité et non-discrimination, d) durabilité, e) respect de la vie privée et protection des données, f) surveillance et décision humaines, g) transparence et explicabilité, h) responsabilité et redevabilité, i) sensibilisation et éducation, j) gouvernance et collaboration multipartites et adaptatives), rendus opérationnels par des actions stratégiques dans 11 domaines. L'on y retrouve notamment l'évaluation de l'impact éthique (EIE) tout le long du cycle de vie d'un système d'IA afin de mettre en évidence « les répercussions sur les droits de l'homme et les libertés fondamentales », confortant le principe européen d'une évaluation préalable des risques et d'une mise en conformité. Même s'il doit être relevé que les droits de

[29] Accessible sur : https://www.oecd.ai, consulté le 26 mai 2021
[30] Les travaux du PMIA sont accessibles sur : https://gpai.ai/fr/, consulté le 26 mai 2021

l'homme ne relèvent naturellement pas de l'éthique, au risque d'en diminuer la portée, le projet de Recommandation demeure le document le plus exhaustif de cette étude, notamment au travers de mesures à caractère sociétal recouvrant la culture, l'éducation, le travail, la santé. La mise en œuvre effective des dispositions énumérées ne bénéficie d'aucun mécanisme dédié, même si les mesures de suivi et d'évaluation comprennent un certain nombre d'offres de services comme l'élaboration d'une méthodologie d'évaluation de l'impact éthique mondialement reconnue, la mesure de l'évaluation de l'efficacité et de l'efficience des politiques relatives à l'éthique de l'IA ou l'analyse fondée sur la recherche et sur des éléments factuels des politiques relatives à l'éthique de l'IA ainsi que l'établissement de rapports à ce sujet, par exemple la publication d'un indice comparatif. L'UNESCO a d'ailleurs soutenu l'ouverture en Slovénie un centre de catégorie 2 dénommé « International Research Center on Artificial Intelligence » (IRCAI)[31]. Ce choix géographique confirme d'ailleurs le fort engagement de la Slovénie dans l'élaboration de politiques de régulation de l'IA, dont les experts se retrouvent en responsabilité dans la plupart des organisations internationales. Parmi leurs ambitions figure celle de créer un pont entre l'Union européenne et l'Afrique[32].

La complémentarité des textes produits par les organisations internationales. – La juxtaposition des mécanismes des cinq textes étudiés révèle donc plutôt une bonne distribution entre les organisations internationales, et ce même sans coordination formelle[33]. Si l'Europe s'est engagée dans une logique d'élaboration de textes juridiques plus contraignants, ceux-ci s'appuient assez notablement sur l'OCDE dans un objectif, nous l'avons vu, de sécurisation du marché et de croissance économique. Sans entrer en concurrence avec Bruxelles, le Conseil de l'Europe chemine vers des mesures s'appuyant sur ses acquis conventionnels (droits de l'homme et protection des données notamment), même si leur portée reste à discuter entre les États membres. En d'autres termes, l'Union européenne s'est placée dans un cadre de sécurité pour les produits d'IA, alors que le Conseil de l'Europe cherche à introduire des garanties telles que la définition de droits substantiels ainsi que des évaluations d'impact sur les droits de l'homme et la possibilité d'un recours effectif. Les mesures de l'UNESCO, qui auront en toute

[31] Les instituts et centres de catégorie 2 placés sous l'égide de l'UNESCO contribuent à la mise en œuvre des priorités et programmes de l'UNESCO, et des programmes de développement mondiaux, par la coopération internationale et régionale, la recherche, la production de connaissances, l'aide à la formulation de politiques et au renforcement des capacités.

[32] Voir la newsletter AI:Decoded du 26 mai 2021, accessible sur : https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-automated-disinformation-ai-treaty-negotiations-slovenia-loves-ai/, consulté le 27 mai 2021

[33] Un site internet, globalpolicy.ai, devrait être mis en ligne le 14 juin 2021 regroupant une coalition d'organisations internationales, afin d'exposer en un lieu unique leurs mesures relatives à la gouvernance de l'IA – Accessible sur : https://globalpolicy.ai

