

VERS UNE RÉGULATION DES SYSTÈMES D'IA

Perspectives internationales sur l'élaboration
d'un cadre juridique fondé 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



Compilation de contributions
DGI (2020)16

Préparée par
le Secrétariat du CAHAI

VERS UNE REGULATION DES SYSTEMES D'IA

Perspectives internationales sur l'élaboration d'un
cadre juridique pour les systèmes d'Intelligence Artificielle (IA)
fondé 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

Compilation de contributions
préparée par le Secrétariat du CAHAI
Décembre 2020

Cette publication a été financée par
une contribution volontaire du Japon

Auteurs:

Isaac BEN-ISRAEL
Jorge CERDIO
Arisa EMA
Leehe FRIEDMAN
Marcelo IENCA
Alessandro MANTELERO
Eviatar MATANIA
Cathelijne MULLER
Hideaki SHIROYAMA
Effy VAYENA

Etude de Conseil de l'Europe
DGI (2020)16

Les opinions exprimées dans ce document relèvent de la responsabilité des auteurs et ne reflètent pas nécessairement la ligne officielle du Conseil de l'Europe.

Toute demande concernant la reproduction ou la traduction de tout ou partie de ce document doit être adressée à la Direction de la Communication (F-67075 Strasbourg Cedex ou publication@coe.int). Toute autre correspondance concernant ce document doit être adressée à la Direction générale Droits de l'homme et État de droit.

Mises en page et page de garde:
Département de la société de l'information,
Conseil de l'Europe,

Images: Shutterstock

Cette publication n'a pas été révisée par l'unité éditoriale du SPDP pour corriger les erreurs typographiques et grammaticales.

@ Conseil de l'Europe, décembre 2020

Table des matières

| | |
|---|-----------|
| MOTS D'OUVERTURE | 5 |
| SOMMAIRE | 7 |
| | |
| TITRE I. PERSPECTIVE INTERNATIONALE DE LA REGULATION DE L'IA BASEE SUR LES STANDARDS DU CONSEIL DE L'EUROPE..... | 8 |
| | |
| CHAPITRE PRELIMINAIRE. Rapport d'état d'avancement des travaux du Comité Ad Hoc sur l'intelligence artificielle (CAHAI)..... | 8 |
| I. Résumé | 8 |
| II. Introduction. Contexte et mandat du CAHAI | 9 |
| III. État d'avancement des travaux..... | 10 |
| IV. L'impact de la pandémie de COVID-19 sur les activités du CAHAI | 12 |
| V. Méthodes de travail, documents et feuille de route | 13 |
| VI. Synergie et complémentarité du travail du CAHAI avec celui d'autres organisations internationales | 14 |
| VII. Implications budgétaires | 14 |
| VIII. Conclusions et propositions concrètes d'action..... | 14 |
| Annexe I. Projet de table des matières de l'étude de faisabilité..... | 16 |
| Annexe II. Feuille de route du CAHAI | 18 |
| Annexe III. Mandats des Groupes de travail | 19 |
| | |
| CHAPITRE I. Impact de l'intelligence artificielle sur les droits de l'homme, la démocratie et l'État de droit..... | 22 |
| I. Remarques liminaires | 22 |
| II. L'intelligence artificielle : définition | 22 |
| i. Définition de l'IA à des fins juridiques..... | 24 |
| III. Impact de l'IA sur les droits de l'Homme, la démocratie et l'État de droit | 24 |
| i. IA et respect de l'être humain | 24 |
| ii. IA et liberté individuelle..... | 26 |
| iii. IA et égalité, non-discrimination et solidarité | 28 |
| iv. IA et droits sociaux et économiques | 28 |
| v. IA et démocratie..... | 29 |
| vi. IA et État de droit | 31 |

| | | |
|------|--|----|
| IV. | Que faire face à l'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit ?..... | 31 |
| i. | Les droits de l'homme dans le contexte de l'IA | 31 |
| ii. | Mesures visant la conformité, la réparation et l'obligation de rendre des comptes.... | 32 |
| iii. | Protection des structures démocratiques et de l'État de droit | 33 |
| V. | Et si les normes actuelles de droits de l'homme, de démocratie et d'État de droit ne nous protégeaient pas assez ? | 34 |
| i. | Question zéro | 34 |
| ii. | Lignes rouges | 34 |
| iii. | Des droits de l'homme nouveaux ou adaptés | 35 |
| iv. | Scénarios d'avenir | 35 |

CHAPITRE II. Lignes directrices sur l'éthique en matière d'IA : Situation en Europe et dans le monde..... 36

| | | |
|------|---|----|
| I. | Résumé | 37 |
| II. | Principaux constats..... | 37 |
| III. | Principales incidences sur l'élaboration des politiques : | 38 |
| IV. | Introduction..... | 39 |
| V. | Méthodologie | 40 |
| i. | Criblage..... | 40 |
| ii. | Évaluation de la pertinence | 41 |
| iii. | Analyse du contenu | 44 |
| iv. | Analyse de l'éthique normative et des politiques..... | 44 |
| VI. | Constatations..... | 45 |
| i. | Insuffisances | 52 |
| ii. | Débat et analyse de l'éthique normative | 53 |
| iii. | Incidences sur l'élaboration des politiques | 56 |
| | Remerciements | 58 |
| | Bibliographie..... | 58 |

CHAPITRE III. Analyse des instruments internationaux juridiquement contraignants..... 60

| | | |
|------|--|----|
| I. | Résumé | 60 |
| II. | Méthodologie | 62 |
| i. | Le scénario | 63 |
| ii. | Axe de recherche et méthodologie | 63 |
| iii. | Analyse et résultats escomptés | 64 |

| | | |
|------|---|-----|
| III. | Analyse..... | 67 |
| | i. Vue d'ensemble | 68 |
| | ii. Protection des données..... | 70 |
| | iii. Santé..... | 74 |
| | iv. Démocratie..... | 78 |
| | v. Justice..... | 84 |
| IV. | Harmonisation des principes identifiés..... | 90 |
| V. | Conclusions | 91 |
| | Bibliographie..... | 92 |
| | Annexe 1. Instruments juridiques..... | 99 |
| | Annexe 2. Domaines concernés | 105 |
| | Annexe 3. Principes..... | 110 |
| | Annexe 4. Protection des données | 116 |

TITRE II. PERSPECTIVES NATIONALES DE LA REGULATION DE L'IA 122

CHAPITRE I. Exploiter l'innovation : Perspectives israéliennes sur l'éthique et la gouvernance de l'IA..... 122

| | | |
|------|--|-----|
| I. | Résumé | 122 |
| | Remerciements | 123 |
| II. | Introduction..... | 123 |
| | L'initiative nationale israélienne pour des systèmes intelligents sécurisés..... | 123 |
| III. | Les applications de l'IA en Israël - une opportunité pour les politiques publiques..... | 125 |
| | i. Le secteur privé | 125 |
| | ii. Initiatives et politiques du gouvernement | 127 |
| IV. | Risques et défis posés par l'IA dans les domaines des droits de l'Homme, de la démocratie et de l'État de droit | 132 |
| | i. Quelles sont les nouveautés et les particularités de l'IA ? | 132 |
| | ii. Risques et défis éthiques..... | 133 |
| V. | L'approche d'Israël pour relever les défis..... | 138 |
| | i. Six principes éthiques pour l'IA..... | 138 |
| | ii. Une réglementation équilibrée pour favoriser l'innovation..... | 139 |
| | iii. Outil original d'évaluation du risque éthique..... | 143 |
| | iv. Activité et coopération internationales..... | 144 |
| | Annexe I. Le paysage israélien des start-ups d'IA segmenté par secteurs et applications . | 145 |
| | Annexe II. Carte de fréquence des défis éthiques dans le processus de développement de l'IA | 146 |

| | |
|--|------------|
| CHAPITRE II. La gouvernance de l'IA au Japon | 151 |
| I. Introduction..... | 151 |
| II. Gouvernance de l'IA au Japon..... | 152 |
| i. Le rôle de chaque acteur | 152 |
| ii. Discussions sur l'IA dans les ministères et les agences | 157 |
| iii. Résultats | 159 |
| III. Comparaison de la gouvernance de l'IA au Japon et à l'étranger | 166 |
| i. Comparaison des rôles des différents acteurs | 166 |
| ii. Tendances et problèmes dans les technologies et les domaines ciblés de l'IA..... | 168 |
| iii. Comment créer un forum de discussion et ses enjeux | 170 |
| IV. Conclusion..... | 171 |
| Remerciements | 171 |
| Annexes..... | 172 |

| | |
|--|------------|
| CHAPITRE III. Applications de l'AI au Mexique. Une vue de l'intérieur | 180 |
| I. Introduction..... | 180 |
| II. Observations des applications publiques et privées de l'IA au Mexique | 181 |
| i. Les applications d'IA d'initiatives publiques au Mexique | 182 |
| ii. Les applications de l'IA d'initiatives privées au Mexique | 187 |
| III. Comptabilisation de la (dé)réglementation de l'IA au Mexique | 193 |
| i. Cadre juridique et politiques publiques concernant l'IA | 194 |
| ii. Un cadre collectif concernant la réglementation de l'IA..... | 198 |
| IV. Conclusions | 199 |
| Références | 199 |

MOTS D'OUVERTURE



Claudia Luciani

Directrice – Direction de la dignité humaine, de l'égalité et de la gouvernance, Conseil de l'Europe



Jan Kleijssen

Directeur – direction de la société de l'information et de la lutte contre la criminalité, Conseil de l'Europe

Les systèmes d'intelligence artificielle (IA) sont de plus en plus utilisés et ce dans presque tous les types d'activités humaines. Les bénéfices de cette technologie sont reconnus et font déjà parti de notre quotidien. Dans le cadre même de la lutte contre la Covid-19, de nombreuses applications ont été déployés pour accélérer la recherche, mieux détecter les cas et mesurer la pandémie. Cependant, l'essor de cette technologie soulève toutefois des inquiétudes dans l'opinion publique et il revient aux Etats d'assurer que, dans cette nouvelle ère technologique, les droits de l'homme, à la démocratie et à l'Etat de droit continuent d'être pleinement protégés.

Le Conseil de l'Europe est la principale organisation de défense des droits de l'homme du continent qui compte 47 États membres et il protège et promeut les droits de l'homme, la démocratie et l'État de droit au niveau paneuropéen. Le Conseil de l'Europe a donc une claire vocation à traiter la question du développement et de l'utilisation de l'intelligence artificielle. L'organisation a d'ailleurs déjà produit des normes mondiales pionnières et juridiquement contraignantes dans des environnements technologiques complexes, comme pour les données à caractère personnel, la bioéthique ou la cybercriminalité, en conciliant la protection des individus et l'innovation. En outre, le Conseil de l'Europe a élaboré de nombreux instruments juridiques traitant de l'impact des systèmes d'IA sur les droits de l'homme, la démocratie et l'État.

Une étape supplémentaire a été franchie avec la création du Comité ad hoc sur l'intelligence artificielle (CAHAI), qui est chargé d'examiner, sur la base de larges consultations multipartites, la faisabilité et les éléments potentiels d'un cadre juridique pour le

développement, la conception et l'application de l'intelligence artificielle, fondée des normes du Conseil de l'Europe dans le domaine des droits de l'homme, de la démocratie et de l'État de droit. Le Comité rassemble de façon unique des États membres et des États observateurs, ainsi que des observateurs de la société civile, du monde universitaire et du secteur privé. Convaincu de l'importance d'une réflexion globale et de d'un effort collectif dans ce domaine, le CAHAI travaille en étroite coopération avec d'autres institutions internationales, telles que l'UNESCO, l'OCDE et l'Union Européenne.

Cette publication vise à soutenir les réflexions en cours au sein du CAHAI sur l'analyse des défis posés par les systèmes d'IA et les réponses réglementaires envisageables. Tout d'abord, elle vise à informer le lecteur de l'avancement des travaux du CAHAI et elle présente plusieurs études produites dans le cadre du CAHAI sur l'impact des systèmes d'IA sur les droits de l'homme, l'État de droit et la démocratie, ainsi que sur les instruments internationaux juridiquement contraignants et les directives éthiques existants en matière d'IA. Deuxièmement, elle apporte des perspectives nationales de différents États afin de soutenir le développement d'un cadre juridique international portant sur certaines utilisations de l'intelligence artificielle basé sur les normes établies par le Conseil de l'Europe.

Nous sommes convaincus que cette publication apportera une perspective précieuse sur le débat actuel sur la réglementation des systèmes d'IA, et mettra en évidence la contribution unique que le Conseil de l'Europe et le CAHAI apportent.

SOMMAIRE

TITRE I. PERSPECTIVE INTERNATIONALE

Le chapitre préliminaire présente ce rapport soumis par le CAHAI au Comité des Ministres et détaille les progrès réalisés à ce jour, en tenant compte de l'impact des mesures de lutte contre la pandémie de la COVID-19. Il comprend également des réflexions sur les méthodes de travail, la synergie et la complémentarité avec d'autres parties prenantes concernées et des propositions d'actions futures de la CAHAI au moyen d'une feuille de route solide et claire.

Chapitre 1 décrit l'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit. Il identifie les droits de l'homme, tels qu'énoncés dans la Convention européenne des droits de l'homme (« CEDH »), ses Protocoles et la Charte sociale européenne (« CSE »), qui sont actuellement les plus touchés par l'IA ou les plus exposés.

Chapitre 2 analyse le corpus de documents de droit souple et autres cadres éthico-juridiques définis par des institutions gouvernementales et des entités non gouvernementales du monde entier, et ce à deux fins. Tout d'abord, il étudie l'ensemble d'instruments de gouvernance non contraignants qui est en constante évolution. Ensuite, il analyse les incidences que l'IA pourrait avoir sur les principes éthiques, les droits de l'homme, l'État de droit et la démocratie.

Chapitre 3 a pour objectif de contribuer à l'élaboration de la future réglementation de l'IA en s'appuyant sur les instruments contraignants existants, en contextualisant leurs principes et en fournissant les principales orientations réglementaires d'un futur cadre juridique, afin de continuer à assurer l'harmonisation du cadre juridique actuel dans le domaine des droits de l'homme, de la démocratie et de l'État de droit

TITRE II. PERSPECTIVES INTERNATIONALES

Chapitre 1 présente l'état actuel du développement de la politique israélienne, en ce qui concerne les opportunités et les défis présentés par l'intelligence artificielle (IA) en relation avec les droits de l'Homme et l'éthique. Il est basé, en grande partie, sur le rapport de l'Initiative nationale israélienne pour des systèmes intelligents sécurisés, qui a été récemment soumis au gouvernement israélien.

Chapitre 2 résume la nature de la gouvernance de l'IA et les caractéristiques des discussions au Japon. Étant donné que ce rapport traite principalement des tentatives et des discussions de gouvernance jusqu'en 2018, il est facile d'imaginer que les détails qu'il contient changeront avec les développements technologiques futurs et les changements des conditions sociales. Cependant, l'organisation de discussions à un point fixe est utile pour les futures discussions sur la gouvernance de l'IA et pour la recherche comparative avec d'autres gouvernances de technologies de pointe.

Chapitre 3 présente certaines des utilisations les plus répandues et les plus étendues des applications d'intelligence artificielle (IA) au Mexique ainsi que le cadre réglementaire applicable aux applications d'IA. Nous nous efforcerons de représenter avec précision le contexte dans lequel chaque système d'intelligence artificielle fonctionne. Le contexte, ainsi qu'une brève description du système, devraient fournir au lecteur suffisamment d'informations pour permettre des comparaisons avec d'autres juridictions et pour tenter d'éclaircir les complexités de la réglementation potentielle des applications d'IA au Mexique et ailleurs.

TITRE I. PERSPECTIVE INTERNATIONALE DE LA REGULATION DE L'IA BASEE SUR LES STANDARDS DU CONSEIL DE L'EUROPE

CHAPITRE PRELIMINAIRE. Rapport d'état d'avancement des travaux du Comité Ad Hoc sur l'intelligence artificielle (CAHAI)¹

I. Résumé

1. Le 11 septembre 2019, le Comité des Ministres a adopté le mandat du Comité ad hoc sur l'intelligence artificielle (CAHAI), chargeant le Comité d'examiner, sur la base de larges consultations multipartites, la faisabilité et les éléments potentiels, d'un cadre juridique pour le développement, la conception et l'application de l'intelligence artificielle, fondés sur les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit. Le présent rapport est soumis par le CAHAI au Comité des Ministres et détaille les progrès réalisés à ce jour, en tenant compte de l'impact des mesures de lutte contre la pandémie de la COVID-19. Il comprend également des réflexions sur les méthodes de travail, la synergie et la complémentarité avec d'autres parties prenantes concernées et des propositions d'actions futures de la CAHAI au moyen d'une feuille de route solide et claire.

2. Le rapport souligne que le Conseil de l'Europe a un rôle crucial à jouer pour assurer que l'intelligence artificielle (IA) respecte les normes de l'Organisation en matière de droits de l'homme, de démocratie et d'État de droit. Le CAHAI a concentré ses travaux en particulier sur la cartographie des cadres juridiques internationaux et nationaux pertinents et des lignes directrices éthiques, ainsi que sur l'analyse des risques et des opportunités découlant de l'intelligence artificielle, notamment leur impact sur les droits de l'homme, l'État de droit et la démocratie.

3. L'analyse préliminaire effectuée jusqu'à présent confirme l'importance d'approfondir les réflexions sur la faisabilité et le développement des éléments d'un cadre juridique « horizontal » et transversal pour réglementer l'utilisation et les effets des applications de l'IA, qui s'appuierait sur l'expertise unique de l'Organisation et sur les travaux réalisés à un niveau « vertical » et sectoriel. Sa pertinence évidente a également été confirmée et renforcée par la récente pandémie de COVID-19. Un tel cadre juridique pourrait consolider les normes existantes dans ce domaine ou développer des normes supplémentaires nécessaires à l'ère numérique. Il pourrait être fondé sur une approche d'évaluation de l'impact sur les droits de l'homme et sur des mécanismes opérationnels pratiques. Enfin, il pourrait également représenter le socle sur lequel s'appuierait d'autres d'initiatives et d'instruments sectoriels, ce qui demeure indispensable pour traiter de manière exhaustive les défis posés par les applications de l'intelligence artificielle dans les domaines d'activité pertinents du Conseil de l'Europe.

4. Le travail entrepris par le CAHAI constitue également une opportunité de contribuer et de compléter d'autres initiatives internationales dans ce domaine (notamment celles de l'OCDE, l'Union européenne - en particulier la Commission européenne - l'UNESCO et les Nations Unies en général) en adoptant un instrument concret basé sur les droits de l'homme, dans le cadre d'un mécanisme juridique mondial de régulation des technologies numériques, domaine dans lequel le Conseil de l'Europe a une réelle valeur ajoutée. Une coordination étroite et des synergies avec les travaux de ces organisations dans le domaine de l'intelligence artificielle se poursuivront en permanence.

¹ 23 Septembre 2020, CM(2020)90-final.

5. Enfin, le rapport comprend la version de travail à ce jour de la table des matières de la future étude de faisabilité et une feuille de route claire fixant les principaux délais du processus. De nouvelles méthodes de travail ont été envisagées, notamment la création de trois groupes de travail thématiques chargés de tâches spécifiques relatives à l'étude de faisabilité du CAHAI, à savoir la préparation de propositions sur des aspects politiques spécifiques, la préparation de consultations multipartites et l'élaboration de cadres juridiques - ce qui devrait permettre de réaliser des progrès substantiels dans l'accomplissement du mandat du CAHAI au cours des prochains mois et jusqu'en décembre 2021.

II. Introduction. Contexte et mandat du CAHAI

6. Depuis plusieurs années, le Conseil de l'Europe évalue et anticipe l'impact des technologies numériques sur les droits de l'homme, la démocratie et l'État de droit. Il a élaboré des instruments juridiques pertinents et deux Stratégies pour la gouvernance de l'Internet² afin que notre société et les individus tirent pleinement parti des bénéfices de pratiques innovantes. Parmi ces technologies, celles fondées sur l'intelligence artificielle (IA) se démarquent des traditionnelles applications de l'informatique par leur autonomie. De plus en plus utilisées dans un large éventail de services publics ou privés, ces technologies particulières offrent de très grandes opportunités de développement mais soulèvent également des questions importantes et complexes d'ordre éthique, juridique, politique et économique.

7. Le Conseil de l'Europe s'est engagé à encadrer leur portée et leurs implications dans la plupart de ses secteurs d'activités spécialisés, comme la justice, la protection des données, l'égalité et la non-discrimination³. Les impacts de l'IA sur les individus et la société ont également été examinés sous différents angles par le Comité des Ministres, l'Assemblée parlementaire et la Commissaire aux droits de l'homme⁴. L'Organisation s'inscrit désormais dans une démarche complémentaire aux initiatives mises en œuvre à ce jour, en ayant adopté une approche stratégique transversale qui s'appuie sur l'examen de la faisabilité de cadres contraignants et non contraignants dans le but d'assurer que la conception, le développement et l'application de l'IA soient conformes aux normes européennes des droits de l'homme, de la démocratie et de l'État de droit. Une conférence de haut niveau, organisée par la présidence finlandaise du Comité des Ministres et par le Conseil de l'Europe à Helsinki (Finlande) les 26-27 février 2019, a relevé l'importance et l'urgence de réponses politiques aux questions d'incidence de l'IA sur les droits de l'homme, la démocratie et l'État de droit, et a donné la nécessaire impulsion à la création du Comité ad hoc sur l'Intelligence Artificielle (CAHAI), dont le mandat a été adopté par le Comité des Ministres le 11 septembre 2019.

8. Le CAHAI est chargé 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. Une large consultation multipartite et une coordination étroite avec d'autres organisations internationales seront assurées dans le cadre de son mandat.

9. Les avantages économiques et les autres bénéfices découlant de l'application de l'IA, tout comme l'importance pour l'Europe de suivre la même cadence que le reste du monde, font partie des questions que le CAHAI pourra examiner dans le cadre de ses travaux.

² Gouvernance de l'Internet – [Stratégie](#) du Conseil de l'Europe 2016 - 2019 et 2012 - 2015.

³ Commission européenne pour l'efficacité de la justice (CEPEJ), [Charte éthique européenne d'utilisation de l'intelligence judiciaire dans les systèmes judiciaires et leur environnement](#); Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement des données à caractère personnel (Convention 108), [Lignes directrices sur l'intelligence artificielle et la protection des données](#); Commission européenne contre le racisme et l'intolérance (ECRI), [étude. Discrimination, intelligence artificielle et décisions algorithmiques](#).

⁴ [Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques](#), [Recommandation de l'Assemblée Parlementaire du Conseil de l'Europe sur la convergence technologique, l'intelligence artificielle et les droits de l'homme](#) et Recommandation de la Commissaire aux droits de l'homme « [Décoder l'IA : 10 mesures pour protéger les droits de l'homme](#) »

III. État d'avancement des travaux

10. Le CAHAI a tenu à ce jour une première réunion plénière, les 18, 19 et 20 novembre 2019⁵, et une deuxième réunion plénière a eu lieu les 6-8 juillet 2020⁶. Au cours de sa première réunion plénière, le Comité a élu son Président (Gregor Strojín, Slovaquie), sa Vice-Présidente (Peggy Valcke, Belgique) ainsi que son Bureau (composé, en plus du Président et de la Vice-Présidente, de cinq membres provenant d'Allemagne, d'Estonie, de France, d'Italie et de Suisse). Le Bureau a également tenu trois réunions, respectivement le 19 novembre 2019⁷, les 23 et 24 janvier 2020⁸ et une réunion en ligne le 27 mars 2020⁹. Le Comité a également nommé la rapporteure pour l'égalité de genre (Jana Novohradská, République slovaque) lors de sa première réunion plénière.

11. Au cours de sa première réunion plénière, le CAHAI a précisé que l'étude de faisabilité devrait inclure une cartographie des instruments juridiques nationaux et internationaux (à la fois du Conseil de l'Europe et d'autres organisations internationales) et des cadres éthiques relatifs aux applications d'IA, tout comme une cartographie des risques et opportunités découlant du développement, de la conception et de l'application de l'IA, notamment son impact sur les droits de l'homme, la démocratie et l'État de droit. Les résultats de ces cartographies doivent être pris en compte pour décider de l'opportunité d'une définition de l'IA et pour définir un cadre juridique approprié pour la conception, le développement et l'application de l'IA sur la base des normes du Conseil de l'Europe. Trois experts indépendants ont été mandatés pour cadrer l'étude de faisabilité, en vue de soutenir le processus décisionnel du CAHAI sur cette question, et ont présenté leurs rapports au CAHAI lors de sa deuxième réunion plénière.

12. S'agissant de la cartographie des instruments juridiques internationaux applicables à l'IA, l'approche choisie est celle d'identifier, dans les nombreux secteurs d'activité du Conseil de l'Europe et avec l'appui des comités intergouvernementaux concernés, les instruments internationaux pertinents, les principes juridiques directeurs et les principales valeurs émergent de ces instruments, ainsi que les potentielles lacunes de ceux-ci. Il est notamment question d'examiner la possible émergence de principes communs spécifiques à l'IA, « horizontaux » et « transversaux », fondateurs de principes spécialisés « verticaux » définis (ou en cours de définition) dans les différents domaines d'activité du Conseil de l'Europe.

13. Les résultats des consultations en ligne menées parmi les États membres du Conseil de l'Europe en novembre 2019 - février 2020 sont utilisés pour cartographier les initiatives nationales en matière d'IA.

14. La cartographie des cadres éthiques et non contraignants analyse la vaste production de chartes, déclarations et principes relatifs à l'IA développés au cours des dernières années par des acteurs privés, scientifiques et de la société civile, dans l'objectif de recenser ceux qui pourraient contribuer à établir un possible cadre juridique sur l'IA conforme aux normes du Conseil de l'Europe.

15. La cartographie des risques et opportunités découlant du développement, de la conception et de l'application de l'IA, et notamment de l'impact de celle-ci sur les droits de l'homme, la démocratie et l'État de droit, est menée spécifiquement à l'égard des droits protégés par la Convention de sauvegarde des droits de l'homme et libertés fondamentales (ci-après : la Convention européenne des droits de l'homme) et vise à déterminer dans quelle mesure leur exercice pourrait être impacté par certaines applications de l'IA et quelles stratégies adopter afin d'y remédier. Les réponses reçues dans le cadre de la consultation en ligne peuvent aussi apporter une certaine aide

⁵ Voir le [rapport abrégé de la première réunion plénière](#) du CAHAI.

⁶ Voir le [rapport de la deuxième réunion plénière du CAHAI](#). La réunion prévue à l'origine les 11-13 mars 2020 a dû être reportée en raison de la situation internationale concernant le virus COVID-19 et notamment les différentes restrictions mises en place au sein du Conseil de l'Europe et des États membres.

⁷ Voir le [rapport abrégé de la première réunion du Bureau](#) du CAHAI.

⁸ Voir le [rapport abrégé de la deuxième réunion du Bureau](#) du CAHAI.

⁹ Voir le [rapport abrégé de la troisième réunion du Bureau](#) du CAHAI.

en ce qui concerne les stratégies mises en place dans les États membres pour limiter l'impact de l'IA sur les droits protégés par cette Convention.

16. Ces cartographies devraient viser à identifier les cadres juridiques applicables en matière de droits de l'homme, qui sont importants à l'ère numérique. Néanmoins, l'analyse préliminaire réalisée jusqu'à présent relève, d'une part, les limites des régimes juridiques existants, rédigés antérieurement au développement d'applications de l'IA, ce qui réduit leur efficacité à répondre de manière adéquate aux défis actuels ; et, d'autre part, l'impact significatif de l'IA sur plusieurs normes définies dans les instruments du Conseil de l'Europe et notamment sur les droits protégés par la Convention européenne des droits de l'homme. À cet égard, la question du rôle et des responsabilités de toutes les parties prenantes, y compris les États et les entreprises, les instituts de recherche et la société civile, dans le développement et le déploiement de technologies d'intelligence artificielle qui respectent les droits de l'homme devrait être approfondie.

17. Face à ces défis, le Conseil de l'Europe a un rôle déterminant à jouer pour assurer que le développement, la conception et l'application des technologies d'IA soient conformes aux normes de l'Organisation. Il est donc important d'approfondir la réflexion sur l'élaboration des éléments d'un cadre juridique visant à réglementer l'utilisation et les effets des applications de l'IA, en s'appuyant sur l'expertise unique de l'Organisation et sur une approche d'évaluation de l'impact sur les droits de l'homme. En outre, ce cadre juridique pourrait examiner l'opportunité de consolider les normes existantes dans ce domaine (telles que celles de la Convention européenne des droits de l'homme ou de la Convention 108+) ou de développer des normes supplémentaires nécessaires à l'ère numérique. Enfin, il pourrait également représenter le socle sur lequel s'appuierait un certain nombre d'initiatives et d'instruments sectoriels, ce qui demeure indispensable pour traiter de manière exhaustive les défis posés par les applications de l'intelligence artificielle dans les domaines d'activité pertinents du Conseil de l'Europe.

18. Il s'agit également d'une opportunité d'apporter un soutien à d'autres initiatives internationales (OCDE, Union européenne - en particulier la Commission européenne - UNESCO et Nations unies de manière générale) en édictant un instrument concret fondé sur les droits de l'homme, s'insérant dans un mécanisme mondial juridique d'encadrement des technologies numériques dans lequel le Conseil de l'Europe a une réelle valeur ajoutée.

19. Par ailleurs, il serait important d'examiner aussi, dans le cadre de l'étude de faisabilité, les mécanismes et les actions pouvant être menées pour garantir l'effectivité du cadre juridique proposé, tant au niveau du Conseil de l'Europe que des États membres. Ceci inclurait notamment les questions relatives à la création de mécanismes collaboratifs de rédaction des normes (en particulier l'utilisation, dans certains cas spécifiques, de « bacs à sable »¹⁰), la vérification et/ou certification *ex ante* des applications d'IA ou des mécanismes de contrôle par des autorités indépendantes, ainsi que l'opportunité d'une réglementation des professions d'experts en mégadonnées (par exemple l'opportunité d'établir des ordres professionnels pour les *data scientists*, de développer des chartes déontologiques fondées sur les principes éthiques, un « serment d'Hippocrate » pour les professionnels de l'IA ou créer un « permis de conduire » pour l'IA, etc.).

20. Par le processus de préparation de l'étude de faisabilité, y compris dans les activités des groupes de travail et dans le cadre de la consultation multipartite, le CAHAI visera à intégrer une perspective d'égalité de genre, à contribuer à la construction de sociétés cohésives et à promouvoir et protéger les droits des personnes handicapées.

21. Pendant la deuxième réunion plénière du CAHAI, les États membres ont exprimé leurs opinions sur le contenu des différents éléments du projet de table des matières de l'étude de faisabilité, qui figure à l'Annexe I de ce rapport.

¹⁰ Le processus dans lequel les régulateurs permettent l'utilisation et l'essai d'une nouvelle technologie dans un environnement fermé ou limité et en dialogue étroit avec les décideurs politiques. En plus de l'évaluation de la technologie, ce mécanisme permet également aux régulateurs d'essayer de nouvelles règles et d'observer leur impact sur la technologie dans un environnement où les dommages ou le danger pour le public sont limités.

IV. L'impact de la pandémie de COVID-19 sur les activités du CAHAI

22. L'IA fait partie des nombreux moyens employés par les États pour endiguer et lutter contre la pandémie de COVID-19. Plusieurs applications d'IA ont été développées dans des domaines tels que l'appui à la recherche pour l'élaboration d'un vaccin, la formulation de diagnostics en soutien du personnel soignant et l'élaboration de modèles prévisionnels sur la possible évolution de la pandémie. L'IA a été utilisée pour faciliter l'analyse de milliers d'articles de recherche publiés sur la pandémie et pour assurer un meilleur partage de connaissances scientifiques, ainsi qu'à des fins de dépistage de la maladie. En outre, l'IA pourrait être utile pour lutter contre la désinformation sur les coronavirus, pour autant que les droits en matière de protection des données et la liberté d'expression soient correctement protégés.

23. D'autres utilisations visant à maîtriser les risques épidémiques ont également été recensées : par exemple, des dispositifs de reconnaissance faciale et biométriques, ou encore de géolocalisation ainsi que des drones, ont été utilisés pour s'assurer du respect des mesures de confinement par des individus infectés ou des populations à risques. Des applications pour avertir les utilisateurs qu'ils ont été en contact avec des personnes potentiellement porteuses du virus ont également été développées. Ce type d'utilisation impacte de manière significative certains droits et libertés protégés par des instruments internationaux de protection des droits de l'homme, y compris la Convention européenne des droits de l'homme, comme le droit au respect de la vie privée et familiale (droit à la vie privée en particulier) et la liberté de circulation, et devrait être examiné par le CAHAI. L'utilisation actuelle ou envisagée d'applications d'IA dans certains pays pour la surveillance et le contrôle des épidémies ainsi que pour la prévention de nouveaux foyers de maladies pourrait conduire à la banalisation du recours à la surveillance de masse des populations en l'absence de garanties adéquates.

24. Les mesures de confinement du fait de la COVID-19 ont involontairement entraîné la discrimination des enfants et des jeunes qui n'ont pas accès à des ordinateurs portables et ne fréquentent pas des écoles qui proposent des cours en ligne. Cela a entraîné une interruption de leur éducation et un obstacle à l'avancement de leurs compétences numériques, y compris en matière d'IA en tant que domaine d'étude interdisciplinaire. Par ailleurs, en période de crise économique, autre conséquence possible de la crise de la COVID-19, il existe un modèle historique de discrimination des filles et des jeunes femmes lorsque les familles prennent des décisions sur l'opportunité d'investir dans l'éducation supérieure ou technique des enfants. Les filles et les jeunes femmes sont souvent désavantagées et reléguées au second plan par rapport aux garçons et aux jeunes hommes. Cela va encore aggraver le statu quo de la sous-représentation des femmes dans les rôles de développement de l'IA et dans les rôles de prise de décision concernant l'IA. Le CAHAI, par l'intermédiaire de son groupe d'élaboration des politiques (CAHAI-PDG) pourrait accorder une attention à cette question et veiller à ce que l'accent soit mis sur des mesures visant à atténuer ce schéma historique de discrimination des filles et des jeunes femmes en période de crise économique. L'application de mesures positives temporaires d'intégration de la dimension de genre dans les politiques d'IA pourrait être envisagée pour protéger et soutenir les femmes et les jeunes filles, afin qu'elles ne soient pas laissées pour compte et qu'elles aient accès à des rôles décisionnels au sein de l'écosystème de l'IA.

25. Dans ce contexte, la pertinence d'un cadre juridique sur la conception, le développement et l'application de l'IA fondé sur les principes et les valeurs du Conseil de l'Europe se trouve renforcée. Le CAHAI pourrait souhaiter examiner les implications et, le cas échéant, la nécessité d'un cadre juridique pour répondre à des situations telles que l'utilisation d'applications d'IA dans des situations spécifiques comme c'est le cas pour d'autres traités du Conseil de l'Europe et conformément aux principes et valeurs du Conseil de l'Europe.

V. Méthodes de travail, documents et feuille de route

26. Au cours de sa première réunion plénière, le CAHAI a chargé son Bureau, en conformité avec la Résolution [CM/Res\(2011\)24](#)¹¹, de procéder aux cartographies nécessaires à l'étude de faisabilité. Le CAHAI, en ayant pris note de l'avis juridique du Jurisconsulte du Conseil de l'Europe, a également autorisé l'association ponctuelle d'autres États membres aux travaux du Bureau, selon les modalités définies par le CAHAI et par son Bureau¹².

27. Afin d'assurer une participation plus active des États membres au CAHAI et son approche transparente et équilibrée de la prise de décision, le CAHAI a tenu un échange de vues sur une proposition de la Fédération de Russie visant à augmenter le nombre de membres du Bureau de six (6), pour atteindre un nombre total de treize (13) membres. Il était proposé que les nouveaux membres pourraient participer aux réunions du Bureau sans que leurs frais soient pris en charge par le Conseil de l'Europe¹³.

28. Suite au report de la réunion prévue du 11 au 13 mars 2020 en raison de la pandémie de COVID-19, des propositions concrètes d'adaptation des méthodes de travail ont été discutées par le Bureau afin de s'assurer que le CAHAI puisse continuer son activité de manière efficace et dans le respect des objectifs assignés, à moins qu'il ne s'avère difficile de le faire et jusqu'à ce que la tenue de la prochaine réunion plénière soit possible, notamment dans des conditions de nature à garantir la sécurité et la participation des membres et observateurs. Le Bureau du CAHAI a chargé le Secrétariat d'organiser une consultation écrite des membres du CAHAI sur plusieurs documents de travail et questions, y compris l'adoption du présent rapport d'avancement et la considération de nouvelles méthodes de travail du CAHAI, tels que la mise en place de groupes de travail restreints. Sur la base de cette consultation, les membres du CAHAI ont convenu de créer trois groupes de travail, en charge respectivement de l'élaboration de politiques, des cadres juridiques, de la consultation et de la sensibilisation (en lien avec la consultation multipartite), dont les descriptions des tâches figurent à l'annexe III de ce document. De plus, le CAHAI a adopté un document sur les méthodes de travail et les modalités de fonctionnement des groupes de travail lors de sa deuxième réunion plénière¹⁴.

29. Comme relevé précédemment, l'étude de faisabilité et les éléments potentiels du futur cadre juridique doivent reposer sur une large consultation multipartite. À cet égard, les membres du CAHAI ont indiqué par le biais de commentaires écrits et dans le cadre d'une première consultation en ligne organisée à la suite de la première réunion plénière, que celle-ci devrait s'adresser en priorité à des représentants du secteur privé, de la société civile et du milieu scientifique, ainsi qu'à d'autres organisations internationales y compris des organisations d'établissement de normes techniques. La feuille de route du CAHAI envisage qu'un projet préliminaire de l'étude de faisabilité comprenant les éléments principaux d'un futur cadre juridique soit examiné par le CAHAI, lors de sa troisième réunion plénière (30 novembre - 2 décembre 2020). Cela permettrait d'avoir un premier livrable sur lequel les différentes parties prenantes pourraient échanger de manière ouverte et transparente, et sur lequel s'appuieraient par la suite les consultations multipartites qui démarreraient en 2021.

¹¹ Voir notamment l'Article 13 de l'Annexe 1 à la Résolution [CM/Res\(2011\)24](#).

¹² Voir à cet égard les paragraphes 19 et 20 du rapport de la 1^{er} réunion plénière du CAHAI. Voir aussi l'avis juridique du Jurisconsulte du Conseil de l'Europe, [DD\(2020\)16](#) 24/01/2020, paragraphe 2, distribué à la demande du Secrétariat lors de la réunion tenue par le GR-J le 16 janvier 2020 : « La Résolution [CM/Res\(2011\)24](#) concernant les comités intergouvernementaux et les organes subordonnés, leur mandat et leurs méthodes de travail n'interdit pas à un comité d'associer exceptionnellement d'autres membres aux travaux du bureau, pour autant que les conditions suivantes soient remplies :

- le CAHAI adopte cette décision par consensus (aucune objection ne doit être exprimée) ;
- la possibilité d'assister aux réunions du bureau sera offerte à tous les membres du CAHAI sur un pied d'égalité ;
- les membres supplémentaires ne deviendraient pas membres du bureau et n'auraient pas droit à la prise en charge de leurs frais ;
- le bureau se réserve le droit de réserver certaines parties de ses réunions à la seule participation des membres élus du bureau ;
- la résolution [CM/Res\(2011\)24](#) demeure applicable dans toutes ses dispositions, en particulier aucune décision sur des questions de fond ne serait prise par le bureau ».

¹³ Voir les paragraphes 86 et 87 du rapport de la deuxième réunion plénière du CAHAI.

¹⁴ Voir [CAHAI \(2020\) 10ADD rev1](#).

L'année 2021 serait également dédiée à la finalisation des éléments du cadre juridique susmentionné si possible.

30. Le projet de feuille de route du CAHAI figure à l'Annexe II de ce rapport.

VI. Synergie et complémentarité du travail du CAHAI avec celui d'autres organisations internationales

31. La coordination avec les travaux réalisés ou en cours au sein d'autres organisations internationales sur l'IA, en particulier l'Union européenne (comme le processus en cours lancé par la Commission européenne sur le Livre blanc sur l'IA) et l'OCDE, a été activement recherchée afin de promouvoir les synergies et d'éviter tout doublon. En plus de la participation aux réunions du CAHAI et des contacts ayant régulièrement lieu entre les institutions, le Président du CAHAI et le secrétariat ont participé au lancement de l'Observatoire sur les politiques en matière d'IA de l'OCDE le 27 février 2020. Des échanges avec des représentants de l'OCDE, de l'Union européenne et de l'UNESCO, d'UIT et d'ISO sur les travaux en cours au sein de chaque Organisation sont prévus lors de manière régulière lors des réunions plénières du CAHAI. L'objectif est de promouvoir des synergies dans l'élaboration d'un cadre juridique sur l'IA, dans laquelle la contribution et l'expertise spécifique à chaque organisation (s'agissant du CAHAI, une expertise en matière de droits de l'homme, démocratie et État de droit notamment) puissent être mises en valeur et se compléter mutuellement.

VII. Implications budgétaires

32. Le budget du CAHAI, tel qu'adopté par le Comité de Ministres, permet de couvrir uniquement la tenue des réunions plénières et du Bureau et partiellement la tenue d'un nombre limité de réunions de groupes de travail. Le budget alloué ne permet pas d'envisager des actions complémentaires¹⁵, telles que le renforcement de la dimension des consultations multipartites, l'implication d'experts de haut niveau, et/ou l'organisation d'actions de nature à renforcer la visibilité des travaux du Conseil de l'Europe dans ce domaine.

33. C'est pourquoi les États membres et les États observateurs du CAHAI sont appelés à envisager de contribuer financièrement au renforcement des activités du CAHAI, par le biais de contributions volontaires ou de tout autre moyen qu'ils jugeront utile.

VIII. Conclusions et propositions concrètes d'action

34. Le Conseil de l'Europe a un rôle crucial à jouer aujourd'hui afin d'assurer que des applications d'IA soient développées d'une manière conforme à la protection des droits de l'homme et respectent les normes de l'Organisation lorsqu'elles existent. Ainsi, l'initiative du CAHAI demeure unique et complémentaire à celles entreprises par d'autres organisations internationales. Il est important que le travail sur l'élaboration des éléments potentiels d'un cadre juridique puisse démarrer dès janvier 2021 et que le CAHAI puisse disposer de moyens financiers supplémentaires dans la mise en œuvre de son mandat.

35. Le Comité des Ministres est invité à mandater le CAHAI pour :

¹⁵ L'annulation de la réunion plénière du 11-13 mars 2020 et l'impossibilité de récupérer certains frais déjà engagés ont aussi un impact sur l'enveloppe budgétaire disponible pour le CAHAI en 2020.

(i) faire des progrès substantiels dans la rédaction de son étude de faisabilité d'un cadre juridique en novembre 2020, afin de démarrer dès janvier 2021 une réflexion sur les éléments d'un cadre juridique qui feraient l'objet d'une large consultation multipartite ; ce cadre juridique pourrait réglementer la conception, le développement et l'application de l'IA ayant un effet significatif sur les droits de l'homme, la démocratie et l'État de droit. Il pourrait par ailleurs considérer l'opportunité de consolider les normes existantes par une interprétation des normes, principes et valeurs déjà édictés en la matière ou de développer des nouvelles normes rendues nécessaires à l'ère numérique. Enfin, il jetterait les bases sur lesquelles un certain nombre d'initiatives et d'instruments pourraient être développés par la suite dans les différents secteurs d'activité du Conseil de l'Europe, ce qui demeure indispensable pour traiter de manière exhaustive les défis posés par les applications de l'IA dans les domaines d'activité pertinents du Conseil de l'Europe ;

(ii) proposer, en même temps, des mesures complémentaires afin d'opérationnaliser le cadre juridique susmentionné : il pourrait notamment être évoqués la procédure d'étude d'impact préalable sur les droits de l'homme, les moyens de validation ou de certification des algorithmes et des systèmes d'IA ou la formation et l'organisation de certaines professions impliquées dans l'application des outils d'IA ;

36. Le Comité des Ministres est également invité à prendre note de la feuille de route du CAHAI et à proposer aux États membres et aux États observateurs qui le souhaiteraient de contribuer financièrement au renforcement des activités du CAHAI, par le biais de contributions volontaires ou de tout autre moyen qu'ils jugeront utile.

Annexes

- I. Projet de table des matières de l'étude de faisabilité
- II. Feuille de route du CAHAI
- III. Mandats des Groupes de travail

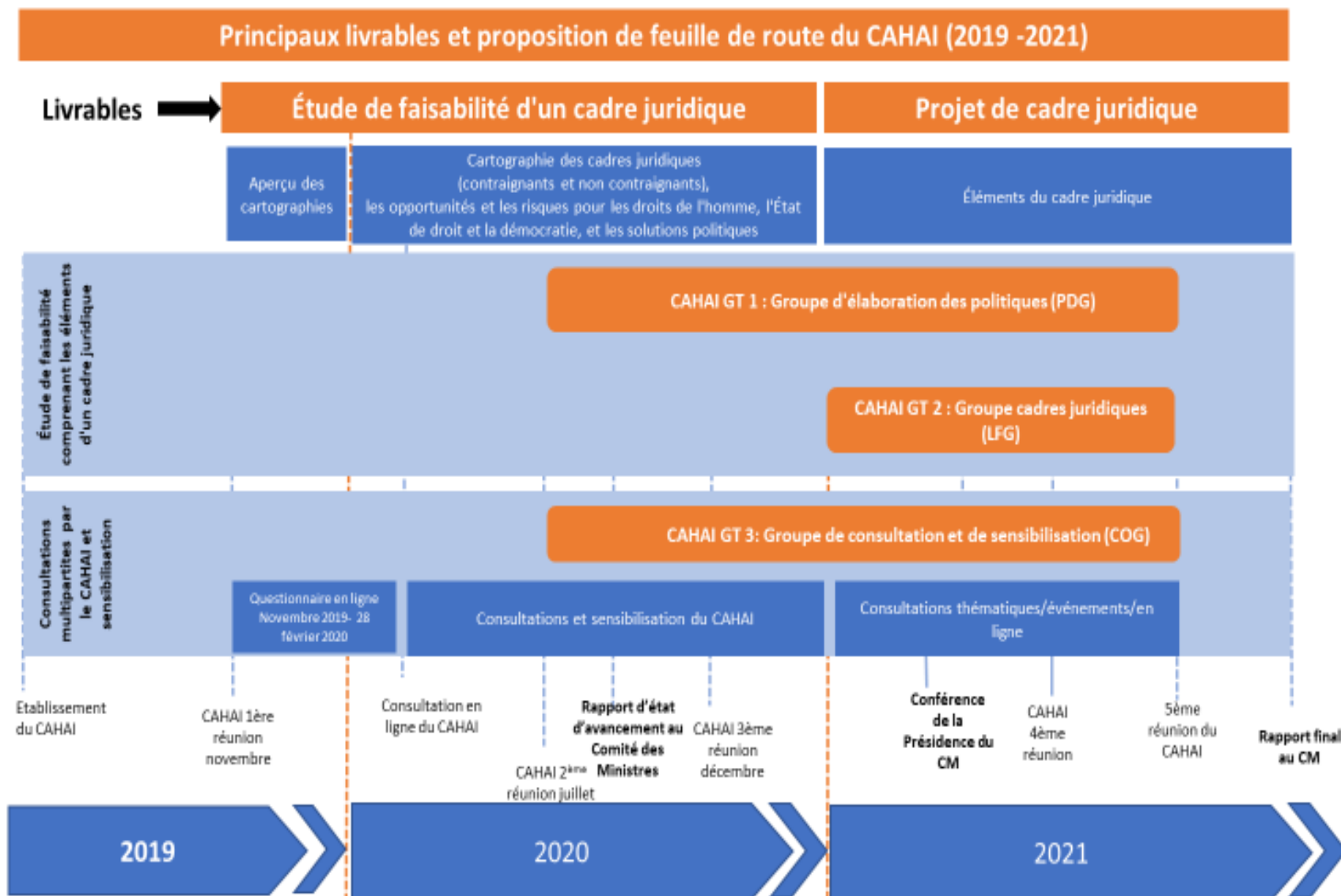
Annexe I. Projet de table des matières de l'étude de faisabilité

(version de travail du 16 juin 2020)

1. **Introduction générale**
2. **Champ d'application d'un cadre juridique du Conseil de l'Europe sur l'intelligence artificielle**
3. **Opportunités et risques** découlant de la conception, du développement et de l'application de l'intelligence artificielle sur les droits de l'homme, l'État de droit et la démocratie. « Zones vertes » et « rouges » - c'est-à-dire des exemples respectivement positifs et problématiques d'applications de l'intelligence artificielle du point de vue des droits de l'homme, de l'État de droit et de la démocratie, tout en tenant compte du contexte spécifique à la conception, au développement et à l'application de l'intelligence artificielle en Europe et des développements au niveau mondial.
4. **Les travaux du Conseil de l'Europe en matière d'intelligence artificielle à ce jour**
5. **Cartographie des instruments applicables à l'intelligence artificielle**
 - i. Instruments juridiques internationaux applicables à l'intelligence artificielle
 - ii. Lignes directrices éthiques applicables à l'intelligence artificielle
 - iii. Aperçu des instruments, politiques et stratégies nationales liés à l'intelligence artificielle
 - iv. Avantages, inconvénients et limites des instruments internationaux et nationaux existants et des lignes directrices éthiques en matière d'intelligence artificielle
 - v. Instruments juridiques internationaux, lignes directrices éthiques et acteurs privés
6. **Principales conclusions des consultations multipartites**
 - i. Table des matières de l'étude de faisabilité
 - ii. Principales conclusions sur le type et le contenu d'un cadre juridique pour la conception, le développement et l'application de l'intelligence artificielle, sur la base des normes du Conseil de l'Europe sur les droits de l'homme, la démocratie et l'État de droit
7. **Principaux éléments d'un cadre juridique pour la conception, le développement et l'application de l'intelligence artificielle**
 - i. Valeurs, droits et principes clés découlant - dans une perspective ascendante - d'approches sectorielles et de lignes directrices éthiques ; dans une perspective descendante - des exigences en matière de droits de l'homme, de démocratie et d'État de droit.
 - ii. Rôle et responsabilités des États membres et des acteurs privés dans le développement d'applications conformes à ces exigences
 - iii. Responsabilité pour les dommages causés par l'intelligence artificielle

- 8. Options possibles** d'un cadre juridique du Conseil de l'Europe pour la conception, le développement et l'application de l'intelligence artificielle fondé sur les droits de l'homme, la démocratie et l'État de droit (pour chaque option : contenu, destinataires, valeur ajoutée, rôle des acteurs privés, attentes des États membres découlant des commentaires écrits présentés)
- i. Mise à jour des instruments juridiquement contraignants existants
 - ii. Convention
 - iii. Convention-cadre
 - iv. Instrument(s) de droit souple
 - v. Autre type de soutien aux États membres comme l'identification de bonnes pratiques
 - vi. Complémentarité possible entre les éléments horizontaux et transversaux qui pourraient faire partie d'un instrument de type conventionnel et les travaux verticaux et sectoriels qui pourraient donner lieu à des instruments spécifiques de nature différente.
- 9. Mécanismes pratiques** possibles pour assurer la conformité et l'efficacité du cadre juridique (comme par exemple la création d'un mécanisme de vérification et/ou de certification ex ante, la surveillance par des autorités indépendantes, le « bac à sable », etc.)
- 10. Considérations finales**

Annexe II. Feuille de route du CAHAI



Annexe III. Mandats des Groupes de travail

Groupe d'élaboration des politiques (CAHAI-PDG)

Composition

Le CAHAI-PDG est composé d'un maximum de 20 experts désignés par les États membres, et de participants et d'observateurs qui expriment leur intérêt à participer à ce groupe de travail et contribuent de manière professionnelle et continue à ses travaux.