hypothèse une portée non-contraignante et morale, seront à même à influencer des États hors de la sphère d'influence occidentale avec une logique de minimisation des risques dès le début de la conception et tout au long du cycle de vie. Des points où les organisations se trouvent en concurrence demeurent, comme le modèle d'étude d'impact préalable, sur lequel travaillent notamment le Conseil de l'Europe, la Commission européenne et l'UNESCO. La question du cadre d'emploi de la reconnaissance faciale dans les espaces publics à des fins de prévention de la criminalité ne semble pas, à ce stade, faire l'objet d'une totale convergence entre le Conseil de l'Europe et la Commission européenne. Alors que les lignes directrices sur la reconnaissance faciale, écrites par le Comité consultatif de la Convention sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel du Conseil de l'Europe, se positionnent pour une certaine temporisation[34], la proposition de règlement de la Commission européenne en autorise l'utilisation sous réserve d'un certain nombre de règles strictes (autorisation judiciaire préalable ou, en cas d'urgence, contrôle a posteriori). En toute hypothèse, l'on voit se dessiner un possible mécanisme de gouvernance global, au sein duquel des mesures contraignantes seront associées à des mesures plus souples afin de chercher assurer un encadrement, rassurant le public, sans nuire à l'innovation et aux perspectives substantielles de croissance. La question de l'autorité coordonnant cette gouvernance sera en revanche tout autre.

[34] « [...] le recours aux technologies de reconnaissance faciale à la volée devrait être soumis à un débat démocratique comprenant la possibilité d'un moratoire en attendant une analyse complète du fait de leur nature intrusive pour la vie privée et la dignité des personnes, ajouté à un risque d'impact préjudiciable sur d'autres droits de l'Homme et libertés fondamentales. », in Lignes directrices sur la reconnaissance faciale, 28 janvier 2021, T-PD(2020)03rev4, accessible sur : https://rm.coe.int/lignes-direectrices-sur-la-reconnaissance-faciale/1680a134f4, consulté le 27 mai 2021

En conclusion : des textes ambitieux, dont l'effectivité reste à garantir

Une régulation s'insérant dans le plan d'action de coopération numérique des Nations-Unies. – La présente analyse a omis, par souci de concision, de faire référence à d'autres travaux internationaux relatifs à la gouvernance de l'IA, dont ceux des Nations-Unies avec le plan d'action de coopération numérique qui s'insère dans la réalisation des objectifs de développement durables[35]. Les cinq textes étudiés s'incorporent toutefois dans ce mécanisme global en cours de constitution et annoncent chercher à garantir, selon différentes modalités, le même niveau de protection des droits humains dans l'espace numérique que dans la vie réelle.

Des textes étant la traduction de compromis entre les États membres. – Même s'il ne faut pas minimiser l'important progrès résultant de chacune de ces initiatives, la réalité des mécanismes juridiques mis en œuvre (ou allant être mis en œuvre) reflète l'intensité des discussions et les compromis ayant dû être effectués, dans un contexte où les instruments multilatéraux peuvent encore être perçus par certains comme une atteinte à la souveraineté des États. De plus, des positions stratégiques bien différentes irriguent en fait les discussions européennes. L'Allemagne, qui a eu en 2020 et 2021 la présidence simultanée du Comité des ministres du Conseil de l'Europe et du Conseil de l'Union européenne, a fait usage de toute son influence pour que les autres États membres s'engagent dans une réglementation contraignante de l'IA, essentiellement dans l'objectif d'aligner tous les acteurs internationaux sur des règles définies et maîtrisées par l'Europe. C'est ainsi que la Commission européenne a pu produire ce projet de règlement et que le Conseil de l'Europe discute d'un possible traité transversal et généraliste, et non d'une simple recommandation. Précisions qu'un traité permet à des pays observateurs du Conseil de l'Europe, tels que les États-Unis, le Canada, le Mexique, le Japon et le Saint-Siège, de s'asseoir à la table de négociation alors qu'une recommandation ne concernerait que les seuls États membres du Conseil de l'Europe. Toutefois, un groupe de pays (notamment le Royaume-Uni, les Pays-Bas, la Suède et la Lettonie) paraissent souahiter une approche bien plus prudente, arguant que la réglementation pourrait entraver l'innovation et notre

[35] Plan d'action accessible sur : https://www.un.org/techenvoy/fr/content/roadmap-digital-cooperation, consulté le 25 mai 2021

compétitivité face à des pays comme la Chine[36]. Ces tensions se retrouvent également à Bruxelles, puisque les « Digital 9 » (et même Digital 9+ avec 14 États membres de l'Union européenne - ou plus exactement 14 ministères économiques ou en charge du numérique) avaient adressé en octobre 2020 une note blanche dans laquelle ils mettaient en garde Bruxelles contre « la mise en place de barrières et d'exigences lourdes qui peuvent constituer un obstacle à l'innovation[37] ».