Tâches

Le CAHAI - PDG doit :

- 1) Contribuer à l'élaboration de l'étude de faisabilité d'un cadre juridique sur les applications de l'intelligence artificielle, en s'appuyant sur le travail de cartographie déjà entrepris par le CAHAI et en se concentrant en particulier sur les questions suivantes :
 - a. identifier et analyser l'impact des applications IA des états et acteurs privés sur les droits de l'homme, l'État de droit et la démocratie, y compris les risques, les menaces et les opportunités que représente ou apporte l'IA, y compris lorsque nécessaire au travers d'une évaluation des propositions de réglementation effectuées par les États membres et d'autres parties prenantes, et prenant en compte l'importance de garder la cadence avec les développements en matière d'IA au niveau mondial ;
 - b. sur la base des conclusions de la cartographie, préparer les conclusions principales et propositions sur les mesures politiques et autres mesures, afin de garantir que les normes internationales et les instruments juridiques internationaux dans ce domaine sont à jour et efficaces et de préparer des propositions pour un instrument juridique spécifique réglementant l'intelligence artificielle ;
 - c. identifier les applications d'IA à haut risque et celles particulièrement prometteuses, examiner et développer des mesures politiques et autres mesures pour faire face aux risques qu'elles présentent et assurer la protection des droits de l'homme, de l'État de droit et de la démocratie ;
 - d. dans le cadre des travaux menés au titre des points b) et c), préparer des propositions visant à faciliter la mise en œuvre des obligations internationales et européennes pertinentes en matière de droits de l'homme, et destinées à soutenir la mise en œuvre d'évaluations efficaces des risques en matière de droits de l'homme et de toute autre action nécessaire pour répondre, le cas échéant, aux opportunités, menaces et risques nouveaux, émergents et importants ;
- 2) Élaborer des propositions pour engager et consulter les parties prenantes concernées sur les éléments susmentionnés, en étroite coordination et en consultation avec le groupe - Consultations et sensibilisation (CAHAI-COG).

Méthodes de travail

Le CAHAI-PDG tient 2 réunions en 2020, et sous réserve d'une décision plénière à cet effet, toute autre réunion supplémentaire requise en 2021. Il effectue ses travaux par le biais des technologies, tel que le courrier électronique ou d'autres moyens similaires de communication électronique, les réunions virtuelles et/ou des réunions physiques. Le groupe désigne un(e) Président(e) et un(e) co-président(e) parmi ses membres lors de sa première réunion.

Le budget du Conseil de l'Europe couvre les coûts de participation des présidents (pour les réunions physiques) et ceux de l'organisation des réunions par des moyens technologiques, afin de permettre au groupe de travail de remplir ses fonctions et ses responsabilités par des échanges en ligne.

La participation des membres aux réunions physiques est en principe prise en charge par les institutions et organisations d'envoi sans remboursement des frais.

Durée Juillet 2020-décembre 2021

Groupe de consultation et de sensibilisation (CAHAI-COG)

Composition

Le CAHAI - COG est composé d'un maximum de 20 experts désignés par les États membres, et de participants et d'observateurs qui expriment leur intérêt à participer à ce groupe de travail et contribuent de manière professionnelle et continue à ses travaux.

Tâches

Le CAHAI - COG fait le point sur l'analyse des réponses à la première consultation en ligne entreprise par le secrétariat et élabore pour la considération et l'approbation du CAHAI :

- 1) Une analyse et une cartographie des parties prenantes (échéance : septembre 2020)
- 2) Sur la base d'un travail de sensibilisation auprès des pays ayant exprimé leur intérêt pour la tenue de consultations nationales, une méthodologie et un ou plusieurs modèles à utiliser par les États membres dans le cadre de ces consultations (échéance : novembre 2020)
- 3) Un calendrier de consultations thématiques (prévu pour novembre 2020) et un premier document de consultation (échéance : décembre 2020)
- 4) Une analyse des développements en cours et des rapports qui sont directement pertinents pour les tâches des groupes de travail du CAHAI (échéance : régulier - octobre 2020) ainsi qu'une analyse des contributions des répondants aux consultations en ligne pour examen par le CAHAI (échéance : 2021)

Méthodes de travail

Le CAHAI-COG travaille principalement en ligne, avec le soutien du secrétariat du CAHAI. Le groupe devra désigner un(e) Président(e) et un (e) co-président(e) parmi ses membres lors de sa première réunion.

Durée Juillet 2020-Janvier 2021

Groupe sur les cadres juridiques (CAHAI-LFG)

Composition

Le CAHAI-LFG est composé d'un maximum de 20 experts désignés par les États membres, et de participants et d'observateurs qui expriment leur intérêt à participer à ce groupe de travail et à contribuer de manière substantielle à ses travaux.

Tâches

Le CAHAI - LFG doit :

- 1) Préparer les principales constatations et propositions relatives aux éléments et dispositions éventuels d'un cadre juridique en vue d'élaborer, le cas échéant, des projets d'instruments juridiques, pour examen et approbation par le CAHAI, en tenant compte de l'examen des instruments juridiques existants applicables à l'intelligence artificielle et des options politiques exposées dans l'étude de faisabilité approuvée par le CAHAI ;
- 2) Élaborer des propositions spécifiques de réglementation pour le développement, la conception et l'application de l'IA dans les domaines identifiés comme étant à risque par les États membres et les autres parties prenantes, en tenant compte des approches réglementaires des États membres.

Méthodes de travail

Le CAHAI-LFG tient 2 réunions en 2021, et sous réserve d'une décision plénière à cet effet, toute autre réunion supplémentaire nécessaire. Il effectue ses travaux par le biais des technologies, tel que le courrier électronique ou d'autres moyens similaires de communication électronique, les réunions virtuelles et/ou des réunions physiques. Le groupe désigne un(e) président(e) et un (e) co-président(e) parmi ses membres lors de sa première réunion.

Le budget du Conseil de l'Europe couvre les coûts de participation des présidents (pour les réunions physiques) et ceux de l'organisation des réunions par des moyens technologiques, afin de permettre au groupe de travail de remplir ses fonctions et ses responsabilités par des échanges en ligne.

La participation des membres aux réunions physiques est en principe prise en charge par les institutions et organisations d'envoi sans remboursement des frais.

Durée Janvier 2021 - décembre 2021

CHAPITRE I. Impact de l'intelligence artificielle sur les droits de l'homme, la démocratie et l'État de droit¹⁶

Catelijne Muller¹⁷

I. Remarques liminaires

1. L'intelligence artificielle (IA), comme technologie généraliste, a des incidences sur tout le tissu social. En 2017, le Comité économique et social européen, dans ce qui est largement considéré comme le "rapport initial" sur l'impact sociétal plus large de l'IA, identifie, les principaux domaines touchés : la sécurité, l'éthique, le droit et réglementations, la démocratie, la transparence, la vie privée, le travail, l'éducation et l'(in)égalité.¹⁸ Cela signifie que l'IA a un impact sur les droits de l'homme, la démocratie et l'État de droit, à savoir les fondements mêmes des sociétés européennes.

2. En 2019, le Groupe d'experts de haut niveau sur l'IA a présenté ses Lignes directrices en matière d'éthique pour une IA digne de confiance¹⁹. Ces lignes directrices affirment qu'une IA « digne de confiance » doit être licite, éthique et robuste sur le plan technique et social. S'agissant de l'aspect éthique²⁰, elles axent expressément leur approche sur les droits. Bien que certains éléments mentionnés dans ces lignes directrices découlent directement de droits (de l'homme) existants, ils n'ont pas en eux-mêmes de caractère contraignant. Or, on assiste depuis peu à une montée des appels à créer des instruments juridiquement contraignants en matière d'IA, ou à renforcer ceux qui existent. La Commission européenne a annoncé dans son Livre blanc sur l'IA²¹ les éléments potentiels d'un cadre législatif, et souligne l'importance d'une IA respectueuse des droits fondamentaux de l'UE et des lois qui en assurent l'application.

3. Le présent document souligne l'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit. Il identifie les droits de l'homme, tels qu'énoncés dans la Convention européenne des droits de l'homme (« CEDH »), ses Protocoles et la Charte sociale européenne (« CSE »), qui sont actuellement les plus touchés par l'IA ou les plus exposés (chapitre II). Aux chapitres III et IV, notre rapport avance plusieurs stratégies possibles, applicables simultanément si nécessaire. Le chapitre III s'intéresse aux mesures adoptées face à cet impact dans le cadre existant des droits de l'homme, de la démocratie et de l'État de droit. Le chapitre IV examine les stratégies envisageables si ce cadre ne s'avérait pas assez protecteur. Compte tenu de l'évolution rapide des technologies et de la société, le présent document ne saurait prétendre à l'exhaustivité, mais il insiste sur les impacts les plus importants parmi ceux que l'on peut identifier à ce jour.

II. L'intelligence artificielle : définition

4. Une multitude d'applications de l'IA ont déjà cours au sein de nos sociétés : reconnaissance biométrique (y compris faciale), reconnaissance des objets, prédiction des risques et des réussites, (aide à la) prise de décision algorithmique, traduction automatique, systèmes de recommandations, etc. Ces applications ont trouvé leur place dans des secteurs comme les forces de l'ordre, la justice, la gestion des ressources humaines, les services financiers, les transports, la santé, les services publics, etc.

5. L'IA reste une notion très controversée, sans définition universellement reconnue. Néanmoins, on peut grouper ses définitions en deux familles : rationalistes d'une part, centrées sur l'humain

¹⁶ Rapport préparé pour le CAHAI, Strasbourg 2020, CAHAI(2020)06-fin

¹⁷ Présidente ALLAI, Membre du groupe d'experts de haut-niveau sur l'IA de l'Union Européenne, Rapporteuse EESC sur l'IA

¹⁸ Avis du CESE sur l'IA et la société (INT/806, 2017).

¹⁹ Lignes directrices en matière d'éthique pour une IA digne de confiance, présentées à la Commission européenne par le Groupe d'experts de haut niveau sur l'IA, 2019.

²⁰ Les lignes directrices décrivent plus loin sept exigences que devrait respecter une IA digne de confiance (licite, éthique et robuste) : 1) Action humaine et contrôle humain ; 2) Robustesse technique et sécurité ; 3) Respect de la vie privée et gouvernance des données ; 4) Transparence ; 5) Diversité, non-discrimination et équité ; 6) Bien-être sociétal et environnemental ; 7) Responsabilité.

²¹ Livre blanc sur l'intelligence artificielle, COM(2020)65 final.

d'autre part. L'une des définitions rationalistes les plus connues présente l'IA comme « un agent, créé par des êtres humains, qui prend des décisions et accomplit des actes sur la base de sa perception²² ». L'une des définitions centrées sur l'humain les plus connues est la réussite au test de Turing, destiné aux systèmes informatiques accomplissant une tâche qui requiert habituellement une intelligence humaine (en particulier l'aptitude à la conversation). Le Groupe d'experts de haut niveau sur l'IA a proposé en 2019 une définition de l'IA²³ :

Les systèmes d'intelligence artificielle (IA) sont des systèmes logiciels (et éventuellement matériels) conçus par des êtres humains et qui, ayant reçu un objectif complexe, agissent dans le monde réel ou numérique en percevant leur environnement par l'acquisition de données, en interprétant les données structurées ou non structurées collectées, en appliquant un raisonnement aux connaissances, ou en traitant les informations dérivées de ces données et en décidant de la/des meilleure(s) action(s) à prendre pour atteindre l'objectif donné. Les systèmes d'IA peuvent soit utiliser des règles symboliques, soit apprendre un modèle numérique. Ils peuvent également adapter leur comportement en analysant la manière dont l'environnement est affecté par leurs actions antérieures.

6. L'IA est souvent décrite comme un ensemble de technologies combinant données, algorithmes et capacité de traitement informatique. Bien que ce soit vrai des systèmes d'IA les plus répandus aujourd'hui, cette description ne concerne qu'une petite partie de l'IA. Le terme d'intelligence artificielle s'étend à de nombreuses applications informatiques ; certaines associent des données et des algorithmes, mais il en existe d'autres non fondés sur les données, comme les systèmes experts ou les systèmes de connaissance, raisonnement et représentation, de planification réactive, d'argumentation, etc.

7. La plupart des systèmes d'IA récemment apparus dans nos sociétés sont des exemples d'IA fondée sur les données et ont des incidences particulières sur les droits de l'homme, la démocratie et l'État de droit.

8. Les termes les plus pertinents sont définis ci-après :

- IA restreinte : systèmes d'IA uniquement capables d'accomplir des tâches restreintes, très spécifiques ;
- Données massives (archivées) : systèmes d'IA nécessitant un grand volume de données (archivées) pour fonctionner correctement. La qualité, le volume et le contenu des données influencent le fonctionnement des systèmes et entraînent souvent la reproduction et l'amplification des biais, erreurs et lacunes présents dans les données ;
- Corrélations : beaucoup de systèmes d'IA n'explorent que les relations entre données. Ils n'établissent (ne « voient ») pas de lien de cause à effet entre un cas et une décision, mais effectuent des prédictions sur la base des traits communs avec d'autres cas ;
- Boîtes noires : beaucoup de systèmes d'IA sont ce qu'on appelle des boîtes noires, c'est-à-dire que les processus (décisionnels) qui s'y déroulent ne peuvent être entièrement expliqués en termes humains ;
- Bon sens : les systèmes d'IA n'ont pas de « bon sens ». Même capables de reconnaître un chat ou une cellule cancéreuse, ils ne conçoivent pas ce qu'est un chat ou une cellule cancéreuse, mais se contentent d'étiqueter un phénomène spécifique. Ces systèmes sont incapables, par ailleurs, d'utiliser ce qu'ils « savent » des chats ou des cellules cancéreuses pour reconnaître un chien ou une céphalée.

Toutes ces caractéristiques peuvent rendre l'IA actuelle fragile, instable et imprévisible – mais aussi populaire, et très largement adoptée.

9. Enfin et surtout, les systèmes d'IA sont plus que la somme de leurs composants logiciels. Ils comprennent aussi tout le système sociotechnique qui les entoure. La question de leur gouvernance ne devrait donc pas se focaliser sur la technologie, mais s'étendre aux structures

²² S. J. Russell, P. Norvig et E. Davis (2010), *Artificial intelligence: A modern approach* (3^e éd.), Prentice Hall.

²³ Groupe d'experts de haut niveau de l'UE sur l'IA, A definition of AI, main capabilities and scientific disciplines, 2019.

sociales, c'est-à-dire aux organisations, personnes et institutions qui imaginent, développent, déploient, utilisent et contrôlent les systèmes d'IA et aux personnes que ces systèmes touchent – les citoyens dans leurs relations avec l'administration, les consommateurs, les travailleurs, voire la société tout entière.

i. Définition de l'IA à des fins juridiques

10. Les définitions juridiques sont encore plus difficiles à établir que les définitions purement scientifiques. Elles doivent en effet satisfaire à plusieurs critères²⁴ (être par exemple inclusives, précises, compréhensibles, pratiques et permanentes), dont certains sont obligatoires et d'autres considérés comme des bonnes pratiques²⁵.

III. Impact de l'IA sur les droits de l'Homme, la démocratie et l'État de droit

11. Il est important d'aborder tout le « cycle de vie de l'IA », c'est-à-dire non seulement sa phase de conception, mais aussi son déploiement et son utilisation. Il faut garder à l'esprit, par ailleurs, que la plupart des applications de l'IA actuellement en usage pourraient avoir d'importants impacts sur les droits de l'homme, la démocratie et l'État de droit ou exacerber et amplifier les impacts existants, affectant simultanément de plus en plus de personnes et des pans de plus en plus vastes de la société.

12. L'IA a des répercussions sur quatre « familles de droits de l'homme » en vertu de la CEDH, de ses Protocoles et de la CSE :

- i) Le respect de l'être humain ;
- ii) La liberté individuelle ;
- iii) L'égalité, la non-discrimination et la solidarité ;
- iv) Les droits sociaux et économiques.

En outre, l'IA a de forts impacts sur :

- v) La démocratie ;
- vi) L'État de droit.

Il est important de noter que beaucoup de systèmes d'IA, ou leurs usages, peuvent affecter simultanément la démocratie, l'État de droit et plusieurs droits de l'homme différents. Ils peuvent aussi avoir des effets négatifs sur les droits d'une personne, mais positifs sur ceux d'une autre.

i. IA et respect de l'être humain

13. Le respect de l'être humain transparaît dans plusieurs droits affirmés par la CEDH, comme le droit à la liberté et à la sécurité (article 5), le droit à un procès équitable (article 6), le principe « pas de peine sans loi » (article 7) et le droit à la vie privée et à l'intégrité physique et mentale (article 8). Nous allons voir comment l'IA peut affecter ces droits.

²⁴ « A Legal Definition of AI », Jonas Schuett, Université Goethe de Francfort, 4 septembre 2019 (« Les définitions juridiques doivent être : 1) inclusives : les buts de la réglementation ne doivent pas concerner un public trop vaste ou trop limité ». (Julia Black, *Rules and Regulators*, Oxford University Press, 1997[32] ; Robert Baldwin, Martin Cave et Martin Lodge, *Understanding Regulation: Theory, Strategy, and Practice*, Oxford University Press, 2^e édition, 2012) ; « 2) précises : la limite entre les cas qui relèvent de la définition et ceux qui n'en relèvent pas doit être claire ; 3) compréhensibles : ceux auxquels la réglementation s'applique doivent comprendre la définition ; 4) pratiques : les professionnels du droit devraient pouvoir déterminer aisément si un cas relève ou non de la définition ; 5) permanentes : une définition juridique ne devrait pas nécessiter de constantes mises à jour ».

²⁵ L'obligation d'inclusivité découle du principe de proportionnalité, en droit de l'UE (article 5.4 du traité sur l'Union européenne). Toujours en droit de l'UE, les critères de précision et de caractère compréhensible reposent sur le principe de sécurité juridique. La praticité et la permanence sont considérées comme de bonnes pratiques législatives.

Liberté et sécurité, procès équitable, pas de peine sans loi (art. 5, 6 et 7 CEDH)

14. Le fait que l'IA puisse perpétuer ou amplifier les biais existants est particulièrement significatif lorsqu'elle est utilisée par les forces de l'ordre et le pouvoir judiciaire. Lorsque la liberté physique ou la sécurité d'une personne sont en jeu, par exemple lorsque l'IA sert à prédire des infractions, à estimer le risque de récidive ou à fixer une peine, le droit à la liberté et à la sécurité ainsi que le droit à un procès équitable se trouvent exposés. Un système d'IA servant à prédire la récidive ou à fixer une peine peut livrer des résultats biaisés. Or, si ce système est une boîte noire, les professionnels du droit, comme les juges, les avocats et les procureurs, ne peuvent comprendre par quel raisonnement il a abouti à ses résultats, ce qui rend la décision difficile à motiver et donc à contester.

15. L'IA a aussi un impact, bien que moins évident, sur le droit au soupçon raisonnable et sur l'interdiction de l'arrestation arbitraire. Les applications de police prédictive par IA cherchent uniquement des corrélations, sur la base de traits communs avec d'autres cas. Au lieu de soupçonner un crime ou un délit de la part d'un suspect un particulier, les enquêteurs se fondent alors uniquement sur les caractéristiques que le suspect partage avec d'autres (adresse, revenus, nationalité, niveau d'endettement, profession, comportement, comportement de ses amis ou de ses proches, etc.). En outre, les caractéristiques exactes prises en compte par le système d'IA, et leur pondération, demeurent souvent inconnues.

16. Cependant, maniés de façon responsable, certains outils d'IA peuvent s'avérer bénéfiques pour la sécurité. Certaines applications, par exemple, sont capables de « vieillir » les personnes disparues sur les photographies pour augmenter les chances de les retrouver, ou de reconnaître certains objets dans les bagages scannés à l'embarquement dans un aéroport.

Vie privée et familiale : intégrité physique, psychologique et morale (art. 8 CEDH)

17. Parmi les systèmes d'IA et leurs usages, beaucoup ont de vastes et profondes répercussions sur le droit à la vie privée. Actuellement, les débats dans ce domaine tendent à se centrer sur la protection des données privées et sur la non-discrimination dans le traitement des données personnelles (et non personnelles). Cependant, malgré l'importance de ces aspects, l'impact de l'IA sur notre vie privée dépasse largement le seul champ de nos données personnelles. L'article 8 CEDH protège la vie privée de chaque personne sous un large éventail d'aspects, qu'on peut classer en trois grandes catégories : (i) la vie privée (en général), (ii) l'intégrité physique, psychologique ou morale, et (iii) l'identité et l'autonomie²⁶. Plusieurs usages et applications de l'IA susceptibles d'avoir un impact sur ces catégories n'ont guère attiré l'attention à ce jour.

18. La surveillance (massive) fondée sur l'IA, par reconnaissance faciale par exemple, suppose de saisir, de stocker et de traiter des données personnelles (y compris biométriques, comme nos visages²⁷) ; mais elle affecte aussi notre vie privée en général, notre identité et notre autonomie jusqu'à faire de nous des êtres épiés, suivis et identifiés – parfois en permanence. Cela peut avoir un effet « tétanisant », pousser les personnes à se comporter selon une certaine norme et modifier le rapport de pouvoir entre l'État ou l'entité privée qui utilise la reconnaissance faciale d'une part, et l'individu d'autre part²⁸. La doctrine et les précédents juridiques montrent que l'effet tétanisant de la surveillance peut constituer une violation de l'espace privé nécessaire à l'épanouissement de la personne et au débat démocratique²⁹. Même si nos visages sont effacés juste après saisie, cette technologie reste une intrusion dans notre intégrité psychologique.

19. La reconnaissance faciale a un impact particulièrement évident sur l'intégrité psychologique et sur le droit « général » à la vie privée ; cependant, un effet similaire pourrait découler du pistage généralisé – en ligne et hors ligne – de tous les aspects de nos vies (via notre comportement en

²⁶ Conseil de l'Europe, Guide sur l'article 8 CEDH.

²⁷ La [jurisprudence](#) de la Cour européenne des droits de l'homme (Cour eur. DH) dit clairement que le prélèvement, le stockage et le traitement de telles informations, même pour une courte durée, ont des incidences sur l'article 8 CEDH.

²⁸ Examined Lives: Informational Privacy and the Subject as Object, Julie E. Cohen, 2000.

²⁹ Cet « effet tétanisant » se produit lorsque l'exercice légitime d'un droit se trouve inhibé ou découragé. Cet effet a été démontré : les personnes qui se savent surveillées changent de comportement et connaissent une évolution différente. J. Staben (2016), *Der Abschreckungseffekt auf die Grundrechtsausübung: Strukturen eines verfassungsrechtlichen Arguments*, Mohr Siebeck.

ligne, nos données de localisation et toutes les données issues des objets connectés : montres, applications de santé, haut-parleurs, thermostats, voitures, etc.).

20. D'autres formes de reconnaissance biométrique par IA ont d'encore plus fortes répercussions sur notre intégrité psychologique. Reconnaissance des micro-expressions, de la démarche, du timbre ou du ton de la voix, du rythme cardiaque, de la température, etc. : tous ces procédés sont actuellement appliqués pour évaluer, voire prédire notre comportement, notre état d'esprit et nos émotions.

21. Il faut le noter d'emblée : aucune preuve scientifique solide ne corrobore l'idée qu'il serait possible de « lire » les émotions ou l'état d'esprit d'une personne sur son visage ou dans sa démarche, son rythme cardiaque, le ton de sa voix ou sa température – sans parler de prédire ses comportements futurs. Dans une méta-étude récente, un groupe de scientifiques³⁰ a conclu que la reconnaissance des émotions par IA pouvait, tout au plus, détecter comment une personne *interprétait* subjectivement telle ou telle caractéristique d'une autre personne. Cela ne coïncide pas nécessairement avec les sentiments réels, et l'IA ne fait qu'étiqueter cette interprétation, par ailleurs fortement dépendante du contexte et de la culture. Les déclarations fantaisistes sur ce thème – l'IA pourrait, par exemple, prédire la réussite à un entretien d'embauche sur la base des micro-expressions ou du ton de la voix – n'ont absolument aucun fondement scientifique.

22. Plus important encore, l'usage généralisé de telles techniques d'IA, par exemple pour les recrutements, dans les forces de l'ordre ou dans l'enseignement, a un impact sur l'intégrité physique, psychologique ou morale des intéressés, et donc sur des aspects constitutifs de leur vie privée.

23. Il faut noter que le RGPD n'encadre que dans une certaine mesure le traitement des données biométriques. Il définit les « données biométriques » comme suit : « données à caractère personnel résultant d'un traitement technique spécifique, relatives aux caractéristiques physiques, physiologiques ou comportementales d'une personne physique, qui permettent ou confirment son identification unique ». La dernière partie de la phrase est cruciale, car une reconnaissance biométrique ne visant pas des fins d'identification (mais par exemple de catégorisation, de profilage ou de reconnaissance des émotions) pourrait ne pas entrer dans la définition du RGPD. De fait, le considérant n° 51 du RGPD affirme que les photographies ne sont considérées comme des données biométriques que « lorsqu'elles sont traitées selon un mode technique spécifique permettant l'identification ou l'authentification unique d'une personne physique ».

24. Beaucoup de technologies de reconnaissance biométrique ne visent pas uniquement à identifier une personne, mais aussi à évaluer son comportement (par exemple en cours) ou à catégoriser les personnes (par exemple pour déterminer leur prime d'assurance sur la base de leur exposition statistique à tel ou tel problème de santé). De tels usages pourraient ne pas relever de la définition des données biométriques (et de leur traitement) en vertu du RGPD.

25. Pour en revenir à la confidentialité des données, les données personnelles et non personnelles ne servent pas seulement à entraîner les systèmes d'IA, mais aussi à profiler et à noter les personnes à des fins aussi variées que la police prédictive, la signature d'un contrat d'assurance, l'octroi de prestations sociales ou la prédiction des performances dans le cadre de procédures d'embauche ou de licenciement. En outre, une énorme quantité de « points de données » sur notre vie quotidienne sont utilisés non seulement pour nous adresser des publicités ciblées, mais aussi pour nous amener vers certaines informations (via des pressions/influences/incitations/coups de pouce) et peser ainsi sur nos choix, affectant par là notre intégrité morale.

ii. IA et liberté individuelle

26. La liberté individuelle est affirmée dans plusieurs articles de la CEDH, comme l'article 10 (liberté d'expression) et l'article 11 (liberté de réunion et d'association). L'IA peut aussi avoir un effet tétanisant sur ces libertés.

³⁰ L. Barrett, R. Adolphs, S. Marsella, A. M. Martinez et S. D. Pollak (2019), « Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements », *Psychological Science in the Public Interest*, 20(1), 1–68.

Liberté d'expression (art. 10 CEDH)

27. L'article 10 CEDH affirme le droit à la liberté d'expression et d'information, dont la liberté d'avoir des opinions et de recevoir des informations et des idées. L'usage de l'IA pour profiler, étudier, pister et identifier des personnes et pour filtrer, définir, trier ou influencer plus ou moins ouvertement leurs comportements peut avoir non seulement un impact sur le droit à l'intégrité morale, tel que défini plus haut, mais aussi un effet tétanisant sur ces libertés.

28. Appliquée dans l'espace public, la reconnaissance faciale peut porter atteinte à la liberté individuelle d'opinion et d'expression : en ouvrant la possibilité d'identifier tous les membres d'un groupe, elle supprime en effet la protection qu'offre l'« anonymat de la foule ». Cette situation pourrait pousser les intéressés à changer de comportement – à cesser, par exemple, de se rendre à des manifestations pacifiques³¹.

29. Il en va de même lorsque toutes nos données sont exploitées pour rendre possibles, via l'IA, des notations, évaluations et mesures de performances (en vue d'octroyer ou de refuser, par exemple, un crédit, un emprunt immobilier, un emploi, une promotion, etc.). Le risque est alors que les intéressés hésitent à exprimer ouvertement certaines opinions, à lire certains ouvrages ou journaux en ligne ou à consulter certains médias en ligne.

30. Concernant le droit de recevoir et de communiquer des informations et des idées, l'IA pose problème lorsqu'elle sert à sélectionner des informations et à personnaliser de plus en plus les contenus en ligne et les actualités adressées à telle ou telle personne. Les moteurs de recherche, systèmes de recommandation de vidéos et agrégateurs d'informations sont souvent opaques non seulement sur les données qu'ils utilisent pour sélectionner ou hiérarchiser les contenus, mais aussi sur les finalités d'une telle sélection³². Beaucoup de modèles économiques reposent sur les recettes publicitaires en ligne. Pour que les internautes passent le plus de temps possible sur un site ou une application, il est possible que les systèmes mettent en avant des contenus destinés à cette seule fin (les retenir sur le site), peu importe que ces contenus soient objectifs, avérés, divers ou même présentent un quelconque intérêt.

31. Outre ces considérations commerciales, des motivations politiques ou autres pourraient conduire à optimiser les systèmes d'IA pour qu'ils sélectionnent ou mettent en avant certains contenus en vue d'imposer ou de souligner des opinions en particulier, par exemple dans le cadre d'un processus électoral³³.

32. En outre, l'IA est aujourd'hui de plus en plus capable de produire des contenus médiatiques (sons, images, vidéos) qui imitent l'apparence et/ou la voix de personnes réelles – trucages ultra-réalistes connus sous le nom de « *deep fakes* », ce qui permet de tromper le public, dans des buts divers.

33. Tout cela peut engendrer des bulles de filtres et une prolifération d'infoc, de désinformation et de propagande qui affectent la capacité à se forger ses propres opinions et à recevoir et communiquer des informations et des idées, et ont donc un impact sur la liberté d'expression³⁴.

Liberté de réunion et d'association (art. 11 CEDH)

34. On sait qu'internet et les réseaux sociaux aident à exercer le droit à la liberté d'association et de réunion pacifique. Cependant, le recours à l'IA pourrait aussi compromettre ces libertés, si des personnes ou des groupes de personnes se trouvent automatiquement pistés, identifiés et peut-être même écartés des manifestations ou des mouvements de contestation³⁵.

³¹ Privacy Impact Assessment Report for the Utilization of Facial Recognition Technologies to Identify *Subjects in the Field*, 30 juin 2011, p. 18.

³² J. Burrell (2016): « How the machine 'thinks': Understanding opacity in machine learning algorithms », *Big Data & Society*, 3(1)2053951715622512.

³³ *The Great Hack*, documentaire consacré à Cambridge Analytica disponible sur Netflix

³⁴ Rapport du rapporteur spécial des Nations unies sur la promotion et la protection du droit à la liberté d'opinion et d'expression, A/73/348

³⁵ Algorithmes et droits humains – Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires, Conseil de l'Europe, 2018.

35. Comme déjà évoqué, le recours à la reconnaissance faciale, en particulier dans des espaces publics, pourrait décourager la participation à des manifestations et à des rassemblements pacifiques, qui comptent parmi les éléments les plus importants d'une société démocratique. On en a déjà vu des exemples à Hong Kong, où des manifestants se sont mis à porter des masques et à utiliser des lasers pour éviter d'être « pris » par les caméras de reconnaissance faciale.

iii. IA et égalité, non-discrimination et solidarité

Interdiction de la discrimination (art. 14 CEDH, Protocole n° 12)

36. Parmi les droits de l'homme les plus signalés comme menacés par l'IA figurent l'interdiction de la discrimination et le droit à l'égalité de traitement. Comme nous l'avons vu plus haut, dans de nombreux cas, il s'est avéré que l'IA perpétuait ou amplifiait des partis pris inacceptables, voire en créait de nouveaux. Certains de ces partis pris constituent des discriminations.³⁶ En outre, les systèmes fondés sur les données masquent l'existence de biais et marginalisent les mécanismes de contrôle social qui régissent les comportements humains.

37. Par exemple, Amazon a recruté à l'aide d'une IA favorisant les hommes, car entraînée à partir du profil d'employés « modèles » qui se trouvaient être des hommes. Non content d'exclure les femmes, ce système d'IA avait étudié les caractéristiques de ces employés modèles, comme les formules typiquement utilisées dans leurs CV, et écartait les CV ne présentant pas ces caractéristiques³⁷.

38. Lorsque les systèmes d'IA ne recherchent, comme c'est largement le cas aujourd'hui, que des corrélations fondées sur des traits communs avec d'autres « cas », ils suscitent très facilement de multiples biais. Le problème est que ces systèmes, même s'ils excellent dans l'identification de schémas, comme les expressions typiquement employées par des salariés méritants, ne comprennent pas le sens de ces expressions, sans parler du sens du « mérite » ou de la définition d'un « salarié ». Ils sont uniquement capables d'étiqueter un schéma spécifique.

39. Contrairement à une croyance répandue, tous ces biais ne résultent pas de données de faible qualité. La conception de tout artefact résulte en soi d'une série de choix orientés, depuis les entrants jusqu'aux objectifs. Le système a-t-il pour seul but une efficacité optimale, ou tient-il compte des effets sur les travailleurs et sur l'environnement ? Vise-t-il à détecter un maximum de fraudeurs, ou s'attache-t-il à éviter de pointer des innocents ? Tous ces choix sont orientés, d'une manière ou d'une autre, par les partis pris inhérents à la ou aux personne(s) qui les font. En un mot, il serait illusoire de prétendre éliminer tous les partis pris associés à l'IA³⁸.

iv. IA et droits sociaux et économiques

L'IA intégrée ou liée au lieu de travail

40. Appliquée à des travaux dangereux, pénibles, épuisants, salissants, désagréables, répétitifs ou ennuyeux, l'IA peut présenter un intérêt majeur. Cependant, les systèmes d'IA sont aussi de plus en plus utilisés pour surveiller et pister les travailleurs, répartir les tâches sans intervention humaine et apprécier et prédire le potentiel des travailleurs et leurs performances dans des contextes d'embauche ou de licenciement. De telles applications de l'IA pourraient compromettre le droit à des conditions de travail équitables, à la sécurité et à l'hygiène au travail, ainsi que le droit des travailleurs à s'organiser (CSE, articles 2 et 3). Des salariés constamment surveillés par leur employeur risquent d'hésiter à s'associer au sein d'organisations (article 5). Des systèmes d'IA qui évaluent et prédisent les performances des travailleurs pourraient compromettre le droit à l'égalité de chances et de traitement en matière d'emploi et de profession sans discrimination fondée sur le sexe (article 20 CSE), en particulier si ces systèmes perpétuent des partis pris issus des données ou de leurs concepteurs.

³⁶ Profilage d'affinité et discrimination par association dans la publicité comportementale en ligne, Wachter 2020

³⁷ J. Dastin (10 octobre 2018), « Amazon scraps secret AI recruiting tool that showed bias against women », Reuters.

³⁸ Première analyse du Livre blanc sur l'IA, Virginia Dignum, Catelijne Muller et Andreas Theodorou, 2020

41. Si de plus en plus de tâches et de décisions auparavant opérées par des humains sont transférées à des systèmes d'IA, des compétences nécessaires risquent d'être perdues. Cela pourrait non seulement rendre la main-d'œuvre moins qualifiée, mais aussi créer un risque de défaillance systémique, puisque seuls quelques êtres humains seraient capables de travailler sur les systèmes d'IA et de réagir à une panne.

42. On ignore si l'IA fera perdre ou gagner des emplois et, le cas échéant, dans quelles proportions ; mais pendant la période de bouleversements liée à sa mise en place, l'écart entre une main-d'œuvre de plus en plus vulnérable et les compétences recherchées pourrait entraîner une situation de « chômage technologique³⁹».

v. IA et démocratie

43. L'IA peut avoir (et a probablement déjà) un effet négatif sur la démocratie, en particulier sur les aspects suivants : 1) discours social et politique, accès aux informations et influence sur les électeurs, 2) inégalité et ségrégation, et 3) défaillance ou disruption systémique.

Discours social et politique, accès aux informations et influence sur les électeurs

44. Une démocratie en bonne santé suppose des citoyens éclairés, un discours social et politique ouvert et l'absence d'influence souterraine sur les électeurs.

45. À cet égard, la bonne information des citoyens est cruciale. Au sein de nos sociétés de l'information, les citoyens peuvent choisir de ne consulter qu'une petite part des informations disponibles. Les moteurs de recherche, fils d'actualité sur les réseaux sociaux, systèmes de recommandation, et de nombreux sites d'actualités emploient l'IA pour déterminer quel contenu créer et montrer aux utilisateurs (en un mot, les informations sont « personnalisées »). Bien conçu, ce procédé peut aider les citoyens à se repérer dans les flots d'informations disponibles et à améliorer leurs compétences démocratiques, par exemple en leur permettant d'accéder à des ressources en d'autres langues grâce aux outils de traduction.⁴⁰ Cependant, si l'IA décide quelles informations afficher, et quels sujets présenter en boucle ou, au contraire, noyer dans le flux des informations en ligne, la représentation des différentes voix et opinions risque de devenir inégale et biaisée.

Le suivi et le profilage constants de chaque individu rendent possible la **personnalisation des informations par IA**. Cette infrastructure informationnelle, fondée sur les technologies et mue par des considérations commerciales ou politiques, pourrait amplifier les messages hyperpartisans avec lesquels nous sommes susceptibles d'être d'accord, et fournir un outil d'influence individualisée d'une puissance sans précédent⁴¹. Elle pourrait ainsi saper les visions communes, le respect mutuel et la cohésion sociale dont la démocratie a besoin pour s'épanouir. Les prédictions personnalisées par IA, si elles gagnent en puissance et en efficacité, pourraient même menacer la capacité de choix et l'autonomie humaines nécessaires pour que les électeurs prennent de véritables décisions⁴²

46. Troisièmement, l'IA peut nuire à l'équité des processus électoraux. Des responsables de campagne ou des acteurs étrangers peuvent utiliser (et ont utilisé) des publicités personnalisées pour adresser des messages différents à différents groupes d'électeurs, sans avoir à rendre de comptes sur ce procédé.⁴³

³⁹ E. Brynjolfsson et A. McAfee (2014), *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*, W. W Norton & Company

⁴⁰ R. Schroeder (2018), *Social Theory after the Internet*, UCL Press.

⁴¹ T. Wu (2016), *The attention merchants: The epic scramble to get inside our heads* (1re édition) ; S. Zuboff (2019), *The age of surveillance capitalism: The fight for the future at the new frontier of power*.

⁴² M. Taddeo et L. Floridi (2018b), « How AI can be a force for good », *Science*, 361(6404), 751–752.

⁴³ S. Bradshaw et P. Howard (2019), *Social Media and Democracy in Crisis*, Oxford University Press

Toutefois, il faut noter que le doute plane encore sur le degré d'influence exact des publicités micro-ciblées⁴⁴. L'IA peut aussi servir à créer et à répandre de la désinformation et des trucages hyperréalistes, sous forme de textes, d'images, de sons ou de vidéos. Pour les citoyens et les journalistes comme pour les institutions publiques, ces trucages s'avèrent difficiles à repérer. Il devient donc de plus en plus facile de tromper et de manipuler le public, ce qui pourrait dégrader le sérieux et la crédibilité des médias et du discours démocratique.

Inégalités et ségrégation

47. On attend très largement de l'IA qu'elle améliore la productivité de l'économie. Toutefois, il est également probable que ces gains de productivité soient inégalement répartis et bénéficient surtout aux plus riches. De même, il est démontré que la sélection des données et les choix de conception, associés au manque de transparence des algorithmes « boîtes noires », perpétuent les partis pris injustes à l'encontre des groupes sociaux défavorisés, comme les femmes ou les minorités ethniques⁴⁵. L'IA pourrait entraîner des inégalités et une ségrégation, et donc menacer le niveau d'égalité économique et sociale nécessaire à l'épanouissement de la démocratie.

Risques systémiques

48. Le transfert à l'IA de décisions que seuls des humains étaient auparavant capables de prendre fait peser de nouveaux défis sur la sécurité et la résilience des systèmes sociétaux. En particulier, si de multiples décideurs décentralisés se trouvent remplacés par quelques systèmes centralisés pilotés par IA, les risques systémiques augmentent : une seule panne sur un système peut produire des résultats catastrophiques.

49. Comme le montre l'augmentation des krachs éclair, ces nouveaux risques systémiques sont déjà une réalité sur les marchés financiers, où différents agents d'IA interagissent à une vitesse surhumaine⁴⁶. Le fait que les infrastructures énergétiques essentielles, les réseaux de transports ou les hôpitaux dépendent de plus en plus de décisions prises automatiquement par des agents artificiels crée de nouvelles vulnérabilités, puisqu'une atteinte sur un seul point faible peut avoir des effets dévastateurs. Une fois installés de tels systèmes d'infrastructures essentielles, il devient plus difficile⁴⁷ de les remplacer ou de leur trouver des substituts en cas de dysfonctionnement.

La vision du développement de l'IA comme une course avec la concurrence compromet particulièrement la paix et la sécurité internationales. Dans ces conditions, l'IA promet non seulement d'entraîner des effets secondaires indésirables, mais aussi d'aider des acteurs malveillants, des cybercriminels aux États totalitaires, à resserrer leur emprise sur les populations.

Concentration des pouvoirs numériques

50. Beaucoup d'applications d'IA sont développées et déployées pour une poignée de grands groupes privés, parfois surnommés les géants du Net, les Gafam ou même les *Frightful Five* (« les Cinq qui font peur⁴⁸ »). Si quelques acteurs privés, pour lesquels la valeur des actions passe avant le bien commun, concentrent trop de pouvoir politique, l'autorité des États démocratiques peut se trouver menacée.

⁴⁴ R.M. Bond, C.J. Fariss, J.J. Jones, A.D. Kramer, C. Marlow, J.E. Settle et J.H. Fowler (2012), « A 61-million person experiment in social and political mobilization », *Nature*, 489(7415), 295-298

⁴⁵ V. Eubanks (2017), *Automating inequality: How high-tech tools profile, police, and punish the poor* (1re édition), St. Martin's Press ; C. O'Neil (2016), *Weapons of math destruction: How big data increases inequality and threatens democracy* (1re édition).

⁴⁶ M. P. Wellman et U. Rajan (2017), *Ethical Issues for Autonomous Trading Agents*. *Minds and Machines*, 27(4), 609-624.

⁴⁷ M. Brundage, S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, P. Scharre, T. Zeitzoff, B. Filar, H. Anderson, H. Roff, G. C. Allen, J. Steinhardt, C. Flynn, S. Ó. Éigeartaigh, S. Beard, H. Belfield, S. Farquhar, D. Amodei (2018), *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*.

⁴⁸ Google, Apple, Facebook, Amazon et Microsoft. Voir : P. Nemitz (2018), « Constitutional democracy and technology in the age of artificial intelligence », *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180089. A. Webb (2019), *The Big Nine: How the tech titans and their thinking machines could warp humanity*.

vi. IA et État de droit

51. Les institutions démocratiques ont une responsabilité particulière à l'égard des individus et de la société, qui se traduit en principes comme la proportionnalité, l'égalité ou la motivation des décisions. D'un côté, l'IA peut rendre ces institutions plus efficaces ; mais de l'autre, elle peut aussi affaiblir leur légitimité procédurale, la confiance dont elles jouissent, et l'autorité du droit.

52. Les tribunaux, les autorités répressives et les administrations pourraient ainsi devenir plus efficaces, mais aussi plus opaques et laisser moins de place aux qualités humaines (choix, autonomie, supervision⁴⁹).

53. De même, alors qu'auparavant, seuls les tribunaux pouvaient dire ce qui constituait ou non un discours de haine illégal, ce sont aujourd'hui avant tout des systèmes d'IA privés qui désignent les messages à retirer des sites de réseaux sociaux⁵⁰. Ces systèmes d'IA empiètent de fait sur l'autorité des juges et du droit, et en général, l'IA peut favoriser le développement de systèmes judiciaires fonctionnant hors des limites et des protections de l'État de droit.

54. Ainsi, les règlements automatiques des litiges en ligne proposés par des entreprises privées reposent sur leurs conditions d'utilisation, et non sur la loi, et n'offrent pas aux consommateurs les mêmes droits et les mêmes garanties procédurales que les juridictions publiques⁵¹.

55. Dès 2018, la Commission européenne pour l'efficacité de la justice définissait cinq principes pour l'usage de l'IA dans le système judiciaire, dans un texte intitulé « Charte éthique européenne d'utilisation de l'IA dans les systèmes judiciaires et leur environnement ». Le Groupe d'experts de haut niveau sur l'IA a appelé les organismes publics à respecter, lorsqu'ils conçoivent, achètent ou utilisent l'IA, sept exigences pour une IA digne de confiance. Des principes et exigences similaires devraient être imposés aux services répressifs.

56. Toutefois, l'IA ne fait pas que menacer l'État de droit : elle peut aussi le renforcer⁵². Développée et appliquée de manière responsable, elle peut aider les autorités à détecter les cas de **corruption**⁵³. Elle peut aussi servir à détecter les cyberattaques et à s'en protéger⁵⁴.

IV. Que faire face à l'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit ?

57. Depuis peu, l'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit attire de plus en plus l'attention, notamment grâce à la parution du Livre blanc sur l'IA de la Commission européenne. Les réponses possibles à cet impact, cependant, restent à ce jour incertaines. Le présent chapitre décrit les stratégies pouvant être adoptées au sein du cadre existant en matière de droits de l'homme, de démocratie et d'État de droit. La liste n'est pas exhaustive ; le but est de favoriser le passage à la phase de débat suivante.

i. Les droits de l'homme dans le contexte de l'IA

58. Parmi les acteurs (publics et privés) qui conçoivent, déploient et utilisent l'IA, beaucoup ne semblent pas avoir conscience des impacts – potentiels – de l'IA sur les droits de l'homme. Dans un premier temps, il serait très utile de se livrer à un exercice de reformulation, c'est-à-dire de « traduire

⁴⁹ J. Danaher (2016), « The Threat of Algocracy: Reality, Resistance and Accommodation », *Philosophy & Technology*, 29(3), 245–268.
J. Weizenbaum (1976), *Computer Power and Human Reason: From Judgment to Calculation*, W. H. Freeman.

⁵⁰ J. E. Cohen (2019), *Between truth and power: The legal constructions of informational capitalism*.

⁵¹ J. Susskind (2018), *Future politics: Living together in a world transformed by tech*.

⁵² R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark et F. F. Nerini (2020), « The role of artificial intelligence in achieving the Sustainable Development Goals », *Nature Communications*, 11(1), 1–10.

⁵³ J. West et M. Bhattacharya (2016), « Intelligent financial fraud detection: A comprehensive review », *Computers & Security*, 57, 47–66. P. Hajek et R. Henriques (2017), « Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods », *Knowledge-Based Systems*, 128, 139–152.

⁵⁴ M. Taddeo et L. Floridi (2018a), « Regulate artificial intelligence to avert cyber arms race », *Nature*, 556(7701), 296–298.

» dans le contexte de l'IA les droits affirmés dans la CEDH, éventuellement au moyen d'une Convention- cadre.

ii. Mesures visant la conformité, la réparation et l'obligation de rendre des comptes

Pour faire face à l'impact de l'IA sur la démocratie, l'État de droit et les droits de l'homme existants, certains mécanismes actuels visant la conformité, la réparation et l'obligation de rendre des comptes pourraient être étoffés, avec en parallèle la création de nouveaux mécanismes. Cependant, l'utilisation de l'IA est souvent dissimulée ou ignorée, ce qui rend difficile, voire impossible de savoir s'il existe ou non un impact sur les droits de l'homme, la démocratie et l'État de droit. Toute mesure devrait donc commencer par une obligation de transparence sur l'utilisation de systèmes d'IA susceptibles de se répercuter sur les droits de l'homme, la démocratie et l'État de droit. Cela passerait par un répertoire des systèmes d'IA, qui préciserait la catégorie de risque et le degré de transparence et d'obligation de rendre des comptes pour chaque application.

59. La conformité pourrait donc commencer par une nouvelle culture consistant à intégrer, selon une expression récente, « les droits de l'homme, la démocratie et l'État de droit dès la conception⁵⁵ ». Les acteurs qui conçoivent, déploient et utilisent l'IA s'interrogeraient dès le départ sur ses effets possibles sur les droits de l'homme, la démocratie et l'État de droit, et ajusteraient en conséquence la technologie ou ses applications. Cette nouvelle culture pourrait être renforcée par l'obligation (juridique) de réaliser une étude d'impact de l'IA sur les droits de l'homme, la démocratie et l'État de droit.

60. Cette nouvelle culture devrait aussi englober l'obligation de rendre des comptes sur la structure mise en place, mais aussi sur les résultats de cette étude d'impact et sur les décisions prises en matière de conception et de gouvernance à la suite de l'étude d'impact.

61. Dans le contexte de l'impact de l'IA sur les droits de l'homme, la réparation suppose l'accès à la justice et l'existence de recours effectifs. S'agissant de l'accès à la justice, il est peut-être trop tôt pour dire s'il est suffisamment garanti au regard de l'IA et de son impact sur les droits de l'homme. Les tout premiers jugements nationaux concernant l'impact (potentiel) de l'IA sur les droits de l'homme ont été rendus très récemment, et ils concernent un droit en particulier : le respect de la vie privée (article 8 CEDH⁵⁶).

62. Point important, l'accès à la justice est d'autant plus difficile que beaucoup d'applications d'IA ne sont développées et déployées que par une poignée de grands acteurs privés. Ces entreprises dominent à la fois l'élaboration de systèmes d'IA et les écosystèmes au sein desquels ils fonctionnent. Les États sont tenus de protéger les individus et les groupes contre les atteintes aux droits de l'homme commises par des tiers, et la reconnaissance de l'influence des acteurs non étatiques sur les droits de l'homme ne cesse de croître⁵⁷. Aujourd'hui, les géants du numérique sont capables de déterminer, voire de modifier nos structures sociales et même démocratiques ; l'impact de leurs systèmes d'IA, et de l'usage qu'ils en font, sur les droits de l'homme revêt donc une importance considérable. L'IA pourrait offrir une bonne raison de réfléchir à une structure qui contraindrait juridiquement les acteurs privés à respecter les droits de l'homme, et à permettre l'accès à la justice s'ils manquent à cette obligation⁵⁸. La question fondamentale est la suivante : faut-il a) reconnaître le pouvoir des entreprises privées utilisant l'IA et s'assurer qu'elles l'exercent de façon responsable, ou b) le contester et réaffirmer le pouvoir de l'État ?

⁵⁵ P. Nemitz (2018), *Constitutional democracy and technology in the age of artificial intelligence*.

⁵⁶ Un tribunal néerlandais a jugé qu'une loi ouvrant aux institutions publiques la possibilité d'utiliser l'IA pour prédire les fraudes aux prestations sociales violait le droit à la vie privée, inscrit dans l'article 8 CEDH ; au Royaume-Uni, la Haute Cour de justice, siégeant à Cardiff, a admis que la reconnaissance faciale affectait l'article 8 CEDH, puisqu'elle permet de recueillir des informations « par définition privées », mais a jugé son utilisation légale au regard du principe de proportionnalité. En France, un tribunal a considéré que l'utilisation de la reconnaissance faciale dans des établissements scolaires violait l'article 8 CEDH.

⁵⁷ *Business and Human Rights, A Handbook for Legal Practitioners*, Claire Methven O'Brien, Conseil de l'Europe.

⁵⁸ Cela suppose d'aller au-delà des préconisations de la Recommandation CM/Rec(2016)3 du Comité des Ministres du Conseil de l'Europe sur les droits de l'homme et les entreprises (et des Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme).

63. Concernant les recours effectifs, Dinah Shelton observe que lorsqu'il s'agit d'IA, ils ne consistent « pas seulement à indemniser la victime ; ils expriment l'opprobre de toute la société sur le fautif » et, par-là, « affirment, renforcent et concrétisent les valeurs fondamentales de la société⁵⁹ ». Dans son arrêt *Broniowski*, la Cour européenne des droits de l'homme a souligné qu'en vertu du droit international, il était souhaitable que réparation individuelle et réparation générale « aillent de pair⁶⁰ ».

64. Pour assurer un recours effectif en cas d'atteinte aux droits de l'homme du fait de l'IA, il faut donc étudier **à la fois** les moyens de réparation individuels et généraux. En outre, les recours devraient être adaptés aux applications de l'IA visées, qui sont multiples : surveillance et identification, profilage, influence, aide à la décision. Pour être effectifs, ils devraient comprendre l'arrêt de la conduite illicite et des garanties pour qu'elle ne se reproduise pas; les États pourraient, par exemple, être tenus d'adopter et d'appliquer une législation contraignante pour éviter de futurs impacts de l'IA sur les droits de l'homme. Il devrait y avoir obligation de réparer les dommages ou les préjudices entraînés par la violation, soit pour un individu, soit pour une communauté. Pour certaines applications de l'IA, assurer simplement un recours effectif pourrait ne pas suffire à traiter l'impact de ces applications sur les droits de l'homme. Des mesures plus larges pourraient s'imposer, comme une interdiction ou des restrictions d'utilisation (voir le chapitre IV).

iii. Protection des structures démocratiques et de l'État de droit

65. Pour éviter les défaillances ou interruptions systémiques dues aux prises de décisions centralisées par IA au sein de structures vitales, et donc prévenir les risques de pannes catastrophiques, il convient de répartir les processus décisionnels au lieu de le centraliser. Ces processus devraient intégrer de solides mécanismes de supervision humaine⁶¹

66. La supervision d'un système d'IA par des êtres humains garantit que ce système ne sape pas, entre autres effets adverses, l'autonomie humaine. Cette supervision peut être assurée par plusieurs mécanismes de gouvernance, comme ceux des humains « dans la boucle » (*human-in-the-loop*, HITL), « sur la boucle » (*human-on-the-loop*, HOTL) ou « aux commandes » (*human-in-command*, HIC). L'approche HITL désigne la possibilité d'une intervention humaine à tous les cycles décisionnels du système, intervention qui n'est dans de nombreux cas ni possible ni souhaitable. L'approche HOTL désigne la possibilité d'une intervention humaine au cours du cycle de conception du système et du suivi de son fonctionnement. L'approche HIC désigne la possibilité de superviser l'activité globale du système d'IA (y compris son impact économique, sociétal, juridique et éthique au sens large) et celle de décider quand et comment utiliser le système dans telle ou telle situation. Cela peut englober la décision de ne pas utiliser de système d'IA dans une situation donnée, de prévoir certains niveaux de latitude humaine lors de l'utilisation du système ou de permettre de passer outre une décision prise par le système.

67. Pour remédier au risque d'inégalité, les pouvoirs publics doivent activement enrayer l'usage des applications d'IA qui aggravent les inégalités.

68. Prévenir les influences sur les élections ou les manipulations du public au travers des informations personnalisées par IA n'est pas une tâche facile. On pourrait envisager une réglementation des campagnes en ligne, soit pour les médias en ligne et sites de réseaux sociaux, soit pour les partis politiques. Cela soulève évidemment des questions de liberté d'expression. La présence d'êtres humains « dans/sur la boucle » et « aux commandes » (voir plus haut) pourrait aider à détecter et à éliminer les influences indésirables sur les électeurs.

69. Pour assurer un usage responsable de l'IA dans les services publics, les passations de marchés publics représentent un outil crucial. La mise à jour des exigences légales associées aux commandes **publiques**, pour y intégrer des critères comme l'équité, l'obligation de rendre des comptes et la transparence sur l'IA, pourrait servir deux fins. Premièrement, elle garantirait que les pouvoirs publics n'utilisent que des systèmes compatibles avec l'État de droit ; deuxièmement, elle

⁵⁹ Dinah Shelton, « The Right to Reparations for Acts of Torture: What Right, What Remedies? », 17(2) *Torture* 96 (2007), 96.

⁶⁰ Cour eur. DH, *Broniowski c. Pologne*.

⁶¹ Groupe d'experts de haut niveau sur l'IA, *Lignes directrices en matière d'éthique pour une IA digne de confiance*, 2019.

inciterait le secteur privé, par intérêt économique, à développer et utiliser des systèmes conformes aux principes de l'État de droit. En outre, l'utilisation de l'IA par les pouvoirs publics devrait être soumise à des mécanismes de supervision, dont des ordonnances judiciaires et la possibilité de saisir des médiateurs.

V. Et si les normes actuelles de droits de l'homme, de démocratie et d'État de droit ne nous protégeaient pas assez ?

i. Question zéro

70. Vu le caractère intrusif de certaines applications ou utilisations de l'IA, il pourrait arriver que le cadre actuel en matière de droits de l'homme, de démocratie et d'État de droit ne nous protège pas assez, ou pas à temps. Nous devons alors peut-être prendre le temps de la réflexion, et chercher la bonne réponse à ce qu'on pourrait appeler la « question zéro » : allons-nous autoriser ce système ou cet usage et si oui, dans quelles conditions ? On ne peut y répondre sans examiner le système d'IA ou son usage sous tous les angles, avec pour résultat plusieurs « solutions » possibles :

- Le système d'IA ou son usage est placé sous moratoire, interdit (temporairement ou définitivement) ou soumis à restrictions (« lignes rouges ») ;
- Pour parer aux nouveaux impacts négatifs de l'IA, de nouveaux droits de l'homme sont adoptés ;
- Les droits de l'homme existants sont adaptés pour permettre le développement et l'utilisation responsables de l'IA ;
- Un système ou un usage particulier est soumis à un mécanisme de contrôle démocratique spécifique ;
- Les propriétaires privés des puissants systèmes d'IA sont obligés d'aligner leurs structures de développement et de gouvernance de l'IA sur les intérêts de ceux que les systèmes affectent et de l'ensemble de la société, ce qui peut appeler des mesures pour associer les parties concernées (travailleurs, consommateurs, clients, citoyens, responsables politiques, etc.).

71. D'abord et avant tout, l'impact de l'IA doit être étudié à la fois au niveau individuel et au niveau sociétal/collectif, puisqu'il peut toucher aussi bien une personne en particulier que de larges pans de notre société. Deuxièmement, pour décider quelle action est pertinente et proportionnée à la situation, il est important de connaître le contexte, la finalité, la gravité, l'ampleur et la probabilité de l'impact. Pour les applications d'IA qui engendrent des risques inacceptables ou créent la menace d'un préjudice ou d'une défaillance systémique, il convient d'adopter une approche réglementaire fondée sur des principes, dont le principe de précaution. Pour les autres applications d'IA, une approche fondée sur les risques pourrait s'avérer plus adéquate.

ii. Lignes rouges

72. Des lignes rouges pourraient être définies, et appliquées aux systèmes ou usages de l'IA dont les conséquences sont trop lourdes pour qu'ils restent sans contrôle ni réglementation – ou même pour qu'ils soient autorisés. Les applications suivantes pourraient rendre nécessaires une interdiction, un moratoire et/ou de strictes limites ou conditions déterminant des usages exceptionnels et/ou contrôlés :

- Le recours généralisé à la reconnaissance faciale et à d'autres formes de reconnaissance biométrique par des acteurs publics ou privés ;
- La surveillance de masse fondée sur l'IA (via la reconnaissance faciale/biométrique, mais aussi d'autres formes de pistage et/ou d'identification par IA comme la géolocalisation, le comportement en ligne, etc.) ;
- Le pistage, l'évaluation, le profilage, la notation et les incitations, à caractère personnel,

physique ou psychologique, via la reconnaissance biométrique et la reconnaissance des comportements ;

- La notation sociale assistée par IA ;
- Les systèmes d'IA dissimulés et les trucages hyperréalistes (deep fakes) ;
- Les interfaces IA-êtres humains.

73. L'usage exceptionnel de telles technologies, par exemple à des fins de sécurité nationale, de diagnostic ou de traitement médical, devrait être fondé sur des preuves, nécessaire et proportionné, et n'être autorisé que dans des environnements contrôlés et (si possible) pour une durée limitée.

iii. Des droits de l'homme nouveaux ou adaptés

74. Outre la définition de lignes rouges, on pourrait envisager de créer ou d'adapter les droits de l'homme suivants (liste non exhaustive) :

- Droit des êtres humains à l'autonomie, à la liberté d'action et à l'exercice d'une supervision sur l'IA ;
- Droit à la transparence/au caractère explicable des résultats de l'IA, dont le droit à des explications sur la façon dont l'IA fonctionne, la logique qu'elle suit et les effets de son utilisation sur les intérêts de la personne concernée, même si le système d'IA ne traite pas de données personnelles – s'il traite de telles données, ce droit à l'information existe déjà en vertu du RGPD⁶².
- Droit distinct à l'intégrité physique, psychologique et morale au regard du profilage et de la reconnaissance des émotions par IA ;
- Droit renforcé à la vie privée, protégeant contre la surveillance de masse assistée par IA ;
- Adaptation du droit à la confidentialité des données pour protéger contre le pistage en ligne des individus, généralisé et à l'échelle de toute une société, utilisant des données personnelles ou non personnelles (ces dernières livrant souvent des indices permettant l'identification).

75. Il ne devrait être permis de déroger à ces droits que dans des circonstances exceptionnelles – à des fins de sécurité par exemple –, dans des conditions strictes et dans le respect du principe de proportionnalité.

iv. Scénarios d'avenir

76. Dans une perspective à plus long terme, des problèmes majeurs susceptibles de se poser à l'avenir font actuellement l'objet de recherches ; les inconnues qui subsistent, et la possibilité d'imprévus catastrophiques, appellent une approche fondée sur les risques. Parmi les problèmes à redouter (mais n'existant pas aujourd'hui), certains rangent l'intelligence artificielle généralisée, la conscience artificielle, les agents moraux artificiels ou la superintelligence, tandis que beaucoup d'autres les jugent irréalistes. Quoi qu'il en soit, il est nécessaire de suivre de près les évolutions, afin de déterminer dans quelle mesure les systèmes qui encadrent actuellement les droits de l'homme, la démocratie et l'État de droit demandent à être adaptés

⁶² P. Nemitz (2018), Constitutional democracy and technology in the age of artificial intelligence.

ALLAI : Stichting ALLAI Nederland, fondation de droit néerlandais

Les informations figurant dans ce rapport provisoire n'ont qu'un caractère général et indicatif. Ni l'ALLAI ni les membres de son conseil d'administration et de son conseil consultatif, ses employés et agents, les organisations ou personnes qui leur sont associées ou les entités liées (« réseau ALLAI ») n'offrent, à travers ce rapport, des conseils ou services professionnels. Avant de prendre toute décision ou d'entreprendre toute action susceptible d'affecter votre organisation ou votre situation financière, veuillez solliciter un consultant professionnel qualifié. Aucune perte ou dommage de quelque nature que ce soit, directe ou indirecte, subie par une entité ou une personne du fait de son usage du présent rapport, ne peut être imputée à une entité ou personne du réseau ALLAI, dont la responsabilité ne saurait être engagée.

CHAPITRE II. Lignes directrices sur l'éthique en matière d'IA : situation en Europe et dans le monde⁶³

Marcello Ienca⁶⁴ et Effy Vayena⁶⁵

I. Résumé

Ces dernières années, des entreprises du secteur privé, des instituts de recherche et des organismes du secteur public ont énoncé des principes et diffusé des lignes directrices et d'autres instruments non contraignants au sujet de l'utilisation éthique de l'intelligence artificielle (IA). Cependant, malgré un consensus manifeste sur le fait que l'IA doit être « éthique », il y a débat sur ce qu'il faut entendre par « IA éthique » et sur les exigences éthiques, les normes techniques et les meilleures pratiques nécessaires à cet effet. Il s'agit ici d'analyser le corpus de documents de droit souple et autres cadres éthico-juridiques définis par des institutions gouvernementales et des entités non gouvernementales du monde entier, et ce à deux fins.

Tout d'abord, nous souhaitons étudier cet ensemble d'instruments de gouvernance non contraignants qui est en constante évolution.

Ensuite, nous voulons analyser les incidences que l'IA pourrait avoir sur les principes éthiques, les droits de l'homme, l'État de droit et la démocratie. Le rapport s'appuie sur un protocole d'examen de la portée, qui a été adapté et préalablement validé et dont l'objet était d'avoir un aperçu complet et actualisé des dispositions non contraignantes en vigueur. En tout, nous avons examiné 116 documents, entre autres publiés par des institutions publiques, des entités non gouvernementales, des établissements universitaires et des entreprises du secteur privé.

Notre analyse fait ressortir cinq grands groupes de principes éthiques, dont elle examine la place dans le discours actuel sur la gouvernance. En revanche, elle montre qu'en l'état actuel des choses, le droit souple est lacunaire. Nous avons en outre établi un lien entre les principes éthiques et les droits de l'homme, plus spécifiquement les droits et libertés inscrits dans la Convention européenne des droits de l'homme (CEDH), pour évaluer dans quelle mesure les cadres de gouvernance non contraignants en vigueur offrent une protection intégrale des droits de l'homme.

Enfin, nous avons dressé de façon empirique la liste des incidences sur l'élaboration des politiques, et ce à l'intention des scientifiques, des établissements de recherche, des organismes de financement, des organismes gouvernementaux et intergouvernementaux, et autres acteurs concernés par la promotion d'innovations éthiquement responsables en matière d'IA.

II. Principaux constats

- De plus en plus d'institutions gouvernementales et d'entités non gouvernementales (notamment des entreprises du secteur privé et des établissements universitaires) élaborent des lignes directrices éthiques et d'autres instruments non contraignants au sujet de l'IA.
- Ces documents de droit souple sont principalement rédigés en Europe, en Amérique du Nord et en Asie. L'hémisphère sud est à l'heure actuelle sous-représenté parmi les organismes proposant des lignes directrices éthiques sur l'IA.
- À l'heure actuelle, les lignes directrices sur l'IA s'accordent le plus souvent sur certains principes généraux mais absolument pas sur les mesures concrètes à prendre. En outre, aucun des différents principes éthiques n'est commun aux 116 documents sur l'IA éthique que nous avons examinés.

⁶³ Rapport préparé pour le CAHAI. Strasbourg, 15 juin 2020, intitulé CAHAI(2020)07-fin.

⁶⁴ Chaire de bioéthique, Laboratoire d'éthique et de politique de la santé, Département des sciences et technologies de la santé, ETH Zurich.

⁶⁵ Chaire de bioéthique, Laboratoire d'éthique et de politique de la santé, Département des sciences et technologies de la santé, ETH Zurich.

- Nous avons constaté qu'un consensus se dégagait de plus en plus autour des principes éthiques suivants : transparence, justice, non-malfaisance, responsabilité et respect de la vie privée. Les notions éthiques de durabilité, de dignité et de solidarité semblent nettement moins traitées.
- Si la plupart des lignes directrices s'accordent à dire que l'IA doit être *transparente* pour éviter d'éventuels problèmes, on ne sait en revanche pas clairement si la transparence doit passer par la publication du code source, par les bases de données sous-jacentes ou par divers autres moyens.
- Un peu plus de la moitié des textes de droit souple qui ont été examinés recommandent explicitement de promouvoir les droits de l'homme ou mettent en garde contre leur violation, lors de la conception, du développement et du déploiement de systèmes d'IA.
- Les expressions courantes tirées des codes montrent d'importantes différences dans les thèmes qu'abordent les documents issus des pays membres du Conseil de l'Europe et ceux que traitent les documents élaborés ailleurs. Les documents de droit souple établis dans les pays membres du Conseil de l'Europe semblent mettre l'accent sur les principes éthiques de solidarité, de confiance et de fiabilité, contrairement aux documents établis dans des pays du reste du monde. Ils semblent en revanche faire plus rarement référence aux principes de bienfaisance et de dignité.
- Les principes de respect de la vie privée, de justice et de loyauté sont ceux qui varient le moins entre les pays membres du Conseil de l'Europe, les pays observateurs du Conseil de l'Europe et le reste du monde, et donc ceux qui ont le plus haut degré de constance transgéographique et transculturelle.

III. Principales incidences sur l'élaboration des politiques :

- Les instruments non contraignants que diffusent les institutions gouvernementales et les entités non gouvernementales (en ce compris les entreprises du secteur privé et les établissements universitaires) sont de précieux outils permettant d'exercer une influence sur les décisions publiques relatives à l'IA et d'orienter le développement des systèmes d'IA en faveur du bien-être social et du respect des valeurs éthiques et des normes légales. Il ne faut toutefois pas considérer qu'ils remplacent les instruments de gouvernance contraignants. Pour cause de conflit d'intérêts, il est particulièrement à craindre que les mesures d'autoréglementation prises par les acteurs du secteur privé de l'IA ne servent à contourner ou à écarter les mesures de gouvernance contraignantes édictées par les autorités gouvernementales ou intergouvernementales.
- Par souci d'inclusivité, de pluralisme culturel et de participation équitable à la prise de décisions collective au sujet de l'IA, il faudrait encourager l'élaboration de documents de droit souple par des organismes situés dans les régions du monde qui sont actuellement sous-représentées, tout particulièrement l'Afrique et l'Amérique du Sud.
- Les cinq principes éthiques génériques – transparence, justice, non-malfaisance, responsabilité et respect de la vie privée – autour desquels convergent les instruments non contraignants en vigueur sont à surveiller en priorité et pourraient faire l'objet de mesures de gouvernance obligatoires prises aux niveaux gouvernemental et intergouvernemental.
- Afin que ces principes éthiques puissent être dûment traduits en règles de gouvernance, ils doivent être précisés sur le plan conceptuel. Les décideurs se doivent de remédier aux ambiguïtés sémantiques de ces principes et à leurs caractéristiques antagonistes.
- Les profondes divergences entre les textes de droit souple en vigueur au sujet de l'interprétation et de la mise en œuvre concrète de ces principes montrent que le public sera probablement divisé au sujet des solutions de gouvernance contraignantes et qu'un débat démocratique transparent sera donc nécessaire.
- Les notions éthiques qui sont très peu traitées, par exemple la durabilité, la dignité et la solidarité, doivent être analysées de très près pour que les lacunes conceptuelles et normatives que présente le droit souple ne soient pas transposées dans les mesures obligatoires.
- Comme près de la moitié des documents de droit souple que nous avons étudiés ne recommandent pas expressément de promouvoir les droits de l'homme – pas plus qu'ils ne mettent pas en garde contre leur violation – lors de la conception, du développement et du déploiement des

systèmes d'IA, il est urgent de se focaliser davantage sur les incidences de l'IA sur les droits de l'homme.

➤ Les pays membres du Conseil de l'Europe sont bien placés pour orienter la gouvernance internationale de l'IA vers la promotion des droits de l'homme.

IV. Introduction

L'intelligence artificielle (IA) est la conception et le développement de systèmes informatiques capables d'exécuter des tâches nécessitant normalement une intelligence humaine. En règle générale, les systèmes informatiques sont réputés intelligents (c'est pour cela qu'ils sont appelés « agents intelligents » ou « machines intelligentes ») dès lors qu'ils sont capables de percevoir leur environnement et d'agir de façon autonome pour atteindre un but. Si depuis l'antiquité la littérature scientifique et philosophique regorge de réflexions d'ordre général sur la logique mécanique, le domaine de l'IA stricto sensu est né dans les années 1940, dans le prolongement des progrès réalisés simultanément en logique mathématique (par ex. la thèse Church-Turing), en théorie de l'information, en neurobiologie et en cybernétique. Le domaine de l'IA englobe tout un éventail de démarches informatiques complexes permettant d'accomplir ou d'imiter des fonctions cognitives telles que l'apprentissage, la mémoire, le raisonnement, la vision et le traitement du langage naturel. La plus courante de ces démarches – l'apprentissage automatique – permet, grâce à des algorithmes développés à cet effet, de réaliser des tâches sans instructions explicites d'opérateurs humains. Contrairement aux programmes informatiques conventionnels, les algorithmes d'apprentissage automatique créent des modèles mathématiques sur la base des données d'apprentissage et s'appuient exclusivement sur l'inférence et l'identification de modèles pour faire des prédictions et prendre des décisions en toute autonomie¹. Aujourd'hui, l'IA est l'un des principaux catalyseurs de transformation technologique. En ce début des années 2020, les systèmes d'IA sont intégrés à un nombre incalculable de systèmes et de dispositifs dont les humains se servent régulièrement : les téléphones mobiles, les médias sociaux, les voitures, les avions, les logiciels d'analyse, les systèmes de communication par courrier électronique, les appareils électroménagers, etc. L'IA fait partie intégrante d'un vaste éventail d'activités humaines, dont, notamment, les télécommunications, les transports, la fabrication, la santé, la banque, les assurances, le maintien de l'ordre et l'armée.

De par sa nouveauté technologique et sa capacité d'agir et de tendre vers un but de façon autonome, l'IA a le potentiel de transformer les sociétés humaines à un rythme plus élevé et de façon plus marquée que n'importe quelle autre technologie. Le potentiel transformateur de l'IA a été qualifié de « révolutionnaire » par des experts², et divers auteurs ont comparé le développement de l'IA à une « révolution en marche » qui « transformera presque tous les métiers »³. C'est pourquoi il est capital et urgent d'évaluer les incidences de l'IA sur les valeurs et principes fondamentaux de la vie humaine, l'avenir des sociétés humaines et les systèmes régissant celles-ci, à commencer par la démocratie et l'État de droit⁴⁻⁶.

Ces dernières années, plusieurs organismes gouvernementaux et intergouvernementaux aussi bien que des acteurs non étatiques ont établi des principes et diffusé des lignes directrices, des recommandations, des cadres de gouvernance et d'autres instruments non contraignants en matière d'IA. Les instruments de droit souple sont des textes normatifs qui ne sont pas juridiquement contraignants ou n'ont aucune force exécutoire mais dont la nature persuasive peut avoir une influence concrète sur la prise de décision, comparable à celle des actes contraignants (droit dur). Ils ont pour but d'orienter le développement de l'IA en faveur du bien-être social et du respect des valeurs éthiques ainsi que des normes juridiques. Cependant, malgré un consensus manifeste sur le fait que l'IA doit être « éthique », il y a débat sur ce qu'il faut entendre par « IA éthique » et sur les exigences éthiques, les normes techniques et les meilleures pratiques nécessaires à cet effet. Par ailleurs, vu la prolifération rapide des textes de droit souple relatifs à l'IA et la grande diversité des auteurs, il est difficile de suivre de façon approfondie et rigoureuse l'évolution constante de ce train de mesures non contraignantes et d'en comprendre tous les tenants et aboutissants.

Le présent rapport dresse une cartographie complète du corpus actuel de principes et de lignes

directrices au sujet de l'IA éthique, et en fait une méta-analyse. Celle-ci a pour but d'informer les scientifiques, les établissements de recherche, les organismes de financement, les organismes gouvernementaux et intergouvernementaux, et autres acteurs concernés par les progrès de l'innovation éthiquement responsable en matière d'IA. Elle s'intéresse en outre à la façon dont ces principes éthiques, coutumes morales et pratiques sociales recommandées peuvent se traduire en règles de gouvernance obligatoires, notamment en instruments juridiques contraignants à l'échelon international.

Le présent rapport accorde une large place à l'examen du lien entre gouvernance de l'IA et droits de l'homme, et il évalue les répercussions que la technologie de l'IA pourrait avoir sur les droits de l'homme et les libertés⁷⁻⁹. Les droits de l'homme sont des droits inhérents à la personne, sans distinction de race, de sexe, de nationalité, d'origine ethnique, de langue, de religion ou de toute autre condition¹⁰. Ils renvoient aussi bien à des principes moraux qu'à des normes juridiques du droit national et international dont une personne peut se prévaloir par essence, en tant qu'être humain. Ils sont par conséquent inaliénables, inviolables et universels. Ils sont inaliénables car personne ne peut en priver qui que ce soit, inviolables, car personne ne devrait en aucun cas les enfreindre, et universels car ils s'appliquent partout et à tout moment. Les droits de l'homme et les libertés sont protégés par des conventions internationales. L'un des instruments fondamentaux de l'espace européen est la Convention européenne des droits de l'homme (CEDH), qui a été rédigée en 1950 par le Conseil de l'Europe et qui est entrée en vigueur le 3 septembre 1953. La CEDH consacre une série de droits et libertés fondamentaux qu'il faut protéger, et ses parties s'engagent juridiquement à se conformer à des normes de conduite respectueuses de ces droits et libertés¹¹.

V. Méthodologie

En février 2020, nous avons mené un examen de la portée du corpus de textes de droit souple existant au sujet de l'IA. Les méthodes employées à cet effet permettent de faire une synthèse et de cartographier la documentation existant dans tel ou tel domaine en suivant une démarche exploratoire, et sont donc particulièrement adaptées au criblage et à l'évaluation de domaines de recherche complexes et hétérogènes. Comme il n'existait pas de base de données unifiée des instruments non contraignants, nous avons défini un protocole de recherche et de sélection en suivant les normes PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*). Le protocole, qui a été testé et calibré avant la collecte de données, comprend trois phases séquentielles liées par itération : criblage, évaluation de la pertinence et analyse du contenu. Cette méthodologie visait à établir une procédure formelle, fondée sur les résultats, pour cartographier, étudier et évaluer de façon itérative les mesures de gouvernance non contraignantes prises dans le domaine de l'IA.

i. Criblage

Durant la phase de criblage, nous avons procédé à l'examen rétrospectif des répertoires existants dans le cadre d'une recherche intentionnelle et non structurée du web. Nous avons commencé par examiner les quatre répertoires de données et sources textuelles ci-après pour en extraire les entrées relatives aux documents de droit souple sur l'IA :

- *AI Policy Initiatives List*, de l'Agence des droits fondamentaux de l'Union européenne (11 décembre 2019).
- Fjeld et Nagy (janvier 2020), *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, Harvard Berkman Klein Center Research Publication.
- Jobin, Ienca et Vayena (septembre 2019). *The Global Landscape of AI Ethics Guidelines. Nature Machine Intelligence*.
- Wong et CASBS (juin 2019). *Fluxus Landscape: An Expansive View of AI Ethics and Governance*.

Conformément au cadre qu'Arksey et O'Malley ont défini pour les examens de la portée¹², la

recherche structurée dans les bases de données a été complétée par une recherche non structurée dans la littérature grise pour ne rater aucun instrument non contraignant. Nous avons ensuite vérifié si les éléments trouvés étaient pertinents (voir 1.2) et, lorsque c'était confirmé, nous les avons inclus manuellement dans la synthèse finale et sélectionnés pour la seconde phase.

Nous avons ensuite étudié la liste des 45 plus grandes entreprises du secteur de l'IA, dressée en mai 2019 par Datamation (« *Top-45 AI companies* »), un magazine américain de science informatique spécialisé dans l'analyse des technologies. Pour chacune de ces 45 entreprises, nous avons examiné leurs sites web et cherché manuellement et par mots-clés des déclarations au sujet de l'éthique ou des orientations générales en matière d'IA. Pour terminer, nous avons procédé à une recherche non structurée sur le web pour extraire des informations que notre stratégie de recherche n'avait peut-être pas décelées. Les éléments pertinents ont été inclus manuellement dans la synthèse finale et sélectionnés pour la seconde phase.

ii. Evaluation de la pertinence

Au cours de cette phase, nous avons examiné de façon approfondie tous les éléments extraits pour vérifier s'il était pertinent de les inclure dans la synthèse finale. Les décisions en la matière ont été guidées par les critères d'inclusion et d'exclusion énumérés dans le tableau 1.

Tableau 1 – Critères de sélection

| Criblage | |
|------------------------------|--|
| | - Types : sites web, articles et autres documents écrits publiés en ligne ou éléments les composant, comme des pages web spécifiques, des articles de blog, des rapports et déclarations de diverses institutions ainsi que les références sur lesquelles ils s'appuient. |
| Sources incluses : | - Émetteurs : organismes à but lucratif du secteur privé (entreprises, multinationales, holdings, etc. et alliances du secteur privé) ; établissements universitaires et de recherche (universités, associations professionnelles, fondations scientifiques, etc.) ; organismes gouvernementaux nationaux (ministères, autorités de protection des données, autorités de réglementation de la concurrence, etc.) ; organisations non gouvernementales (organisations à but non lucratif, organisations caritatives, etc.). |
| | - Langues : anglais, allemand, français, espagnol, néerlandais, italien et grec (langues parlées par les chercheurs). |
| Sources exclues : | - Types : vidéos, images et audio/podcasts (sauf descriptions écrites), livres, articles de journaux, articles universitaires, plans de cours, législation, normes officielles, synthèses de conférences. |
| | - Émetteurs : organisations intergouvernementales et supranationales. |
| | - Langues : autres que celles qui sont susmentionnées. |
| Critères de sélection | |
| Sources incluses : | - font référence aux termes « intelligence artificielle » et/ou « IA » de façon explicite, dans leur titre ou dans le corps du document (par exemple : Google, « principes de l'IA ») ; ou |
| | - ne contiennent pas les termes susmentionnés mais « robot », « robotique », « mégadonnées » (« big data »), « apprentissage automatique » et font explicitement référence à l'IA ou l'intelligence artificielle comme faisant partie intégrante des robots et/ou de la robotique ; ou |
| | - ne contiennent pas les termes susmentionnés dans leur titre mais des termes en lien (« algorithmes », « analyse prédictive », « informatique cognitive », « apprentissage automatique », « mégadonnées » (« big data »), « apprentissage profond », « autonome » ou « automatisé »). |
| | ET |
| | - décrivent un principe, des lignes directrices, une norme (notamment avec les termes « éthique », « principes », « préceptes », « déclaration », « politiques », « lignes directrices », « valeurs », etc.), une stratégie interne (par ex. création d'un conseil consultatif) ou autre type d'initiative. |
| | ET |

| | |
|-------------------|--|
| | <ul style="list-style-type: none"> - sont exprimées dans un langage normatif ou impératif (c'est-à-dire avec l'emploi de modaux ou de termes indiquant une obligation, par ex. « responsable », « juste », « confiance/fiable », etc.). |
| | <ul style="list-style-type: none"> - reposent sur des principes ou des valeurs (c.-à-d. indiquent une préférence et/ou une adhésion à une certaine vision éthique ou ligne de conduite). |
| | <ul style="list-style-type: none"> - font référence à des actions/visions/engagements/lignes d'action s'appliquant à l'acteur qui les énonce ou à d'autres acteurs du secteur privé. |
| Sources exclues : | <ul style="list-style-type: none"> - les sites web et les documents concernant la robotique n'indiquant pas que l'IA fait partie intégrante des robots/de la robotique. |
| | <ul style="list-style-type: none"> - les sites web et documents concernant les données ou l'éthique des données n'indiquant pas que l'IA fait partie intégrante des données. |
| | <ul style="list-style-type: none"> - les sites web et documents sur l'éthique de l'IA s'adressant directement à des acteurs n'appartenant pas au secteur privé (par ex. services de conseil pour le secteur public). |
| | <ul style="list-style-type: none"> - les sites web et documents sur l'éthique dont le sujet principal n'est pas l'intelligence artificielle (par ex. : l'éthique dans les affaires). |

iii. Analyse du contenu

Dans la seconde phase, les éléments inclus dans la synthèse finale ont été examinés à l'aide de la version étendue d'un protocole d'analyse de contenu précédemment validé, que les auteurs avaient mis au point^{13,14}. Ce protocole prévoit une analyse quantitative et une analyse qualitative. Sur le plan quantitatif, les éléments ont été classés en fonction du type d'instrument, de l'émetteur et de l'emplacement géographique de ce dernier. Par ailleurs, la fréquence relative des données quantitatives pertinentes a été mesurée et illustrée par une représentation graphique. Enfin, nous avons procédé à l'examen de l'intégralité des textes à l'aide d'une recherche par mots-clés pour repérer les documents qui faisaient explicitement référence aux droits de l'homme. Les documents relevant de cette catégorie faisaient explicitement référence soit à la protection et à la promotion des droits de l'homme, soit à la prévention de leur violation au moment de la conception, du développement ou du déploiement des applications d'IA. Ils ont été séparés de ceux qui n'évoquaient pas les droits de l'homme ou le faisaient mais sans formuler de déclaration normative explicite sur leur promotion ou leur non-violation dans le contexte de l'IA.

Sur le plan qualitatif, une analyse du contenu thématique a été menée pour repérer des thèmes récurrents concernant les domaines suivants : i) les principes et valeurs éthiques, et ii) les droits de l'homme. Cette analyse a été réalisée manuellement par les chercheurs, à l'aide d'un logiciel d'analyse de données qualitatives (NVivo/MAXQDA pour Mac). Les thèmes émergents ont été analysés en profondeur, encodés et regroupés dans des catégories éthiques prédéfinies sur la base de la matrice d'éthique mise au point par Jobin, Ienca et Vayena (2019).

Vu la taille de la base de données de la synthèse finale, cette analyse thématique manuelle a été réalisée à l'aide d'un moteur de traitement du langage naturel (TLN). Nous avons extrait des documents et du contenu web automatiquement là où c'était possible, à l'aide du paquet Python wget, et ajouté le reste manuellement. Nous avons ensuite créé des expressions régulières¹⁵ à partir des codes résultant du protocole d'analyse qualitatif créé par Jobin, Ienca et Vayena¹³. Les expressions régulières des codes appartenant à un même thème ont été reliées entre elles en une seule expression régulière par « or » ('|'). Par souci d'exhaustivité et d'inclusion, les codes anglais d'origine ont été traduits dans les langues suivantes : allemand, français, espagnol, italien et néerlandais. Pour déterminer si un thème était traité, nous avons cherché des occurrences de ses expressions régulières dans les documents, et nous avons considéré que s'il y en avait au moins une, cela signifiait que le thème était bien évoqué. Enfin, nous avons regroupé les résultats par catégorie de membre et nous les avons normalisés en fonction du nombre total de lignes directrices dans chaque groupe. Les principes et les valeurs éthiques évoqués par les 47 États membres du Conseil de l'Europe ont été comparés aux principes et valeurs évoqués par les États observateurs ainsi que par les pays du reste du monde.

iv. Analyse de l'éthique normative et des politiques

Dans la quatrième et dernière phase, nous avons procédé à l'analyse empirique de l'éthique normative et des politiques afin de traduire en normes prescriptives les constatations descriptives faites lors des précédentes phases de l'étude. À cet effet, nous avons suivi trois étapes théoriques successives. Premièrement, nous avons examiné les résultats de notre analyse de contenu pour déterminer quels principes et valeurs éthiques revenaient le plus souvent dans le corpus des documents traités. Étant donné que de précédentes recherches avaient montré d'importantes variations dans l'interprétation des groupes thématiques récurrents¹³, nous avons complété l'évaluation de la fréquence relative d'occurrence des divers thèmes par une évaluation détaillée de l'interprétation qui en était faite. Cette dernière a largement contribué à déterminer quelles interprétations des principes considérés étaient les plus effectives et devaient donc être adoptées et reprises par l'ensemble des acteurs. Deuxièmement, nous avons analysé les données recueillies pour déterminer quels principes et valeurs étaient moins souvent traités voire absents des lignes directrices actuelles sur l'éthique en matière d'IA. Cette étape a aidé à repérer d'éventuelles lacunes dans les initiatives internationales de droit souple et, par conséquent, à énoncer des recommandations normatives sur la façon de combler ces lacunes éthiques. Troisième et dernière

étape, nous avons formulé des recommandations normatives au sujet des valeurs et principes éthiques fondamentaux à traiter en priorité dans le cadre de la gouvernance internationale de l'IA. Cette dernière étape a permis d'étayer les futurs cadres normatifs de l'éthique et de définir une feuille de route pour l'élaboration des politiques internationales sur l'IA, l'éthique et les droits de l'homme. À cet effet, nous avons réalisé une synthèse graphique des résultats de l'étude facile à exploiter, et créé une boîte à outils qui permettra d'assurer un suivi et une évaluation (indicateurs) à l'interface entre IA, éthique et droits de l'homme.

VI. Constatations

Notre recherche a permis de trouver 116 documents contenant des instruments de droit souple sur l'IA, diffusés par des organismes non intergouvernementaux jusqu'en février 2020. Il ressort des données que depuis 2016, le nombre de publications a enregistré une hausse de 93,9 %, ce qui est considérable. Le nombre de documents de droit souple publiés à l'échelon international a culminé en 2018 puis enregistré une baisse non négligeable l'année suivante (voir figure 1).

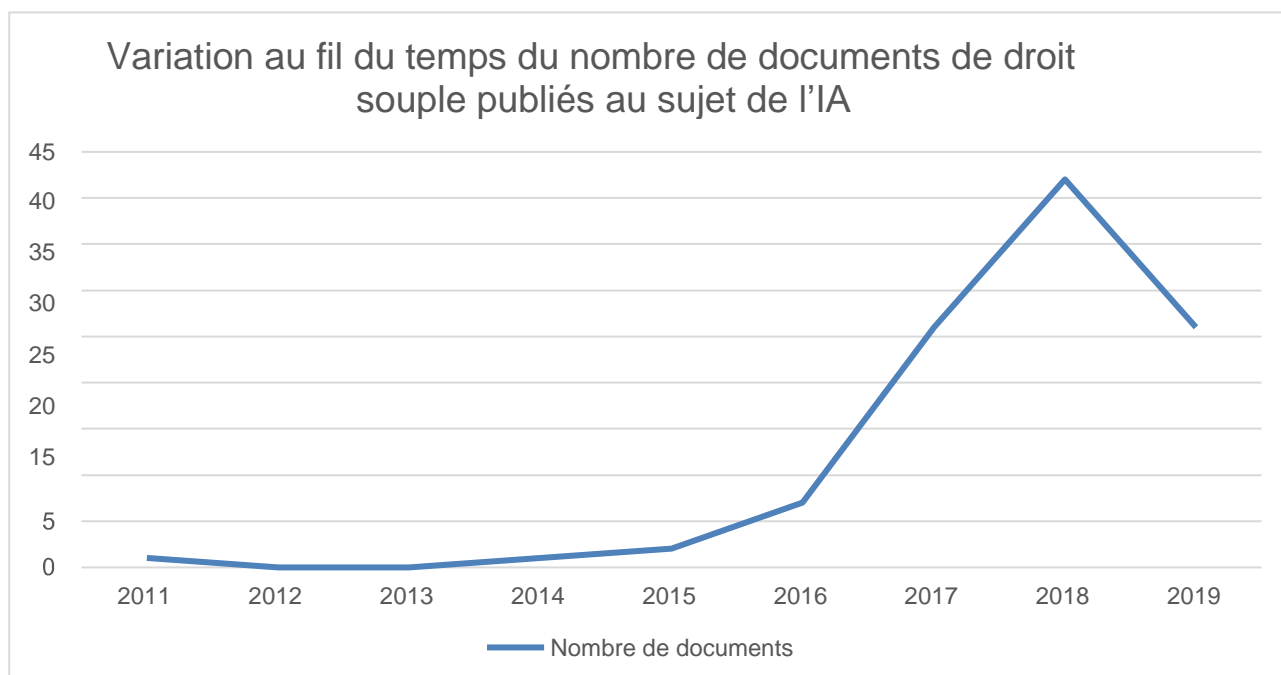


Figure 1 – Variation au fil du temps du nombre de documents de droit souple publiés au sujet de l'IA

Il ressort de la ventilation des données par type d'organisme émetteur que la plupart des documents ont été émis par des autorités publiques (n=39), puis des entreprises et des alliances du secteur privé (n=36), puis des établissements universitaires et de recherche, notamment des fondations scientifiques, des associations professionnelles et des alliances de recherche (n=28) ainsi que par des organisations non gouvernementales (ONG) et notamment des organismes à but non lucratif et des organisations caritatives (n=13). La figure 2 montre la répartition détaillée des organismes émetteurs.

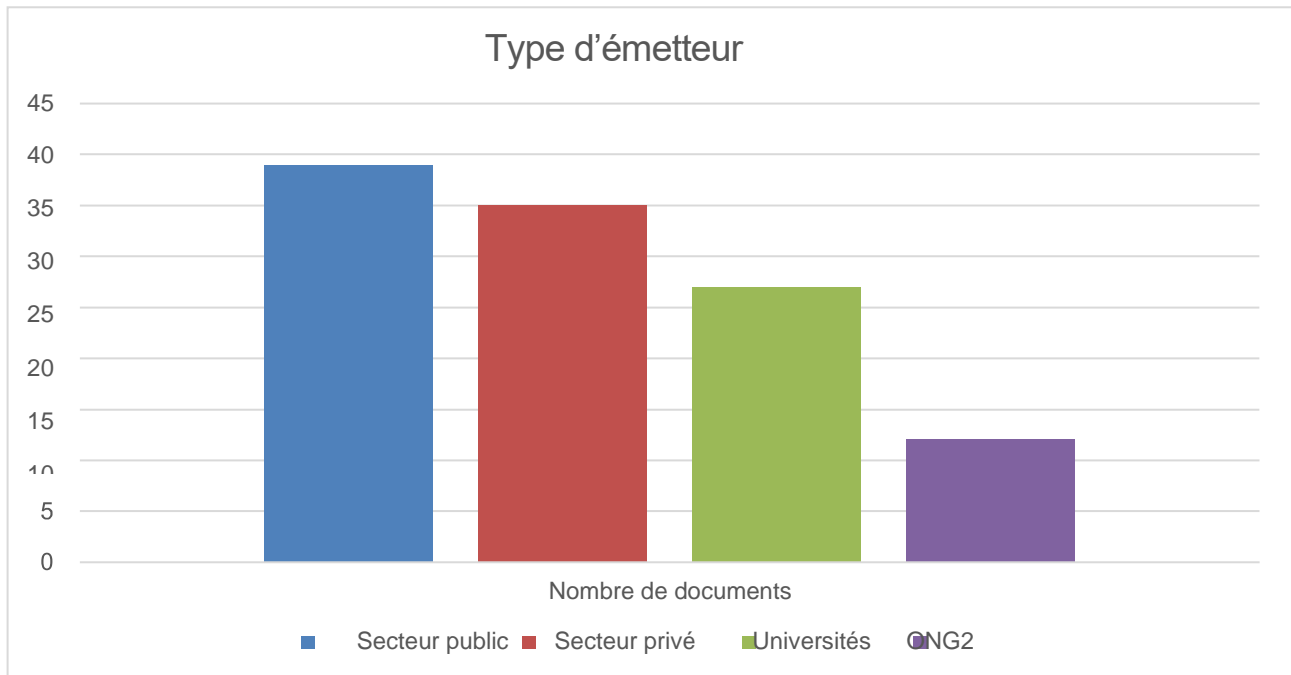


Figure 2 – Types d'organismes émetteurs

La ventilation des données en fonction de la répartition géographique des organismes émetteurs montre que 46 % (n=53,5) des documents de droit souple sont diffusés par des organisations basées dans des pays membres du Conseil de l'Europe ; 32 % (n=37,5) par des organisations basées dans des pays ayant le statut d'observateur auprès du Conseil de l'Europe ; 21 % (n=25) par des organisations basées dans des pays qui ne sont ni membres du Conseil de l'Europe ni observateurs. Dans l'ensemble, il ressort des données recueillies que les organismes émetteurs situés dans des pays économiquement développés prédominent, les États-Unis (n=29,5 ; 25,2 %) et le Royaume-Uni (n=17,5 ; 16 %) totalisant à eux deux plus d'un tiers de tous les principes éthiques énoncés en matière d'IA. Les autres pays sont, par ordre décroissant, l'Allemagne (n=8), le Japon (n=6), la Finlande (n=4), la Belgique, la Chine, la France et les Pays-Bas (n=3), l'Inde, l'Italie, Singapour et l'Espagne (n=2), l'Australie, l'Autriche, la République tchèque, l'Islande, la Lituanie, Malte, le Mexique, la Nouvelle-Zélande, la Norvège, la Russie, la Corée du Sud, la Suède, la Suisse, les EAU et le Vatican (n=1). Treize documents ont été diffusés par des organisations internationales ou des organisations ne pouvant pas être rattachées à un pays précis. Les pays d'Afrique et d'Amérique du Sud n'étaient pas représentés indépendamment d'organisations internationales. La figure 3 donne un aperçu de la répartition géographique des organismes émetteurs.