Pas de moratoire ni de bannissement d'applications de l'IA à des fins de sécurité. – C'est également dans ce contexte que des sujets sensibles comme l'utilisation de l'IA à des fins répressives ou de sécurité semblent avoir bénéficié d'une certaine bienveillance des régulateurs malgré les alertes soulevées par la société civile et les milieux académiques[38]. Les textes étudiés paraissent donc converger pour soutenir la croissance économique et l'innovation numérique, en traduisant certains principes éthiques vers des normes juridiques ou en s'appuyant sur des normes juridiques existantes en vue de les compléter (comme celles relatives à la protection des données), tout en ne s'opposant pas à des applications à caractère répressif ou sécuritaire, pourtant polémiques. N'oublions pas que l'Union européenne elle-même a soutenu la mise en œuvre des applications de l'IA pour le contrôle aux frontières[39].

Un approfondissement nécessaire des modalités de recours et du débat démocratique. – À ce stade de rédaction des différentes initiatives se pose donc la question de leur effectivité : les tribunaux pourront-ils traiter de manière plus aisée des contentieux extrêmement techniques mêlant des algorithmes d'IA à d'autres processus industriels ou commerciaux ? Pourra-t-on contester facilement les mécanismes de prise de décision (ou d'aide à la prise de décision), en facilitant si nécessaire les actions de groupe[40] ? Avec quel régime de preuve ? Sera-t-on en mesure de revenir sur des « innovations », comme la reconnaissance faciale dans les espaces publics à des fins de lutte contre la criminalité, si des abus sont

[36] Voir la newsletter AI:Decoded du 26 mai 2021, accessible sur : https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-automated-disinformation-ai-treaty-negotiations-slovenia-loves-ai/, consulté le 27 mai 2021

[37] Note accessible sur : https://em.dk/media/13914/non-paper-innovative-and-trustworthy-ai-two-side-of-the-same-coin.pdf, consulté le 27 mai 2021

[38] L'utilisation de polygraphes, par exemple, ne fait pas l'objet d'un consensus en matière pénale pour détecter les mensonges et n'est pas admise dans certains pays comme la France. Voir notamment J. Bittle, Lie detectors have always been suspect. AI has made the problem worse. An in-depth investigation into artificial-intelligence-based attempts to recognize deception, MIT Technology Review, 13 mars 2020 – Accessible sur : https://www.technologyreview.com/2020/03/13/905323/ai-lie-detectors-polygraph-silent-talker-iborderctrl-converus-neuroid/, consulté le 30 mai 2021

[39] Projet iBorderCtrl, accessible sur : https://www.iborderctrl.eu/, consulté le 27 mai 2021

[40] M. Tisne, Collective data rights can stop big tech from obliterating privacy, MIT Technology Review, 25 mai 2021 – Accessible sur : https://www.technologyreview.com/2021/05/25/1025297/collective-data-rights-big-tech-privacy/, consulté le 31 mai 2021

constatés ? Ne faut-il bannir une liste plus large d'applications, sans base scientifique sérieuses ? Voilà quelques-uns des sujets de société qui mériteraient un très large débat démocratique, inclusif et participatif, éclairé par des contributions scientifiques multidisciplinaires vulgarisées et de haut niveau, pour déterminer, collectivement, le type de société dans laquelle nous souhaitons vivre. Mais la succession de crises, avec des séries d'attentats ou la pandémie de la COVID-19, laisse supposer un fort désir de protection de la part des populations et donc l'acceptation d'un modèle sécuritaire… modèle pour lequel nous ne devrions pourtant pas abandonner aucune de nos garanties fondamentales, âprement gagnées à l'issue de la Seconde Guerre mondiale.

Annexe

Liste des documents étudiés

- Étude de faisabilité du Comité ad hoc sur l'intelligence artificielle (Conseil de l'Europe)
- Lignes directrices éthiques pour une IA digne de confiance (Groupe d'experts indépendants de haut niveau, Commission européenne)
- Proposition de règlement de l'intelligence artificielle (Commission européenne)
- Recommandation du Conseil sur l'intelligence artificielle (OCDE)
- Projet de Recommandation sur l'éthique de l'intelligence artificielle (UNESCO)

Sites internet sur l'IA des organisations internationales

- Site commun aux organisations internationales : https://globalpolicy.ai
- Conseil de l'Europe : https://www.coe.int/ai
- Groupe d'experts indépendants de haut niveau, Commission européenne : https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai
- Commission européenne : https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
- OCDE : https://oecd.ai
- UNESCO : https://fr.unesco.org/artificial-intelligence