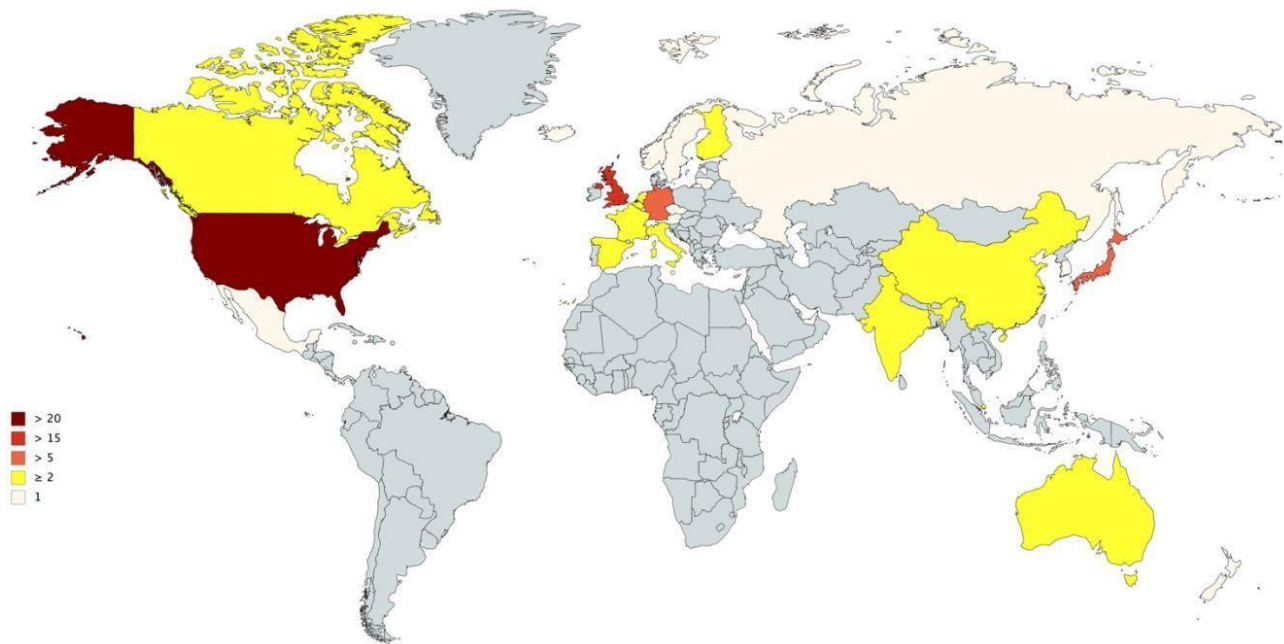


Figure 3 – Répartition géographique des documents de droit souple en fonction du pays d'appartenance de l'organisme émetteur

Plus de la moitié des documents (n=62) font explicitement référence à la promotion et au respect des droits de l'homme ou à la prévention de leur violation. Trente et un de ces documents ont été diffusés par des organismes basés dans des pays membres du Conseil de l'Europe, 14 par des organismes basés dans des pays observateurs et 17 par des organismes basés dans des pays qui ne sont ni l'un ni l'autre. Les documents émis par des organismes basés dans des pays membres du Conseil de l'Europe évoquent les droits de l'homme dans 57,9 % des cas. Ceux qui proviennent de pays non membres du Conseil de l'Europe y font référence dans 49,6 % des cas. Cela montre que les incidences de l'intelligence artificielle sur les droits de l'homme sont plus fréquemment examinées par des organismes basés dans des pays membres du Conseil de l'Europe que dans des pays du reste du monde.

Notre analyse du contenu thématique nous a permis d'extraire divers codes se rapportant à l'éthique, susceptibles d'être tous systématiquement applicables aux onze groupes de principes éthiques définis par Jobin, Ienca et Vayena (2019)¹³. Ce sont, par ordre décroissant de fréquence des sources dans lesquels ils apparaissent : la transparence, la justice et la loyauté, la non-malfaisance, la responsabilité, le respect de la vie privée, la bienfaisance, la liberté et l'autonomie, la confiance, la dignité, la durabilité et la solidarité. Le tableau 2 illustre avec précision la fréquence avec laquelle les principes éthiques et les codes y associés sont évoqués.

| Principes éthiques | Nombre de documents | Codes associés |
|--------------------------|---------------------|---|
| | | Transparence, explicabilité, compréhensibilité, |
| | | interprétabilité, communication, divulgation, |
| Transparence | 101/116 | montrer |
| | | Justice, loyauté, cohérence, inclusion, égalité, |
| | | équité, (im)partialité, (non-)discrimination, |
| | | diversité, pluralité, accessibilité, réversibilité, |
| Justice et loyauté | 97/116 | recours, réparation, défi, accès et distribution, |
| | | impartialité |
| | | Non-malfaisance, sécurité, sûreté, préjudice, |
| | | protection, précaution, prévention, intégrité |
| Non-malfaisance | 84/116 | (physique et mentale), non-subversion |
| | | Responsabilité, obligation de rendre compte, |
| Responsabilité | 79/116 | obligation, agir avec intégrité |
| | | Respect de la vie privée, informations |
| Respect de la vie privée | 74/116 | personnelles ou privées, confidentialité |
| | | Avantages, bienfaisance, bien-être, paix, |
| Bienfaisance | 58/116 | bien-être social, bien commun |
| | | Liberté, autonomie, consentement, choix, |
| Liberté et autonomie | 48/116 | autodétermination, autonomisation |
| Fiabilité | 41/116 | Confiance, fiabilité |
| | | Durabilité, environnement (nature), énergie, |
| Durabilité | 20/116 | ressources (énergie) |
| Dignité | 20/116 | Dignité |
| Solidarité | 10/116 | Solidarité, sécurité sociale, cohésion |

Tableau 2 – Fréquence des thèmes éthiques et codes y associés

Aucun des différents principes éthiques n'est commun à tous les documents du corpus mais une convergence se dégage autour des principes suivants : transparence, justice et loyauté, non-malfaisance, responsabilité et respect de la vie privée. Près de deux tiers de toutes les sources y font référence. Toutefois, l'analyse thématique met en évidence la persistance de nettes divergences sémantiques et conceptuelles, à la fois quant à l'interprétation des onze principes éthiques et aux recommandations spécifiques ou aux questions préoccupantes qui en découlent.

Les expressions régulières tirées des codes montrent des différences perceptibles entre les thèmes abordés dans les documents issus des pays membres du Conseil de l'Europe et ceux dont traitent les documents élaborés ailleurs. Par rapport aux documents élaborés dans les pays ayant le statut d'observateur auprès du Conseil de l'Europe, les textes de droit souple établis au sein des pays membres du Conseil de l'Europe semblent insister sur les principes éthiques ci-après : la transparence, la durabilité la liberté et l'autonomie, la confiance/fiabilité et la solidarité (voir figure 4). Ils semblent en revanche faire moins souvent référence aux principes de justice, de bienfaisance et de dignité. Par rapport aux textes de droit souple établis dans le reste du monde (pays non-membres et non observateurs), ceux qui sont établis dans les pays membres du Conseil de l'Europe paraissent mettre l'accent sur les principes de confiance/fiabilité et de solidarité et traiter tous les autres principes moins fréquemment. Les principes de respect de la vie privée, de justice et de loyauté sont ceux pour lesquels les différences sont les moins marquées, et donc ceux qui ont le plus haut degré de constance transgéographique et transculturelle.

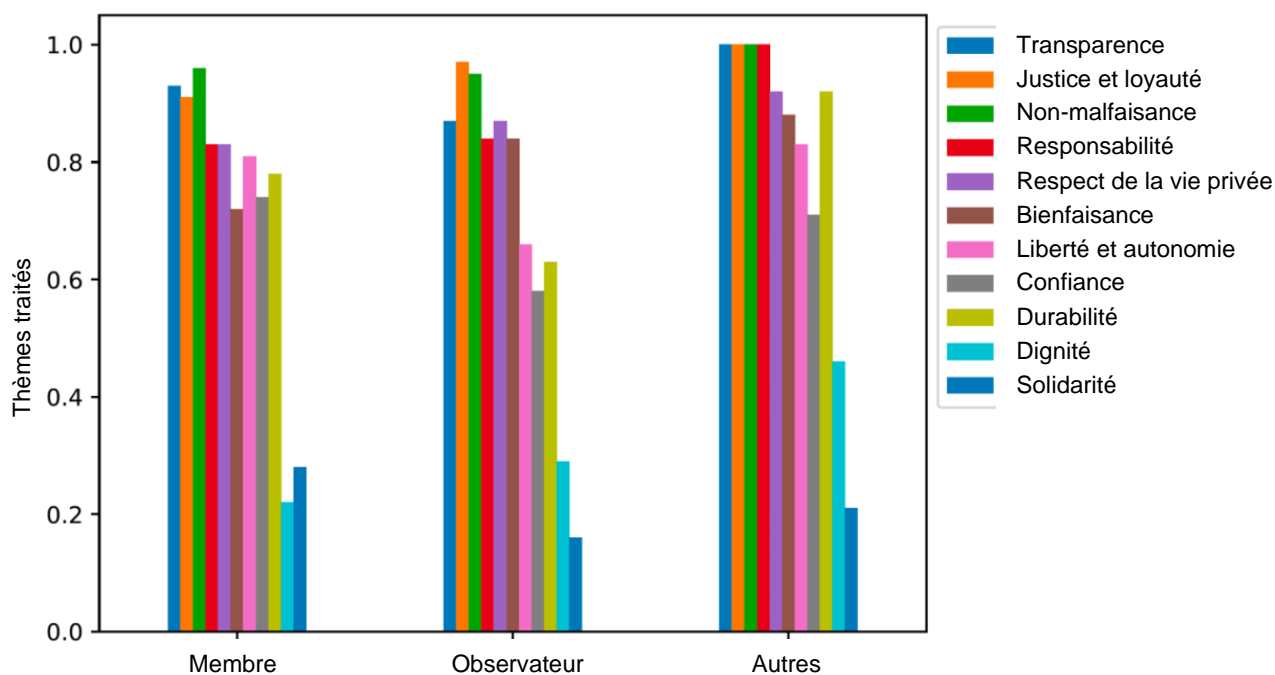


Figure 4 – Variations dans la façon de traiter les différents thèmes entre les documents établis dans des pays membres du Conseil de l'Europe et les documents établis dans le reste du monde.

Voici une évaluation détaillée de chacun des thèmes énumérés plus haut.

Transparence : La transparence, qui est évoquée dans 101 des 116 documents examinés, est le principe éthique le plus fréquemment traité dans les textes non contraignants actuels. L'analyse thématique montre que la façon dont elle est interprétée et les raisons qui la justifient varient considérablement, ce qui, constate-t-on, entraîne d'indéniables divergences dans les stratégies de mise en œuvre proposées pour parvenir à la transparence en matière d'IA. Les références faites à la transparence peuvent être rangées dans deux grandes catégories thématiques : 1) la transparence des algorithmes et des méthodes de traitement des données, 2) la transparence des pratiques humaines relatives à la conception, au développement et au déploiement des systèmes d'IA. La première catégorie implique le plus souvent la promotion des démarches méthodologiques

en faveur de « l'explicabilité de l'IA », à savoir le fait que les produits et les décisions des systèmes d'IA soient compréhensibles par les experts humains. Ces méthodes et techniques sont en opposition avec le mécanisme de « boîte noire » de l'apprentissage automatique, où les étapes par lesquelles un système d'IA arrive à telle ou telle décision sont inintelligibles pour les experts humains, en ce compris les concepteurs du système. Alors que les entreprises du secteur privé, surtout les acteurs de l'IA, ont tendance à réduire la transparence à l'interprétabilité et à l'explicabilité moyennant diverses solutions techniques – par exemple, la méthode LRP (*layerwise relevance propagation*) et l'interprétabilité locale – les autorités publiques, telles que les responsables nationaux de la protection des données, soulignent l'importance des méthodes de surveillance comme les audits. La deuxième forme de transparence ne porte pas sur des algorithmes interprétables mais sur la transparence des pratiques humaines relatives aux données et à l'IA, par exemple divulguer des informations pertinentes aux personnes concernées, éviter l'opacité lors du déploiement des stratégies d'IA, et interdire les conflits d'intérêts entre les acteurs de l'IA et les organes de surveillance. Ce type de transparence est plus couramment évoqué par les acteurs publics et les ONG.

Justice, loyauté, équité : Il est essentiellement fait référence à la justice en termes de loyauté et de prévention (ou d'atténuation) des biais algorithmiques qui peuvent conduire à de la discrimination. La crainte de voir l'IA accroître les inégalités et être à l'origine de discriminations semble moins répandue dans les documents de droit souple émanant du secteur privé que dans ceux qui émanent des organismes publics et du milieu universitaire. Les textes sont divisés sur la question de savoir comment assurer la justice et la loyauté en matière d'IA. Certaines sources sont axées sur le respect de la diversité et favorisent l'inclusion et l'égalité, aussi bien dans la conception des systèmes d'IA (en particulier lors de la compilation des jeux de données de l'apprentissage) que dans leur déploiement dans la société. D'autres sources, en s'appuyant sur le droit de recours et de réparation, préconisent d'offrir la possibilité de contester les décisions. L'accès équitable aux avantages que procure l'IA est aussi un thème récurrent. Les documents émanant d'acteurs publics insistent particulièrement sur les répercussions de l'IA sur le marché du travail, et sur la nécessité d'examiner les défis démocratiques et sociétaux. Nous avons repéré cinq grandes stratégies de mise en œuvre, conciliables entre elles, permettant de préserver et de promouvoir la justice et la loyauté dans l'IA :

- i. appliquer des solutions techniques telles que des normes et des bonnes pratiques ;
- ii. sensibiliser le public aux droits et à la réglementation en vigueur ;
- iii. améliorer les tests, le suivi et le contrôle des systèmes d'IA ;
- iv. développer ou renforcer l'État de droit et instaurer un droit de recours et de réparation ;
- v. modifier le système et introduire des processus, par exemple des mesures ou des contrôles gouvernementaux, renforcer la pluridisciplinarité de la main-d'œuvre, et améliorer l'inclusion de la société civile ou de divers autres acteurs de façon interactive.

Si les solutions II à V semblent être privilégiées par les autorités publiques (notamment les responsables de la protection des données), la première est privilégiée par les acteurs du secteur privé de l'IA.

Non-malfaisance : Les références à la non-malfaisance, bien plus fréquentes que les références à la bienfaisance, sont liées à des recommandations générales de sûreté et de sécurité ou à des déclarations selon lesquelles l'IA ne devrait jamais causer de préjudices prévisibles ou involontaires. Certains documents se concentrent sur des risques précis ou des préjudices potentiels, en particulier le risque de détournement intentionnel dans le cadre d'une cyberguerre et d'un piratage. Les sources de préjudice les plus courantes que citent les divers documents sont la discrimination sociale, la violation de la vie privée et les préjudices corporels ou psychologiques. Les documents de droit souple consacrés à l'atténuation des préjudices recommandent souvent des solutions techniques et des mesures de gouvernance obligatoires relatives à la recherche en matière d'IA, à la conception ainsi qu'au développement et au déploiement technologiques. Les solutions techniques passent entre autres par des évaluations de la qualité des données intégrées, ou des systèmes de sécurité et de respect de la vie privée inclus dans le cadre de conception, tandis que d'autres documents

privilégient la création de normes sectorielles. Les stratégies de gouvernance proposées prévoient notamment une coopération active entre les diverses disciplines et parties prenantes, le respect de la législation en vigueur ou à venir, et la nécessité de mettre en place des processus et des pratiques de surveillance, notamment grâce à des tests, à un suivi, à des contrôles et à des évaluations effectuées par des services internes, les clients, les utilisateurs, des tierces parties indépendantes ou des autorités publiques. Certaines sources désignent expressément l'usage militaire de l'IA – le problème des technologies dites à double usage – comme étant le principal domaine dans lequel des mesures de gouvernance doivent être prises.

Responsabilité et obligation de rendre compte : Si de nombreux textes évoquent le développement d'une « IA responsable », il n'en reste pas moins que les notions de responsabilité et d'obligation de rendre compte sont rarement définies. Divers acteurs sont désignés comme ayant une responsabilité et des comptes à rendre pour les actions et décisions de l'IA. Il s'agit des développeurs et concepteurs de l'IA et du secteur de l'IA dans son intégralité. Les documents ne s'accordent pas sur la question de savoir si la responsabilité de l'IA doit être calquée sur celle des humains ou si les humains doivent toujours être les seuls responsables en dernier ressort des actions des objets technologiques.

Respect de la vie privée : Le respect de la vie privée est généralement considéré comme une valeur à défendre et un droit à protéger. Si les questions de respect de la vie privée sont fréquemment abordées dans les lignes directrices existantes sur l'IA, aucun consensus ne se dégage sur les difficultés spécifiques que pourraient soulever les progrès en matière d'IA par rapport aux autres technologies utilisant de gros volumes de données. L'analyse thématique montre que la plupart des documents évoquent le respect de la vie privée en termes généraux, sans relier explicitement les capacités de l'IA et les nouveaux défis en la matière. Bien qu'il soit mal décrit, le problème de respect de la vie privée que pose l'IA est souvent abordé en lien avec les questions de protection et de sécurité des données. Les stratégies proposées pour protéger la vie privée face à l'IA peuvent être rangées en trois catégories : A) les solutions techniques telles que la confidentialité différentielle, le calcul multipartite sécurisé et le chiffrement homomorphe ; B) les solutions consistant à informer le public, comme les campagnes de sensibilisation des utilisateurs et des personnes concernées par les données ; et C) les solutions réglementaires, consistant par exemple à mieux définir les règles juridiques à respecter (notamment en matière de protection des données) ou à rédiger de nouveaux textes législatifs ou réglementaires portant spécifiquement sur l'IA.

Bienfaisance : Si la pratique du bien (la *bienfaisance*, en termes éthiques) est souvent évoquée, elle est rarement définie, à l'exception de quelques documents encourageant l'adoption de mesures en faveur du bien-être et de l'épanouissement humain, de la paix et du bonheur, de la création d'opportunités socio-économiques et de la prospérité économique. Il n'est pas non plus clair de savoir quels acteurs doivent bénéficier de l'IA : en règle générale les émetteurs du secteur privé mettent en avant leurs clients alors que les sources universitaires et publiques affirment le plus souvent que l'IA doit profiter à « tout le monde », à « l'humanité » et à « la société au sens large ». Les stratégies de bienfaisance prévoient entre autres d'aligner l'IA sur les valeurs humaines, de limiter la concentration du pouvoir, et d'employer les capacités de l'IA pour promouvoir les droits de l'homme.

Liberté et autonomie : Les documents de droit souple établissent un lien entre l'IA et la protection ou la promotion de plusieurs libertés, à savoir, notamment, la liberté d'expression, l'autodétermination informationnelle, le droit au respect de la vie privée et l'autonomie personnelle. Cette dernière notion correspond en règle générale à une liberté positive, et plus précisément à la liberté de s'épanouir et au libre arbitre. Un petit nombre de documents assimilent toutefois l'autonomie à une liberté négative, par exemple la liberté face à l'expérimentation technologique, à la manipulation ou à la surveillance. Les solutions proposées pour protéger la liberté et l'autonomie en matière d'IA consistent notamment à favoriser la transparence et l'explicabilité, à renforcer la maîtrise de l'IA, à veiller à ce que les personnes concernées donnent leur consentement éclairé ou, à l'inverse, à s'abstenir de recueillir et de divulguer des données si les personnes concernées n'y ont pas consenti en toute connaissance de cause.

Confiance et fiabilité : Un peu plus d'un tiers des documents de droit souple prônent la fiabilité des recherches et de la technologie en matière d'IA, ou la promotion d'une culture de la confiance chez les scientifiques et les ingénieurs. Certains documents mettent toutefois expressément en garde contre un excès de confiance dans l'IA et affirment que la confiance ne peut s'instaurer qu'entre pairs et ne doit pas être déléguée à l'IA. Il est notamment suggéré de s'appuyer sur les éléments ci-après pour instaurer ou maintenir la confiance : l'éducation, la fiabilité, la responsabilité, des processus permettant d'assurer le suivi et l'évaluation de l'intégrité des systèmes d'IA au fil du temps, et des outils et des techniques pour veiller au respect des règles et normes.

Durabilité : La durabilité est rarement évoquée, et lorsqu'elle l'est c'est en règle générale dans le cadre de la protection de l'environnement, voire de l'amélioration des écosystèmes et de la biodiversité de la planète. Certains documents exigent que les systèmes d'IA traitent les données de façon durable et optimisent leurs performances énergétiques pour réduire au maximum leur empreinte écologique^{4,7}. Un plus petit nombre de documents sont axés sur la durabilité sociale, c'est-à-dire veiller à ce que les responsables rendent des comptes en cas de pertes d'emplois, et accroître les possibilités d'innovation.

Dignité : Si la dignité n'est pas définie dans les lignes directrices existantes, les documents de droit souple précisent que c'est une prérogative des humains et non des robots. Les références faites à la dignité sont étroitement liées à la protection et à la promotion des droits de l'homme. L'IA ne devrait pas réduire ou détruire la dignité humaine mais plutôt la respecter, la préserver, voire la renforcer. La dignité est présumée protégée si, en premier lieu, les développeurs de systèmes d'IA la respectent et si elle est favorisée par une nouvelle législation, des initiatives de gouvernance ou des lignes directrices techniques et méthodologiques édictées par les pouvoirs publics.

Solidarité : La solidarité est le thème éthique qui revient le moins souvent, et lorsqu'il est évoqué, c'est principalement en lien avec les incidences de l'IA sur le marché du travail. Les sources documentaires recommandent un renforcement des protections sociales face aux conséquences à long terme de l'IA sur le travail humain. Elles soulignent la nécessité de redistribuer les bénéfices de l'IA afin de ne pas mettre en danger la cohésion sociale^{6,5} et de respecter les personnes et groupes potentiellement vulnérables. Enfin, elles mettent en garde contre les pratiques de collecte et de traitement des données axées sur les personnes et susceptibles de fragiliser la solidarité en faveur d'un « individualisme radical ».

i. Insuffisances

Cette étude affiche plusieurs insuffisances. Tout d'abord, du point de vue bibliographique, les lignes directrices et les documents de droit souple font partie de la littérature grise et ne sont donc pas indexés dans les bases de données conventionnelles de recherche. Leur extraction est inévitablement moins reproductible et impartiale que s'ils se trouvaient dans des bases de données systématiques de documents soumis à un examen collégial. En suivant les meilleures pratiques relatives à l'examen de la littérature grise, nous avons réussi à atténuer ce problème en créant un protocole de recherche et de sélection que nous avons testé avant de recueillir les données. Comme les résultats des recherches effectuées sur des moteurs de recherche sont personnalisés, il a fallu atténuer le risque d'influence en élargissant aussi bien les mots-clés que la sélection des résultats. Comme le choix de l'anglais était susceptible de fausser les résultats en ne renvoyant qu'à des documents en anglais, nous avons limité le problème en incluant également des éléments rédigés dans les langues suivantes : allemand, français, italien, espagnol et néerlandais. Les mots-clés et les codes dans ces langues ont été traduits en anglais et inclus dans l'analyse. Notre analyse du contenu présente les insuffisances caractéristiques des méthodes d'analyse qualitative. Conformément aux meilleures pratiques en matière d'analyse de contenu, nous avons atténué le biais subjectif grâce à une stratégie de codage inductif appliquée séparément par deux évaluateurs. Enfin, vu le rythme soutenu auquel sont publiés les documents directifs relatifs à l'IA, de nouveaux documents d'orientation risquaient d'être diffusés après la fin de nos travaux. Aussi avons-nous surveillé en permanence la littérature jusqu'au 1^{er} mars 2020, tout en procédant à l'analyse des données.

ii. Débat et analyse de l'éthique normative

Nous avons constaté que le nombre et la diversité des documents de droit souple sur l'IA augmentaient rapidement, ce qui montre que la communauté internationale participe de plus en plus à l'élaboration de mesures non contraignantes portant sur ce domaine technologique. Les organismes qui établissent des lignes directrices, des principes et d'autres instruments non contraignants en matière d'IA appartiennent à des secteurs très diversifiés. Le fait que le secteur public (c.-à-d. les organismes gouvernementaux) et le secteur privé (entreprises et alliances) publient quasiment autant de documents l'un que l'autre montre que les difficultés que soulève l'IA sur le plan éthique préoccupent aussi bien les organismes publics que les entreprises privées. Il existe toutefois de nettes différences dans les solutions proposées face aux problèmes éthiques que soulève l'IA, les acteurs publics privilégiant des solutions techniques telles que l'explicabilité et l'interprétabilité de l'IA plutôt qu'une réglementation obligatoire et une réflexion éthique approfondie. Par ailleurs, le fait que certaines zones géographiques, par exemple l'Afrique et l'Amérique du Sud, soient relativement sous-représentées montre que le débat sur l'IA éthique n'a pas la même intensité partout. Les pays économiquement développés façonnent ce débat davantage que les autres, ce qui fait craindre que le savoir local, le pluralisme culturel et l'équité mondiale ne soient laissés pour compte. Ces observations confirment les inégalités de représentation et de répartition géographiques des acteurs de l'éthique de l'IA que de précédentes études avaient constatées¹³. Par rapport à celles-ci, notre rapport permet toutefois de constater que de nouveaux acteurs, issus de pays qui étaient auparavant absents du débat, prennent désormais part à l'élaboration de mesures de gouvernance non contraignantes à l'échelon international. Ils viennent de grandes puissances du secteur de l'IA, à savoir de pays leaders mondiaux dans ce secteur, comme la Chine, mais aussi de pays à revenus moyens situés dans des régions autrefois non représentées, comme la Russie et le Mexique.

La prolifération de textes non contraignants peut être interprétée comme une réponse sur le plan de la gouvernance aux recherches de pointe concernant l'IA, dont la production et le marché ont considérablement augmenté ces dernières années¹⁶. Notre analyse montre que les différents acteurs semblent commencer à converger vers la promotion des principes éthiques de transparence, de justice, de non-malfaisance, de responsabilité et de respect de la vie privée. Elle fait néanmoins apparaître de profondes divergences sur quatre grands points : i) la façon d'interpréter les principes éthiques, ii) la raison pour laquelle ils sont jugés importants, iii) les questions, domaines et acteurs auxquels ils se rapportent, et iv) la façon de les mettre en œuvre. Par ailleurs, on ne sait toujours pas quels principes éthiques doivent être prioritaires, comment résoudre les incompatibilités entre les principes, qui doit assurer la surveillance éthique de l'IA et de quelle façon les chercheurs et les institutions peuvent respecter les lignes directrices. Ces observations donnent à penser qu'il existe, au niveau de la formulation des principes et de leur mise en œuvre concrète, des lacunes que l'expertise technique ou des approches descendantes ne peuvent guère combler.

Bien qu'aucun des différents principes éthiques ne soit expressément approuvé par toutes les lignes directrices existantes, la transparence, la justice et la loyauté, la non-malfaisance, la responsabilité et le respect de la vie privée apparaissent tous dans plus de la moitié des documents. Cela pourrait indiquer qu'à l'échelon mondial, les politiques sur l'IA éthique sont en train de converger vers ces principes. En particulier, la fréquence des appels à la transparence, à la justice et à la loyauté montre qu'il devient primordial, sur le plan de l'éthique, d'exiger des processus transparents à toutes les étapes de l'IA (du développement et de la conception des algorithmes jusqu'aux modalités d'utilisation de l'IA) et de mettre en garde la communauté mondiale face au risque que l'IA renforce les inégalités si les notions de justice et de loyauté ne sont pas dûment prises en considération. Ces thèmes apparaissent étroitement liés à celui de la responsabilité, car la promotion de la transparence et de la justice semble aller de pair avec le renforcement de la responsabilité et de l'obligation de rendre compte qui incombent aux fabricants d'IA et à ceux qui la déploient.

D'aucuns ont affirmé que la transparence n'est pas un principe éthique en tant que tel mais plutôt un prérequis éthique facilitant ou compromettant d'autres pratiques ou principes éthiques¹⁷. Cette définition de la transparence en tant que prérequis éthique à d'autres principes est décelable dans la Déclaration de conformité du fournisseur d'IBM, qui donne des informations sur les quatre grands piliers de la fiabilité de l'IA. L'hypothèse selon laquelle la transparence serait un prérequis éthique

pourrait expliquer en partie pourquoi elle est plus fréquemment évoquée que d'autres principes éthiques. Il est à noter que les lignes directrices existantes accordent une grande importance à la promotion de la responsabilité et de l'obligation de rendre compte alors que peu de textes insistent sur le devoir d'agir avec intégrité qui incombe à tous les acteurs du développement et du déploiement de l'IA. Ce décalage est probablement dû au fait que, comme nous l'avons constaté, les lignes directrices existantes ne mettent pas pleinement en correspondance les principes et les conditions concrètes à remplir, plusieurs principes n'étant d'ailleurs ni définis ni reliés aux critères indispensables à leur réalisation.

Alors que les codes liés à la non-malfaisance sont plus nombreux que les codes liés à la bienfaisance, il semble qu'à l'heure actuelle, dans la communauté de l'IA, l'obligation morale de prévenir les dommages l'emporte sur la promotion des bonnes pratiques. Cela pourrait être en parti interprété comme un exemple de ce que l'on appelle un biais négatif, à savoir un biais cognitif général qui amène à donner plus de poids aux éléments négatifs^{18,19}, hypothèse récemment mise en avant par le psychologue cognitif Steven Pinker dans une analyse approfondie réalisée par l'Unité de la prospective scientifique (STOA) du Parlement européen²⁰. Ce biais négatif est accentué par le fait que les lignes directrices existantes se concentrent avant tout sur la façon de protéger la vie privée, la dignité, l'autonomie et la liberté individuelle *face aux* progrès de l'IA, et qu'elles ne se posent guère la question de savoir si ces principes pourraient être activement encouragés par des innovations responsables en matière d'IA.

La question de la confiance dans l'IA, traitée dans moins d'un tiers de toutes les sources, aborde un dilemme éthique essentiel de la gouvernance de l'IA : est-il souhaitable sur le plan de l'éthique de favoriser la confiance du public dans l'IA ? Si plusieurs sources, en particulier celles qui émanent du secteur privé, soulignent qu'il importe d'encourager la confiance dans l'IA par l'éducation et la sensibilisation, un petit nombre d'entre elles affirment que la confiance dans l'IA risque en fait d'entraîner un allègement des contrôles et d'aller à l'encontre de certaines des obligations sociétales des producteurs d'IA²¹. Cette hypothèse remet en question le point de vue dominant en matière d'éthique de l'IA, à savoir que renforcer la confiance du public dans l'IA est un critère essentiel de la gouvernance éthique²². S'agissant de la confiance, nous avons relevé d'autres difficultés conceptuelles. Tout d'abord, les documents actuels ne semblent clarifier ni le sens de la notion de confiance ni la dynamique de la confiance. La plupart des sources omettent de préciser qui est de quel côté dans la relation de confiance qu'elles décrivent, et font donc abstraction du fait qu'il s'agit d'une relation très complexe impliquant au moins deux acteurs, qui ne doivent pas douter du fait que l'autre fera ou ne fera pas quelque chose. Tout un ensemble d'éléments, par exemple la culture, les croyances, le contexte ainsi que les caractéristiques des acteurs, viennent influencer la relation de confiance. Or la littérature actuelle semble ne pas tenir compte de ces facteurs contextuels. En outre, la « fiabilité », une caractéristique, et la « confiance », un concept relationnel, paraissent fréquemment confondues ou utilisées de façon interchangeable par les acteurs de l'IA que nous avons étudiés. Il en découle non seulement une confusion entre deux notions mais aussi de faux espoirs pour les utilisateurs de l'IA et les décideurs. La confiance et la fiabilité sont deux notions différentes, la fiabilité n'étant pas à elle seule gage de confiance. Aux fins de la gouvernance, il faudrait éclaircir ces différences notionnelles capitales et exiger plus de précisions sur les impératifs d'une relation de confiance.

Le fait que les thèmes de la durabilité et de la solidarité sont relativement peu abordés pourrait signifier qu'ils ne sont actuellement pas pris en compte dans le discours éthique dominant sur l'IA. Le fait que les principes liés à la durabilité ne sont que très peu traités est particulièrement problématique, car le déploiement de l'IA exige des ressources informatiques considérables, qui entraînent à leur tour une consommation énergétique élevée²³. L'impact environnemental de l'IA ne se réduit toutefois pas aux effets négatifs de l'empreinte écologique élevée de ses infrastructures numériques : l'IA peut aussi être exploitée dans l'intérêt des écosystèmes et de la biosphère dans son intégralité. Ce dernier point, mis en avant dans un rapport du Forum économique mondial, et non dans ses lignes directrices sur l'IA, devra être approuvé à plus grande échelle pour pouvoir être incorporé dans le discours sur l'IA éthique²⁴. Le principe éthique de solidarité est peu abordé et lorsqu'il l'est, c'est le plus souvent en lien avec le développement de stratégies inclusives visant à éviter des pertes d'emploi et un partage inéquitable des fardeaux. Peu de sources documentaires évoquent le fait que l'on pourrait encourager la solidarité en s'appuyant sur la possibilité, récente,

d'utiliser l'expertise en matière d'IA pour résoudre des problèmes humanitaires, une mission que mènent actuellement, entre autres, diverses organisations intergouvernementales, comme le Bureau des Nations Unies pour les services d'appui aux projets (UNOPS) ou l'Organisation mondiale de la santé (OMS), ainsi que des entreprises du secteur privé comme Microsoft. Alors que le coût humanitaire du changement climatique causé par l'homme augmente rapidement²⁵, les principes de durabilité et de solidarité paraissent étroitement liés, mais ils sont très peu évoqués par rapport à d'autres principes.

Si les données numériques montrent qu'un consensus se forme autour de la promotion de certains principes éthiques, une analyse thématique approfondie dépeint un tableau bien plus complexe : il existe en effet des différences cruciales dans la *façon* dont ces principes sont interprétés et au sujet des critères jugés indispensables à leur réalisation. Nous avons constaté que des mesures différentes et souvent incompatibles étaient proposées pour mettre en œuvre l'IA éthique. C'est ainsi que la nécessité de disposer de jeux de données plus vastes et plus diversifiés dans un souci d'impartialité de l'IA semble difficilement conciliable avec l'obligation de donner davantage de contrôle aux personnes sur leurs données et l'utilisation qui en est faite, aux fins du respect de leur vie privée et de l'autonomie. Des contrastes similaires se font jour entre le besoin d'éviter les préjudices à tout prix et celui de trouver un équilibre entre risques et avantages. Par ailleurs, il faut noter que l'évaluation du rapport risques-avantages donnera des résultats différents selon les personnes pour le bien-être desquelles il sera optimisé et les acteurs qui s'en occuperont. Si ces questions ne sont pas résolues, des décalages et des tensions pourraient compromettre l'élaboration d'un programme mondial pour l'IA éthique.

Bien que tous les acteurs s'accordent à dire que l'IA doit être éthique, d'importantes divergences se font jour dans les lignes directrices en faveur de l'IA éthique. En outre, nul ne sait encore précisément comment les principes et lignes directrices éthiques devraient être mis en œuvre. Ces difficultés ont des incidences sur les politiques scientifiques, la gouvernance des technologies et l'éthique de la recherche. Au niveau des pouvoirs publics, il faut absolument que les organismes publics renforcent leur coopération afin d'harmoniser et de hiérarchiser les priorités de leurs programmes d'IA, tâche pour laquelle les organisations intergouvernementales peuvent jouer un rôle d'intermédiaires et de facilitatrices. Si une harmonisation est souhaitable, il ne faut pas qu'elle se fasse au détriment du pluralisme culturel et éthique en matière d'IA. Le premier défi consistera donc, pour pouvoir établir un programme mondial sur l'IA, à trouver un équilibre entre la nécessaire harmonisation entre pays et le respect de la diversité culturelle et du pluralisme éthique. Il faudra, pour ce faire, mettre en place des mécanismes de délibération pour trancher les désaccords entre les différents acteurs de chaque région du monde au sujet des valeurs et des incidences des progrès de l'IA. En matière de gouvernance technologique, l'harmonisation passe en règle générale par la normalisation. Des efforts ont été faits dans ce sens, notamment par l'Institut des ingénieurs électriciens et électroniciens (*Institute of Electrical and Electronics Engineers*, IEEE), dans le cadre de l'initiative « Ethically Aligned Design » (conception respectueuse de l'éthique)²⁶. Enfin, il sera de plus en plus souvent fait appel à des mécanismes de gouvernance souples tels que les comités d'examen indépendant (*Independent Review Boards*, IRB) pour évaluer la valeur éthique des applications de l'IA dans la recherche scientifique, tout particulièrement dans le milieu universitaire. Toutefois, les applications de l'IA par les pouvoirs publics ou des entreprises privées échapperont probablement à la surveillance des IRB, à moins que les compétences de ces derniers ne soient considérablement élargies.

Dans l'ensemble, nous avons constaté que la communauté internationale n'est pas d'accord sur ce qui constitue l'IA éthique ni sur les critères à réunir impérativement pour sa mise en œuvre. Néanmoins, des signes de convergence apparaissent autour des notions de transparence, de non-malfaisance, de responsabilité et de respect de la vie privée. Enrichir le débat actuel sur l'IA éthique en améliorant l'analyse des principes éthiques essentiels – mais peu évoqués – que sont la dignité humaine, la solidarité et la durabilité, permettrait vraisemblablement de mieux définir le cadre éthique de l'IA. Il serait par ailleurs souhaitable que les travaux, qui sont pour l'instant axés sur la formulation des principes, soient réorientés vers la réalisation de ces derniers. Un programme mondial pour l'IA éthique devrait concilier la nécessaire harmonisation entre pays et entre domaines avec le respect de la diversité culturelle et du pluralisme moral. Globalement, notre analyse offre un précieux point

de départ pour comprendre la diversité intrinsèque des principes et lignes directrices actuels sur l'IA éthique et elle donne un aperçu des difficultés qui attendent la communauté internationale.

iii. Incidences sur l'élaboration des politiques

Les nombreuses initiatives internationales visant à élaborer des documents de droit souple sur l'IA fournissent indirectement des informations utiles sur la façon dont l'humanité réagira aux multiples difficultés de gouvernance que pose l'IA. La communauté internationale semble commencer à s'accorder sur l'importance de la transparence, de la non-malfaisance, de la responsabilité et du respect de la vie privée dans le cadre du développement et du déploiement de l'IA éthique. Toutefois, enrichir l'actuel débat sur l'IA éthique en améliorant l'analyse des principes essentiels, pourtant peu évoqués, que sont la dignité humaine, la solidarité et la durabilité, permettrait vraisemblablement de mieux définir le cadre éthique de l'IA. De plus, les travaux, qui sont pour l'instant axés sur la formulation des principes, doivent être réorientés vers la réalisation de ces derniers. Un programme mondial pour l'IA éthique devrait concilier la nécessaire harmonisation entre pays et entre domaines avec le respect de la diversité culturelle et du pluralisme moral.

Ces éléments ont des incidences sur les politiques publiques, la gouvernance des technologies et l'éthique de la recherche. Au niveau des pouvoirs publics, il faut renforcer la coopération entre les parties prenantes pour harmoniser les différents programmes relatifs à l'éthique de l'IA et faire converger les procédures non seulement au sujet des principes éthiques mais aussi de leur mise en œuvre. S'il est souhaitable qu'un consensus se dégage à l'échelon mondial, il ne faut pas que ce soit au détriment du pluralisme culturel et éthique ; aussi faudrait-il peut-être mettre en place des mécanismes de délibération pour trancher les désaccords entre les différents acteurs de chaque région du monde. Dans ce cadre, des organisations intergouvernementales comme le Conseil de l'Europe peuvent jouer un rôle d'intermédiaires et de facilitatrices, en complément de quoi, des approches ascendantes associant toutes les parties prenantes concernées sur un pied d'égalité peuvent être adoptées.

Les pouvoirs publics devraient intervenir pour préciser de quelle façon les lignes directrices sur l'éthique de l'IA s'articulent avec les textes de droit national et international en vigueur. Malgré la prétendue originalité sociotechnique de l'IA, les documents de droit souple sur l'IA ne s'appliquent pas dans un vide éthico-juridique. Au contraire, les lignes directrices sur l'éthique et les autres instruments non contraignants devront au bout du compte s'appliquer dans un cadre dans lequel il existe déjà une multitude de règles, notamment de droit dur (règles de gouvernance obligatoires). Ne pas tenir compte du cadre dans lequel s'inscrivent ces règles pourrait compromettre la traduction des principes ici évoqués en mesures concrètes et efficaces de gouvernance internationale. La transparence, principe éthique le plus largement traité, en est un exemple. Le principe de transparence est certes fréquemment évoqué, mais le plus souvent sans lien explicite avec les règles sous-jacentes qui ont force contraignante. De nos jours, les établissements qui se servent de la technologie de l'IA sont déjà soumis à de nombreuses règles de transparence au titre des mécanismes juridiques en vigueur, par exemple, aux États-Unis, la loi intitulée *Fair Credit Reporting Act*, et, en Europe, les obligations concrètes des responsables du traitement et des sous-traitants qui sont énoncées aux articles 12 à 14 du Règlement général sur la protection des données (RGPD) de l'UE. De la même manière, il est indispensable que les décideurs précisent la distinction entre « confiance » et « fiabilité ».

Non seulement il faudra relever le défi de faire concorder le droit dur et le droit souple, mais aussi celui de mettre les principes éthiques en pratique tout cherchant à harmoniser des codes divergents sur l'éthique de l'IA. Au niveau de la gouvernance technologique, des tentatives d'harmonisation prometteuses ont eu lieu dans le cadre d'initiatives de normalisation telles que celles de l'IEEE, qui est la plus grande association professionnelle de métiers techniques au monde, dont les travaux sont consacrés aux progrès de l'innovation technologique. L'IEEE mène des travaux en matière d'éthique de l'IA aussi bien au sujet de systèmes généraux autonomes et intelligents, dans le cadre de l'initiative sur la conception respectueuse de l'éthique²⁶, que de systèmes spécifiques, avec, par exemple, la feuille de route qu'elle a établie en neurotechnologie au sujet des normes relatives à l'interface cerveau-machine.

L'IA a par ailleurs des incidences sur la supervision de la recherche. Il sera de plus en plus fait appel à des mécanismes d'éthique de la recherche, par exemple les comités d'examen indépendant (*Independent Review Boards*, IRB), pour évaluer la valeur éthique des applications de l'IA dans la recherche scientifique, tout particulièrement dans le milieu universitaire. Toutefois, les applications de l'IA par les pouvoirs publics ou des entreprises privées échapperont probablement à la surveillance des IRB, à moins que les compétences de ces derniers ne soient considérablement élargies.

Dans l'ensemble, la diversité thématique et la richesse informationnelle des documents que nous avons analysés conduisent à penser que les instruments non contraignants publiés par des institutions gouvernementales et des entités non gouvernementales (en ce compris les entreprises privées et les établissements universitaires) sont de précieux outils pour exercer une influence concrète sur l'élaboration des politiques concernant l'IA par les pouvoirs publics. Si elles sont adéquatement conçues et ébauchées, les initiatives de droit souple peuvent orienter le développement des systèmes d'IA en faveur du bien-être social et dans le respect des valeurs éthiques et des normes juridiques.

Il ne faut toutefois pas considérer qu'elles remplacent les instruments de gouvernance contraignants. Il est particulièrement à craindre que les acteurs du secteur privé de l'IA n'adoptent des mesures d'autoréglementation destinées à contourner ou à écarter les mesures de gouvernance contraignantes édictées par des autorités gouvernementales ou intergouvernementales. Ce risque a été mis en avant par le philosophe allemand Thomas Metzinger, membre du groupe d'experts de haut niveau de l'UE sur l'IA, qui a fait observer que le débat sur l'éthique de l'IA est en grande partie façonné par le secteur privé²⁰.

Le fait que les organismes diffusant des lignes directrices éthiques sur l'IA représentent certaines zones géographiques bien plus que d'autres doit être surveillé et examiné de près par des organisations internationales, notamment intergouvernementales. Par souci d'inclusivité, de pluralisme culturel et de participation équitable à la prise de décisions collective au sujet de l'IA, il faudrait encourager l'élaboration de documents de droit souple par des organismes situés dans les régions du monde qui sont actuellement sous-représentées, tout particulièrement l'Afrique et l'Amérique du Sud. Les organisations intergouvernementales comme le Conseil de l'Europe peuvent jouer un rôle capital dans la création de plateformes internationales d'échanges et de débats sur l'éthique et la gouvernance de l'IA.

Les instruments non contraignants en vigueur convergent vers cinq principes éthiques génériques – transparence, justice, non-malfaisance, responsabilité et respect de la vie privée – qui sont autant de domaines à surveiller en priorité, au sujet desquels des mesures de gouvernance obligatoires pourraient être adoptées aux niveaux gouvernemental et intergouvernemental. Se consacrer en priorité à la réalisation de ces principes pourrait faciliter l'élaboration d'un train de normes essentielles, fondées sur des préceptes éthiques faisant l'unanimité. Par ailleurs, comme un très grand nombre d'acteurs privés et publics adhèrent à ces principes, ils seront probablement plus largement respectés. Cela étant, afin que ces principes éthiques puissent être dûment traduits en règles de gouvernance, ils doivent être précisés sur le plan conceptuel. Il incombe aux décideurs de remédier aux ambiguïtés sémantiques de ces principes et à leurs caractéristiques antagonistes. Les profondes divergences entre les textes de droit souple en vigueur au sujet de l'interprétation et de l'application concrète de ces principes montrent que le public sera probablement divisé au sujet des solutions de gouvernance contraignantes et qu'un débat démocratique transparent sera donc nécessaire.

Parallèlement, les notions éthiques qui sont relativement peu traitées, par exemple la durabilité, la dignité et la solidarité, doivent être analysées de très près pour que les lacunes conceptuelles et normatives que présente le droit souple ne soient pas transposées dans le droit dur. Des mesures de gouvernance obligatoires doivent compléter les approches non contraignantes et combler leurs lacunes au lieu de reproduire celles-ci. Pour dûment faire face aux difficultés que pose l'IA en termes de durabilité et de solidarité, peut-être faudra-t-il renforcer la coopération entre les autorités chargées de la protection de l'environnement, les ministères du Travail et de l'Emploi ainsi que les ministères de la Technologie et de l'Innovation.

Étant donné que près de la moitié des textes de droit souple qui ont été examinés ne recommandent pas explicitement de favoriser le respect des droits de l'homme – pas plus qu'ils ne mettent pas en garde contre leur violation – lors de la conception, du développement et du déploiement des systèmes d'IA, il est urgent de se focaliser davantage sur les incidences de l'IA sur les droits de l'homme. Les pays membres du Conseil de l'Europe sont bien placés pour orienter la gouvernance internationale de l'IA vers la promotion des droits de l'homme. Les incidences de l'IA sur les droits de l'homme doivent être examinées attentivement sous divers angles. Tout d'abord sous l'angle des droits et des obligations au sens philosophique, car ils existent indépendamment de toute loi en tant que normes morales justifiées. Ensuite, sous l'angle du droit international des droits de l'homme.

À cet égard, la Convention européenne des droits de l'homme (CEDH) peut jouer un rôle essentiel dans la recherche doctrinale et la réflexion sur l'IA à l'échelon international. Son respect est une condition essentielle au développement et à l'adoption socialement responsables d'une nouvelle technologie. Il est donc de la plus haute importance d'évaluer les effets de la transformation sociotechnique qu'entraîne l'IA sur les droits et libertés fondamentaux consacrés par la CEDH.

Cette évaluation doit avoir un double objectif :

- i) examiner si l'IA aura des répercussions ou fera peser de nouveaux risques sur les droits de l'homme et les libertés, et de quelle façon ;
- ii) examiner si le développement responsable de l'IA et la consultation du public à cet égard pourraient contribuer à la promotion de ces droits et libertés, et de quelle façon.

Il faut souligner que comme les technologies ne sont pas développées dans le néant mais dans un contexte sociohistorique de pratiques, de coutumes et de normes humaines, les stratégies d'évaluation des effets de l'IA devront, pour être efficaces, s'inscrire dans le contexte des pratiques et normes actuelles et non en faire abstraction²⁷.

Enfin, il est important d'examiner l'interface entre l'IA et les droits de l'homme non seulement d'un point de vue général, mais aussi et surtout sous l'angle de la prépondérance des droits de l'homme dans des domaines d'application spécifiques de l'IA, entre autres la robotique^{8,28}, les mégadonnées^{29,30}, les armes autonomes^{31,32} et les interfaces cerveau-machine.

Remerciements

Les auteurs tiennent à remercier Mmes Anna Jobin et Karolina Ignatiadis et M. Manuel Schneider, dont les travaux ont contribué à la réalisation du présent rapport.

Bibliographie

1. Michie D, Spiegelhalter DJ, Taylor C. Machine learning. *Neural and Statistical Classification*. 1994;13.
2. Appenzeller T. The AI revolution in science. *Science*. 2017;357:16-17.
3. Harari YN. Reboot for the AI revolution. *Nature News*. 2017;550(7676):324.
4. Helbing D, Frey BS, Gigerenzer G, et al. Will democracy survive big data and artificial intelligence? In: *Towards Digital Enlightenment*. Springer; 2019:73-98.
5. Livingston S, Risse M. The Future Impact of Artificial Intelligence on Humans and Human Rights. *Ethics & International Affairs*. 2019;33(2):141-158.
6. Nemitz P. Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2018;376(2133):20180089.
7. Ashrafian H. Intelligent robots must uphold human rights. *Nature*. 2015;519(7544):391-391.
8. Van Est R, Gerritsen J, Kool L. Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality—Expert report written for the Committee on Culture. *Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE)(Rathenau Institute)*, extrait de <https://www.rathenau.nl> (5 janvier 2019), 2017.
9. Raso FA, Hilligoss H, Krishnamurthy V, Bavitz C, Kim L. Artificial Intelligence & Human Rights:

- Opportunities & Risks. *Berkman Klein Center Research Publication*. 2018(2018-6).
10. Assembly UG. Déclaration universelle des droits de l'Homme. *Assemblée générale des Nations Unies*1948;302(2).
 11. Mowbray A. The European Convention on Human Rights. In: *International Human Rights Law*. Routledge; 2016:287-304.
 12. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International journal of social research methodology*. 2005;8(1):19-32.
 13. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*. 2019;1(9):389-399.
 14. Ienca M, Ferretti A, Hurst S, Puhon M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. *PloS one*. 2018;13(10):e0204937.
 15. Li Y, Krishnamurthy R, Raghavan S, Vaithyanathan S, Jagadish H. Regular expression learning for information extraction. Document présenté lors de la conférence de 2008 sur les méthodes empiriques de traitement du langage naturel.
 16. Shoham Y, Perrault R, Brynjolfsson E, et al. The AI Index 2018 annual report. *AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA*. 2018.
 17. Turilli M, Floridi L. The ethics of information transparency. *Ethics and Information Technology*. 2009;11(2):105-112.
 18. Rozin P, Royzman EB. Negativity bias, negativity dominance, and contagion. *Personality and social psychology review*. 2001;5(4):296-320. Ito TA, Larsen JT, Smith NK, Cacioppo JT. Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of personality and social psychology*. 1998;75(4):887.
 19. Peter J. Bentley, Miles Brundage, Olle Häggström, Thomas Metzinger. Should we fear artificial intelligence? *European Parliamentary Research Service*. 2018:Scientific Foresight Unit (STOA).
 20. Bryson J. No one should trust artificial intelligence. *Science & Technology: Innovation, Governance, Technology*. 2018;11:14.
 21. Winfield AF, Jirotko M. Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering sciences*. 2018;376(2133):20180085.
 22. Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:190602243*. 2019.
 23. Forum WE. Harnessing Artificial Intelligence for the Earth.
 24. Scheffran J, Brzoska M, Kominek J, Link PM, Schilling J. Climate change and violent conflict. *Science*. 2012;336(6083):869-871.
 25. IEEE. Ethically aligned design. *IEEE Standards v1*. 2016(Global Initiative).
 26. Rasmussen T. *Social theory and communication technology*. Routledge; 2019.
 27. Liu H-Y, Zawieska K. From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence. *Ethics and Information Technology*. 2017:1-13.
 28. Mantelero A. AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review*. 2018;34(4):754-772.
 29. Vayena E, Tasioulas J. The dynamics of big data and human rights: The case of scientific research. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2016;374(2083):20160129.
 30. Heyns C. Human rights and the use of autonomous weapons systems (AWS) during domestic law enforcement. *Hum Rts Q*. 2016;38:350.
 31. Asaro P. On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross*. 2012;94(886):687-709.
 32. Ienca M, Andorno R. Towards new human rights in the age of neuroscience and neurotechnology. *Life Sci Soc Policy*. 2017;13(1):5.

CHAPITRE III. Analyse des instruments internationaux juridiquement contraignants⁶⁶

Alessandro Mantelero⁶⁷

I. Résumé

Les développements les plus récents de l'intelligence artificielle (IA) ont une action transformatrice de plus en plus marquée sur la société et amènent de nouvelles interrogations dans des domaines aussi divers que la médecine prédictive, la modération de contenus, le « quantified self » (automesure connectée) et les systèmes judiciaires, sans oublier les questions relatives à leur impact environnemental.

Une analyse des instruments juridiques internationaux contraignants est donc le point de départ obligé pour définir le cadre juridique existant, en recenser les valeurs fondamentales et vérifier si ce cadre et ses principes répondent bien à l'ensemble des problématiques soulevées par l'IA.

Afin de continuer à assurer l'harmonisation du cadre juridique actuel dans le domaine des droits de l'homme, de la démocratie et de l'État de droit, la présente étude a pour objectif de contribuer à l'élaboration de la future réglementation de l'IA en s'appuyant sur les instruments contraignants existants, en contextualisant leurs principes et en fournissant les principales orientations réglementaires d'un **futur cadre juridique**.

Cette approche a pour fondement théorique l'idée que les principes généraux fournis par les instruments internationaux relatifs aux droits de l'homme devraient sous-tendre l'ensemble des activités humaines, y compris l'innovation basée sur l'intelligence artificielle. Par ailleurs, seul le cadre des droits de l'homme peut offrir **une référence universelle à la réglementation de l'IA**, tandis que les autres sphères (par exemple l'éthique) n'ont pas la même dimension internationale, sont davantage liées au contexte et se caractérisent par une diversité d'approches théoriques.

L'analyse des instruments juridiques contraignants actuels contenue dans ce document ne se limite pas à une étude d'harmonisation, qui consiste à extraire des valeurs et des principes communs d'un ensemble donné de règles relatives à l'IA. Elle comprend également un examen plus structuré en plusieurs étapes.

Après une première analyse sectorielle visant à recenser les grands principes directeurs dans quatre domaines essentiels (protection des données, santé, démocratie et justice), ces principes sont contextualisés compte tenu des mutations opérées par l'IA au sein de la société. Ce faisant, nous tirons profit des instruments non contraignants existants qui appliquent avec un plus grand niveau de

⁶⁶ Rapport préparé pour le CAHAI, Strasbourg, 15 juin 2020, intitulé CAHAI(2020)08-fin.

⁶⁷ Maître de conférences en droit privé et en éthique et protection des données, université polytechnique de Turin (Politecnico di Torino). Les opinions exprimées dans cette analyse ne reflètent pas nécessairement la position du CAHAI ou du Conseil de l'Europe

détail les principes contenus dans les instruments juridiques contraignants et donnent également, dans certains cas, des orientations spécifiques sur l'IA.

Cette **contextualisation des principes directeurs et des valeurs juridiques** permet de les formuler de manière plus fine, en tenant compte de la nature spécifique des produits et services d'IA. En outre, elle permet de mieux répondre aux défis soulevés par l'IA. Par conséquent, il est possible de définir **un premier ensemble de dispositions pour la future réglementation de l'IA**, en mettant l'accent sur les points les plus problématiques de chaque secteur examiné.

Compte tenu du nombre important de documents adoptés par divers organismes internationaux et intergouvernementaux et de l'étude parallèle sur les instruments d'éthique actuellement menée par le CAHAI, ce document se limite aux instruments juridiquement contraignants et aux instruments non contraignants adoptés pour leur mise en œuvre.

L'étude comporte deux parties. La première identifie la portée et la méthodologie de cette analyse, tandis que la seconde présente les résultats de l'analyse sectorielle des principes directeurs.

Dans l'analyse sectorielle, les deux premiers domaines essentiels examinés sont la **santé** et la **protection des données**. L'intersection entre ces deux sphères est intéressante pour la présente étude, compte tenu du grand nombre d'applications de l'IA relatives aux données de santé et de leurs dénominateurs communs que l'on retrouve dans plusieurs dispositions de la Convention d'Oviedo, de la Convention 108+ et des instruments non contraignants. L'autre question qui occupe une place centrale aussi bien dans le domaine de la protection des données que dans celui de la biomédecine est celle de l'autodétermination de l'individu ; les problématiques liées à l'IA présentent donc un intérêt tout particulier à cet égard et suscitent des préoccupations communes du point de vue de la complexité et de l'opacité des opérations de traitement.

Les quatrième et cinquième sous-parties s'intéressent à la **démocratie** et à la **justice**. Ici, le champ d'investigation est plus large et il n'existe pas d'instruments juridiques complets pouvant offrir des principes sectoriels spécifiques comme la Convention 108+ ou la Convention d'Oviedo. Par conséquent, l'analyse se concentre davantage sur les principes essentiels et leur contextualisation ; en comparaison avec les sous-parties précédentes moins de dispositions clés sont élaborées.

La dernière sous-partie donne une vue d'ensemble des principes directeurs recensés et suggère un cadre d'harmonisation qui met en avant les corrélations existantes et les dénominateurs communs entre ces principes, tout en soulignant les contributions uniques de chaque secteur à la future réglementation de l'IA.

L'objectif principal de cette étude n'est pas d'ajouter une nouvelle liste de principes directeurs à ceux déjà fournis par diverses instances et entités, mais d'atteindre un résultat différent sur le plan méthodologique et matériel.

Premièrement, **l'analyse effectuée et la solution proposée trouvent leurs racines dans les droits de l'homme et les libertés**, adoptant une approche concrète centrée sur les instruments juridiques internationaux existants. Les autres études sont souvent sectorielles et s'appuient sur différents ensembles de références normatives (nationales ou régionales) ou adoptent une approche théorique énonçant des principes ou se référant aux droits de l'homme de manière générale et abstraite. Bien que ces travaux fassent avancer le débat éthique et juridique sur l'IA, leur impact en termes de contribution au cadre réglementaire est limité et n'est pas spécialement contextualisé dans le cadre des normes du Conseil de l'Europe relatives aux droits de l'homme, à la démocratie et à l'État de droit.

Deuxièmement, le résultat de cette analyse des instruments juridiquement contraignants, y compris les instruments non contraignants adoptés pour les mettre en œuvre, n'est pas simplement une liste de principes aussi précis soient-ils. **Il est important de recenser les principes directeurs communs, mais cela ne suffit pas à constituer une feuille de route pour la future**

réglementation de l'IA. La transparence, l'obligation de rendre compte, le contrôle humain et de nombreux autres principes déjà énumérés dans plusieurs chartes sur l'IA sont des concepts abstraits qui ne sont pas bien contextualisés.

Tel est l'objectif principal de cette étude, fournir précisément cette **contextualisation eu égard au cadre juridique et aux problématiques de l'IA.** Si ce document parvient à suggérer des **moyens concrets et efficaces de formuler et de codifier les principes directeurs de l'IA** et à concrètement **ancrer les normes du Conseil de l'Europe relatives aux droits de l'homme, à la démocratie et à l'État de droit dans le projet de réglementation de l'IA,** il aura atteint son objectif d'aider à structurer la relation entre les humains et l'IA d'un point de vue juridique.

II. Méthodologie

À l'instar d'Internet, de l'électricité ou de la propulsion à vapeur, l'intelligence artificielle (IA) englobe une multitude de technologies – dont les interactions humain-robot – qui ont des répercussions importantes sur diverses activités humaines et sur la société.

Dans ce contexte, de nombreuses questions se posent. Par exemple, quand un système d'IA devrait-il prendre une décision ? Quels critères le système devrait-il appliquer ? Qui est responsable des décisions susceptibles d'avoir des effets négatifs sur les personnes et la société ? À la lumière de ces questions et de nombreuses questions émergentes, les réglementations existantes devraient être réexaminées.

Une analyse des instruments juridiques internationaux contraignants est donc le point de départ obligé pour définir le cadre juridique existant, en recenser les valeurs fondamentales et vérifier si ce cadre et ces valeurs répondent bien à l'ensemble des problématiques soulevées par l'IA, afin de continuer à assurer l'harmonisation du cadre juridique actuel dans le domaine des droits de l'homme, de la démocratie et de l'État de droit.

Le but n'est pas de créer un tout nouveau cadre de référence complet, la réglementation devant mettre l'accent sur les changements que l'IA apportera à la société plutôt que de repenser tous les domaines dans lesquels elle peut s'appliquer⁶⁸. Cette approche ciblée est rendue possible en utilisant les instruments contraignants existants et en contextualisant leurs principes directeurs, puis en définissant les principales règles d'un futur cadre juridique relatif à l'IA qui pourra couvrir les domaines non encore réglementés par les instruments contraignants existants.

Il est important de souligner à cet égard la différence entre les instruments juridiquement contraignants existants et d'autres documents comme les instruments non contraignants ou les chartes d'éthique relatives à l'IA. Les instruments juridiquement contraignants existaient avant l'essor actuel de l'IA. Ils n'ont pas été rédigés en pensant à l'intelligence artificielle et ne fixent pas un ensemble spécifique de règles applicables à ce domaine, tandis que les instruments non contraignants et les documents d'éthique relatifs à l'IA portent sur des points précis, mais les abordent sous des angles différents.

L'analyse des instruments juridiques contraignants existants ne se limite donc pas à une étude d'harmonisation (qui consisterait à extraire des valeurs et principes communs d'un ensemble donné de règles relatives à l'IA), mais nécessite un processus plus structuré qui englobe l'harmonisation sans s'y limiter. Ce processus, décrit dans les parties suivantes, peut être divisé en trois étapes distinctes : (i) inventaire et une identification des concepts clé; (ii) une contextualisation ; et (iii) une harmonisation

⁶⁸ voir par exemple l'approche de l'UE en matière de réglementation du commerce électronique

i. Le scénario

Les développements les plus récents de l'IA ont une action transformatrice de plus en plus marquée sur la société et amènent de nouvelles interrogations dans des domaines aussi divers que la médecine prédictive, la modération de contenus, le *quantified self* (automesure connectée) et les systèmes judiciaires, sans oublier les questions relatives à leur impact environnemental.

L'évolution rapide de l'IA appliquée ces dernières années est incompatible avec l'adoption d'une réponse juridique spécifique sous la forme d'instruments internationaux juridiquement contraignants centrés sur l'IA. C'est pourquoi deux stratégies opérationnelles différentes ont été mises en œuvre pour traiter ces questions : (i) un travail considérable d'interprétation du cadre juridique existant à la lumière des problématiques liées à l'IA (voir par exemple le débat en cours sur les dispositions du RGPD relatives à transparence et aux décisions automatisées) ; (ii) le recours à des instruments non contraignants pour contextualiser les principes contenus dans les instruments contraignants existants (par exemple Lignes directrices T-PD(2019)01 sur l'intelligence artificielle et la protection des données ; CEPEJ. 2019. Charte éthique européenne d'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement⁶⁹).

La future réglementation de l'IA devrait par conséquent prendre appui sur ces initiatives, en mettant l'accent d'une part sur les principes directeurs et valeurs tirés des instruments contraignants existants et d'autre part, sur leurs instruments de mise en œuvre non contraignants, qui envisagent déjà dans certains cas le nouveau scénario de l'IA.

ii. Axe de recherche et méthodologie

Cette étude vise principalement à définir les principes de base de la future réglementation de l'intelligence artificielle en analysant le cadre juridique existant. La méthodologie sera donc nécessairement déductive et consistera à extraire les principes fondamentaux des diverses réglementations applicables aux domaines dans lesquels des solutions d'intelligence artificielle pourraient être adoptées.

Cette approche a pour fondement théorique l'idée que les principes généraux fournis par les instruments internationaux relatifs aux droits de l'homme devraient sous-tendre l'ensemble des activités humaines, y compris l'innovation basée sur l'intelligence artificielle⁷⁰. Par ailleurs, seul le cadre des droits de l'homme peut offrir une référence universelle pour la réglementation de l'IA, tandis que les autres sphères (par exemple l'éthique) n'ont pas la même dimension internationale, sont davantage liées au contexte et se caractérisent par une diversité d'approches théoriques.

Dans ce contexte, de nombreuses questions se posent. Par exemple, quand un système d'IA devrait-il prendre une décision ? Quels critères le système devrait-il appliquer ? Qui est responsable des décisions susceptibles d'avoir des effets négatifs sur les personnes et la société ? À la lumière de ces questions et de nombreuses questions émergentes, les réglementations existantes devraient être réexaminées.

Afin de fournir un cadre réglementaire harmonisé pour relever les défis de l'IA, des orientations communes de haut niveau sur les principes et les valeurs à consacrer devraient être tirées des chartes internationales des droits de l'homme (par exemple, la Déclaration universelle des droits de

⁶⁹ Commission européenne pour l'efficacité de la justice (CEPEJ). 2018. Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement.

⁷⁰ Voir Comité directeur sur les médias et la société de l'information (CDMSI). 2019. Projet de Recommandation du Comité des Ministres aux États membres concernant les impacts des systèmes algorithmiques sur les droits de l'homme (préparé et finalisé par le Comité d'experts sur la dimension droits de l'homme des traitements automatisés de données et différentes formes d'intelligence artificielle (MSI- AUT) ; Conseil de l'Europe - Comité d'experts sur les intermédiaires d'Internet (MSI-NET). 2018

l'homme, le Pacte international relatif aux droits civils et politiques, la Convention de sauvegarde des droits de l'homme et des libertés fondamentales, la Charte des droits fondamentaux de l'Union européenne).

Les principes directeurs doivent être envisagés dans le cadre d'un scénario de transformation basé sur l'IA, qui, dans de nombreux cas, nécessitera leur adaptation. Ces principes restent valables, mais leur fonctionnement doit être reconsidéré à la lumière des changements sociaux et techniques induits par l'IA (par exemple, la liberté de choix en cas de "boîtes noires"). Cela permettra une application plus contextualisée et plus granulaire des principes afin qu'ils puissent apporter une contribution concrète à la forme de la future réglementation sur l'IA.

Pour mener cette étude, il faut commencer par définir les principaux sujets d'étude, en considérant à la fois les effets potentiels de l'IA et les domaines d'action du Conseil de l'Europe. À cet égard, quatre domaines essentiels ont été sélectionnés : les données, la santé, la démocratie et la justice.

iii. Analyse et résultats escomptés

L'étude adopte une approche descendante en vue de contribuer au futur cadre réglementaire relatif à l'IA qui sera mis en œuvre au moyen d'instruments contraignants complétés par des instruments non contraignants, sur le modèle de ce qui a été fait dans le domaine de la biomédecine. Il devrait en ressortir un ensemble de dispositions relatives aux domaines étudiés et les principes directeurs communs essentiels, reposant sur une analyse globale de tous les instruments, contraignants ou non, adoptés.

Première étape : recensement des principes essentiels. Des principes directeurs seront recensés dans les domaines étudiés. L'analyse commence par les différentes thématiques, car les instruments contraignants sont spécifiques à des secteurs et non basés sur des droits. Les deux tableaux suivants donnent un premier aperçu de cet exercice de recensement, qui part d'une vue d'ensemble des sphères de la protection des données et de la justice afin d'identifier les principes directeurs de la future réglementation de l'IA.

Figure 1 : Protection des données

| | |
|--|--|
| Instrument contraignants | Convention 108+ Convention sur la cybercriminalité |
| Zones touchées | Systèmes de prise de décision Vie privée et dimension collective du groupe Profilage |
| Instrument non contraignants connexes | CdE. 2019. Lignes directrices sur les implications de l'intelligence artificielle en matière de protection des données |

| | |
|--|---|
| | <p>CdE. 2017. Lignes directrices sur la protection des personnes physiques à l'égard du traitement des données à caractère personnel dans un monde de "Big Data"</p> <p>CdE. 2010. Recommandation sur la protection des personnes physiques à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage [en cours de révision].</p> <p>UNESCO. 2019. Étude préliminaire sur un éventuel instrument normatif sur l'éthique de l'intelligence artificielle</p> <p>OCDE. 2019. Recommandation du Conseil sur l'intelligence artificielle</p> <p>40e Conférence internationale des commissaires à la protection des données et de la vie privée. 2018</p> |
| Principes directeurs et valeurs juridiques | <p>Responsabilité</p> <p>Une approche fondée sur les risques Le principe de précaution</p> <p>Qualité et sécurité des données Transparence</p> <p>Équité</p> <p>Approche contextuelle Rôle des experts</p> <p>Participation / Inclusion / Liberté de choix / Autonomie / Contrôle / Supervision humaine / Sensibilisation / Alphabétisation</p> <p>Innovation responsable</p> <p>Coopération entre les autorités de contrôle</p> |

Figure 2 : Justice

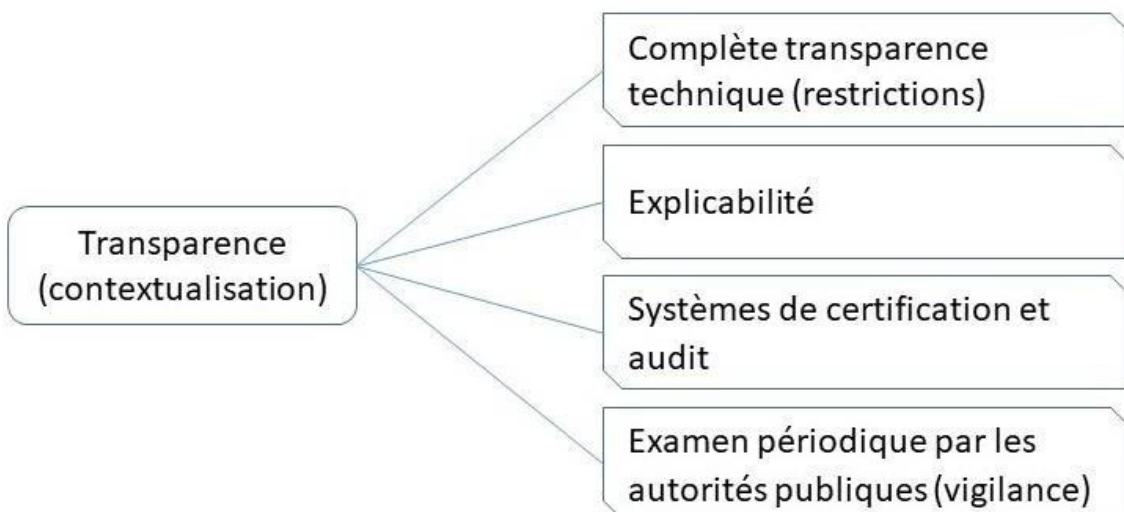
| | |
|----------------------------|---|
| Instrumentes contraignants | <p>Déclaration universelle des droits de l'homme</p> <p>Pacte international relatif aux droits civils et politiques</p> <p>Convention internationale sur l'élimination de toutes les formes de discrimination raciale</p> <p>Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes</p> <p>Convention de sauvegarde des droits de l'homme et des libertés fondamentales</p> <p>Charte des droits fondamentaux de l'Union européenne</p> |
| Zones touchées | <p>Traitement des décisions judiciaires et des données La police prédictive</p> |

| | |
|--|--|
| Instruments non contraignants connexes | CEPEJ. 2019. Charte éthique européenne sur l'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement . |
| Principes directeurs et valeurs juridiques | Non-discrimination Qualité et sécurité des données Transparence Impartialité Équité Approche contextuelle Liberté de choix/ Indépendance des juges (processus décisionnel) Contrôle/supervision humain Garanties du droit à un procès équitable |

Deuxième étape : Contextualisation. Les valeurs directrices identifiées dans l'exercice de cartographie doivent être contextualisées à la lumière des changements de société produits par l'IA. Cette phase bénéficiera des instruments non contraignants existants qui fournissent des applications plus granulaires des principes inscrits dans les instruments contraignants, en fournissant également dans certains cas des orientations spécifiques sur l'IA.

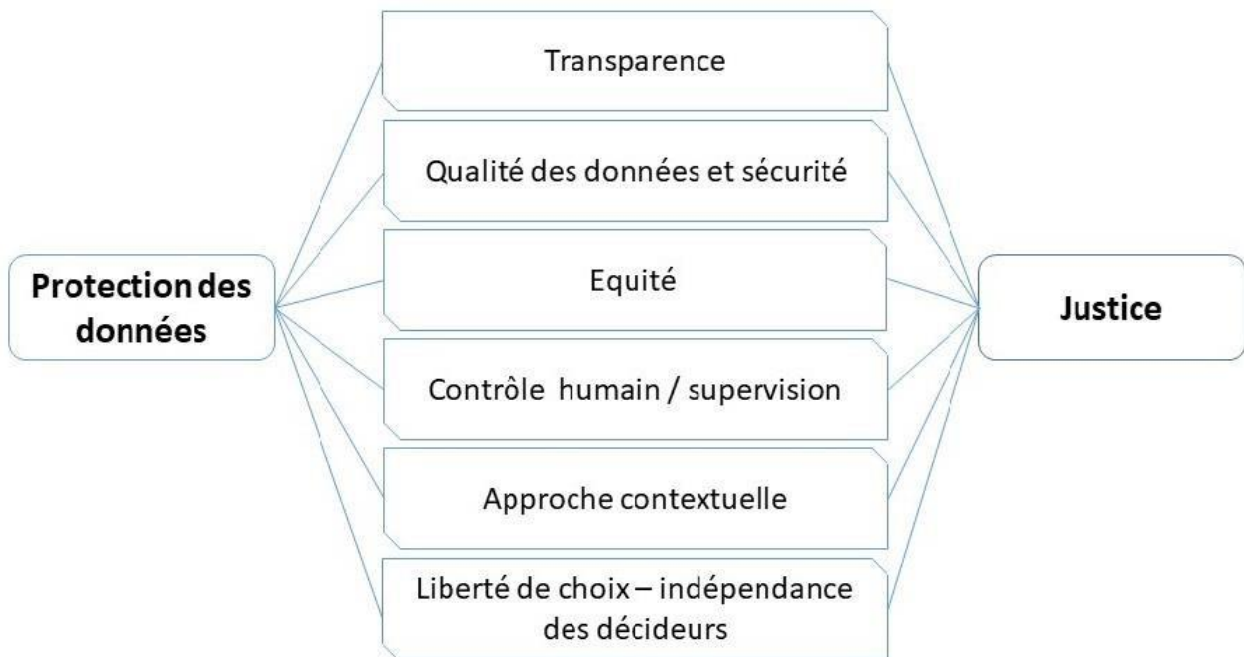
Cette contextualisation des principes directeurs et des valeurs juridiques permettra de les formuler de manière plus précise et plus élaborée, compte tenu de la nature spécifique des produits et services de l'IA. À ce stade, il sera donc possible de formuler une première série de dispositions pour la future réglementation sur l'IA en se concentrant sur les questions les plus difficiles dans chaque secteur.

Figure 3 : Mise en œuvre adaptée au contexte du principe de transparence



Troisième étape : harmonisation (analyse intersectorielle). À partir de l'analyse sectorielle menée dans cette étude, une liste de principes directeurs essentiels communs aux différentes sphères sera établie dans la dernière sous-partie (figure 4). Ces principes constitueront le fondement commun des futures dispositions relatives à l'IA.

Figure 4 : Valeurs fondamentales communes aux domaines de la protection des données et de la justice



III. Analyse

Compte tenu du nombre important de documents adoptés par divers organismes internationaux et intergouvernementaux et de l'étude parallèle sur les instruments d'éthique actuellement menée par le CAHAI, cette partie se limite aux instruments juridiquement contraignants.

Les codes d'éthique ne seront donc pas examinés à ce stade et les documents relatifs aux futures stratégies réglementaires (par ex. livres blancs) ne seront pris en compte qu'à titre d'informations contextuelles.

Cette partie se subdivise en six sections, suivies de conclusions provisoires. La première sous-partie contient une vue d'ensemble des instruments existants adoptés par le Conseil de l'Europe et des principes/valeurs sur lesquels ils reposent. Cet inventaire aidera à définir les principes fondamentaux d'une future réglementation de l'IA et à assurer sa cohérence avec le cadre existant.

Les deuxième et troisième sous-parties mettent l'accent sur deux domaines essentiels allant de pair : la santé et la protection des données. L'intersection entre ces deux sphères est intéressante pour la présente étude sur les fondements de la future réglementation de l'IA, compte tenu du grand nombre d'applications de l'IA relatives aux données de santé et de leurs dénominateurs communs que l'on retrouve dans plusieurs dispositions de la Convention d'Oviedo, de la Convention 108+ et d'autres instruments non contraignants⁷¹. L'autre question qui occupe une place centrale aussi bien dans le domaine de la protection des données que dans celui de la biomédecine est celle de l'autodétermination de l'individu ; les problématiques liées à l'IA présentent donc un intérêt tout particulier à cet égard et suscitent des préoccupations communes du point de vue de la complexité et de l'opacité des opérations de traitement.

Les quatrième et cinquième sous-parties s'intéressent à la démocratie et à la justice. Ici, le champ d'investigation est plus large et il n'existe pas d'instruments juridiques généraux pouvant offrir des principes sectoriels, comme la Convention 108+ ou la Convention d'Oviedo. Par conséquent,

⁷¹ Voir Recommandation CM/REC(2019)2 sur la protection des données de santé

l'analyse se focalise sur les principes essentiels et leur contextualisation ; en comparaison avec les sous-parties précédentes moins de dispositions clés sont élaborées.

La sixième sous-partie donne une vue d'ensemble des principes directeurs recensés et suggère un cadre d'harmonisation qui met en avant à la fois les corrélations existantes entre ces principes et la contribution unique de chaque secteur à la future réglementation de l'IA.

Comme il a été souligné dans les commentaires reçus lors de l'exercice de veille décrit dans la prochaine sous-partie, les technologies d'IA ont un impact sur de très nombreux secteurs et posent des questions qui mettent en jeu un vaste ensemble d'instruments réglementaires.⁷² Cette première étude constitue donc un point de départ axé sur les quatre domaines essentiels cités. Malgré les limites de l'analyse, les résultats valident la méthodologie proposée et donnent un certain nombre d'indications pour les futures dispositions de la réglementation de l'IA.

i. Vue d'ensemble

L'IA s'appliquant à de nombreuses situations⁷³ couvertes par différents instruments contraignants qui portent sur divers domaines, il convient d'analyser l'ensemble des données disponibles pour définir les principes fondamentaux et les valeurs communes à prendre en considération dans la future réglementation.

Un travail de veille a été mené en ce sens entre le 12 et le 28 février 2020 auprès des différents secteurs du Conseil de l'Europe pour tirer parti de l'expertise spécifique des services qui travaillent depuis de nombreuses années sur divers sujets liés aux droits de l'homme, à la démocratie et à l'Etat de droit.

Au moyen d'un questionnaire ouvert, il leur a été demandé de donner des informations sur les points suivants : (i) instruments contraignants, (ii) domaines concernés (applications), (iii) instruments non contraignants connexes, (iv) principes directeurs et valeurs juridiques et (v) principes/questions non abordés. Une grande diversité d'informations a pu être recueillie grâce à la participation active des différents secteurs.

D'un point de vue méthodologique, la structure de cette enquête préliminaire, basée sur des questions ouvertes, a nécessairement un impact sur les résultats de l'analyse quantitative. Les principales limites rencontrées sont dues à l'utilisation de catégories générales qui se chevauchent parfois et au niveau de précision des réponses.

Néanmoins, par agrégation au niveau macro et prise en compte des similarités (fréquence) entre les principes et valeurs recensés, il a été possible de mettre en perspective les résultats et d'obtenir une cartographie plus détaillée des instruments non contraignants existants adoptés par le Conseil de l'Europe, pouvant servir à établir le cadre juridique d'une future réglementation (voir annexe 1).

S'agissant des domaines concernés (voir annexe 2), l'exercice suggère de bâtir la future réglementation de l'IA autour de deux grands axes : l'utilisation de l'IA et le développement de l'IA. Dans les deux cas, différents droits de l'homme et libertés fondamentales pourraient être menacés ou au contraire, jouer un rôle important dans les futurs scénarios de l'IA⁷⁴.

Pour ce qui est de l'utilisation de l'IA, on recense quatre grands domaines d'application et donc de réglementation : l'analyse prédictive et les systèmes d'aide à la décision, l'analyse prédictive et les systèmes de prise de décision automatisée, la collecte de preuves/informatique judiciaire et la production de contenus.

⁷² Voir Annexe 1

⁷³ Voir également UNESCO, « Étude préliminaire concernant un éventuel instrument normatif sur l'éthique de l'intelligence artificielle », bibliothèque numérique de l'Unesco, consultée le 21 novembre 2019, <https://unesdoc.unesco.org/ark:/48223/pf0000369455>

⁷⁴ Voir Conseil de l'Europe- Comité d'experts sur les intermédiaires d'Internet (MSI-NET) 2018

Les deux premiers domaines sont bien connus et étudiés, car ils recouvrent un ensemble extrêmement vaste d'applications (voir annexe 2). Cela étant, la distinction entre systèmes d'aide à la décision et systèmes de décision automatisée est essentielle du point de vue de la conception centrée sur les valeurs et du rôle de l'humain dans les processus décisionnels : ces deux types de systèmes diffèrent dans leur nature et devront donc faire l'objet de garanties procédurales et matérielles distinctes dans la réglementation de l'IA.

Les deux autres domaines d'application, bien qu'ils relèvent de secteurs spécifiques, devraient être examinés séparément car ils ne concernent pas directement le processus décisionnel mais fournissent les éléments qui en constituent le fondement (collecte de preuves et informatique judiciaire) ou ont une influence sur les processus de création (production de contenus). Dans ce cas, l'enjeu principal semble résider davantage dans les aspects procéduraux et leur cohérence avec les approches traditionnelles (non fondées sur l'IA).

Bien que la plupart des travaux et recommandations existants se concentrent sur les systèmes d'IA et leurs conséquences potentielles, deux autres aspects - le développement de l'IA et la prestation de services d'IA – ont également un impact important sur les droits de l'homme et les libertés fondamentales. C'est pourquoi la future réglementation devrait également prendre en considération les questions liées aux conditions de travail des personnes impliquées dans l'ensemble de la chaîne d'approvisionnement de produits et services d'IA⁷⁵.

Le deuxième ensemble d'informations fourni par l'exercice de veille concerne les principes directeurs et valeurs juridiques qui devraient sous-tendre le développement et les usages futurs de l'IA (voir annexe 3). La diversité des notions employées par les services interrogés plaide en faveur d'une agrégation des principes et des valeurs. Le résultat de cette analyse a permis de regrouper les principes directeurs et valeurs fondamentales autour d'un certain nombre d'éléments dont la répartition était la suivante en termes de fréquence :

Non-discrimination (15)

Diversité, inclusion et pluralisme (13) Protection des données et de la vie privée (11)

Transparence (9)

Égalité (8)

Accès à la justice, droit à un procès équitable (7) Contrôle humain (7)

Impartialité (6)

Accès à l'information (5) Sécurité (5)

Traitement équitable (5)

Participation (5) Liberté de choix (5)

Liberté d'expression et de création (5)

Obligation de rendre compte (3) Compétence et capacité (2) Indépendance (3)

Autonomie individuelle (3)

Coopération culturelle (2)

Durabilité (2)

Malgré les limites de l'analyse mentionnée, il apparaît clairement que les trois premiers principes sont considérés comme des éléments clés de la future réglementation de l'IA et qu'ils en constitueront donc les axes principaux. Ce constat est confirmé par le deuxième ensemble de principes/valeurs, qui est étroitement lié au premier : la transparence et le contrôle humain sont des facteurs importants dans la lutte contre les discriminations et la protection des données, tandis que

⁷⁵ Voir également Crawford et Joler 2018

l'accès à la justice est essentiel pour réagir aux éventuels manquements à ces valeurs. De la même manière, bien que plus substantiellement, l'égalité est à de nombreux égards liée aux trois premiers grands principes/valeurs. Il existe de plus grandes différences entre les autres valeurs/principes, qui se rapportent à des questions spécifiques liées à la mise en œuvre de l'IA.

Cet exercice a permis de recenser les grands principes directeurs de la réglementation de l'IA, déjà codifiés dans des instruments juridiques contraignants et non contraignants, mais nécessitant une adaptation au contexte de l'IA. Lors de l'analyse sectorielle, cette contextualisation sera réalisée sur la base d'une analyse approfondie des instruments internationaux juridiquement contraignants, en faisant le point sur les éventuelles lacunes dans le cadre réglementaire existant, secteur par secteur.

L'IA étant une technologie intersectorielle, cette analyse devrait laisser entrevoir des possibilités d'interventions réglementaires du même type dans d'autres domaines, comme indiqué dans la partie consacrée à la méthodologie⁷⁶. À l'issue de l'analyse sectorielle, ces interventions potentielles seront classées pour éviter tout doublon, puis réunies dans un cadre cohérent reposant sur des valeurs fondamentales qui devraient concorder avec les résultats de l'exercice de veille.

ii. Protection des données

Ces dix dernières années, le cadre réglementaire international dans le domaine de la protection des données a subi une vaste refonte. Les instruments juridiques établis sur la base de principes définis dans les années 1970 et 1980⁷⁷ n'étaient plus adaptés au nouveau paysage social et technique apparu avec l'augmentation de la largeur de bande disponible pour le transfert de données, la conservation de données et les ressources informatiques (informatique en nuage), la mise en données progressive de parts importantes de nos vies et de notre environnement (Internet des objets), l'analyse de données à grande échelle et l'analyse prédictive reposant sur les mégadonnées et l'apprentissage-machine.

En Europe, les principales réponses à ce changement sont la version modernisée de la Convention 108 (Convention 108+) et le RGPD. Une redéfinition du cadre réglementaire est également en cours ou a été menée dans d'autres contextes internationaux comme l'OCDE⁷⁸ – ou par certains pays.

Cela dit, la dernière vague de développement de l'IA a été si rapide que certaines problématiques spécifiques à l'IA n'ont pas pu être traitées directement par ces nouveaux instruments contraignants ; plusieurs instruments non contraignants ont donc été adoptés pour combler ces lacunes et les futures stratégies réglementaires sont en cours d'examen⁷⁹. L'analyse menée aux fins de la présente étude a porté sur les instruments juridiques non contraignants suivants⁸⁰ : T-PD(2019)01, Lignes directrices sur l'intelligence artificielle et la protection des données (GAI) ; T-PD(2017)1, Lignes directrices sur la protection des personnes à l'égard du traitement des données à caractère personnel à l'ère des mégadonnées ; Recommandation CM/Rec(2019)2 du Comité des Ministres du Conseil de l'Europe aux États membres en matière de protection des données relatives à la santé⁸¹ ; Recommandation CM/Rec(2010)13 du Comité des Ministres du Conseil de l'Europe aux États membres sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage ; [UNESCO]. 2019. Étude préliminaire concernant un

⁷⁶ Voir ci-dessus partie I.3

⁷⁷ Voir also Mayer-Schönberger, 1997; González Fuster, 2014.

⁷⁸ Voir OCDE. 2013. Recommandation du Conseil de l'OCDE relative aux Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel, C(80)58/FINAL, telle que modifiée le 11 juillet 2013 par la C(2013)79.

⁷⁹ Commission européenne. 2020. Rapport sur les conséquences de l'intelligence artificielle, de l'internet des objets et de la robotique sur la sécurité et la responsabilité, COM(2020) 64 final ; Commission européenne. 2020. Livre blanc sur l'intelligence artificielle – une approche européenne axée sur l'excellence et la confiance, COM(2020) 65 final. Voir également Commission européenne. 2020. Une stratégie européenne pour les données, COM(2020) 66 final

⁸⁰ Voir annexe 4

⁸¹ Cette recommandation remplace la Recommandation Rec(97)5 du Comité des Ministres aux États membres sur la protection des données médicales. Voir également Rec(2016)8 sur le traitement des données à caractère personnel relatives à la santé à des fins d'assurance, y compris les données résultant de tests génétiques et son exposé des motifs.

éventuel instrument normatif sur l'éthique de l'intelligence artificielle⁸² ; [OCDE]. 2019. Recommandation du Conseil sur l'intelligence artificielle ; 40^e Conférence internationale des commissaires à la protection des données et de la vie privée. 2018 [ICDPPC]. Déclaration sur l'éthique et la protection des données dans le secteur de l'intelligence artificielle.

Ces instruments sont de nature différente : tandis que les instruments du Conseil de l'Europe contiennent des exigences et dispositions particulières, les autres établissent un certain nombre de principes mais ne fournissent pas, ou en partie seulement, d'indications plus détaillées quant à des exigences spécifiques. Les paragraphes suivants illustrent les principes essentiels tirés de ces différents instruments et la façon dont ils peuvent être contextualisés dans le cadre de l'IA.

Parmi les principes classés dans le domaine de la protection des données à caractère personnel (par ex. qualité des données), certains peuvent être étendus aux données non personnelles, principalement pour ce qui est de l'impact de l'utilisation de données non personnelles (par ex. données agrégées) sur les personnes et les groupes dans le cadre des processus décisionnels (par ex. données sur la mobilité ou sur la consommation d'énergie).

Primauté de l'être humain

Les systèmes d'IA devraient être conçus pour servir l'humanité ; toute création, évolution et utilisation de systèmes d'IA doit respecter pleinement les droits de l'homme, la démocratie et l'Etat de droit.⁸³

Contrôle humain

Les applications de l'IA devraient permettre l'exercice d'un contrôle humain significatif de leurs effets sur les individus et la société⁸⁴.

Transparence et extensibilité

Toute personne qui interagit directement avec un système d'IA a le droit de recevoir des informations adéquates et facilement compréhensibles sur l'objectif et les effets de ce dernier, y compris l'existence de décisions automatisées, pour vérifier qu'il reste conforme aux attentes des personnes concernées, pour assurer un contrôle humain global sur les systèmes d'IA et pour permettre aux personnes qui subissent les effets néfastes d'un système d'IA de contester son résultat.⁸⁵

Toute personne a le droit d'obtenir, à sa demande, connaissance du raisonnement qui sous-tend un processus décisionnel reposant sur l'IA lorsque les résultats de ce processus lui sont appliqués.⁸⁶

Les États devraient soutenir la recherche scientifique sur l'intelligence artificielle explicable et les meilleures pratiques en matière de transparence et de vérifiabilité des systèmes d'IA.⁸⁷

Approche de précaution

Lorsque les risques potentiels des applications d'IA sont inconnus ou incertains, le développement de l'IA doit reposer sur le principe de précaution.⁸⁸

⁸² Bien que le titre ne mentionne que l'éthique, l'objet de cette étude est présenté comme suit : « Le présent document contient l'étude préliminaire sur les aspects techniques et juridiques liés à l'opportunité d'un instrument normatif sur l'éthique de l'intelligence artificielle, ainsi que les commentaires et observations du Conseil exécutif à ce sujet ».

⁸³ Voir CM/Rec(2019)2, ICDPPC, GAI II.1, UNESCO ; voir également considérant n° 4 RGPD.

⁸⁴ Voir GAI I.6

⁸⁵ Voir ICDPPC, CM/Rec(2019)2, OCDE, UNESCO. Voir également Comité d'experts sur la dimension « droits de l'homme » des traitements automatisés de données et des différentes formes d'intelligence artificielle (MSI-AUT). 2019.

⁸⁶ Voir Convention 108+, GAI II.11

⁸⁷ Voir ICDPPC

⁸⁸ Voir GAI II.2

Gestion des risques

Les développeurs, fabricants et prestataires de service en IA devraient évaluer et répertorier les éventuelles conséquences négatives des applications d'IA sur les droits de l'homme et les libertés fondamentales et adopter des mesures appropriées de prévention et de réduction des risques dès la conception (approche centrée sur les droits de l'homme dès la conception) et pendant toute leur durée de vie.⁸⁹ Les conséquences préjudiciables incluent celles qui résultent de l'utilisation de données décontextualisées et de modèles algorithmiques décontextualisés.⁹⁰

Les développeurs, fabricants et prestataires de service en IA devraient consulter les autorités de contrôle compétentes dès lors que les applications de l'IA peuvent avoir un impact significatif sur les droits de l'homme et les libertés fondamentales des individus.⁹¹

Risque de réidentification

Des mesures appropriées devraient être mises en place pour éviter que des données anonymes et agrégées ne puissent déboucher sur une réidentification des personnes concernées⁹²

Qualité et minimisation des données

Les développeurs d'IA devraient évaluer de manière critique la qualité, la nature, l'origine et la quantité des données à caractère personnel utilisées, en limitant les données inutiles, redondantes ou marginales lors des phases de conception et d'apprentissage de l'IA puis en vérifiant l'exactitude du modèle à mesure qu'il est alimenté par de nouvelles données. Le recours à des données synthétiques peut être considéré comme une solution possible pour minimiser la quantité de données personnelles traitées par des applications de l'IA.⁹³

Rôle des experts

Les développeurs, fabricants et prestataires de service en IA sont encouragés à créer et à consulter des comités indépendants composés d'experts issus de différents domaines ainsi qu'à collaborer avec des institutions universitaires indépendantes pouvant contribuer à concevoir des applications de l'IA fondées sur les droits de l'homme et orientées de façon éthique et sociale, ainsi qu'à détecter des biais potentiels. Ces comités peuvent jouer un rôle particulièrement important dans les secteurs où la transparence et la mobilisation des parties prenantes peuvent être plus difficiles à assurer en raison d'intérêts et de droits concurrents, par exemple dans les domaines de la justice prédictive, de la prévention du crime et de la détection des infractions.⁹⁴

Des mécanismes adéquats devraient être mis en place pour assurer l'indépendance de ces comités d'experts.⁹⁵

Participation et contrôle démocratique du développement de l'IA

Des formes participatives d'évaluation des risques, reposant sur l'engagement actif des personnes et groupes potentiellement concernés par les applications de l'IA, devront être mises en place.

⁸⁹ Voir GAI II.2 et II.3, ICDPPC, OCDE, UNESCO Voir également Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems.

⁹⁰ Voir GAI II.5

⁹¹ Voir GAI III.5

⁹² Voir CM/Rec(2010)13

⁹³ Voir GAI II.4, OCDE. Voir également CM/Rec(2020)1

⁹⁴ Voir Section II.3.

⁹⁵ Voir GAI II.6 et II.7, ICDPPC ; voir également article 11, UNESCO. Déclaration universelle sur le génome humain et les droits de l'homme (11 novembre 1997).

Les personnes, les groupes et les autres parties prenantes devraient être informés et impliqués de façon active dans le débat sur le rôle que l'IA devrait jouer dans la formation des dynamiques sociales et dans les processus décisionnels qui les concernent.⁹⁶

Des dérogations peuvent être prévues dans l'intérêt général, lorsqu'elles constituent des mesures proportionnées dans une société démocratique et sont assorties de garanties adéquates.

Contrôle humain

Les produits et services de l'IA doivent être conçus de manière à garantir le droit des personnes de ne pas être soumises à des décisions qui les affectent de manière significative, prises uniquement sur le fondement d'un traitement automatisé de données, sans que leur point de vue soit pris en compte. Les produits et services de l'IA doivent permettre un contrôle humain global.⁹⁷

Le rôle de l'intervention humaine dans les processus décisionnels reposant sur l'IA et la liberté des décideurs humains de ne pas suivre les résultats de recommandations fondées sur l'utilisation de l'IA devraient être préservés.⁹⁸

Vigilance sur les algorithmes

Les développeurs, fabricants et prestataires de service en IA devraient adopter diverses formes de vigilance sur les algorithmes contribuant à la responsabilisation de toutes les parties prenantes par l'évaluation et la description des impacts attendus sur les individus et la société tout au long du cycle de vie du système d'IA, afin d'assurer le respect des droits de l'homme, de l'État de droit et de la démocratie.⁹⁹ Les gouvernements devraient présenter des rapports réguliers sur leur utilisation de l'IA dans la police, le renseignement et la sécurité.¹⁰⁰

Liberté de choix

Afin d'accroître la confiance des utilisateurs, les développeurs, fabricants et prestataires de services en IA sont encouragés à concevoir leurs produits et services de manière à préserver la liberté de choix des utilisateurs concernant l'usage de l'IA, en proposant des alternatives réalistes aux applications de l'IA.¹⁰¹

Droit d'opposition

Le droit d'opposition quant aux systèmes d'IA basés sur des technologies qui influencent les opinions et le développement personnel des individus devrait être garanti.¹⁰²

Interopérabilité

L'interopérabilité entre les systèmes d'IA doit être mise en œuvre conformément aux principes de licéité, de nécessité et de proportionnalité, en mettant en place des garanties adéquates concernant les droits de l'homme, la démocratie et l'État de droit.¹⁰³

⁹⁶ Voir GAI II.7 et III.8, ICDPPC. Voir également CM/Rec(2020)1

⁹⁷ Voir Convention 108+, GAI II.8, ICDPPC, UNESCO

⁹⁸ Voir GAI III. 4

⁹⁹ Voir GAI II.10, OCDE, ICDPPC . Voir également CM/Rec(2020)1

¹⁰⁰ Voir UNESCO

¹⁰¹ Voir GAI II.9

¹⁰² Voir GAI II.12 Voir également Section II.4

¹⁰³ Voir CM/Rec(2019)2

Coopération

Une coopération doit être encouragée entre les autorités de contrôle ayant des compétences liées à l'IA.¹⁰⁴

Éducation, éducation au numérique et formation professionnelle

Les responsables politiques devraient investir des ressources dans l'éducation au numérique et l'éducation afin que les personnes concernées connaissent et comprennent mieux les applications de l'IA et leurs effets. Ils devraient également encourager la formation professionnelle des développeurs en IA pour les sensibiliser aux effets potentiels de l'IA sur les personnes et la société. Ils devraient soutenir la recherche sur l'IA orientée vers les droits de l'homme.¹⁰⁵

Intégrité de la recherche scientifique

Lorsqu'une personne décide de se retirer d'un projet de recherche scientifique, le retrait du consentement ne compromet pas la licéité du traitement fondé sur le consentement effectué avant ce retrait ; ses données à caractère personnel devraient être détruites ou anonymisées de manière à ne pas compromettre la validité scientifique de la recherche et la personne concernée devrait en être informée.¹⁰⁶

iii. Santé

Le cadre réglementaire européen en matière de santé se compose de quelques instruments contraignants du Conseil de l'Europe et d'un certain nombre d'instruments sectoriels adoptés au niveau de l'UE, conformément à la nature, aux domaines d'activité et aux compétences réglementaires respectives de ces deux entités.

La Convention européenne des droits de l'homme, la Convention 108+ et la Charte sociale européenne contiennent plusieurs dispositions générales sur la protection de la santé et les droits connexes. Ces dispositions et principes, déjà énoncés dans d'autres instruments généraux au niveau international¹⁰⁷, sont toutefois contextualisés plus largement dans la Convention d'Oviedo.

La Convention d'Oviedo – seul instrument contraignant multilatéral entièrement consacré à la biomédecine – et ses protocoles additionnels constituent donc la principale référence pour définir les principes fondamentaux applicables dans ce domaine¹⁰⁸, principes qui devront être précisés et si nécessaire étoffés pour réglementer les applications de l'IA. La Convention est complétée par deux instruments non contraignants, la Recommandation sur les données relatives à la santé¹⁰⁹ et la Recommandation sur la recherche utilisant du matériel biologique d'origine humaine¹¹⁰. La première de ces deux recommandations illustre bien le lien étroit, examiné ci-après, entre la biomédecine (et plus généralement les soins de santé) et le traitement de données.

La réglementation en vigueur dans le domaine de la santé concerne principalement les traitements médicaux, la recherche (y compris les essais médicaux) et les dispositifs/produits médicaux. L'IA

¹⁰⁴ Voir ICDPPC, GAI III.6

¹⁰⁵ Voir ICDPPC, OCDE, GAI III.9, UNESCO. Voir également CM/Rec(2020)1

¹⁰⁶ Voir Convention 108+, CM/Rec(2019)2

¹⁰⁷ Certains principes généraux consacrés par cette convention étaient déjà présents dans de précédents instruments internationaux de droits de l'homme, tels que le Pacte international relatif aux droits civils et politiques, le Pacte international relatif aux droits économiques, sociaux et culturels et la Convention internationale des droits de l'enfant du 20 novembre 1989.

¹⁰⁸ Andorno, 2005 ; Seatzu, 2015.

¹⁰⁹ Voir la Recommandation Rec(2019)2 du Comité des Ministres aux États membres en matière de protection des données relatives à la santé.

¹¹⁰ Voir la Recommandation Rec(2016)6 du Comité des Ministres aux États membres sur la recherche utilisant du matériel biologique d'origine humaine.

peut avoir un impact sur tous ces domaines, compte tenu de ses applications dans la médecine de précision¹¹¹, le diagnostic et les dispositifs et services médicaux.

Bien que la Convention d'Oviedo et les instruments non contraignants connexes aient été adoptés avant l'avènement de l'IA, ils apportent des garanties spécifiques concernant l'autodétermination, le traitement du génome humain et la recherche impliquant des êtres humains, non modifiées par l'application de l'IA à ce secteur.

Les problématiques qui se posent quant à l'autodétermination dans le domaine de la biomédecine sont les mêmes que celles qui ont déjà été évoquées dans le domaine du traitement des données. Malgré les différences de nature entre le consentement au traitement médical et le consentement au traitement de données, le haut niveau de complexité – et souvent, une certaine obscurité – des applications de l'IA peuvent compromettre dans les deux cas l'exercice effectif de l'autonomie individuelle¹¹².

À cet égard, la principale contribution de la Convention d'Oviedo à la future réglementation de l'IA ne provient pas des garanties qu'elle apporte dans un secteur donné, mais réside dans l'ensemble important de principes et de valeurs généraux qui peuvent en être tirés pour servir de fondement à cette réglementation.

Les principaux apports du Conseil de l'Europe dans le domaine de la médecine sont au nombre de huit et concernent la dignité humaine, la primauté de l'être humain, les normes professionnelles, la règle générale sur le consentement éclairé, la vie privée et le droit à l'information, la non-discrimination, la protection des personnes se prêtant à une recherche et le débat public. La contribution de la Convention d'Oviedo au débat sur la future réglementation de l'IA va au-delà de la biomédecine car plusieurs de ses dispositions, centrées sur le juste équilibre entre technologie et droits de l'homme, peuvent être étendues de manière générale au domaine de l'IA, comme expliqué dans les paragraphes suivants¹¹³.

Primauté de l'être humain

Dans un contexte géopolitique et économique caractérisé par une course au développement de l'IA, la primauté de l'être humain doit être affirmée de façon générale comme un élément essentiel de l'approche européenne : l'amélioration des performances des systèmes reposant sur l'IA et leur rentabilité ne doivent pas l'emporter sur les intérêts et le bien-être des personnes.

Ce principe devrait s'appliquer aussi bien à la conception des systèmes d'IA (par ex. mise au point de systèmes qui violent les droits de l'homme et les libertés fondamentales) qu'à leurs usages.¹¹⁴

Accès équitable aux soins de santé

Le principe de l'accès équitable peut être étendu à l'accès aux effets bénéfiques de l'IA. Cela suppose l'adoption de mesures adéquates pour traiter les risques liés à la fracture numérique, à la discrimination, à la marginalisation des personnes vulnérables ou des minorités culturelles et aux restrictions d'accès à l'information.¹¹⁵

Normes professionnelles

Le développement de l'IA recouvre une multitude de domaines d'expertise et lorsque la mise au point de systèmes d'IA peut avoir un impact sur les personnes et sur la société, elle doit être effectuée

¹¹¹ Voir Azencott, 2018 ; Ferryman et Pitcan, 2018.

¹¹² Voir ci-dessus, section 2.1 de cette partie

¹¹³ La dignité humaine et le consentement éclairé ne figurent pas dans le tableau car la première est une valeur commune aux instruments adoptés par le Conseil de l'Europe dans le domaine des droits de l'homme, de la démocratie et de l'Etat de droit (voir paragraphe 1, partie II) et le second est un principe également applicable dans le cadre du traitement de données.

¹¹⁴ See also Oviedo Convention, Article 2.

¹¹⁵ See also Oviedo Convention, Article 3

conformément aux obligations et aux normes professionnelles applicables dans chaque domaine d'expertise concerné. Les normes et compétences professionnelles requises doivent se fonder sur l'état actuel des connaissances.¹¹⁶

Les États devraient encourager la formation des professionnels pour améliorer leur connaissance et leur compréhension de l'IA et de ses effets potentiels sur les personnes et sur la société. Ils devraient soutenir la recherche relative à l'IA axée sur les droits de l'homme.

Les États devraient coopérer en vue de l'élaboration de programmes de formation communs et de normes communes pour les professionnels concernés par l'IA et la société.

Lorsque l'IA est mise en œuvre dans le secteur médical, une attention particulière doit être portée à la confiance du patient envers son médecin et à la confiance mutuelle, qui ne doit pas être compromise par le recours à l'IA.

La décision de recourir à l'IA dans les processus décisionnels concernant les personnes et la société doit reposer sur le principe de proportionnalité, en prenant en compte les intérêts et le bien-être des personnes.

Protection des personnes n'ayant pas la capacité de consentir et qui n'ont pas la capacité de consentir à une recherche

Le respect du principe du bénéfice doit être considéré comme une obligation lorsque le consentement individuel est soumis à des restrictions compte tenu de la complexité ou de l'opacité des traitements reposant sur l'IA et ne peut constituer le fondement exclusif du traitement.¹¹⁷

Vie privée et droit à l'information

Conformément à l'article 10 de la Convention d'Oviedo, les applications d'IA en matière de santé doivent garantir le droit à l'information et respecter la volonté des personnes de ne pas être informées, hormis lorsque le respect de cette volonté constitue un risque grave pour la santé d'autrui¹¹⁸.

Non-discrimination

Le principe de non-discrimination dans le domaine de la santé doit être complété en interdisant toute forme de discrimination contre une personne ou un groupe fondée sur des prédictions de problèmes de santé futurs¹¹⁹.

Rôle des experts

L'expérience des comités d'éthique dans le domaine de la biomédecine doit être prise en compte en associant des comités d'experts multidisciplinaires à l'évaluation des applications de l'IA.¹²⁰

Débat public

Les évolutions de l'IA posent des questions fondamentales qui doivent être soumises à un débat public justifié au regard de leurs incidences sur le plan médical, social, économique, éthique et juridique, et leurs applications éventuelles doivent faire l'objet de consultations appropriées.¹²¹

Ces considérations montrent que le cadre juridique en vigueur dans le domaine de la biomédecine contient des principes et des éléments importants qui peuvent être étendus à la future réglementation

¹¹⁶ Voir Oviedo Convention, Article 4; Recommendation CM/Rec(2019)2 on the protection of health-related data.

¹¹⁷ Voir Oviedo Convention, Articles 6 and 17.

¹¹⁸ Voir Oviedo Convention, Article 10

¹¹⁹ Voir article 5, CM/Rec(2016)6. Voir également Oviedo Convention, Article 11.

¹²⁰ Voir Oviedo Convention, Article 16.

¹²¹ Voir Oviedo Convention, Article 28.

de l'IA, au-delà du secteur de la santé. Cela dit, vu l'impact de l'IA, il reste encore un certain nombre de lacunes à combler dans les domaines suivants.

a) Systèmes de prise de décision [approche contextuelle, équité, qualité des données, surveillance/contrôle humains]

Ces dernières années, un nombre croissant d'applications de l'IA ont été développées et sont utilisées dans le secteur médical à des fins de diagnostic, grâce à des solutions d'analytique et d'apprentissage machine. De vastes ensembles de données sont ainsi créés et l'analyse prédictive est utilisée pour tenter de trouver des solutions à des cas cliniques en s'appuyant sur les connaissances et les pratiques existantes. De la même manière, en ce qui concerne la reconnaissance d'images, les applications d'apprentissage machine sembleraient pouvoir améliorer les capacités de dépistage du cancer. Pour ce qui est de la médecine de précision, la collecte et l'analyse à grande échelle de sources de données multiples (données médicales mais également données non médicales, comme la qualité de l'air et la qualité du logement) servent à obtenir des indications individualisées sur la santé et la maladie.

L'utilisation des données cliniques, des connaissances et pratiques médicales mais aussi de données non médicales n'est pas nouveau en soi dans le secteur de la médecine et des études de santé publique. Cela dit, l'ampleur de la collecte de données, le niveau de détail de l'information recueillie, la complexité (et dans certains cas l'opacité) du traitement de données et la nature prédictive des résultats de l'analyse font s'interroger sur les systèmes de prise de décisions et leurs éventuelles faiblesses.

La plupart de ces questions ne se limitent pas au secteur de la santé car les biais potentiels (notamment le manque de diversité et l'exclusion des valeurs aberrantes et des populations de plus petite taille), la qualité des données, la décontextualisation, la nature contextuelle de l'étiquetage des données et la réutilisation des données¹²² sont des caractéristiques communes à de nombreuses applications de l'IA et concernent les données en général¹²³. En cohérence avec la méthodologie adoptée¹²⁴, les lignes directrices existantes dans le domaine de la protection des données¹²⁵ peuvent aussi s'appliquer dans ce cas et les aspects relatifs à la qualité des données peuvent être étendus aux données non personnelles.

b) Autodétermination [liberté de choix/autonomie, sensibilisation]

L'opacité des applications d'IA et l'utilisation novatrice des données dans des analyses à grande échelle bouleversent la notion traditionnelle de consentement, que ce soit pour le traitement de données¹²⁶ ou les traitements médicaux, ce qui suggère l'adoption de nouvelles formes de consentement comme le consentement large¹²⁷ ou dynamique, lesquelles ne contribueraient toutefois qu'en partie à résoudre le problème.

¹²² Ferryman et Pitcan, 2018, 19-20 (« les désignations des maladies, par exemple la notion d'infection, n'étant pas clairement définies, elles peuvent décrire des réalités cliniques très différentes » et « ces enregistrements n'ont pas été conçus à des fins de recherche mais de facturation, ce qui peut être une source d'erreurs et de biais systématiques »)

¹²³ Voir sous-partie II.2 ci-dessus

¹²⁴ Voir par ex. ci-dessus figure 3 : Valeurs fondamentales communes aux domaines de la protection des données et de la justice

¹²⁵ Voir Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatique des données à caractère personnel 2017. Lignes directrices sur la protection des personnes à l'égard du traitement des données à caractère personnel à l'ère des mégadonnées. T-PD(2017)1 ; Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatique des données à caractère personnel 2019. Lignes directrices sur l'intelligence artificielle et la protection des données. T-PD(2019)01. Voir également les études préliminaires connexes : Mantelero, A. 2019 ; Rouvroy, A. 2016..

¹²⁶ Voir également la Recommandation Rec(2019)2 du Comité des Ministres du Conseil de l'Europe aux Etats membres en matière de protection des données relatives à la santé.

¹²⁷ Voir également le rapport explicatif de la Convention 108+, par. 43 (« En matière de recherche scientifique, il arrive fréquemment qu'il ne soit pas possible de cerner entièrement la finalité du traitement des données à caractère personnel à des fins de recherche scientifique au moment de la collecte des données. Par conséquent, les personnes concernées devraient pouvoir donner leur consentement en ce qui concerne certains domaines de la recherche scientifique, dans le respect des normes éthiques reconnues en matière de recherche scientifique. Les personnes concernées devraient pouvoir donner leur consentement uniquement pour ce qui est de certains domaines de la recherche ou de certaines parties de projets de recherche, dans la mesure où la finalité visée le permet ») et la Recommandation CM/Rec(2019)2 du Comité des Ministres du Conseil de l'Europe aux Etats membres en matière de protection des données relatives à la santé, 15.6 (« Dans la mesure où il n'est pas toujours possible de définir de façon préalable les finalités des différents projets de recherche au moment de la collecte des données, les personnes concernées devraient pouvoir

c) Relation médecin-patient

Plusieurs aspects du diagnostic par IA, comme la perte des informations qui ne peuvent être encodées en données¹²⁸, le recours excessif à l'IA dans les décisions médicales, les effets des pratiques locales sur les ensembles de données d'entraînement et le risque de déqualification dans le secteur médical¹²⁹ peuvent avoir une incidence sur la relation entre les soignants et le patient¹³⁰ et devraient être évalués lorsque l'IA est adoptée dans ce domaine.

d) Gestion des risques [approche axée sur les risques, obligation de rendre compte]

Les dispositifs médicaux¹³¹ constituent un exemple intéressant sur le plan de la gestion des risques car leur utilisation peut avoir d'importantes conséquences sur les personnes. L'Union européenne a adopté en la matière un modèle de classification basé sur les risques¹³² qui prévoit des garanties progressives selon la classe de risque de chaque dispositif (procédures d'évaluation de conformité placées sous la seule responsabilité du fabricant ou avec l'intervention d'un organisme notifié, inspection par un organisme notifié ou, pour les risques les plus élevés, mise sur le marché soumise à autorisation préalable).

Un modèle reposant sur de telles garanties progressives pourrait être généralisé à la future réglementation de l'IA et adopté également en dehors du domaine des dispositifs médicaux, en prêtant attention à l'impact sur les droits de l'homme et les libertés fondamentales. Cela dit, il est plus difficile de procéder à une classification des produits/services de l'IA en raison de leur diversité et de leurs différents domaines d'application : il faudrait alors établir plusieurs classifications sectorielles, ou adopter des critères généraux sur la base de procédures d'évaluation des risques.

Des dispositions spécifiques sur la vigilance dans le domaine de l'IA et l'adoption du principe de précaution pour le développement de l'IA, dont il a été question précédemment¹³³, peuvent contribuer à résoudre ces problèmes.

iv. Démocratie

La démocratie couvre un très large éventail de défis juridiques et sociétaux¹³⁴ dont la plupart sont susceptibles d'être relevés avec le soutien des TIC¹³⁵. En effet, l'IA peut jouer un rôle important dans le développement présent et futur de la démocratie numérique au sein d'une société de l'information.

En comparaison avec les autres domaines examinés (protection des données et santé), ce sujet est si vaste qu'il est difficile de se référer à un seul instrument juridique contraignant spécifique au secteur. Plusieurs instruments internationaux traitent de la démocratie et de ses différents aspects, à commencer par la Déclaration des droits de l'homme des Nations Unies et le Pacte international relatif aux droits civils et politiques. De la même manière, dans le contexte européen, les principes essentiels de la démocratie sont présents dans plusieurs sources internationales.

donner un consentement uniquement pour certains domaines de recherche ou certaines parties de projets de recherche, dans la mesure où la finalité visée le permet et en tenant compte des normes éthiques reconnues »).

¹²⁸ Caruana et autres, 2015.

¹²⁹ Cabitza, Rasoini et Gensini, 2017.

¹³⁰ Voir aussi Déclaration d'Helsinki de l'AMM – Principes éthiques applicables à la recherche médicale impliquant des êtres humains, 9 juillet 2018, <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>.

¹³¹ Commission européenne, 2014.

¹³² Directive 93/42/CEE

¹³³ Voir ci-dessus paragraphe 2.1 de cette partie.

¹³⁴ Voir par ex. Conseil de l'Europe. Direction générale de la démocratie - Comité européen sur la démocratie et la gouvernance. 2016. Recueil des textes les plus pertinents du Conseil de l'Europe dans le domaine de la démocratie.

¹³⁵ Voir par ex. Direction générale de la démocratie et des affaires politiques - Direction des institutions démocratiques. 2009. Projet « Bonne gouvernance dans la société de l'information », CM(2009)9 Addendum 3. Guides indicatifs et Glossaire relatifs à la Recommandation Rec(2009) 1 du Comité des Ministres aux États membres sur la démocratie électronique (e-démocratie), élaborés par le Comité ad hoc pour la démocratie électronique (CAHDE) du Conseil de l'Europe ; Protocole additionnel à la Charte européenne de l'autonomie locale sur le droit de participer aux affaires des collectivités locales, 2009, article 2.2.iii.

Sur la base de l'article 25 du PIDCP, nous pouvons identifier deux principaux domaines d'intervention : i) la participation¹³⁶ et la bonne gouvernance et ii) les élections. Il est sans conteste difficile, voire impossible de tracer une ligne rouge entre ces domaines du fait qu'ils sont interconnectés de diverses manières. L'IA peut avoir un impact sur tous : la participation (par ex. l'engagement des citoyens, les plateformes de participation), la bonne gouvernance (par ex. l'e-gouvernance, les processus décisionnels, les villes intelligentes), la phase préélectorale (par ex. le financement, le ciblage et le profilage, la propagande), les élections (par ex. la prévision des résultats électoraux, le vote électronique) et la période postélectorale (par ex. le règlement des litiges électoraux).

Comme dans toute classification, cette distinction se caractérise par une marge directionnelle. Il convient de souligner ici qu'il s'agit d'une classification fonctionnelle basée sur les différents impacts de l'IA, sans aucune intention de donner une représentation juridique ou politique de la démocratie et de ses divers éléments clés. La relation entre la participation, la bonne gouvernance et les élections peut ainsi être considérée sous différents angles et façonnée de plusieurs façons, uniformisant certains domaines ou les divisant davantage.

La participation s'exprime tant par la participation au débat démocratique que par le processus électoral. Cependant, dans ces deux cas, les outils d'IA interagissent différemment avec la participation et il existe des instruments juridiques internationaux distincts spécifiques au processus électoral.

Participation et bonne gouvernance

Le droit de participer aux affaires publiques (article 25 du Pacte), étroitement lié à la liberté d'expression, de réunion et d'association,¹³⁷ repose sur un vaste concept d'« affaires publiques »¹³⁸ qui englobe à la fois le débat public et le dialogue entre les citoyens et leurs représentants. À cet égard, l'IA est pertinente à deux points de vue : comme moyen de participation et comme objet des décisions participatives.

Si l'on considère l'IA comme un moyen, les obstacles techniques et éducatifs peuvent compromettre l'exercice du droit de participer. Les outils de participation fondés sur l'IA devraient donc tenir compte des risques de sous-représentation et de manque de transparence dans les processus participatifs (par ex. les plateformes de rédaction des projets de loi. Parallèlement, l'IA fait également l'objet de décisions participatives étant donné qu'elles incluent des décisions portant sur le développement de l'IA en général et sur son utilisation dans les affaires publiques.

Les plateformes participatives basées sur l'IA (Consul,¹³⁹ Citizenlab,¹⁴⁰ Decidim¹⁴¹) peuvent contribuer de manière significative au processus démocratique, en facilitant les échanges entre citoyens, la définition des objectifs prioritaires et les approches collaboratives en matière de prise de décisions¹⁴² sur des sujets d'intérêt général à différents niveaux (quartier, municipalité, métropole, région, pays).¹⁴³ Ces plateformes étant utilisées dans un environnement social et dans le cadre de la collecte d'informations, les mêmes aspects déjà examinés concernant la protection des données,

¹³⁶ Pour une analyse plus détaillée, voir Faye Jacobsen, 2013. Voir également Maisley, 2017

¹³⁷ Voir Haut-Commissariat des Nations Unies aux droits de l'homme. 1996. Observation générale n° 25 : Le droit de participer aux affaires publiques, le droit de vote et le droit d'accéder, dans des conditions d'égalité, aux fonctions publiques (article 25). CCPR/C/21/Rev.1/Add.7

¹³⁸ Voir également Haut-Commissariat des Nations Unies aux droits de l'homme. 1981. CESCR, Observation générale n° 1 : Rapports des États parties, paragraphe 5 (« de faciliter l'évaluation, par l'opinion publique, des politiques nationales en matière de droits économiques, sociaux et culturels, et d'encourager la participation des divers secteurs économiques, sociaux et culturels de la société à la formulation de ces politiques, à leur mise en œuvre et à leur réexamen »).

¹³⁹ Voir <<https://consulproject.org/en/>>, consulté le 29.12.2019. 73

¹⁴⁰ Voir <<https://www.citizenlab.co/>>, consulté le 29.12.2019.

¹⁴¹ <https://decidim.org/>>, consulté le 29.12.2019

¹⁴² Voir également Conseil de l'Europe. Lignes directrices relatives à la participation civile aux décisions politiques CM(2017)83-final. Adoptées par le Comité des Ministres le 27 septembre 2017 lors de la 1295e réunion des Délégués des Ministres.

¹⁴³ Voir également la Recommandation CM/Rec(2009)2 sur l'évaluation, l'audit et le suivi de la participation et des politiques de la participation aux niveaux local et régional.

notamment la sécurité, peuvent être rappelés ici en étendant à ces applications les lignes directrices sur les données abordées dans la sous-partie précédente.

Cependant, d'autres questions plus spécifiques se posent en ce qui concerne les outils d'IA pour la participation démocratique (notamment ceux pour la prévention et la lutte contre la corruption¹⁴⁴) ; elles concernent les quatre principaux domaines suivants : la transparence, **l'obligation de rendre compte**, **l'inclusivité**, et **l'ouverture**. À cet égard, les principes généraux énoncés dans les instruments internationaux contraignants trouvent une application importante dans la Recommandation CM/Rec(2009)1 du Comité des Ministres aux États membres sur la démocratie électronique (e-démocratie), qui jette les bases d'une élaboration plus précise des principes directeurs en matière de démocratie dans le domaine de l'IA.

La **transparence** est une exigence pour l'utilisation d'applications technologiques à des fins démocratiques.¹⁴⁵ Ce principe est commun aux sujets analysés ci-dessus, les données et la santé. Il s'agit, toutefois, d'une notion fondée sur le contexte. Si dans ces domaines la transparence est étroitement liée à l'autodétermination, elle prend ici une signification plus large. Dans un processus démocratique, la transparence n'est pas seulement une condition de l'autodétermination des citoyens relativement à un outil technique, c'est aussi un élément du processus participatif démocratique.¹⁴⁶ La transparence n'a plus de dimension individuelle, mais revêt une dimension collective comme garantie du processus démocratique.

Dans ce contexte, les solutions de démocratie électronique fondées sur l'IA doivent être utilisées en toute transparence eu égard à leur logique et à leur fonctionnement (par ex. la sélection du contenu des plateformes participatives), grâce à la communication d'informations claires, facilement accessibles, intelligibles et actualisées sur les outils d'IA concernés.¹⁴⁷

De plus, la mise en œuvre de cette notion de transparence devrait aussi tenir compte de la grande diversité d'utilisateurs de ces outils. L'adoption d'une approche **accessible**¹⁴⁸ dès les premiers stades de la conception des outils d'IA permettrait de garantir une transparence effective à l'égard des groupes vulnérables et des personnes handicapées, apportant une valeur ajoutée à l'accessibilité dans ce contexte.

La transparence et l'accessibilité sont étroitement liées à la nature de l'architecture utilisée pour élaborer les systèmes d'IA. Les solutions **à code source ouvert et les normes ouvertes**¹⁴⁹ peuvent ainsi contribuer au contrôle démocratique des applications d'IA les plus critiques.¹⁵⁰ Dans certains cas, l'ouverture est compromise par des restrictions dues à la nature de l'application d'IA spécifique (par ex. la prévention de la criminalité). Dans de tels cas, les audits et les mécanismes de certification jouent un rôle encore plus important qu'ils ne le font déjà pour les systèmes d'IA en général.¹⁵¹

Dans le contexte des applications d'IA visant à favoriser la participation démocratique, **l'interopérabilité**¹⁵² peut également jouer un rôle majeur en facilitant l'intégration entre les différents

¹⁴⁴ Voir l'article 13 de la Convention des Nations Unies contre la corruption, 2003.

¹⁴⁵ Voir le paragraphe 6 de la Recommandation CM/Rec(2009)1 sur la démocratie électronique

¹⁴⁶ Voir également les Lignes directrices relatives à la participation civile aux décisions politiques. CM(2017)83-final, IV.

¹⁴⁷ Voir le paragraphe 6 de la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e- démocratie) (« faciliter et améliorer l'accès, l'accessibilité [...] à l'aide, si possible, de moyens transparents [...] ») et le paragraphe P.57 de l'annexe à la Recommandation CM/Rec(2009)1. Voir également les paragraphes 2.1.3 et 3.2 de l'Annexe à la Recommandation CM/Rec(2016)5 sur la liberté d'internet.

¹⁴⁸ Voir également le paragraphe B.IV de l'annexe à la Recommandation CM/Rec(2018)4 relative à la participation des citoyens à la vie publique locale.

¹⁴⁹ Voir également le paragraphe 6 de la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie) et le paragraphe P.54 de l'annexe.

¹⁵⁰ Voir également le paragraphe G.58 de l'annexe à la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie).

¹⁵¹ Il convient de souligner que les audits et les mécanismes de certification jouent également un rôle important dans le cas d'une architecture d'IA à code source ouvert, étant donné que cette nature ne suppose pas en soi l'absence de partialité ni de tout autre défaut. Voir également les paragraphes P.55 et G.57 de l'annexe à la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e- démocratie), (« les logiciels de la démocratie électronique devraient être à code source ouvert et pouvoir être inspectés ou, alternativement, être homologués par un organisme indépendant »).

¹⁵² Voir également les paragraphes P.56, G.56, 59 et 60 de l'annexe à la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie).

services/plateformes de démocratie électronique à différents niveaux géographiques. Déjà pertinent pour la démocratie électronique en général,¹⁵³ cet aspect devrait être étendu à la conception de systèmes fondés sur l'IA.

Un autre principe clé de la démocratie électronique, comme dans les secteurs des données et de la santé, est **l'obligation de rendre des comptes**. Contrairement aux principes examinés ci-dessus, l'obligation de rendre des comptes garde ici la même signification. Par conséquent, elle ne semble pas nécessiter une application sectorielle dans le contexte de l'IA, autre que son application générale.

Enfin, étant donné le rôle des médias dans le contexte de la participation démocratique, et conformément à la Recommandation CM/Rec(2016)4 du Comité des Ministres du Conseil de l'Europe,¹⁵⁴ les applications d'IA ne doivent pas compromettre la confidentialité ou la sécurité des communications ni la protection des sources journalistiques et des lanceurs d'alerte.¹⁵⁵

Lorsque l'on aborde les différents aspects liés au développement des solutions d'IA pour la participation démocratique, la première idée qui vient à l'esprit est qu'une approche démocratique est incompatible avec une approche techno-déterministe. Les solutions sont le résultat d'un processus inclusif. Ainsi, les valeurs telles que la protection des minorités, le pluralisme et la diversité devraient absolument être examinées dans le cadre du développement de ces solutions.

D'un point de vue démocratique, la première question que nous devrions poser est la suivante : devons-nous vraiment privilégier une solution fondée sur l'IA plutôt qu'une autre possibilité pour résoudre un problème donné,¹⁵⁶ étant donné l'impact potentiel de l'IA sur les droits et les libertés ? Dans l'affirmative, l'étape suivante consiste à examiner **l'intégration des valeurs** dans le développement de l'IA.¹⁵⁷

Les solutions d'IA proposées doivent être conçues sous l'angle des droits de l'homme. Elles doivent garantir le plein respect des droits de l'homme et des libertés fondamentales, et notamment adopter des **outils et des procédures d'évaluation** à cette fin.¹⁵⁸ Dans le cas des applications d'IA ayant un fort impact sur les droits de l'homme et les libertés, telles que les processus électoraux, il convient **préalablement d'évaluer** la conformité avec la législation. En outre, les systèmes d'IA prévus pour les tâches publiques devraient **pouvoir faire l'objet d'un audit** et, lorsque cela n'est pas exclu par des intérêts premiers concurrents, les audits devraient être publiés.

Un autre aspect important à prendre en considération est le **partenariat public-privé** qui caractérise souvent les services d'IA aux citoyens ; il convient d'évaluer le meilleur choix entre les solutions internes et tierces, y compris les nombreuses combinaisons de ces deux extrêmes. À cet égard, lorsque des solutions d'IA sont totalement ou partiellement développées par des sociétés privées, **la transparence des contrats** et **la clarté des règles sur l'accès aux données des citoyens et leur utilisation** ont une valeur essentielle en termes de contrôle démocratique.

Il est pertinent de restreindre l'accès aux données des citoyens et leur utilisation non seulement du point de vue de la protection des données (principes de la minimisation des données et de la limitation des finalités), mais plus généralement eu égard au volume de données générées par une communauté, parmi lesquelles on trouve également des données à caractère non personnel et des

¹⁵³ Voir également le paragraphe 6 de la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie)

¹⁵⁴ Voir le paragraphe 2 de l'annexe à la Recommandation CM/Rec(2016)4 sur la protection du journalisme et la sécurité des journalistes et autres acteurs des médias, Conseil de l'Europe, Assemblée parlementaire. 2019. Résolution 2254 (2019) La liberté des médias en tant que condition pour des élections démocratiques.

¹⁵⁵ Voir également la Résolution 2300 (2019)1 de l'Assemblée parlementaire, Améliorer la protection des lanceurs d'alerte partout en Europe ; Recommandation CM/Rec(2014)7 sur la protection des lanceurs d'alerte.

¹⁵⁶ Voir également le paragraphe 5.7 de l'annexe à la Recommandation CM/Rec(2020)1.

¹⁵⁷ 90 Voir également la Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques (adoptée par le Comité des Ministres le 13 février 2019 lors de la 1337^e réunion des Délégués des Ministres), paragraphe 7.

¹⁵⁸ Voir la Recommandation CM/Rec(2009)1, paragraphes 5 et 6, et Annexe à la Recommandation CM/Rec(2009)1, paragraphe G.67. Voir également la sous-partie II.2 sur les données et le rôle des comités d'experts et A Mantelero, 'AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment' (2018) 34 Computer Law & Security Review 754.

données agrégées. Dans l'environnement numérique, cette question devrait être considérée comme faisant partie de la démocratie. En outre, la **dimension collective** des ressources numériques générées par une communauté devrait comporter des formes de surveillance et de contrôle par les citoyens, comme c'est le cas pour les autres ressources d'un territoire/d'une communauté (par exemple, l'environnement).

Les considérations exprimées ci-dessus sur le caractère essentiel de l'ouverture dans les outils de participation démocratique devraient être rappelées ici, étant donné leur impact sur la conception des systèmes d'IA. De plus, la conception, le développement et le déploiement de ces systèmes devraient également prendre en compte l'adoption d'une stratégie écologique et durable.¹⁵⁹

Enfin, il convient de noter que si la conception de l'IA est un élément clé de ces systèmes, elle n'est pas neutre. Des valeurs peuvent être intégrées dans les composants technologiques,¹⁶⁰ notamment les systèmes d'IA. Ces valeurs peuvent être choisies de manière intentionnelle ; dans le contexte de la démocratie électronique, ce choix doit reposer sur un processus démocratique. Cependant, certaines valeurs peuvent aussi être intégrées involontairement dans les solutions d'IA, du fait de la composition culturelle, sociale et hommes-femmes des équipes de développeurs d'IA. D'où la valeur ajoutée du **caractère inclusif** ici, à la fois en termes d'inclusion et de diversité¹⁶¹ dans le développement de l'IA.

En ce qui concerne la bonne gouvernance,¹⁶² les principes examinés pour la démocratie électronique peuvent être repris ici.¹⁶³ Tel est le cas avec les villes intelligentes et la gestion environnementale par capteurs, dans lesquels les processus décisionnels ouverts, transparents et inclusifs jouent un rôle majeur. De même, l'utilisation de l'IA pour superviser les activités des collectivités locales,¹⁶⁴ à des fins de contrôle et de lutte contre la corruption,¹⁶⁵ devrait être fondée sur l'**ouverture** (logiciels à code source ouvert), la **transparence** et l'**auditabilité**.

Plus généralement, l'IA peut être utilisée dans les échanges entre les gouvernements et les citoyens afin d'automatiser les requêtes et les demandes d'informations des citoyens.¹⁶⁶ Cependant, dans ce cas, il est important de garantir le droit de savoir que nous interagissons avec une machine¹⁶⁷ et d'avoir un interlocuteur humain. En outre, l'accès aux services publics ne doit pas dépendre de la communication de données inutiles et inappropriées.

Il convient d'accorder une attention particulière à l'utilisation potentielle de l'IA dans l'interaction homme-machine afin de mettre en œuvre des stratégies d'incitation.¹⁶⁸ Ici, en raison de la complexité et de l'opacité des solutions techniques adoptées, l'IA peut accroître le rôle passif des citoyens et avoir des répercussions négatives sur le processus décisionnel démocratique. Il convient, au contraire, de privilégier une approche dynamique fondée sur une participation consciente et active aux objectifs de la communauté, qui sera mieux gérée par les outils de participation basés sur l'IA. Lorsqu'elles sont adoptées, les stratégies d'incitation devraient toutefois suivre une approche fondée sur des données probantes.

Enfin, l'utilisation des systèmes d'IA dans les tâches de gouvernance soulève des questions épineuses sur la relation entre les décideurs humains et le rôle de l'IA dans le processus

¹⁵⁹ Voir également la Recommandation CM/Rec(2009)1, Annexe, paragraphe P. 58.

¹⁶⁰ Voir également P-P Verbeek, 2011, 41-65

¹⁶¹ Voir également la Recommandation CM/Rec(2020)1, Annexe, paragraphe 3.5.

¹⁶² Voir la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie), paragraphe P.4 de l'annexe (« [...] bonne gouvernance, qui se caractérise par un exercice démocratique du pouvoir, efficient, efficace, participatif, transparent et responsable, s'appuyant sur des moyens électroniques, et inclut débat politique informel et intervention d'acteurs non gouvernementaux »).

¹⁶³ Voir également la Recommandation Rec(2004)15 sur la gouvernance électronique (« e-gouvernance ») ; Conseil de l'Europe. 2008. Les 12 Principes de bonne gouvernance ancrés dans la Stratégie sur l'innovation et la bonne gouvernance au niveau local, adoptée par une décision du Comité des Ministres du Conseil de l'Europe en 2008.

¹⁶⁴ Voir également la Recommandation CM/Rec(2019)3 sur le contrôle des actes des collectivités locales, Annexe, Lignes directrices concernant l'amélioration des systèmes de contrôle de l'action des collectivités locales, paragraphes 4 et 9.

¹⁶⁵ Voir aussi Savaget, Chiarini et Evans, 2019, examen de l'affaire brésilienne « Operação Serenata de Amor » (OSA).

¹⁶⁶ Voir Mehr. 2017.

¹⁶⁷ Voir également GAI 2.11.

¹⁶⁸ Voir, ex multis, Sunstein, 2015a ; Sunstein, 2015a ; Sunstein et Thaler, 2003 ; Thaler et Sunstein, 2008.

décisionnel.¹⁶⁹ Ces questions concernent davantage les fonctions qui ont un fort impact sur les droits et les libertés individuels, comme c'est le cas des décisions juridictionnelles. Pour cette raison, les préoccupations relatives à la transparence (notamment l'explicabilité) du raisonnement de l'IA et à la relation entre l'utilisation de l'IA et la liberté des décideurs seront analysées dans la cinquième sous-partie.

Élections

Comme dans d'autres domaines, l'impact de l'IA sur les processus électoraux est vaste et touche la phase préélectorale, les élections et la phase postélectorale de diverses manières. Cependant, il ne suffit pas d'analyser les étapes du processus électoral pour mettre en évidence les différentes interactions avec les solutions d'IA.

L'influence de l'IA est mieux représentée par la distinction suivante : l'IA pour le processus électoral d'une part (vote électronique, prévision des résultats et règlement des litiges électoraux) et l'IA pour les campagnes électorales (micro-ciblage ou profilage, propagande et fausses nouvelles). Si dans le premier domaine l'IA consiste principalement en une amélioration technologique d'un processus existant, dans le domaine des campagnes électorales le profilage et la propagande basés sur l'IA suscitent de nouvelles préoccupations qui ne sont pas totalement prises en compte par le cadre juridique existant. Par ailleurs, plusieurs documents ont mis en avant le rôle actif des États dans la création d'un environnement propice à la liberté d'expression.¹⁷⁰

En ce qui concerne la mise en œuvre technologique de la démocratie électronique (vote électronique, prédiction des résultats et règlement des litiges électoraux), certains des principes clés mentionnés concernant la participation démocratique sont également pertinents ici. **L'accessibilité¹⁷¹, la transparence¹⁷², l'ouverture¹⁷³, la gestion des risques et la responsabilité** (y compris l'adoption de procédures de certification et d'audit)¹⁷⁴ sont des éléments fondamentaux des solutions technologiques adoptées à ces étapes du processus électoral.

En ce qui concerne l'utilisation de l'IA dans les campagnes électorales (micro-ciblage et profilage, propagande et fausses nouvelles), plusieurs questions soulevées concernent le traitement des données à caractère personnel en général. Les principes énoncés dans la Convention 108+ peuvent donc être appliqués et correctement contextualisés.¹⁷⁵

Dans le cas de la propagande et de la désinformation, il convient d'apporter des réponses nouvelles et plus spécifiques.¹⁷⁶ Ici, les instruments contraignants et non contraignants existants ne définissent pas de conditions précises. En effet, cette désinformation repose sur de nouvelles formes de

¹⁶⁹ Voir également Calo et Citron, 2020, à venir.

¹⁷⁰ Voir également la Recommandation CM/Rec(2018)1 sur le pluralisme des médias et la transparence de leur propriété ; la Déclaration conjointe sur la liberté d'expression et les fausses nouvelles (« fake news »), la désinformation et la propagande, le Rapporteur spécial des Nations Unies (UN) sur la liberté d'expression, la Représentante de l'Organisation pour la sécurité et la coopération en Europe (OSCE) pour la liberté des médias, le Rapporteur spécial de l'Organisation des États américains (OEA) sur la liberté d'expression et le Rapporteur spécial de la Commission africaine des droits de l'homme et des peuples (CADHP) sur la liberté d'expression et à l'accès à l'information (3 mars 2017). Voir également la Recommandation CM/Rec(2016)5 sur la liberté d'internet, Annexe, paragraphes 1.5, 2.1 et 3 ; Commission européenne pour la démocratie par le droit (Commission de Venise). 2019. Rapport conjoint de la Commission de Venise et de la Direction de la société de l'information et de la lutte contre la criminalité de la Direction Générale des droits de l'homme et de l'État de droit (DGI) sur les technologies numériques et les élections, paragraphe 151.E ; OSCE, 2020. Voir également Bychawska-Siniarska, 2017

¹⁷¹ Voir la Recommandation CM/Rec(2017)5 sur les normes relatives au vote électronique, Annexe I, Normes relatives au vote électronique, paragraphes 1 et 2.

¹⁷² Voir la Recommandation CM/Rec(2017)5, Annexe I, paragraphes 1 et 2. 32. Voir aussi Conseil de l'Europe. Direction générale de la démocratie et des affaires politiques - Direction des institutions démocratiques. 2011. Lignes directrices sur la transparence des élections par voie électronique.

¹⁷³ Voir Recommandation CM/Rec(2017)5, Annexe I, paragraphe. 35.

¹⁷⁴ Voir Recommandation CM/Rec(2017)5, Annexe I, paragraphes 36, 37, 38, 39 et 40.

¹⁷⁵ Recommandation CM/Rec(2010)13 sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage et son analyse en cours. Voir Conseil de l'Europe, Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel. 2019. Profilage et la Convention 108+ : Pistes pour une actualisation. T-PD(2019)07BISrev.

¹⁷⁶ Voir Manheim et Kaplan, 2019 ; Commission européenne - Direction générale des réseaux de communication, - contenu et des technologies, 'A Multi-Dimensional Approach to Disinformation Report of the Independent High Level Group on Fake News and Online Disinformation' (2018). Voir également l'affaire Stoll c. Suisse [GC], Requête n° 69698/01, § 104.

communication, telles que les réseaux sociaux, qui diffèrent des médias traditionnels¹⁷⁷ et contournent souvent la médiation professionnelle des journalistes.

Cependant, les principes généraux, tels que le **principe de non-ingérence** par les pouvoirs publics dans les activités médiatiques visant à influencer les élections,¹⁷⁸ peuvent être étendus à ces nouvelles formes de propagande et de désinformation. L'IA étant utilisée pour automatiser la propagande, sa future réglementation devrait élargir le champ d'application des principes généraux de la non-ingérence aux systèmes d'IA utilisés pour diffuser des informations fausses, trompeuses et préjudiciables. En outre, afin d'éviter une telle ingérence, les États¹⁷⁹ et les fournisseurs de médias sociaux devraient adopter une **approche dès la conception** pour accroître leur résilience à la désinformation et à la propagande.

De même, en vertu de l'obligation de couvrir les campagnes électorales de manière **équitable, équilibrée et impartiale**¹⁸⁰, les médias et les opérateurs de médias sociaux devraient assumer les obligations relatives à la transparence de la logique des algorithmes utilisés pour la sélection des contenus,¹⁸¹ garantissant ainsi le pluralisme et la diversité des voix,¹⁸² notamment les voix critiques.¹⁸³

De plus, les États et les intermédiaires devraient promouvoir et faciliter l'accès aux outils permettant de détecter la désinformation et les agents non humains, et soutenir les recherches indépendantes sur l'impact de la désinformation et les projets proposant aux utilisateurs des services de vérification des informations (« fast-checking »).¹⁸⁴

Étant donné le rôle majeur de la publicité dans la désinformation et la propagande, les critères utilisés par les solutions basées sur l'IA pour la publicité politique devraient être **transparents**,¹⁸⁵ **pouvoir faire l'objet d'audits** et offrir des **conditions équitables** à tous les partis et candidats politiques.¹⁸⁶ En outre, les intermédiaires devraient revoir leurs modèles publicitaires pour s'assurer qu'ils ne nuisent pas à la **diversité des opinions et des idées**.¹⁸⁷

v. Justice

Comme indiqué dans la sous-partie précédente, le domaine de la justice est vaste. Il serait trop ambitieux d'envisager l'analyse de l'ensemble des conséquences de l'IA sur la justice et de leurs effets sur la démocratie. Conformément à la portée de cette étude, cette sous-partie décrit les principaux enjeux liés à l'utilisation de l'IA et les principes qui, sur la base des instruments internationaux juridiquement contraignants, peuvent contribuer à sa réglementation future.

¹⁷⁷ Voir également la Recommandation CM/Rec(2011)7 sur une nouvelle conception des médias.

¹⁷⁸ Voir la Recommandation CM/Rec(2007)15 sur des mesures concernant la couverture des campagnes électorales par les médias, paragraphe I.1.

¹⁷⁹ Voir également la Déclaration conjointe sur la liberté d'expression et les fausses nouvelles (« fake news »), la désinformation et la propagande, paragraphe 2.c.

¹⁸⁰ Voir également le paragraphe II.1 de la Recommandation CM/Rec(2007)15.

¹⁸¹ Voir également la Déclaration conjointe sur la liberté d'expression et les fausses nouvelles (« fake news »), la désinformation et la propagande ; la Recommandation CM/Rec(2018)2 sur les rôles et les responsabilités des intermédiaires d'internet, Annexe, paragraphes 2.1.3 et 2.3.5 (« Du fait de la capacité actuellement limitée des moyens automatisés à évaluer le contexte, les intermédiaires devraient mesurer minutieusement les incidences qu'une gestion automatisée des contenus peut avoir sur le plan des droits de l'homme et procéder à un contrôle humain si nécessaire. Ils devraient tenir compte du risque de blocages insuffisants ou au contraire excessifs dus à des systèmes d'algorithmes inexacts, et des effets potentiels de ces algorithmes sur les services fournis sur le débat public »).

¹⁸² Voir également le Code de bonnes pratiques contre la désinformation de l'UE, 2018

¹⁸³ Voir aussi le paragraphe 15 de l'annexe à la Recommandation CM/Rec(2016)4.

¹⁸⁴ Voir également la Déclaration conjointe sur les fausses nouvelles (« fake news »), la désinformation et la propagande, paragraphe 4.E ; Commission européenne pour la démocratie par le droit. 2019, paragraphe 151.D.

¹⁸⁵ Voir également Conseil de l'Europe. Assemblée parlementaire. Résolution 2254 (2019) La liberté des médias en tant que condition pour des élections démocratiques, paragraphes 9.2 et 11.1 ; Commission européenne pour la démocratie par le droit (Commission de Venise). 2019. Rapport sur les technologies numériques et les élections, établi conjointement par la Commission de Venise et la Direction de la société de l'information et de la lutte contre la criminalité (Direction générale Droits de l'homme et État de droit, DGI), paragraphes 151.A et 151.B.

¹⁸⁶ Voir également la Recommandation CM/Rec(2007)15, paragraphe II.5.

¹⁸⁷ Voir également la Déclaration conjointe sur les fausses nouvelles (« fake news »), la désinformation et la propagande, paragraphe 4.e.

En l'absence d'instruments contraignants spécifiques et dédiés, tels que la Convention 108+ et la Convention d'Oviedo, la justice se distingue des domaines de la protection des données et de la santé. Par conséquent, cette analyse est davantage centrée sur la contextualisation des principes directeurs généraux que sur des instruments juridiques précis.

Cet exercice est facilité par la Charte éthique européenne d'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement, adoptée par la CEPEJ en 2019, qui traite directement de la relation entre justice et IA. Bien que cet instrument non contraignant soit considéré comme une charte éthique, dans une large mesure il concerne les principes juridiques ancrés dans les instruments internationaux.

Les principes directeurs du développement de l'IA dans le domaine de la justice peuvent être tirés des instruments contraignants suivants : la Déclaration universelle des droits de l'homme, le Pacte international relatif aux droits civils et politiques, la Convention pour la protection des droits de l'homme et des libertés fondamentales, la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes et la Convention pour la protection des droits de l'homme et des libertés fondamentales.¹⁸⁸

Compte tenu de la diversité des types et des objectifs des opérations dans ce domaine et de la variété des professionnels et des procédures, cette sous-partie établit une distinction fonctionnelle entre deux thèmes : i) les décisions judiciaires et les modes alternatifs de règlement des conflits et ii) la prévention/prédiction de la criminalité. Avant d'analyser et de contextualiser les principes clés relatifs à ces deux domaines, il convient de formuler quelques observations générales, qui peuvent également s'appliquer à l'action de l'administration publique dans son ensemble.¹⁸⁹

Premièrement, il convient de noter que – en comparaison avec les décisions humaines, et plus spécifiquement les décisions judiciaires – la logique qui sous-tend les systèmes d'IA ne ressemble pas à un raisonnement juridique. Ces systèmes exécutent simplement des codes basés sur une approche mathématique/statistique centrée sur les données.

Deuxièmement, les taux d'erreur de l'IA sont proches voire inférieurs à ceux du cerveau humain dans des domaines tels que l'étiquetage d'images, mais les processus décisionnels plus complexes affichent des taux d'erreur supérieurs. C'est le cas du raisonnement juridique en matière de résolution de problèmes.¹⁹⁰ Cela étant, si la classification erronée de l'image d'un chat peut avoir des effets négatifs limités, un taux d'erreur dans les décisions juridiques aura une incidence importante sur les droits et les libertés individuels.

Il convient de signaler que la différence entre les erreurs de décisions humaines et automatisées a une conséquence importante en termes d'échelle : si l'erreur humaine ne se répercute que sur des cas individuels, la mauvaise conception et la partialité des solutions d'IA produisent inévitablement des effets sur toutes les personnes qui se trouvent dans des circonstances identiques ou similaires puisque les outils d'IA sont appliqués à toute une série de cas. Cela peut créer une discrimination de groupe portant atteinte à des personnes appartenant à différentes catégories.

¹⁸⁸ Voir également, pour ce qui est de la zone UE, la Charte des droits fondamentaux de l'Union européenne.

¹⁸⁹ Voir ci-dessus sous-partie II.4

¹⁹⁰ Voir Dupont et al., 2018, 148 (« L'apprentissage profond n'a aucun moyen naturel de gérer la structure hiérarchique, ce qui signifie que toutes les variables disponibles sont considérées au même niveau, comme « plates » ou non hiérarchiques. Cela constitue un obstacle majeur lorsque les décisions ont un poids moral ou juridique important qui doit l'emporter sur d'autres éléments »). Voir également Osoba et Welsler, 2017, 18 (« Si l'on examine le problème sous un autre angle, on constate que les jugements dans l'espace du comportement social sont souvent flous et non des critères binaires bien définis [...]. Nous sommes capables d'apprendre à nous orienter dans des relations complexes et floues, comme celles avec les gouvernements et les lois, en nous fiant souvent à des évaluations subjectives pour ce faire. Les systèmes qui reposent sur un raisonnement quantifié (comme la plupart des agents artificiels) peuvent reproduire l'effet mais exigent souvent une conception soignée. Pour saisir cette nuance, il faudra peut-être plus que des informaticiens et des spécialistes des données. »). Voir également Cummings et al., 2018, 13.

Au vu de la nature textuelle des documents juridiques, le traitement du langage naturel joue un rôle important dans les applications d'IA de la sphère juridique. Cela pose plusieurs questions cruciales autour des solutions commerciales développées spécifiquement pour le marché anglophone, moins efficaces dans un environnement juridique qui utilise une autre langue que l'anglais.¹⁹¹ De plus, les décisions juridiques se caractérisent souvent par un raisonnement implicite non exprimé, qui peut être perçu par des systèmes experts, mais pas par des outils d'apprentissage automatique des langues. Enfin, la présence de clauses générales exige de bien comprendre l'interprétation juridique pertinente et les mises à jour continues qui ne peuvent être tirées de l'analyse de textes.

Toutes ces contraintes suggèrent d'adopter l'IA de manière prudente et plus critique dans le domaine de la justice par rapport aux autres domaines. Par ailleurs, en ce qui concerne les décisions de justice et les modes alternatifs de règlement des litiges, il convient de faire la distinction entre les affaires caractérisées par des évaluations factuelles régulières et les affaires caractérisées par une importante marge de raisonnement juridique et de discrétion.¹⁹²

Décisions de justice et modes alternatifs de règlement des litiges

Plusieurs produits d'IA dits technico-juridiques n'ont pas d'impact direct sur les processus décisionnels au sein des tribunaux ou sur les modes alternatifs de règlement des litiges. Ils facilitent plutôt la gestion des connaissances et des contenus, la gestion organisationnelle et l'évaluation des performances.¹⁹³ Parmi ces applications figurent, par exemple, des outils pour la catégorisation des contrats, la détection des clauses contractuelles divergentes ou incompatibles, la preuve électronique, l'aide à la rédaction, l'extraction des dispositions juridiques et l'examen de conformité assisté. En outre, certaines applications peuvent offrir des fonctions basiques de résolution des problèmes fondées sur des questions et des situations types (par exemple des chatbots juridiques).

Bien que l'IA ait, dans de tels cas, un impact sur la pratique et les connaissances juridiques qui soulève plusieurs questions éthiques,¹⁹⁴ les éventuelles conséquences négatives sur les droits de l'homme, la démocratie et l'État de droit sont limitées. Elles sont principalement liées à des défauts et des lacunes de ces systèmes.

Dans le cas de la gestion de contenus et de connaissances, y compris les recherches et l'analyse de documents, ces défauts peuvent engendrer des représentations incomplètes ou inexacts des faits ou des situations. Néanmoins, cela concerne les méta-produits, les résultats d'un outil de recherche qui doivent être interprétés et dûment justifiés en cas d'utilisation au sein d'un tribunal. Les règles de la responsabilité, dans le cas de la responsabilité du fait des produits par exemple, peuvent résoudre ces problèmes.

Par ailleurs, il est possible de neutraliser la partialité (mauvaise sélection des affaires, mauvaise classification, etc.) des outils de recherche assistée par ordinateur fondés sur des textes standards utilisés pour l'analyse de la législation, de la jurisprudence et des documents,¹⁹⁵ grâce à l'**éducation** et à la **formation** adéquates des professionnels du droit. La **transparence** des systèmes d'IA (à savoir la description de leur logique, de leurs biais potentiels et de leurs limites) peut également réduire les conséquences négatives.

L'utilisation de l'IA par les tribunaux, pour les recherches juridiques et l'analyse de documents, devrait également être définie par la transparence. Les juges doivent être transparents en ce qui concerne les décisions qui dépendent de l'IA et la façon dont les résultats fournis par l'IA sont utilisés

¹⁹¹ Voir Conseil des barreaux européens, 2020, 29.

¹⁹² Voir la sous-partie suivante sur la distinction entre justice codifiée et justice équitable.

¹⁹³ Voir Commission européenne pour l'efficacité de la justice (CEPEJ). 2018. Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, Annexe II.

¹⁹⁴ Voir également Nunez, 2017

¹⁹⁵ Voir la notion de justice électronique dans la Recommandation CM/Rec(2009)1 sur la démocratie électronique (e-démocratie), Annexe, paragraphe 38.

afin de contribuer aux motifs, conformément aux **principes de procès équitable et d'égalité des armes**.¹⁹⁶

Enfin, la transparence peut jouer un rôle important eu égard aux chatbots juridiques fondés sur l'IA, en faisant connaître aux utilisateurs leur logique et les ressources utilisées (la liste des affaires analysées, par exemple). Une transparence totale devrait également inclure les sources utilisées pour entraîner ces algorithmes et accéder à la base de données utilisée pour fournir les réponses. Lorsque ces bases de données sont privées, des **audits** de tierces parties devraient être disponibles afin d'évaluer la qualité des ensembles de données et la façon dont les biais potentiels ont été traités, notamment le risque de sous- ou de sur-représentation de certaines catégories (**non-discrimination**).

D'autres problèmes majeurs touchent les applications d'IA conçues pour automatiser les modes alternatifs de règlement des litiges ou pour soutenir les décisions de justice. Ici, la distinction entre justice codifiée et justice équitable¹⁹⁷ suggère que l'IA devrait être limitée, à des fins de prise de décisions, aux affaires marquées par des évaluations factuelles régulières. D'où l'importance de mener des recherches supplémentaires sur la classification des différents types de processus décisionnels afin d'identifier les applications de raisonnement juridique systématiques qui peuvent être demandées à l'IA, en gardant dans tous les cas la surveillance humaine qui garantit également la créativité juridique des décideurs.¹⁹⁸

En ce qui concerne la justice équitable, comme l'indiquent les ouvrages,¹⁹⁹ sa logique est plus compliquée que la simple conclusion d'affaires individuelles. Les valeurs et les considérations exprimées et non exprimées, à la fois juridiques et non juridiques, caractérisent le raisonnement des tribunaux et ne peuvent être reproduites par la logique de l'IA. Les systèmes d'apprentissage automatique ne sont pas capables de produire un raisonnement juridique. Ils tirent des conclusions en identifiant des modèles dans les ensembles de données juridiques, ce qui n'est pas la même chose que d'élaborer un raisonnement juridique.

Étant donné le vaste contexte du rôle social des tribunaux, la jurisprudence est un système évolutif, ouvert aux nouveaux enjeux sociétaux et politiques. Les outils d'IA dépendants du parcours suivi pourraient donc faire obstacle à ce processus évolutif : la nature déductive et dépendante du parcours suivi de certaines solutions d'apprentissage automatique fondées sur l'IA peut remettre en question le rôle important des décideurs humains dans l'évolution du droit dans la pratique et le raisonnement juridique.

De plus, au niveau individuel, la dépendance au parcours suivi peut générer le risque d'« analyses déterministes »,²⁰⁰ provoquant la résurgence de doctrines déterministes au détriment des doctrines favorisant l'individualisation de la sanction et portant atteinte au principe de réhabilitation et d'individualisation de la peine.

Par ailleurs, dans plusieurs cas, notamment les modes alternatifs de règlement des litiges, la médiation entre les exigences des parties mais aussi l'analyse de la composante psychologique des actions humaines (faute, intention) nécessitent une intelligence émotionnelle que les systèmes d'IA n'ont pas.

¹⁹⁶ Voir également Commission européenne pour l'efficacité de la justice (CEPEJ). 2018. Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, Annexe II.

¹⁹⁷ Voir Re et Solow-Niederman, 2019, 252-254 (« La justice équitable englobe à la fois la réflexion sur les valeurs mises en place par le système juridique et l'application raisonnée de ces valeurs, dans le contexte [...] La justice codifiée désigne l'application routinière de procédures normalisées à un ensemble de faits [...] En résumé, la justice codifiée voit les vices de la discrétion, tandis que la justice équitable voit ses vertus »).

¹⁹⁸ Voir également Clay, 2019. À cet égard, par exemple, un système juridique qui prévoit des indemnités en cas de blessures corporelles, sur la base des dommages patrimoniaux effectifs, pourrait être automatisé. Il ne sera, toutefois, pas capable de réévaluer le fondement du raisonnement juridique et d'étendre les indemnités aux dommages non personnels et existentiels

¹⁹⁹ Voir Re et Solow-Niederman, 2019

²⁰⁰ Voir Commission européenne pour l'efficacité de la justice (CEPEJ). 2018. Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement.

Le cadre juridique existant, fourni par les instruments juridiques internationaux, illustre ces préoccupations. La Déclaration universelle des droits de l'homme (articles 7 et 10), le PIDCP (article 14), la Convention pour la protection des droits de l'homme et des libertés fondamentales (article 6) et la Charte des droits fondamentaux de l'Union européenne (article 47) insistent sur les principales exigences liées à l'exercice du pouvoir judiciaire : le traitement équitable devant la loi, l'impartialité, l'indépendance et la compétence. Les outils d'IA ne possèdent pas ces qualités, ce qui limite leur contribution au processus décisionnel tel qu'il est suivi par les tribunaux.

Comme indiqué par la Commission européenne pour l'efficacité de la justice, « la neutralité des algorithmes est un mythe, leur créateur transférant, de manière consciente ou non, leurs propres systèmes de valeurs. » De nombreux cas de partialité relatifs à des applications d'IA confirment que ces systèmes donnent trop souvent - bien que cela ne soit pas conscient dans bien des cas - une représentation partielle de la société et des situations personnelles, ce qui n'est pas compatible avec les principes du **traitement équitable devant la loi** et de la **non-discrimination**.²⁰¹ La **qualité des données** et d'autres formes d'**évaluation** de la qualité (évaluation de l'impact, audits, etc.) peuvent réduire ce risque.²⁰² Cependant, compte tenu de l'importance des intérêts potentiellement concernés en cas de décisions biaisées, les risques restent élevés dans le cas d'une justice équitable et semblent disproportionnés par rapport aux avantages, principalement en termes d'efficacité du système judiciaire.²⁰³

Les **principes du procès équitable et de l'égalité des armes**²⁰⁴ suscitent d'autres préoccupations lorsque les décisions de justice reposent sur les résultats d'algorithmes propriétaires dont les données et la structure de formation ne sont pas accessibles au public.²⁰⁵ Une notion de **transparence** plus large pourrait y remédier en ce qui concerne l'utilisation de l'IA dans les décisions judiciaires, mais la transparence de l'IA - un objectif ambitieux en soi - ne peut répondre aux autres objections structurelles et fonctionnelles précitées.

Par ailleurs, les spécialistes des données peuvent façonner les outils d'IA de différentes façons au cours des phases de conception et de formation, de sorte que si les outils d'IA devaient nécessairement faire partie du processus décisionnel, les gouvernements qui sélectionnent les outils à utiliser par les tribunaux risqueraient d'interférer indirectement avec l'**indépendance** des juges.²⁰⁶

Ce risque subsiste même si le juge reste libre d'ignorer les décisions d'IA en invoquant un motif spécifique. Bien que le **contrôle humain** constitue un élément important,²⁰⁷ son impact effectif peut être diminué par la propension psychologique ou utilitaire (économique) des décideurs humains à tirer partie de la solution proposée par l'IA.²⁰⁸

Prévention de la criminalité

La complexité de la détection et de la prévention de la criminalité a stimulé la recherche sur les applications d'IA afin de faciliter les activités humaines. Ces dernières années, plusieurs solutions²⁰⁹ et un nombre croissant d'ouvrages ont été élaborés dans le domaine de la police prédictive, une approche proactive de la prévention de la criminalité fondée sur les données. Les solutions

²⁰¹ Voir également CEPEJ, 2018.

²⁰² Voir également CEPEJ, 2018.

²⁰³ Voir également la Recommandation CM/Rec(2020)1, Annexe, paragraphe 11.

²⁰⁴ Voir également CEPEJ, 2018, Annexe I, paragraphe 138

²⁰⁵ Voir également CEPEJ, Annexe I, paragraphe 131 (« l'opacité des processus de fonctionnement des algorithmes par les entreprises privées (qui revendiquent leur propriété intellectuelle) a été une autre source d'inquiétude. Si l'on tient compte du fait qu'elles tiennent leurs données sources des autorités étatiques elles-mêmes, leur absence d'esprit de responsabilité vis-à-vis des citoyens pose un problème démocratique majeur [...] par exemple, lorsque ProPublica a révélé les failles de l'algorithme COMPAS après le refus de la société propriétaire de le partager »).

²⁰⁶ Voir également CEPEJ, 2018.

²⁰⁷ Voir également CEPEJ, 2018

²⁰⁸ Voir également Mantelero, 2019 (« En effet, la nature soi-disant fiable des solutions d'IA fondées sur les mathématiques peut pousser les personnes qui s'appuient sur des algorithmes pour prendre leurs décisions à faire confiance au tableau des individus et de la société suggéré par les procédés analytiques »).

²⁰⁹ Voir Završnik, 2019 ; Agence des droits fondamentaux de l'Union européenne, 2018, 98-100; Osoba et Welser, 2017.

disponibles poursuivent essentiellement deux objectifs distincts : prédire où et quand des crimes pourraient être commis ou prédire qui pourrait commettre un crime.²¹⁰

Ces deux objectifs n'ont pas le même impact potentiel sur les droits de l'homme et la liberté : cet impact est plus marqué lorsque l'IA est utilisée pour prédire les auteurs de crimes. Cependant, dans les deux cas, nous pouvons reprendre ici les considérations concernant les enjeux généraux de l'IA (opacité, droits de propriété intellectuelle, collecte de données à grande échelle²¹¹, etc.) examinés dans les sous-parties précédentes et partiellement traités par la transparence, la **qualité des données**, la **protection des données**, les **audits** et les autres mesures. Il convient de noter que le rôle de la **transparence**²¹² dans le contexte judiciaire pourrait être limité afin de ne pas entraver l'effet dissuasif de ces outils. La transparence totale pourrait donc être remplacée par des audits et un contrôle assuré par des organismes indépendants.

Abstraction faite des aspects organisationnels relatifs aux limites de l'autodétermination des policiers dans le cadre de leurs fonctions, les principales questions concernant l'utilisation de l'IA pour prévoir la criminalité sur une base géographique et temporelle concernent l'impact de ces outils sur le **droit à la non-discrimination**.²¹³ Les préjugés qui se confirment d'eux-mêmes, les préjugés communautaires²¹⁴ et les préjugés historiques²¹⁵ peuvent créer des formes de stigmatisation à l'égard de certains groupes et des lieux où ils résident habituellement.

Lorsque l'analyse des données est utilisée pour classer les crimes et déduire des preuves sur des réseaux criminels, les solutions propriétaires posent problème en termes de respect des **principes de procès équitable et d'égalité des armes** dans le cadre de la collecte et de l'utilisation des preuves. De plus, si les opérations quotidiennes des forces de police sont guidées par des logiciels prédictifs, cela pose le problème de la **responsabilité** des stratégies adoptées, étant donné qu'elles sont en partie déterminées par les logiciels, et donc par les sociétés qui les développent, et non par la police.

Les outils de police prédictive, qui utilisent le profilage pour soutenir les prévisions concernant les auteurs de crimes, sont encore plus en contradiction avec les droits de l'homme. Indépendamment de la question du traitement des données et du profilage,²¹⁶ ces solutions peuvent également nuire au principe de la **présomption d'innocence**,²¹⁷ de **l'équité** procédurale et au droit à la **non-discrimination**.²¹⁸

Bien que les questions de non-discrimination puissent être partiellement traitées, il semble plus difficile de résoudre les autres conflits. Du point de vue des droits de l'homme et en termes de proportionnalité (notamment le droit au respect de la vie privée et familiale²¹⁹), le risque de porter atteinte à ces principes semble élevé et n'est pas suffisamment compensé par la preuve que ces solutions apporteront des bénéfices en matière de droits et de libertés individuels et collectifs.²²⁰ À la lumière de la future réglementation de l'IA, cela devrait inciter à examiner attentivement ces

²¹⁰ Pour une taxonomie des méthodes prédictives, voir Perry et al., 2013, qui recense les quatre catégories suivantes : les méthodes de prévision des crimes (axées sur les lieux et les horaires des crimes), la méthode de prévision des auteurs de crimes (axée sur les personnes), les méthodes de prévision des identités des auteurs de crimes (axées sur les personnes), et les méthodes de prévision des victimes de crimes (axées sur les groupes et, dans certains cas, les personnes).

²¹¹ Voir également la Recommandation Rec(2001)1 sur le Code européen d'éthique de la police, Annexe, paragraphe 42

²¹² Voir également Barrett, 2017, 361-62.

²¹³ Voir Agence des droits fondamentaux de l'Union européenne, 2018, 10.

²¹⁴ Voir également Barrett, 2017, 358-59 (« Pour certains, l'objectif de la sécurité collective mérite un sacrifice unilatéral d'un certain degré des droits individuels dans ce contexte particulier. Cependant, ce calcul peut changer si le sacrifice n'est pas collectif mais limité à des groupes minoritaires, ou devient fondamentalement arbitraire en vertu d'un degré d'erreur inacceptable. »)

²¹⁵ Voir Bennett Moses et Chan, 2018.

²¹⁶ Voir ci-dessus sous-partie II.2

²¹⁷ Voir également la Recommandation Rec(2001)1 sur le Code européen d'éthique de la police, Annexe, paragraphe 47.

²¹⁸ Voir également la Recommandation Rec(2001)1 sur le Code européen d'éthique de la police, Annexe, paragraphe 49.

²¹⁹ Voir van Brakel et De Hert, 2011, 183. Voir également Szabó et Vissy c Hongrie [2016] Cour européenne des droits de l'homme quatrième section. Requête n° 37138/14

²²⁰ Voir Meijer et Wessels, 2019.

questions, en établissant une distinction entre les possibilités techniques des solutions d'IA et leurs avantages concrets pour la protection et la consolidation des droits de l'homme et des libertés.

Enfin, dans une perspective plus large et plus globale des droits de l'homme, le fait que les outils d'IA basés sur les données mettent l'accent sur la criminalité conduit à une approche factuelle à court terme. Celle-ci sous-estime les problèmes sociaux, qui sont souvent liés à la criminalité et nécessitent des stratégies sociales à long terme impliquant le renforcement effectif des droits et des libertés individuels et sociaux.²²¹

IV. Harmonisation des principes identifiés

Les sous-parties précédentes ont identifié **plusieurs principes directeurs de la future réglementation de l'IA**. Ces principes ont été **contextualisés eu égard aux enjeux associés à l'IA** dans les différents domaines examinés, mais il est intéressant d'étudier le degré d'harmonisation qui existe entre ces principes.

Les conclusions de la présente étude indiquent que dans un certain nombre de cas **il existe des principes communs** (la primauté de l'être humain, l'autodétermination, la non-discrimination, le contrôle humain). Cela s'explique par plusieurs facteurs.

Premièrement, **certains principes sont sectoriels**. C'est le cas, par exemple, de l'indépendance des juges ou des principes du procès équitable et de l'égalité des armes, qui concernent uniquement la justice.²²²

Deuxièmement, certains principes directeurs sont les mêmes dans différents domaines, mais avec **différentes nuances** selon le contexte. C'est le cas de la transparence, souvent considérée comme cruciale dans la réglementation de l'IA, mais qui revêt diverses significations en fonction des cadres réglementaires.

Dans les domaines de la santé et des données à caractère personnel, la transparence a trait aux informations communiquées aux personnes sur le traitement de ces données ; elle met particulièrement l'accent sur les processus et les risques associés et revêt une forte connotation d'autodétermination. Cependant, la transparence est également pertinente en matière de protection des données afin de contrôler l'exercice du pouvoir sur les données aux mains des entités publiques et privées. Cette autre facette de la transparence est examinée eu égard aux applications d'IA pour la participation démocratique et la bonne gouvernance. Encore une fois, dans le contexte de la justice, la transparence a une importance plus complexe. En effet, elle est essentielle pour protéger les droits et les libertés fondamentaux (par exemple lorsque l'IA est utilisée dans les tribunaux), mais elle exige aussi des limites afin d'éviter de nuire à des intérêts concurrents (par exemple, dans le cas de la détection et de la prévention de la criminalité en matière de police prédictive).

Nous pouvons donc conclure que la transparence est une valeur fondamentale. Néanmoins, nous devons aller au-delà d'une simple exigence de transparence en tant que principe clé de la réglementation de l'IA. Comme pour les autres principes essentiels (la participation, l'inclusion, le contrôle démocratique et l'ouverture), il convient d'adopter des dispositions qui tiennent compte des différents contextes dans lesquels elles sont appliquées et donc de procéder à une contextualisation en bonne et due forme.

Troisièmement, certains principes sont différents mais appartiennent au **même domaine conceptuel** ; ils prennent différentes nuances selon le contexte. C'est le cas de l'obligation de rendre compte et des principes directeurs en matière de gestion des risques en général. Ici, le niveau de détail et les exigences correspondantes peuvent être plus ou moins élaborés. Par exemple, dans le

²²¹ Voir également Rosenbaum, 2006, 245–266.

²²² Voir également les principes de l'accès équitable et de la bienfaisance dans le secteur de la santé, ou les principes de la non-ingérence par les pouvoirs publics dans les activités médiatiques visant à influencer des élections et de l'obligation d'offrir des conditions de publicité équitables à tous les partis et candidats politiques.

domaine de la protection des données, plusieurs dispositions mettent en œuvre ces principes avec un degré de détail important, tandis que dans le cas de la démocratie et de la justice, ces principes sont moins développés eu égard aux applications à usage intensif de données telles que l'IA.

Enfin, la stratégie de réglementation de l'IA comprend certains éléments qui ne sont pas des principes mais des **approches et des solutions opérationnelles** ; celles-ci sont communes aux différents domaines mais nécessitent un développement fondé sur le contexte. C'est le cas du rôle important joué par l'éducation et la formation, l'interopérabilité et les comités d'experts.

Ces considérations suggèrent que seule une harmonisation partielle est possible. L'approche réglementaire de l'IA devrait donc reposer sur un **instrument juridiquement contraignant qui prévoit à la fois des dispositions générales** - axées sur des principes communs et des solutions opérationnelles - et **des dispositions sectorielles plus spécifiques**, couvrant les principes uniquement pertinents dans un domaine donné ou les cas où le même principe est contextualisé différemment selon les domaines.

V. Conclusions

Cette analyse a confirmé la validité de l'approche méthodologique adoptée, axée sur la **contextualisation** des principes directeurs tirés des instruments internationaux juridiquement contraignants ou non. Dans le même temps, elle a également mis en lumière la difficulté à classer les dispositions d'un vaste ensemble d'instruments, qui diffèrent non seulement du point de vue de leur caractère contraignant mais également en termes d'objectifs, d'approche et de structure.

Les résultats ont également confirmé que le cadre existant fondé sur les droits de l'homme, la démocratie et l'État de droit pouvait fournir un contexte commun et approprié pour l'élaboration d'un **instrument contraignant plus spécifique visant à réglementer l'IA conformément aux principes et aux valeurs ancrés dans les instruments juridiques internationaux, capable de répondre de manière plus efficace aux questions soulevées par l'IA.**

Ce cadre international nous amène nécessairement à réaffirmer le rôle central de la dignité humaine dans le contexte de l'IA, où les solutions automatisées ne sauraient être autorisées à déshumaniser les personnes. En outre, il pourrait être prévu de limiter spécifiquement les technologies IA lorsqu'elles sont conçues ou utilisées d'une manière qui ne respecterait pas la dignité humaine,²²³ les droits de l'homme, la démocratie et l'État de droit.

En vue d'une future réglementation de l'IA, ce résultat méthodologique et concret positif n'exclut pas l'existence de certaines lacunes. Elles concernent principalement de vastes domaines, tels que la démocratie et la justice, où il existe différentes possibilités et interprétations selon la vision politique et sociétale de la future relation entre les humains et les machines.

Des recherches supplémentaires dans le domaine des droits de l'homme et de l'IA, ainsi que le débat continu à l'échelle internationale et régionale, contribueront à combler ces lacunes. Toutefois, étant donné la nature évolutive de l'IA, il est souhaitable d'adopter une **approche de corégulation.**

Un instrument contraignant établissant le cadre juridique de l'IA, notamment les principes communs généraux mais aussi les dispositions précises répondant à des questions spécifiques, pourrait donc compléter les règles détaillées énoncées dans des **instruments sectoriels non contraignants supplémentaires.** Ce modèle offrirait à la fois un cadre réglementaire clair et la souplesse nécessaire pour prendre en compte le développement technologique.

²²³ Voir également UNESCO. 1997. Déclaration sur le génome humain et les droits de l'homme, article 11

Bibliographie

- Andorno, R. 2005. The Oviedo Convention: A European Legal Framework at the Intersection of Human Rights and Health Law. *Journal of International Biotechnology Law*, January 2005. <https://doi.org/10.1515/jibl.2005.2.4.133>, consulté le 20.02.2020.
- Azencott, C.-A. 2018. Machine Learning and Genomics: Precision Medicine versus Patient Privacy. *Phil. Trans. R. Soc. A* 376, n° 2128 (13 septembre 2018) : 20170350, <https://doi.org/10.1098/rsta.2017.0350> consulté le 14.01.2020
- Barrett, L. 2017. *Reasonably Suspicious Algorithms: Predictive Policing at the United States Border*. 41 *N.Y.U. Rev. Law & Social Change* 327.
- Bennett Moses, L., Chan, J. 2018. *Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability*. 28 *Policing and Society* 806.
- Bychawska-Siniarska, D. 2017. *Protecting the Right to Freedom of Expression under the European Convention on Human Rights*. Conseil de l'Europe.
- Cabitza, F., Rasoini, R. et Gensini, G.F. 2017. Unintended Consequences of Machine Learning in Medicine. *JAMA* 318, no. 6 (8 août 2017): 517 <https://doi.org/10.1001/jama.2017.7797>. consulté le 18.12.2019.
- Calo, R., Citron, D.K. 2020. *The Automated Administrative State: A Crisis of Legitimacy*. *Emory Law Journal*, à venir. <https://ssrn.com/abstract=3553590>, consulté le 20.04.2020.
- Caruana, R. et autres 2015. Intelligible models for healthcare : predicting pneumonia risk and hospital 30-day readmission, dans : *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Cham, Switzerland : Springer International Publishing AG)* 721-1730.
- Clay, T. (ed). 2019. *L'arbitrage en ligne. Rapport du Club des Juristes*. Paris. 58 <https://www.leclubdesjuristes.com/les-commissions/larbitrage-en-ligne/>, consulté le 30.05.2020.
- Comité des Ministres. 1999. Recommandation n° R (99) 5 sur la protection de la vie privée sur Internet. Adoptée par le Comité des Ministres le 23 février 1999 lors de la 660^e réunion des Délégués des Ministres.
- Comité des Ministres. 2001. Recommandation Rec(2001)10 du Comité des Ministres aux États membres sur le Code européen d'éthique de la police. Adoptée par le Comité des Ministres le 19 septembre 2001 lors de la 765^e réunion des Délégués des Ministres.
- Comité des Ministres. 2003. Recommandation CM/Rec(2003)4 sur les règles communes contre la corruption dans le financement des partis politiques et des campagnes électorales. Adoptée par le Comité des Ministres le 8 avril 2003 lors de la 835^e réunion des Délégués des Ministres.
- Comité des Ministres. 2004. Recommandation CM/Rec(2004)15 sur la gouvernance électronique (« e-gouvernance »). Adoptée par le Comité des Ministres le 15 décembre 2004, lors de la 909^e réunion des Délégués des Ministres.
- Comité des Ministres. 2007. Recommandation CM/Rec(2007)15 du Comité des Ministres aux États membres relative à des mesures concernant la couverture des campagnes électorales par les médias. Adoptée par le Comité des Ministres le 7 novembre 2010 lors de la 1010^e réunion des Délégués des Ministres.
- Comité des Ministres. 2009. Recommandation CM/Rec(2009)1 du Comité des Ministres aux États membres sur la démocratie électronique (e-démocratie). Adoptée par le Comité des Ministres le 18 février 2009, lors de la 1049^e réunion des Délégués des Ministres.

Comité des Ministres. 2009. Recommandation CM/Rec(2009)2 du Comité des Ministres aux États membres sur l'évaluation, l'audit et le suivi de la participation et des politiques de la participation aux niveaux local et régional. Adoptée par le Comité des Ministres le 11 mars 2009 lors de la 1050^e réunion des Délégués des Ministres.

Comité des Ministres. 2010. Recommandation CM/Rec(2010)13 sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage. Adoptée par le Comité des Ministres le 23 novembre 2010 lors de la 1099^e réunion des Délégués des Ministres.

Comité des Ministres. 2011. Recommandation CM/Rec(2011)7 du Comité des Ministres aux États membres sur une nouvelle conception des médias. Adoptée par le Comité des Ministres le 21 septembre 2011, lors de la 1121^e réunion des Délégués des Ministres.

Comité des Ministres. 2012. Recommandation CM/Rec(2012)3 sur la protection des droits de l'homme dans le contexte des moteurs de recherche. Adoptée par le Comité des Ministres le 4 avril 2012 lors de la 1139^e réunion des Délégués des Ministres.

Comité des Ministres. 2012. Recommandation CM/Rec(2012)4 sur la protection des droits de l'homme dans le cadre des services de réseaux sociaux. Adoptée par le Comité des Ministres le 4 avril 2012 lors de la 1139^e réunion des Délégués des Ministres.

Comité des Ministres. 2014. Recommandation CM/Rec(2014)7 sur la protection des lanceurs d'alerte. Adoptée par le Comité des Ministres le 30 avril 2014 lors de la 1198^e réunion des Délégués des Ministres.

Comité des Ministres. 2016. Recommandation CM/Rec (2016)1 sur la protection et la promotion du droit à la liberté d'expression et du droit à la vie privée en lien avec la neutralité du réseau. Adoptée par le Comité des Ministres le 13 janvier 2020, lors de la 1244^e réunion des Délégués des Ministres.

Comité des Ministres. 2016. Recommandation CM/Rec(2016)4 du Comité des Ministres aux États membres sur la protection du journalisme et la sécurité des journalistes et autres acteurs des médias. Adoptée par le Comité des Ministres le 13 avril 2016 lors de la 1253^e réunion des Délégués des Ministres.

Comité des Ministres. 2016. Recommandation CM/Rec(2016)5 du Comité des Ministres aux États membres sur la liberté d'internet. Adoptée par le Comité des Ministres le 13 avril 2016 lors de la 1253^e réunion des Délégués des Ministres.

Comité des Ministres. 2017. Recommandation CM/Rec(2017)5 du Comité des Ministres aux États membres sur les normes relatives au vote électronique. Adoptée par le Comité des Ministres le 14 juin 2017, lors de la 1289^e réunion des Délégués des Ministres.

Comité des Ministres. 2018. Recommandation CM/Rec(2018)1 du Comité des Ministres aux États membres sur le pluralisme des médias et la transparence de leur propriété. Adoptée par le Comité des Ministres le 7 mars 2008, lors de la 1309^e réunion des Délégués des Ministres.

Comité des Ministres. 2018. Recommandation CM/Rec(2018)2 du Comité des Ministres aux États membres sur les rôles et les responsabilités des intermédiaires d'internet. Adoptée par le Comité des Ministres le 7 mars 2008, lors de la 1309^e réunion des Délégués des Ministres.

Comité des Ministres. 2018. Recommandation CM/Rec(2018)4 du Comité des Ministres aux États membres sur la participation des citoyens à la vie publique au niveau local. Adoptée par le Comité des Ministres le 21 mars 2008, lors de la 1311^e réunion des Délégués des Ministres

Comité des Ministres. 2019. Recommandation CM/Rec(2019)2 du Comité des Ministres aux États membres en matière de protection des données relatives à la santé. Adoptée par le Comité des Ministres le 27 mars 2019, lors de la 1342^e réunion des Délégués des Ministres.

Comité des Ministres. 2019. Recommandation CM/Rec(2019)3 du Comité des Ministres aux États membres sur le contrôle des actes des collectivités locales. Adoptée par le Comité des Ministres le 4 avril 2019 lors de la 1343^e réunion des Délégués des Ministres.

Comité des Ministres. 2020. Recommandation CM/Rec(2020)1 du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme. Adoptée par le Comité des Ministres le 8 avril 2020 lors de la 1373^e réunion des Délégués des Ministres.

Conseil des barreaux européens. 2020. Considérations du CCBE sur les aspects juridiques de l'intelligence artificielle. Bruxelles.

Conseil de l'Europe - Commission de Venise, OSCE/BIDDH. 2011. Lignes directrices sur la réglementation des partis politiques <https://www.osce.org/odihr/77812>, consulté le 20.12.2019.

Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel du Conseil de l'Europe. 2019. Lignes directrices sur l'intelligence artificielle et la protection des données. T- PD(2019)01. <https://rm.coe.int/2018-lignes-directrices>, consulté le 16.11.2019.

Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel du Conseil de l'Europe. 2017. Lignes directrices sur la protection des personnes à l'égard du traitement des données à caractère personnel à l'ère des mégadonnées T-PD(2017)1. <https://rm.coe.int/2018-lignes-directrices>, consulté le 16.11.2019

Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel du Conseil de l'Europe. 2019. Profilage et la Convention 108+ : Pistes pour une actualisation. T- PD(2019)07BISrev.

Conseil de l'Europe, Direction générale de la démocratie - Comité européen sur la démocratie et la gouvernance. 2016. Recueil des textes les plus pertinents du Conseil de l'Europe dans le domaine de la démocratie <https://rm.coe.int/2018-lignes-directrices>, consulté le 18.11.2019.

Conseil de l'Europe, Direction générale de la démocratie et des affaires politiques - Direction des institutions démocratiques. 2011. Lignes directrices relatives à la transparence des élections par voie électronique.

Conseil de l'Europe, Direction générale de la démocratie et des affaires politiques - Direction des institutions démocratiques. 2009. Projet « Bonne gouvernance dans la société de l'information », CM(2009)9.

Conseil de l'Europe, Assemblée parlementaire. 2019. Résolution 2254 (2019) La liberté des médias en tant que condition pour des élections démocratiques.

Conseil de l'Europe. 2008. Les 12 Principes de bonne gouvernance <https://www.coe.int/fr/web/good-governance/12-principles>, consulté le 16.03.2020

Conseil de l'Europe. 2009. Protocole additionnel à la Charte européenne de l'autonomie locale sur le droit de participer aux affaires des collectivités locales, Utrecht, 16.XI.2009.

Conseil de l'Europe. 2017. Lignes directrices relatives à la participation civile aux décisions politiques (CM(2017)83-final). Adoptées par le Comité des Ministres le 27 septembre 2017, lors de la 1295^e réunion des Délégués des Ministres.

Conseil de l'Europe. 2019. Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques. Adoptée par le Comité des Ministres le 13 février 2019, lors de la 1337^e réunion des Délégués des Ministres.

Conseil de l'Europe. 2019. Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques (Adoptée par le Comité des Ministres le 13 février 2019 lors de la 1337^e réunion des Délégués des Ministres).

Conseil de l'Europe. Lignes directrices relatives à la participation civile aux décisions politiques CM(2017)83-final. <https://www.coe.int/fr/web/civil-society/guidelines>, consulté le 15.03.2020.

Comité d'experts du Conseil de l'Europe sur les intermédiaires d'internet (MSI-NET). 2018. Algorithmes et droits humains. Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données (en particulier les algorithmes) et éventuelles implications réglementaires. <https://rm.coe.int/etude-sur-les-algorithmes-version-finale/1680770cc2>, consulté le 28.11.2019.

Cummings, M. L., Roff H. M., Cukier K., Parakilas J. et Bryce H. 2018. *Chatham House Report. Artificial Intelligence and International Affairs Disruption Anticipated*. London: Chatham House. The Royal Institute of International Affairs.

<https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial-intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf>, consulté le 21.03.2020.

Dupont, B. et al. 2018. *L'intelligence artificielle dans le contexte de la criminalité et de la justice pénale*. Institut coréen de criminologie 2018.

<https://www.cyberjustice.ca/publications/lintelligence-artificielle-dans-le-contexte-de-la-criminalite-et-de-la-justice-penale/>, consulté le 30.05.2020.

Code européen de bonnes pratiques contre la désinformation en ligne, 2018

<https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>, 23.03.2020.

Commission européenne pour la démocratie par le droit (Commission de Venise). 2002. Code de bonne conduite en matière électorale. Lignes directrices et rapport explicatif. Adoptés par la Commission lors de ses 51^e et 52^e sessions plénières (Venise, 5-6 juillet et 18-19 octobre 2002).

Commission européenne pour la démocratie par le droit (Commission de Venise). 2019. Rapport sur les technologies numériques et les élections, établi conjointement par la Commission de Venise et la Direction de la société de l'information et de la lutte contre la criminalité (Direction générale Droits de l'homme et État de droit, DGI).

Commission européenne pour l'efficacité de la justice (CEPEJ). 2018. Charte éthique européenne sur l'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, Annexe II. Adoptée par la CEPEJ lors de sa 31^e réunion plénière (Strasbourg, 3-4 décembre 2018) <https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b>, consulté le 04.12.2018.

Commission européenne, Direction générale des réseaux de communications, du contenu et des technologies. 2018. *A Multi-Dimensional Approach to Disinformation Report of the Independent High Level Group on Fake News and Online Disinformation*.

<https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>, consulté le 22.03.2018.

Commission européenne. 2014. Livre vert sur la santé mobile. <https://eur-lex.europa.eu>, consulté le 12.01.2020.

Commission européenne. 2020. Une stratégie européenne pour les données, COM(2020) 66 final. <https://ec.europa.eu/> consulté le 20.02.2020.

Commission européenne. 2020. Rapport sur les conséquences de l'intelligence artificielle, de l'internet des objets et de la robotique sur la sécurité et la responsabilité, COM(2020) 64 final. https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_fr.pdf, consulté le 20.02.2020.

- Commission européenne. 2020. Livre blanc sur l'intelligence artificielle - une approche européenne axée sur l'excellence et la confiance, COM(2020) 65 final. <https://ec.europa.eu> consulté le 20.02.2020.
- Agence des droits fondamentaux de l'Union européenne. 2018. #BigData: *Discrimination in Data-Supported Decision Making*. <https://fra.europa.eu>, consulté le 20.05.2020.
- Agence des droits fondamentaux de l'Union européenne. 2018. *Guide pour la prévention du profilage illicite aujourd'hui et demain*. <https://fra.europa.eu>, consulté le 20.05.2020.
- Faye Jacobsen, A. 2013. *The Right to Public Participation. A Human Rights Law Update*. Document de réflexion. Institut danois pour les droits de l'homme.
- Ferryman, K. et Pitcan, M. 2018. Fairness in Precision Medicine. Data & Society, February, <https://datasociety.net>
- González Fuster, G. 2014. *The Emergence of Personal Data Protection as a Fundamental Right of the EU*. Cham-New York: Springer International Publishing.
- Maisley, N. 2017. *The International Right of Rights? Article 25(a) of the ICCPR as a Human Right to Take Part in International Law-Making*. 28 European Journal of International Law 89.
- Manheim, K., Kaplan, L. 2019. *Artificial Intelligence: Risks to Privacy and Democracy*. 21 Yale J.L. & Tech. 106.
- Mantelero, A. 2018. *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*. 34 Computer Law & Security Review 754.
- Mantelero, A. 2019. Intelligence artificielle et protection des données : Enjeux et solutions possibles. Rapport sur l'intelligence artificielle (Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel : Strasbourg), T-PD(2018)09Rev, <https://rm.coe.int/artificial-intelligence-and-data-protection>
- Mayer-Schönberger, V. 1997. *Generational development of data protection in Europe?* In Agre, P.E., Rotenberg, M. (eds). *Technology and privacy: The new landscape*. Cambridge, MA: MIT Press.
- Mehr, H. 2017. *Artificial Intelligence for Citizen Services and Government*. Harvard Kennedy School. Ash Center for Democracy and Innovation.
- Meijer, A., Wessels, M. 2019. *Predictive Policing: Review of Benefits and Drawbacks*. 42 International Journal of Public Administration 1031.
- Nunez, C. 2017. *Artificial Intelligence and Legal Ethics: Whether AI Lawyers Can Make Ethical Decisions*. 20 Tul. & Tech. & Intell. Prop. 189-204.
- OCDE. 2013. Recommandation du Conseil de l'OCDE concernant les Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel, C(80)58/FINAL, telle que révisée le 11 juillet 2013 par le document C(2013)79.
- OSCE. 2020. *Non-Paper on the Impact of Artificial Intelligence on Freedom of Expression* <https://www.osce.org/representative-on-freedom-of-media/447829>, consulté le 11.06.2020.
- Osoba, O.A., Welsler, W. 2017. *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. RAND Corporation. <https://www.rand.org>, consulté le 20.05.2020.
- Assemblée parlementaire. 2019. Résolution 2254 (2019) La liberté des médias en tant que condition pour des élections démocratiques.
- Assemblée parlementaire. 2019. Résolution 2300 (2019)1 Améliorer la protection des lanceurs d'alerte partout en Europe

Perry, W.L. et al. 2013. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation 2013. https://www.rand.org/pubs/research_reports/RR233.html, consulté le 30.03.2020.

Privacy International. 2017. *Smart Cities: Utopian Vision, Dystopian Reality*. <https://privacyinternational.org/report/638/smart-cities-utopian-vision-dystopian-reality>, consulté le 21.03.2020.

Re, R.M., Solow-Niederman, A. 2019. *Developing Artificially Intelligent Justice*. 22 Stan. Tech. L. Rev. 242.

Rosenbaum, D., 2006. *The limits of hot spots policing*. In: D. Weisburd et A. Braga, eds. *Police innovation: contrasting perspectives*. New York, NY: Cambridge University Press, 245–266.

Rouvroy, A. 2016. « Des données et des hommes » – Droits et libertés fondamentaux dans un monde de données massives (Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel : Strasbourg), T-PD-BUR(2015)09Rev, <http://rm.coe.int> consulté le 04.11.2019.

Savaget, P., Chiarini, T., Evans, S. 2019. *Empowering Political Participation through Artificial Intelligence*. 46 Science and Public Policy 369.

Seatzu, F. 2015. The Experience of the European Court of Human Rights with the European Convention on Human Rights and Biomedicine. 31(81) Utrecht Journal of International and European Law 5, DOI: <http://dx.doi.org/10.5334/ujiel.da>.

Sunstein, C.R. 2015a. *The Ethics of Nudging*. Yale Journal on Regulation 32: 413

Sunstein, C.R. 2015b. *Why Nudge? The Politics of Libertarian Paternalism*. New Haven: Yale University Press.

Sunstein, C.R., Thaler, R. 2003. *Libertarian Paternalism in Not an Oxymoron*. University of Chicago Law Review 70 (4): 1159.

Thaler, R., Sunstein, C.R. 2008. *Nudge*. New Haven, CT: Yale University Press.

Le Rapporteur spécial des Nations Unies (ONU) sur la promotion et la protection du droit à la liberté d'opinion et d'expression, la Représentante de l'Organisation pour la sécurité et la coopération en Europe (OSCE) sur la liberté des médias, le Rapporteur spécial pour la liberté d'expression de l'Organisation des États Américains (OEA) et le Rapporteur spécial sur la liberté d'expression et l'accès à l'information de la Commission africaine des droits de l'homme et des peuples (CADHP). 2017. Déclaration conjointe sur la liberté d'expression et les fausses nouvelles (« fake news »), la désinformation et la propagande. <https://www.law-democracy.org> consulté le 02.02.2020.

T-PD(2019)07BISrev. Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel. 2019. Profilage et la Convention 108+ : Pistes pour une actualisation

Bureau du Haut-Commissaire des Nations Unies aux droits de l'homme 1981. CESCR Observation générale n° 1 : Rapports des États parties. Adoptée lors de la treizième session du Comité des droits économiques, sociaux et culturels, le 27 juillet 1981 (Contenue dans le document E/1989/22).

Bureau du Haut-Commissaire des Nations Unies aux droits de l'homme 1996. Observation générale n° 25 : Le droit de participer aux affaires publiques, le droit de vote et le droit d'accéder, dans des conditions d'égalité, aux fonctions publiques (article 25). CCPR/C/21/Rev.1/Add.7.

UNESCO. 2019. Étude préliminaire concernant un éventuel instrument normatif sur l'éthique de l'intelligence artificielle. <https://unesdoc.unesco.org/ark:/48223/pf0000369455>, consulté le 21.11.2019.

UNESCO. Déclaration universelle sur le génome humain et les droits de l'homme (11 novembre 1997).

Convention des Nations Unies contre la corruption, 2003.

van Brakel, R., De Hert, P. 2011. *Policing, Surveillance and Law in a Pre-Crime Society: Understanding the Consequences of Technology Based Strategies*. 3 Cahiers Politiestudies, Jaargang 163.

Verbeek, P-P. 2011. *Understanding and Designing the Morality of Things*. Chicago- London: The University of Chicago Press.

Završnik, A. 2019. *Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings*. European Journal of Criminology, 1-20, <https://doi.org/10.1177/1477370819876762>, consulté le 20.02.2020.

Annexe 1. Instruments juridiques

| Instruments contraignants | Instruments non contraignants connexes |
|--|---|
| Biomédecine | |
| <p>Convention sur les droits de l'homme et la biomédecine (Convention d'Oviedo)</p> <p>Protocole additionnel, relatif aux tests génétiques à des fins médicales</p> <p>Protocole additionnel, relatif à la recherche biomédicale</p> | <p>Rec(2016)8 sur le traitement des données à caractère personnel relatives à la santé à des fins d'assurance, y compris les données résultant de tests génétiques et son exposé des motifs</p> <p>Recommandation CM/Rec(2016)6 du Comité des Ministres aux États membres sur la recherche utilisant du matériel biologique d'origine humaine</p> <p>Plan d'action stratégique sur les droits de l'homme et les technologies en biomédecine 2020-2025</p> |
| Anti-discrimination | |
| <ul style="list-style-type: none"> - Déclaration universelle des droits de l'homme - Pacte international relatif aux droits civils et politiques - Pacte international relatif aux droits économiques, sociaux et culturels - Convention internationale sur l'élimination de toutes les formes de discrimination raciale - Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes (CEDAW) - Convention des Nations Unies relative aux droits des personnes handicapées - Convention européenne des droits de l'homme (CEDH) et ses protocoles (n° 12 en particulier) - Charte sociale européenne - Convention sur la cybercriminalité et son Protocole additionnel relatif à l'incrimination d'actes de nature raciste et xénophobe commis par le biais de systèmes informatiques - Convention sur la prévention et la lutte contre la violence à l'égard des femmes et la violence domestique (Convention d'Istanbul) - Charte des droits fondamentaux de l'Union européenne | <p>Recommandations de politique générale de l'ECRI, et en particulier les recommandations n°2 (sur les organismes de promotion de l'égalité), 11 (sur la lutte contre le racisme et la discrimination raciale dans les activités de la police) et 15 (sur la lutte contre le discours de haine).</p> <p>Recommandation 2098 (2017) de l'APCE Mettre fin à la cyberdiscrimination et aux propos haineux en ligne</p> <p>Recommandation (2019)1 du Comité des Ministres sur la prévention et la lutte contre le sexisme</p> |
| Cybercriminalité et preuves électroniques | |
| <p>Convention sur la cybercriminalité</p> | <p>Notes d'orientation du Comité de la Convention sur la cybercriminalité sur les attaques DDOS, les attaques visant les infrastructures d'information critiques, les logiciels malveillants, les spams, l'usurpation d'identité, etc.</p> |

| | |
|---|---|
| Justice | |
| <ul style="list-style-type: none"> - Déclaration universelle des droits de l'homme - Pacte international relatif aux droits civils et politiques - Convention internationale sur l'élimination de toutes les formes de discrimination raciale - Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes - Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH) - Charte des droits fondamentaux de l'Union européenne | <p>CEPEJ 2019. Charte éthique européenne d'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement</p> |
| Congrès des pouvoirs locaux et régionaux | |
| <p>Charte européenne de l'autonomie locale</p> | <p>Résolution 435 (2018) et Recommandation 424 (2018) du Congrès :</p> <p>« Transparence et gouvernement ouvert » Résolution 417 (2017) et Recommandation 398 (2017) du Congrès : « Le libre accès aux données = amélioration des services publics » Résolution 394 (2015) du Congrès : « Médias électroniques : une nouvelle donne pour les responsables politiques locaux et régionaux » Résolution 290 (2009) du Congrès : « La démocratie électronique : perspectives et risques pour les collectivités locales ».</p> |
| Démocratie et participation | |
| <ul style="list-style-type: none"> - Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH) - Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (STE n°108) de 1981 et Protocole de 2018 modernisant la Convention | <ul style="list-style-type: none"> - Recommandation Rec(2003)4 du Comité des Ministres sur les règles communes contre la corruption dans le financement des partis politiques et des campagnes électorales - Code de bonne conduite en matière électorale (Commission de Venise) - Lignes directrices conjointes sur la réglementation des partis politiques (Commission de Venise et BIDDH/OSCE) - Recommandation CM/Rec(2007)15 du Comité des Ministres aux États membres sur des mesures concernant la couverture des campagnes électorales par les médias - voir également la Recommandation CM/Rec(2018)1 sur le pluralisme des médias et la transparence de leur propriété, la Recommandation CM/Rec(2018)2 sur les rôles et les responsabilités des intermédiaires d'Internet, la Recommandation CM/Rec(2016)1 sur la protection et la promotion du droit à la liberté d'expression et du droit à la vie privée en lien avec la neutralité du réseau - Recommandation n° R (99) 5 du Comité des Ministres sur la protection de la vie privée sur Internet Recommandation CM/Rec(2010)13 sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage |

| | |
|---|--|
| | Recommandation CM/Rec(2012)3 sur la protection des droits de l'homme dans le contexte des moteurs de recherche |
| | Recommandation CM/Rec(2012)4 sur la protection des droits de l'homme dans le cadre des services de réseaux sociaux |
| Liberté d'expression | |
| <ul style="list-style-type: none"> - Convention européenne des droits de l'homme - Pacte international relatif aux droits civils et politiques - Charte des droits fondamentaux de l'Union européenne | <p>Déclaration universelle des droits de l'homme CM/Rec(2018)2 sur les rôles et les responsabilités des intermédiaires d'Internet CM/Rec(2020)x sur les impacts des systèmes algorithmiques sur les droits de l'homme Déclaration (13/02/2019) sur les capacités de manipulation des processus algorithmiques CM/Rec(2018)1 sur le pluralisme des médias et la transparence de leur propriété CM/Rec(2020)x sur la promotion d'un environnement favorable à un journalisme de qualité à l'ère numérique</p> |
| Elections | |
| <p>Déclaration universelle des droits de l'homme Pacte international relatif aux droits civils et politiques Convention des Nations Unies sur l'élimination de toutes les formes de discrimination raciale Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes Convention des Nations Unies relative aux droits des personnes handicapées Convention des Nations Unies contre la corruption Convention de sauvegarde des droits de l'homme et des libertés fondamentales (STE n° 5) Protocole additionnel à la Convention de sauvegarde des droits de l'homme et des libertés fondamentales (STE n° 9) Charte européenne de l'autonomie locale (STE n° 122) Charte européenne des langues régionales ou minoritaires (STE n° 148) Convention sur la cybercriminalité (STE n° 185) Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (STE n° 108)</p> | <p>Code de bonne conduite en matière électorale, adopté par le Conseil des élections démocratiques du Conseil de l'Europe et la Commission européenne pour la démocratie par le droit (Commission de Venise)</p> <p>Recommandation Rec(2003)3 du Comité des Ministres aux États membres sur la participation équilibrée des femmes et des hommes à la prise de décision politique et publique</p> <p>Convention sur les normes en matière d'élections démocratiques et les droits et libertés électoraux dans les États membres de la Communauté des États indépendants (CDL-EL(2006)031rev)</p> <p>Recommandation Rec(99)5 du Comité des Ministres sur la protection de la vie privée sur Internet</p> <p>Recommandation Rec(2004)15 du Comité des Ministres aux États membres sur la gouvernance électronique (e-gouvernance)</p> <p>Recommandation CM/Rec(2007)15 du Comité des Ministres aux États membres relative à des mesures concernant la couverture des campagnes électorales par les médias</p> <p>Recommandation CM/Rec(2009)1 du Comité des Ministres aux États membres sur la démocratie électronique</p> <p>Recommandation CM/Rec(2017)5 du Comité des Ministres aux États membres sur les normes relatives au vote électronique</p> |

| | |
|---|--|
| <p>Protocole additionnel à la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, concernant les autorités de contrôle et les flux transfrontières de données (STE n° 181)</p> <p>Charte des droits fondamentaux de l'Union européenne</p> <p>Convention-cadre pour la protection des minorités nationales et rapport explicatif</p> | <p>Document de la réunion de Copenhague de la conférence sur la dimension humaine de l'OSCE</p> <p>Rapport sur l'abus de ressources administratives pendant les processus électoraux, adopté par le Conseil des élections démocratiques et par la Commission de Venise (CDL-AD(2013)033)</p> <p>Rapport sur les règles électorales et les actions positives en faveur de la participation des minorités nationales aux processus de décision dans les pays européens, adopté par le Conseil des élections démocratiques et par la Commission de Venise (CDL-AD(2005)009)</p> <p>Code de bonne conduite en matière référendaire, adopté par le Conseil des élections démocratiques et par la Commission de Venise</p> <p>Stratégie du Conseil de l'Europe sur le handicap 2017-2023</p> <p>Résolution 1897 (2012) de l'APCE, Garantir des élections plus démocratiques</p> <p>Code de bonne conduite en matière de partis politiques et rapport explicatif, adoptés par la Commission de Venise (CDL-AD(2009)021)</p> |
| <p>Démocratie (hors questions relatives aux élections et au cycle électoral)</p> | |
| <ul style="list-style-type: none"> - Déclaration universelle des droits de l'homme - Pacte international relatif aux droits civils et politiques - Convention internationale sur l'élimination de toutes les formes de discrimination raciale - Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes - Charte des droits fondamentaux de l'Union européenne - Convention 108+ - Convention de sauvegarde des droits de l'homme et des libertés fondamentales et ses protocoles - Charte européenne de l'autonomie locale - Convention-cadre pour la protection des minorités nationales. | <ul style="list-style-type: none"> - Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques Voir aussi le Recueil Chapitre A (Séparation des pouvoirs / bonne gouvernance) Chapitre B (Diversité et pluralisme des médias ; protection de la liberté d'expression sur Internet) Chapitre C (participation de la société civile) Chapitre E (participation des citoyens) |
| <p>Bonne gouvernance</p> | |
| <ul style="list-style-type: none"> - Déclaration universelle des droits de l'homme - Pacte international relatif aux droits civils et politiques - Convention internationale sur l'élimination de toutes les formes de discrimination raciale - Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes | <ul style="list-style-type: none"> - 12 principes de bonne gouvernance démocratique - Recommandation du Comité des Ministres aux États membres sur le contrôle des actes des collectivités locales CM/Rec(2019)3 Voir aussi le Recueil |

| | |
|---|---|
| <ul style="list-style-type: none"> - Convention de sauvegarde des droits de l'homme et des libertés fondamentales et ses protocoles - Charte des droits fondamentaux de l'Union européenne - Convention 108+ - Charte européenne de l'autonomie locale et ses protocoles - Convention du Conseil de l'Europe sur l'accès aux documents publics | <p>Chapitre A (bonne gouvernance)</p> <p>Chapitre E (Politiques d'intégration - normes et mécanismes)</p> <p>Voir également https://www.coe.int/en/web/good-governance/conventions-recommendations</p> |
| <p>Égalité femmes-hommes, y compris les questions relatives à la violence à l'égard des Femmes</p> | |
| <ul style="list-style-type: none"> - Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH) - Convention du Conseil de l'Europe sur la prévention et la lutte contre la violence à l'égard des femmes et la violence domestique (article 17§1 sur la participation du secteur des technologies de l'information et de la communication à la prévention et à la lutte contre la violence à l'égard des femmes, article 34 sur le harcèlement en ligne) - Charte sociale européenne - Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes - Déclaration universelle des droits de l'homme - Pacte international relatif aux droits civils et politiques - Pacte international relatif aux droits économiques, sociaux et culturels - Convention internationale sur l'élimination de toutes les formes de discrimination raciale - Charte des droits fondamentaux de l'Union européenne | <p>Recommandation (2019)1 du Comité des Ministres Prévention et lutte contre le sexisme</p> <p>Recommandation (2013)1 du Comité des Ministres sur l'égalité entre les femmes et les hommes et les médias</p> <p>Recommandation de politique générale n° 15 de l'ECRI sur le discours de haine</p> |
| <p>Culture, créativité et patrimoine</p> | |
| <p>Déclaration universelle des droits de l'homme Pacte international relatif aux droits économiques, sociaux et culturels Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH)</p> | |
| <p>Convention-cadre pour la protection des minorités nationales</p> | <p>Nombreuses recommandations et résolutions du Comité des Ministres, de l'Assemblée parlementaire et du Congrès des pouvoirs locaux et régionaux du Conseil de l'Europe sur les questions d'identité culturelle et de diversité.</p> |

| | |
|--|---|
| Charte européenne des langues régionales ou minoritaires ²²⁴ | Nombreuses recommandations et résolutions du Comité des Ministres, de l'Assemblée parlementaire et du Congrès des pouvoirs |
| | locaux et régionaux du Conseil de l'Europe sur l'identité culturelle, la diversité culturelle et le dialogue culturel, ainsi que les questions relatives aux minorités |
| Conventions du CdE dans le domaine du patrimoine culturel (Convention de Nicosie, non entrée en vigueur, Convention de Faro, Convention de La Valette, Convention de Grenade) | Nombreuses recommandations du Comité des Ministres et de l'Assemblée parlementaire du Conseil de l'Europe sur le patrimoine culturel |
| Convention de l'Unesco sur la protection et la promotion de la diversité des expressions culturelles | Déclaration du CdE sur la diversité culturelle Recommandation du Comité des Ministres du CdE sur la Convention de l'Unesco |
| Convention du Conseil de l'Europe sur la coproduction cinématographique (révisée) Directive UE « Services de médias audiovisuels » / Directive (UE) 2018/1808 Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE | Résolution (88)15 modifiée instituant un Fonds européen de soutien à la coproduction et à la diffusion des œuvres de création cinématographiques et audiovisuelles Recommandation CM/Rec(2017)9 du Comité des Ministres aux États membres sur l'égalité entre les femmes et les hommes dans le secteur audiovisuel |
| Déclaration universelle des droits de l'homme Pacte international relatif aux droits économiques, sociaux et culturels Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH) | |
| GRECO | |
| Convention civile sur la corruption et Convention pénale sur la corruption ; évaluations du GRECO | Recommandations CM sur le modèle de code de conduite pour les agents publics, les activités de lobbying, la protection des lanceurs d'alerte, la transparence du financement des partis politiques, etc. |
| Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH) | Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH) |
| CSE | |
| Charte sociale européenne (Charte de 1961, protocole de 1988 et Charte révisée de 1996) | |

²²⁴ Étude à venir (printemps 2020) sur l'impact de l'IA sur les domaines couverts par la Charte européenne des langues régionales ou minoritaires

| | |
|--|--|
| Droits contenus dans la Charte sociale européenne | <p>Quelques exemples :</p> <ul style="list-style-type: none"> - Nouvelle stratégie et plan d'action du Conseil de l'Europe pour la cohésion sociale (approuvés par le CM le 7 juillet 2010) - CM Rec(2000)3 proposant un droit individuel, universel et justiciable à la satisfaction des besoins matériels élémentaires - etc. |
| Par ailleurs, de nombreux sujets relevant d'autres domaines de travail du CdE comportent des aspects qui ont trait aux droits sociaux (et droits contenus dans la Charte). | <p>Quelques exemples :</p> <ul style="list-style-type: none"> - CM Rec(93)1 relative à l'accès effectif au droit et à la justice des personnes en situation de grande pauvreté - volet « droits sociaux » des règles pénitentiaires (santé, conditions de vie, emploi, éducation, droits familiaux, etc.) - etc. |

Annexe 2. Domaines concernés

| |
|--|
| Domaines concernés (applications) |
| |
| <p>Biomédecine</p> <ul style="list-style-type: none"> • Surveillance, prévention, diagnostic et interventions dans le domaine des soins de santé, reposant sur l'IA • Surveillance, diagnostic, suivi, financement de traitements (assurance) basés sur la prédiction (par exemple applications de reconnaissance faciale des utilisateurs et services en ligne en dehors du domaine des soins de santé) |
| <p>Anti-discrimination</p> <ul style="list-style-type: none"> • Prise de décisions automatisée dans différents domaines des secteurs public et privé (exemple : candidatures à des postes, prestations sociales, accès aux biens et services, tels que prêts bancaires, assurances) • Police prédictive (avec un risque élevé de profilage racial) • Justice prédictive • Reconnaissance faciale • Technologies de prédiction des comportements, par exemple reconnaissance émotionnelle et détection de mensonges par intelligence artificielle • Assistants personnels (par ex. Siri) • Modération de contenus • Protection des données |

Cybercriminalité et preuves électroniques

- Cybercriminalité et cyberattaques automatisées, par exemple :
 - Attaques par déni de service distribué (attaques DDOS)
 - Attaques contre les infrastructures d'information critiques
 - Attaques « man-in-the-middle »
 - Hameçonnage et techniques similaires d'ingénierie sociale
 - Repérage des vulnérabilités
 - etc.
- Enquêtes sur la cybercriminalité et informatique judiciaire :
 - Collecte et analyse de preuves électroniques (pour toutes les infractions)
 - Attribution
 - Rétro-ingénierie
- Cybersécurité et prévention de la cybercriminalité :
 - Détection de logiciels malveillants, d'intrusions, etc.
 - Correction automatisée des vulnérabilités

Secteur de la justice

- Traitement des décisions et des données judiciaires :
 - pour étayer les décisions judiciaires ou la recherche judiciaire
 - règlement des litiges en ligne
 - conseils juridiques aux parties à un litige
- Police prédictive

Congrès des pouvoirs locaux et régionaux

- Prestation de services publics au niveau local
 - Instruments de promotion de la participation citoyenne
 - Grande diversité d'applications numériques et électroniques dans les villes et les collectivités locales.
 - Application des technologies de l'information et de la communication (TIC) pour améliorer la qualité de vie et les environnements de travail en ville.
 - Gouvernance intelligente des villes
 - Intégration des TIC dans les systèmes du secteur public
- Mise en œuvre locale de pratiques qui rapprochent la population des TIC pour favoriser l'innovation et améliorer les connaissances qu'elles offrent.

Liberté d'expression

- Communication individuelle (par modération et restriction automatisées des contenus ; systèmes de tri, classification, optimisation et recommandation par algorithmes)
- Production et diffusion de médias (robot-journalisme, journalisme de données, TALN, micro-ciblage de la base de lecteurs, fils d'actualité automatisés selon le profil du lecteur)
- Communication sociale et politique, fragmentation/polarisation du discours public, redlining politique (micro-ciblage de la base électorale, recours à des bots pour influencer l'opinion, prolifération de sites de médias locaux automatisés)

Élections

Période préélectorale :

- Planification du calendrier électoral
- Formation des acteurs électoraux
- Délimitation des circonscriptions électorales
- Inscription des électeurs et des candidats
- Accréditation des observateurs (internationaux et nationaux)
- Mise à jour de la liste d'électeurs
- Actualisation du cadre juridique
- Financement des partis politiques
- Propagande électorale par l'administration et par les partis politiques/candidats

Période électorale :

- Financement des campagnes électorales
- Accès aux médias
- Vote
- Décompte des bulletins
- Proclamation des résultats

Période post-électorale :

- Publication des résultats électoraux
- Règlement du contentieux électoral

Démocratie (hors questions relatives aux élections et au cycle électoral)

- Séparation des pouvoirs
- Participation de la société civile
- Participation des citoyens
- Vie privée
- Citoyenneté
- Protection des minorités
- Pluralisme et diversité
- Légitimité

Bonne gouvernance

- Collectivités locales
- Administration régionale
- Prestation de services
- Dotations budgétaires
- Sécurité sociale et mécanismes de prestations sociales
- Police et justice
- Villes intelligentes
- Appels d'offres publics et attribution des marchés
- Capacités institutionnelles

Égalité femmes-hommes, y compris les questions relatives à la violence à l'égard des femmes

Question générale des préjugés sexistes dont héritent les algorithmes des systèmes de données lors de la phase d'entraînement (valable pour de nombreux domaines), pouvant aggraver les inégalités entre femmes et hommes et les inégalités sociales.

Questions générales liées à l'IA en tant que secteur d'emploi

- Le manque de participation/la sous-représentation des femmes accentue les éventuels préjugés sexistes et exclut les femmes d'un secteur influent
- Exploitation des « travailleurs du clic » en Europe et dans le monde (bas salaires, absence de protection sociale, pas de droits des travailleurs, exposition durable à des contenus préjudiciables dans le cas des modérateurs de contenus, etc.)

Problèmes spécifiques

- Discrimination dans la sélection des candidatures
- Décisions automatisées pour des services publics et privés
- Reconnaissance faciale et vocale (fonctionnant moins bien pour les femmes, en particulier certains groupes)
- Surveillance / harcèlement facilités par les outils d'IA, par exemple dans le contexte de la violence domestique
- Les décisions automatisées aggravent encore le risque de discrimination multiple fondée sur le sexe, la race et l'origine sociale en combinant des données secondaires comme le niveau d'études, l'adresse et le niveau de revenu.
- Justice prédictive (par ex. violences faites aux femmes)
- Santé prédictive reposant sur des données empreintes de préjugés sexistes (ex. certaines maladies qualifiées de « féminines » ou « masculines »)
- Préjugés hérités dans la modération de contenus par les machines (tolérance élevée au sexisme, au discours de haine sexiste et aux violences contre les femmes)
- Assistants virtuels/robots genrés perpétuant les stéréotypes fondés sur le sexe
- Marketing genré perpétuant les stéréotypes fondés sur le sexe
- Application de prix différenciés selon le sexe

Impacts positifs

- Utilisation de dispositifs de géolocalisation pour assurer le respect des ordonnances de protection dans le contexte des violences faites aux femmes
- Recours à l'IA par la police et la justice pour évaluer les risques dans les affaires de violence domestique
- Recours à l'IA pour repérer et signaler les préjugés sexistes et bloquer ou supprimer la diffusion de propos haineux (sexistes) sur les plateformes
- Développement d'applications d'information et d'aide aux femmes victimes de violence

Utilisation d'outils d'IA pour analyser les contenus et repérer les préjugés sexistes / analyser les représentations (par ex. dans les films ou autres médias)

Culture, créativité et patrimoine

- Accès et participation à la vie publique/culturelle ;
- Liberté d'expression (y compris liberté d'expression artistique)
- Accès à de l'information impartiale ?
- Décisions automatisées, ciblage, profilage
- Apprentissage de langues menacées pour les préserver/protéger
- Assistance automatisée dans l'administration, la santé, etc. pour les locuteurs de groupes/langues minoritaires
- Géolocalisation, police prédictive, analyse criminelle (destruction, pillage, trafic de biens culturels ; ciblage ; sensibilisation au patrimoine menacé pour contribuer à sa protection)
- Création automatisée de contenus, ciblage, profilage (création culturelle, échanges culturels, consommation culturelle)

- Conception et production de contenus audiovisuels :
 - Analyse prédictive d'audience
 - Analyse automatique de scénarii
 - Écriture de scénarii automatique ou assistée
 - Images de synthèse (effets spéciaux, animation...)
 - Repérage, programmation et budgétisation automatisés (impact à évaluer)
- Diffusion de contenus
 - Algorithmes de recommandation
 - Publicité ciblée
 - Contrôle automatique des contenus (respect des réglementations) / censure (étude de référence « Entering the new paradigm of artificial intelligence and series » commandée par la DG2 et Eurimages)
- Accès et participation à la vie publique/culturelle
- Liberté d'expression (y compris liberté d'expression artistique)
- Accès à de l'information impartiale ?

GRECO

- Lutte contre la corruption
- Responsabilité pénale liée à l'utilisation de véhicules automatisés
- Article 8 : droit au respect de la vie privée et familiale

CSE

Tous les domaines des droits sociaux, sécurité sociale, cohésion sociale, etc. incluant, sans s'y limiter :

- de nombreux aspects de l'emploi (y compris, mais sans s'y restreindre, le contrôle et le suivi, la sélection des candidatures et le travail dans l'économie de plateformes, etc.)
- différents aspects de la santé (droit de jouir du meilleur état de santé possible)
- idem pour l'éducation
- la protection, l'intégration et la participation sociales
- la lutte contre la discrimination
- le logement et la protection contre l'exclusion sociale

par exemple :

- justice (à la fois administration de la justice, justice pénale et milieu carcéral)
- traite des êtres humains (travail forcé et exploitation, etc.)
- migrations et réfugiés
- égalité femmes-hommes et violence à l'égard des femmes
- enfants et jeunes, éducation
- bioéthique
- non-discrimination, Roms et Gens du voyage, orientation sexuelle et identité de genre
- politique en matière de drogues
- participation et culture
- sport

Annexe 3. Principes

| Principes directeurs et valeurs juridiques | Principes manquants |
|--|--|
| Biomédecine | |
| Primauté de l'être humain Respect de la vie privée et confidentialité Consentement éclairé Autonomie Non-discrimination Non-malfaisance / bienfaisance Obligation de rendre compte Transparence et accès équitable Débat public | Principe de précaution Contrôle/surveillance humains Explicabilité Responsabilité quant aux décisions reposant sur l'IA Égalité/équité hommes-femmes |
| Anti-discrimination | |
| Non-discrimination et égalité Diversité et inclusion Intersectionnalité Droit à un recours effectif Droit à un procès équitable Droit au respect de la vie privée Présomption d'innocence et charge de la preuve Transparence Impartialité Équité Contrôle/surveillance humains Accès aux compétences numériques | Explicabilité des systèmes IA Approche inclusive dans la conception, le développement et le déploiement des systèmes IA |
| Cybercriminalité et preuves électroniques | |
| Conduites à ériger en infraction pénale Données à sécuriser dans certaines enquêtes pénales pour être utilisées comme preuves. Pouvoir effectif de sécuriser des preuves électroniques limité par les conditions et garanties de l'État de droit. | Problème des preuves dans le nuage et de la compétence territoriale en matière pénale (question abordée dans le deuxième protocole additionnel à la Convention de Budapest). |

| | |
|--|---|
| Secteur de la justice | |
| <p>Non-discrimination</p> <p>Qualité et sécurité des données</p> <p>Transparence</p> <p>Impartialité Équité</p> <p>Liberté de choix/indépendance des juges (processus décisionnel) Contrôle/surveillance humains</p> <p>Garanties du droit d'accès à un juge</p> <p>Garanties du droit à un procès équitable</p> | <p>Principe de précaution pour les applications ne respectant pas les exigences fondamentales de transparence</p> |
| Congrès | |
| <p>Transparence</p> <p>Contrôle humain (surveillance) Impartialité</p> <p>Droit au respect de la vie privée Sécurité des données Cybersécurité</p> <p>Non-discrimination Villes inclusives</p> | <p>Démocratie et participation – <i>deep fakes</i> (hypertrucage), micro-ciblage et propagande dans le cadre des processus électoraux</p> |
| <p>Pérennité financière Contrôle de la sécurité</p> <p>Organisation rationnelle des services</p> <p>Éducation au numérique</p> | |
| Démocratie et participation | |
| <p>Droit à des élections libres Liberté d'expression</p> <p>Droit d'accès des personnes à Internet Droit au respect de la vie privée Protection des données</p> <p>Égalité des chances pour les partis et les candidats</p> <p>Exigence de neutralité des autorités de l'État vis-à-vis des campagnes électorales, de la couverture médiatique et du financement public des partis politiques et des campagnes</p> <p>Exigence d'un accès minimum à des médias audiovisuels privés pour la campagne et les annonces électorales, pour tous les participants aux élections</p> <p>Transparence du financement des campagnes</p> <p>Prévention de l'influence injustifiée sur les décisions politiques par des dons financiers</p> | <p>Équilibre entre des droits parfois concurrents comme :</p> <ul style="list-style-type: none"> - le droit à des élections libres / la liberté d'expression - le droit d'accès à l'information y compris sur Internet / le droit au respect de la vie privée et à la protection des données <p>Normes qui seraient applicables et appropriées pour les annonces/campagnes numériques, par exemple en ce qui concerne</p> <ul style="list-style-type: none"> - l'égalité des chances pour les partis et les candidats - les campagnes électorales et le contrôle/la transparence de leur financement - la couverture médiatique équitable, le pluralisme des médias - la responsabilité des intermédiaires d'Internet en matière de transparence et d'accès aux données pour améliorer la transparence des dépenses, notamment en ce qui concerne les publicités politiques - la neutralité du réseau la protection des données |

| | |
|--|--|
| <p>Couverture responsable, correcte et équitable des campagnes électorales par les médias ; droit de réponse, modalités de diffusion des sondages d'opinion, exigences de transparence quant aux contenus publicitaires payants ; pluralisme des médias Neutralité du réseau</p> <p>Protection des individus à l'égard de la collecte et du traitement de données à caractère personnel sur les autoroutes de l'information</p> <p>Non-discrimination</p> <p>Qualité et sécurité des données Transparence</p> <p>Impartialité Équité</p> <p>Liberté de choix/indépendance des juges (processus décisionnel) Contrôle/surveillance humains</p> <p>Garanties du droit d'accès à un juge Garanties du droit à un procès équitable</p> | <p>-</p> |
| <p>Liberté d'expression</p> | |
| <p>Autonomie individuelle Égalité</p> <p>Sécurité démocratique</p> <p>Transparence et obligation de rendre compte Indépendance des médias</p> <p>Diversité et pluralisme</p> | <p>Principe de précaution pour les applications ne respectant pas les exigences fondamentales de transparence</p> |
| <p>Élections</p> | |
| <ul style="list-style-type: none"> • élections libres et équitables • liberté de choix/d'opinion/d'expression • suffrage universel • suffrage égal | <ul style="list-style-type: none"> • principe de l'utilisation de systèmes IA dans les processus électoraux (en particulier systèmes de vote électronique, etc.) |
| <ul style="list-style-type: none"> • suffrage libre • suffrage secret • suffrage direct • fréquence des élections • transparence du processus électoral • processus électoral inclusif • participation et représentation équilibrées des femmes et des hommes dans la prise de décisions publiques | <ul style="list-style-type: none"> • possibilités d'améliorer le caractère inclusif des processus électoraux grâce à l'IA (comme outil pour les organes de gestion des élections et les commissions électorales ou comme assistant pour les électeurs). |

| | |
|---|---|
| Démocratie (hors questions relatives aux élections et au cycle électoral) | |
| <p>Transparence Impartialité Équité</p> <p>Liberté de choix Liberté d'expression</p> <p>Liberté de réunion et d'association Accès à l'information Contrôle/surveillance humains Diversité</p> <p>Égalité</p> <p>Non-discrimination</p> <p>Qualité et sécurité des données Protection des données Indépendance</p> | <ul style="list-style-type: none"> - rôle des intermédiaires - éducation aux technologies et au numérique - question de la propriété des médias - contrôle démocratique - données ouvertes et gouvernement ouvert - évaluation du risque |
| Bonne gouvernance | |
| <ul style="list-style-type: none"> - non-discrimination - qualité et sécurité des données - impartialité - équité - participation, représentation, déroulement équitable des élections - capacité d'adaptation - organisation rationnelle et efficacité - ouverture et transparence - état de droit - conduite éthique - compétence et capacité - innovation et ouverture au changement - durabilité (2) - gestion financière saine - droits de l'homme, diversité culturelle et cohésion sociale - obligation de rendre compte - mécanismes de réparation - accès à un recours - indépendance | <ul style="list-style-type: none"> - contrôle démocratique - accès à des mécanismes de recours et de réparation dans le cas d'une prise de décisions automatisée, reposant sur des algorithmes, par les agents de la fonction publique - rôle des intermédiaires - éducation et compétences en matière de technologies - question de la propriété réelle des données - données ouvertes et gouvernement ouvert - responsabilité civile et pénale - évaluations du risque et gestion du risque |
| Égalité femmes-hommes, y compris les questions relatives à la violence à l'égard des Femmes | |
| <p>Égalité et non-discrimination</p> <p>Intégrité / élimination de la violence (à l'égard des femmes)</p> <p>Égalité d'accès à la justice</p> <p>Garanties du droit à un procès équitable et à une réparation</p> | <p>Données non biaisées</p> <p>Caractère inclusif (pour les femmes) de l'IA en tant que secteur d'activité</p> <p>Respect des droits du travail et des droits sociaux dans l'IA en tant que secteur d'emploi</p> <p>Qualité et sécurité des données Transparence et explicabilité Obligation de rendre compte Impartialité</p> <p>Équité</p> <p>Contrôle/surveillance humains</p> |

| | |
|--|---|
| | <p>Éducation au numérique et résorption de la fracture numérique existante (entre femmes et hommes), essentielle pour le droit à réparation : si les citoyens et consommateurs ne comprennent pas l'IA, ils ne seront pas en mesure de faire valoir leurs droits</p> <p>Principe de précaution pour les applications ne respectant pas les exigences fondamentales de transparence</p> <p>Des principes éthiques comme la non- malfaisance ne sont pas respectés car certains logiciels espions sont mis au point et commercialisés dans le seul but de « savoir ce que fait son conjoint ».</p> |
| Culture, créativité et patrimoine | |
| <p>Non-discrimination</p> <p>Accès, liberté d'association, droit de participer à la vie culturelle et de créer et d'apprendre (pacte)</p> <p>Liberté d'expression</p> <p>Accès à de l'information impartiale</p> | <p>Principe de précaution pour les applications ne respectant pas les exigences fondamentales de transparence</p> <p>Nécessité de trouver des paradigmes culturels et des techniques pour traiter la question de l'autonomisation (n'existent que pour l'automatisation)</p> <p>« Éviter une centralisation encore plus forte du savoir et du pouvoir entre les mains de ceux qui les ont déjà et continuent de tenir à l'écart ceux qui en sont dépourvus »</p> <p>(M. Whitaker)</p> <p>Nécessité d'insister sur les règles et droits en matière d'accès aux biens communs et de participation à la vie publique (pratiques centrées sur les citoyens)</p> |
| Non-discrimination – Impartialité (Protection des minorités nationales) | |
| Non-discrimination – Impartialité Protection des minorités et de leurs expressions culturelles (langues / diversité linguistique, patrimoine culturel) ²²⁵ | Propriété et éventuels biais des informations entrées dans les applications d'apprentissage par IA |
| Promotion/protection de l'identité, de la diversité et de la coopération européennes Accès et participation au patrimoine culturel ; protection du patrimoine culturel | Protection de la créativité humaine (et sa nature spécifique) |
| Contrôle/surveillance humains du processus créatif, transparence Protection et promotion de la diversité culturelle Créer les conditions permettant aux cultures de s'épanouir et interagir librement Reconnaître la nature spécifique des activités, biens et services culturels en tant que porteurs d'identité, de valeurs et sens | Gestion des adresses IP et du droit d'auteur Protection de la créativité humaine (et sa nature spécifique) |

²²⁵ Une machine permettant de recréer des langues par IA pourrait combler les lacunes et contribuer à créer une archive globale des langues vivantes, un « Louvre des langues ».

| | |
|---|--|
| Diversité culturelle Coopération culturelle en Europe et au-delà Disponibilité des œuvres Non-discrimination Protection des données | Visibilité des œuvres Transparence de la prise de décision (de concevoir et produire/censurer/recommander une œuvre) |
| Liberté d'expression et de création Surveillance/contrôle humains, transparence | Propriété des adresses IP, questions de droits d'auteur et de droits moraux |
| Non-discrimination Accès, liberté d'association, droit de participer à la vie culturelle et de créer et d'apprendre (Pacte) Liberté d'expression Accès à de l'information impartiale | Principe de précaution pour les applications ne respectant pas les exigences fondamentales de transparence Nécessité de trouver des paradigmes culturels et des techniques pour traiter la question de l'autonomisation (n'existent que pour l'automatisation) « Éviter une centralisation encore plus forte du savoir et du pouvoir entre les mains de ceux qui les ont déjà et continuent de tenir à l'écart ceux qui en sont dépourvus » (M. Whitaker) Nécessité d'insister sur les règles et droits en matière d'accès aux biens communs et de participation à la vie publique (pratiques centrées sur les citoyens) |
| GRECO | |
| Principes directeurs pour la lutte contre la corruption | Rien de spécifique sur : les applications d'IA pour prévenir la corruption ; la nécessité de s'assurer que les algorithmes ne sont pas corrompus |
| | Travail en cours du CDPC sur la responsabilité pénale liée à l'utilisation de véhicules automatisés |
| Article 8 : droit au respect de la vie privée et familiale | Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH) |
| CSE | |
| Plusieurs grands principes peuvent être tirés de la Charte et des activités de suivi menées dans le cadre de la Charte sur la transparence et la participation à la prise de décisions. | Une prise de décision automatisée, assistée par ordinateur ou reposant sur l'IA nécessiterait : <ul style="list-style-type: none"> - un contrôle humain obligatoire pour éviter ou limiter les erreurs dans la gestion, l'attribution ou le retrait de prestations, aides et autres avantages, pouvant accentuer les inégalités et les privations de droits. - des dispositifs efficaces pour protéger les personnes vulnérables contre le dénuement, l'extrême pauvreté ou la perte de logement et contre les blessures graves ou préjudices irréparables découlant de la mise en œuvre de décisions assistées par ordinateur ou reposant sur l'IA dans le domaine des services sociaux. - une approche proactive pour que les personnes concernées par des décisions assistées par ordinateur ou reposant sur l'IA dans le domaine des services sociaux, et en particulier les personnes en situation d'extrême pauvreté ou de vulnérabilité, puissent effectivement faire valoir leurs droits et obtenir réparation. |

Annexe 4. Protection des données

Instruments contraignants et non contraignants dans le domaine de la protection des données

| | |
|---|---|
| <p>Convention 108+</p> <p>Lignes directrices sur les incidences de l'intelligence artificielle sur la protection des données²²⁶</p> <p>Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (Convention 108), 2019.</p> | <p>Contrôle humain</p> <p>I.6 Les applications de l'IA devraient permettre aux personnes concernées d'exercer un contrôle significatif sur le traitement des données et leurs effets connexes tant au niveau individuel que sur la société.</p> <p>Conception centrée sur les valeurs</p> <p>II.1. Les développeurs, fabricants et prestataires de service en IA devraient adopter une approche de conception des produits et services centrée sur les valeurs, conformément à la Convention 108+, notamment son article 10.2, et aux autres instruments pertinents du Conseil de l'Europe.</p> <p>Approche de précaution</p> <p>II.2 Les développeurs, fabricants et prestataires de service en IA devraient évaluer les éventuelles conséquences négatives des applications d'IA sur les droits de l'homme et libertés fondamentales des personnes concernées et au regard de ces conséquences, adopter une approche de précaution basée sur des mesures de prévention et de réduction des risques appropriées.</p> <p>Approche fondée sur les droits de l'homme dès la conception et détection des biais</p> <p>II.3 Les développeurs, fabricants et prestataires de service en IA devraient, à tous les stades du traitement des données, y compris lors de la collecte, adopter une approche des droits de l'homme dès la conception (by-design) et éviter tout biais potentiel, y compris les biais non intentionnels ou cachés, ainsi que les risques de discrimination ou d'autres effets négatifs sur les droits de l'homme et libertés fondamentales des personnes concernées.</p> <p>Qualité et minimisation des données</p> <p>II.4 Les développeurs en IA devraient évaluer de manière critique la qualité, la nature, l'origine et la quantité de données à caractère personnel utilisées, en réduisant les données inutiles, redondantes ou marginales lors des phases de conception et d'apprentissage, puis en vérifiant l'exactitude du modèle lorsqu'il est alimenté par de nouvelles données. Le recours à des données synthétiques pourrait être considéré comme une solution possible pour minimiser la quantité de données personnelles traitées par des applications de l'IA.</p> <p>Risque de décontextualisation</p> <p>II.5 Les risques d'impact négatif sur les personnes et la société inhérents aux données décontextualisées et aux modèles algorithmiques décontextualisés devraient être dûment pris en compte lors du développement et de l'utilisation d'applications de l'IA.</p> <p>Comités d'experts indépendants</p> <p>Les développeurs, fabricants et prestataires de service en IA sont encouragés à recourir à des comités d'experts indépendants issus de différents domaines ainsi qu'à des institutions universitaires indépendantes qui peuvent contribuer à concevoir des applications de l'IA fondées sur les droits de l'homme et orientées de façon éthique et sociale, ainsi qu'à détecter des biais potentiels. Le rôle de ces comités peut être particulièrement important dans les domaines où la transparence et la mobilisation des parties prenantes peuvent être plus difficiles en raison d'intérêts et de droits concurrents, par exemple dans les domaines de la justice prédictive, de la prévention du crime et de la détection des infractions.</p> |
|---|---|

²²⁶ Voir également T-PD(2019)01, Lignes directrices sur l'intelligence artificielle et la protection des données ; T-PD(2017)1, Lignes directrices sur la protection des individus à l'égard du traitement des données à caractère personnel à l'ère des mégadonnées.

III.7 Des mécanismes appropriés visant à garantir l'indépendance des comités d'experts mentionnés à la section II.6 devraient être mis en place.

Participation et contrôle démocratique du développement de l'IA

II.7 Des démarches participatives d'évaluation des risques, reposant sur l'engagement actif des personnes et groupes potentiellement affectés par les applications de l'IA, devraient être encouragées.

III. 8. Les personnes, les groupes et les autres parties prenantes devraient être informés et impliqués de façon active dans le débat relatif au rôle que l'IA devrait jouer dans le modelage des dynamiques sociales, et dans les processus décisionnels les affectant.

Contrôle humain

II.8 Tous les produits et services de l'IA devraient être conçus de manière à garantir le droit des personnes à ne pas être soumises à des décisions qui les affectent de manière significative prises uniquement sur le fondement d'un traitement automatisé de données, sans que leur point de vue soit pris en compte.

Liberté de choix

II.9 Afin d'accroître la confiance des utilisateurs, les développeurs, fabricants et prestataires de services en IA sont encouragés à concevoir leurs produits et services de manière à préserver la liberté de choix de l'utilisateur concernant l'usage de l'IA en proposant des alternatives réalistes aux applications de l'IA.

Vigilance algorithmique

II.10 Les développeurs, fabricants et prestataires de service en IA devraient adopter des formes de vigilance algorithmique qui fassent la promotion de la responsabilité de toutes les parties prenantes et ce tout au long du cycle de vie des applications afin d'assurer leur conformité avec les principes et la législation relatifs à la protection des données personnelles et aux droits de l'homme.

Transparence et extensibilité

II.11 Les personnes concernées devraient être informées si elles interagissent avec des applications de l'IA et ont le droit de connaître du raisonnement qui sous-tend les opérations de traitement des données qui les concernent. Ceci devrait inclure les conséquences de ce raisonnement.

Droit d'opposition

II.12 Le droit d'opposition devrait être garanti par rapport au traitement basé sur des technologies qui influencent les opinions et le développement personnel des individus.

Responsabilité et vigilance

III.2 Sous réserve des secrets protégés par la loi, les procédures de passation des marchés publics devraient imposer aux développeurs, fabricants et prestataires de service en IA des devoirs spécifiques de transparence et d'évaluation préalable de l'impact des traitements de données personnelles sur les droits de l'homme et libertés fondamentales des systèmes d'IA, ainsi que de vigilance à l'égard des effets défavorables potentiels et des conséquences des applications d'IA (ci-après dénommée « vigilance algorithmique »).

Liberté des décideurs humains

III.4. Une dépendance excessive des solutions fournies par les applications de l'IA, de même que la crainte de contester des décisions suggérées par des applications de l'IA risquent d'altérer l'autonomie de l'intervention humaine dans la prise de décision. Le rôle de l'intervention humaine dans le processus décisionnel et la liberté des décideurs humains de ne pas suivre les résultats de recommandations fondées sur l'utilisation de l'IA devraient en conséquence être préservés.

| | |
|--|--|
| | <p>Évaluation préalable</p> <p>III.5. Les développeurs, fabricants et prestataires de service en IA devraient consulter les autorités de contrôle dès lors que les applications de l'IA peuvent avoir un impact significatif sur les droits de l'homme et les libertés fondamentales des personnes concernées.</p> <p>Coopération</p> <p>III.6. La coopération devrait être encouragée entre autorités de contrôle de la protection des données et d'autres instances ayant des compétences liées à l'IA, telles que les autorités de régulation en matière de protection des consommateurs ; concurrence ; anti- discrimination ; médias et autorités sectorielles.</p> <p>Éducation, éducation au numérique et formation professionnelle</p> <p>III.9. Les décideurs devraient affecter des ressources à l'éducation au numérique afin de renforcer la sensibilisation des personnes concernées et la compréhension des applications de l'IA et de leurs effets. Ils devraient également encourager la formation professionnelle des développeurs en IA pour qu'ils soient sensibilisés aux effets potentiels de l'IA sur les personnes et la société. Ils devraient soutenir la recherche en matière d'IA orientée vers les droits de l'homme.</p> |
| <p>Recommandation Rec(2019)2 du Comité des Ministres du Conseil de l'Europe aux États membres en matière de protection des données relatives à la santé.</p> | <p>Le traitement des données relatives à la santé devrait toujours servir la personne concernée ou conduire à améliorer la qualité et l'efficacité des soins de santé, ainsi que les systèmes de santé lorsque cela est possible, tout en respectant les droits fondamentaux de la personne.</p> <p>Interopérabilité</p> <p>1. [...] Elle souligne à cette fin l'importance du développement de systèmes d'information sécurisés interopérables.</p> <p>Normes professionnelles</p> <p>4.4 Les responsables du traitement et leurs sous-traitants qui ne sont pas des professionnels de santé ne devraient traiter des données relatives à la santé que dans le respect de règles de confidentialité et des mesures de sécurité garantissant un niveau de protection équivalant à celui qui incombe aux professionnels de santé.</p> <p>Retrait du consentement</p> <p>5.b Les données relatives à la santé peuvent être traitées dès lors que la personne concernée a donné son consentement, sauf dans les cas où le droit prévoit qu'une interdiction de traiter les données de santé ne peut être levée par le seul consentement de la personne concernée. Lorsque le consentement de la personne concernée au traitement de données relatives à la santé est requis, conformément au droit, celui-ci devrait être libre, spécifique, éclairé et explicite. La personne concernée doit être informée de son droit de retirer son consentement à tout moment et du fait qu'un tel retrait ne compromet pas la licéité du traitement fondé sur le consentement effectué avant ce retrait. Il doit être aussi simple de retirer son consentement que de le donner.</p> <p>Droit de ne pas savoir</p> <p>7.6 La personne concernée a le droit de connaître toute information relative à ses données génétiques, sous réserve des dispositions des principes 11.8 et 12.7. Toutefois, pour des raisons qui lui appartiennent, la personne concernée peut souhaiter ne pas connaître certains éléments relatifs à sa santé et toute personne devrait être informée, préalablement à la réalisation d'une analyse, de la possibilité dont elle dispose de ne pas être informée de résultats, y compris de découvertes inattendues. Le souhait de ne pas savoir peut, dans des circonstances exceptionnelles, faire l'objet de restrictions prévues par la loi, notamment dans l'intérêt de la personne concernée ou au regard de l'obligation de soigner qui incombe aux médecins.</p> |

| | |
|---|---|
| | <p>Transparence</p> <p>11.3. L'information doit, le cas échéant, afin de garantir la loyauté et la transparence du traitement, également porter sur :</p> <ul style="list-style-type: none"> - l'existence de décisions automatisées, y compris le profilage qui n'est acceptable que si la loi le permet et sous réserve de garanties appropriées. <p>Interopérabilité</p> <p>14.1. L'interopérabilité peut permettre de répondre à des impératifs relevant du domaine de la santé et peut apporter des moyens techniques qui facilitent la mise à jour, qui évitent la duplication de données identiques dans de multiples bases de données et qui contribuent à la portabilité.</p> <p>14.2. Il est cependant nécessaire que l'interopérabilité soit mise en œuvre conformément aux principes contenus dans cette recommandation, notamment les principes de licéité, de nécessité et de proportionnalité, et que des mesures de sauvegarde de la protection des données à caractère personnel soient prises lorsque des systèmes interopérables sont utilisés.</p> <p>14.3. Des référentiels fondés sur des normes internationales et offrant un cadre technique qui facilite l'interopérabilité devraient assurer qu'un haut niveau de sécurité est garanti tout en offrant une telle interopérabilité. Leur mise en œuvre peut être suivie au moyen de schémas de certification.</p> <p>Intégrité de la recherche scientifique</p> <p>15.10. Lorsqu'une personne décide de se retirer d'une recherche scientifique, ses données relatives à la santé traitées dans le cadre de cette recherche devraient être détruites ou anonymisées de manière à ne pas compromettre la validité scientifique de la recherche et la personne concernée devrait en être informée.</p> |
| <p>Recommandation CM/Rec(2016)8 du Comité des Ministres aux États membres sur le traitement de données à caractère personnel relatives à la santé à des fins d'assurance, y compris les données résultant de tests génétiques</p> | <p>8. Le traitement à des fins d'assurance de données à caractère personnel relatives à la santé obtenues dans le cadre d'une recherche impliquant l'assuré(e) ne devrait pas être permis.</p> |

| | |
|--|---|
| <p>Recommandation Rec(2010)13 du Comité des Ministres du Conseil de l'Europe aux États membres sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel dans le cadre du profilage</p> | <p>Risque de réidentification</p> <p>8.5. Des mesures appropriées devraient être mises en place pour éviter que des résultats statistiques anonymes et agrégés utilisés dans le cadre du profilage ne puissent déboucher sur une réidentification des personnes concernées²²⁷</p> |
| <p>UNESCO 2019.</p> <p>Étude préliminaire concernant un éventuel instrument normatif sur l'éthique de l'intelligence artificielle.</p> | <p>[Approche fondée sur les principes]</p> <ul style="list-style-type: none"> • Diversité, inclusion et pluralisme (y compris promotion d'une approche multilingue) • Autonomie • Explicabilité • Transparence • Sensibilisation et éducation • Responsabilité • Obligation de rendre compte • Démocratie (l'IA devrait être élaborée, mise en œuvre et utilisée conformément aux principes démocratiques) • Bonne gouvernance («Les gouvernements devraient présenter des rapports réguliers sur leur utilisation de l'IA dans la police, le renseignement et la sécurité») • Durabilité • Contrôle humain • Liberté d'expression (y compris accès universel à l'information ; journalisme de qualité ; médias libres, indépendants et pluralistes ; éviter la diffusion de fausses informations) |
| <p>OCDE. 2019.</p> <p>Recommandation du Conseil sur l'intelligence artificielle</p> | <p>[Approche fondée sur les principes]</p> <ul style="list-style-type: none"> • Valeurs centrées sur l'humain et équité • Transparence et explicabilité (informer les parties prenantes de leurs interactions avec les systèmes IA ; permettre aux personnes concernées d'appréhender le résultat de l'IA ; permettre à ceux qui subissent les effets néfastes d'un système d'IA de contester son résultat sur la base d'informations claires et facilement compréhensibles sur les facteurs et la logique ayant servi à la formulation de prévisions, recommandations ou décisions) • Robustesse, sûreté et sécurité (ne font pas peser un risque de sécurité démesuré ; traçabilité, notamment pour ce qui est des ensembles de données, des processus et des décisions prises au cours du cycle de vie des systèmes d'IA ; approche de gestion des risques à chaque phase du cycle de vie des systèmes d'IA) |

²²⁷ Voir également le rapport explicatif de la Convention 108+ : par. 19 et 20 (« Les données ne peuvent être considérées comme anonymes que lorsque la réidentification de la personne concernée est impossible ou nécessiterait des délais, efforts ou ressources déraisonnables au vu des technologies disponibles au moment du traitement et de l'évolution de celles-ci. [...] Lorsque des données sont rendues anonymes, des moyens appropriés doivent être mis en place pour empêcher toute réidentification des personnes concernées ; en particulier, tous les moyens techniques doivent être mis en œuvre pour garantir que la personne n'est pas ou plus identifiable. Étant donné la rapidité des évolutions techniques, ces moyens techniques devraient être réévalués régulièrement.

| | |
|---|--|
| <p>40^e Conférence internationale des commissaires à la protection des données et de la vie privée (ICDPPC). 2018. Déclaration sur l'éthique et la protection des données dans le secteur de l'intelligence artificielle [ICDPPC]</p> | <p>Obligation de rendre compte [Approche fondée sur les principes]</p> <ul style="list-style-type: none"> • Attention et vigilance permanentes (« mettre en place pour tous les acteurs concernés des processus de gouvernance dont on peut apporter la preuve, par exemple en s'appuyant sur des tiers de confiance ou en créant des comités d'éthique indépendants ») • Transparence et intelligibilité (IA explicable, transparence des algorithmes et vérifiabilité des systèmes, connaissance des interactions avec les systèmes IA ; informations adéquates sur les objectifs et les effets des systèmes IA, contrôle humain global) • Évaluation du risque et application des principes de protection de la vie privée par défaut (<i>privacy by default</i>) et de protection intégrée de la vie privée (<i>privacy by design</i>) (« évaluer et décrire les impacts attendus sur les personnes et la société au début d'un projet d'intelligence artificielle et au cours des développements pertinents durant tout son cycle de vie ») • Participation publique • Réduction et atténuation des préjugés et des discriminations illicites |
|---|--|

TITRE II. PERSPECTIVES NATIONALES DE LA REGULATION DE L'IA

CHAPITRE I. Exploiter l'innovation : Perspectives israéliennes sur l'éthique et la gouvernance de l'IA

Prof. Isaac Ben-Israel²²⁸, Prof. Eviatar Matania²²⁹, Leehe Friedman²³⁰

I. Résumé

Cet article présente l'état actuel du développement de la politique israélienne, en ce qui concerne les opportunités et les défis présentés par l'intelligence artificielle (IA) en relation avec les droits de l'Homme et l'éthique. Il est basé, en grande partie, sur le rapport de l'Initiative nationale israélienne pour des systèmes intelligents sécurisés, qui a été récemment soumis au gouvernement israélien. La présente étude décrit l'approche unique adoptée par Israël pour tenter de tirer parti des possibilités offertes par l'IA tout en relevant les défis qu'elle pose. Cet article décrit comment l'approche de gouvernance d'Israël cherche à équilibrer la nécessité de permettre l'innovation, tant dans le secteur public que dans le secteur privé, en accord les impératifs moraux et des droits de l'Homme qui sont omniprésents dans les développements de l'IA.

Les décideurs politiques israéliens ont tendance à considérer les développements de l'IA non seulement comme un facteur de perturbation mais aussi comme un facteur de transformation : La technologie de l'IA est considérée comme essentielle pour le bien-être, l'économie et la sécurité des citoyens israéliens. Partant de ce constat, la priorité pour Israël a été d'établir un écosystème d'IA sécurisé, holistique et durable, dirigé par le secteur privé mais auquel participent le gouvernement, l'industrie privée et le monde universitaire, et qui soutient l'utilisation de l'IA à tous les niveaux. En gardant cela à l'esprit, cet article souligne les principaux défis qui ont été identifiés par les décideurs politiques, en Israël et à l'étranger, dans les domaines des droits de l'Homme, de la démocratie et de l'État de droit - sécurité, vie privée, autonomie, droits civils et politiques, sécurité, équité - y compris la concurrence loyale - et responsabilité. L'approche proposée par Israël en réponse à ces défis, selon le Rapport de l'initiative nationale, est novatrice. Si elle est fermement ancrée dans les principes de gouvernance établis et les meilleures pratiques de la politique internationale en matière d'IA, elle représente néanmoins une approche de gouvernance novatrice, axée sur une réglementation équilibrée pour favoriser l'innovation. À cet effet, elle propose des outils politiques originaux, tels qu'un outil d'évaluation des risques qui correspond à différentes approches

²²⁸ Le professeur Isaac Ben-Israel est un général à la retraite qui a rejoint l'université de Tel-Aviv en 2002, où il est actuellement directeur du Centre interdisciplinaire de recherche cybernétique de Blavatnik et de l'atelier Yuval Ne'eman pour la science, la technologie et la sécurité. Outre ses activités à l'Université, il est le président de l'Agence spatiale israélienne. En 2010, il a été nommé par le Premier ministre pour diriger un groupe de travail qui a mené la cyber-révolution en Israël. Au cours des deux dernières années, le professeur Ben-Israel a coprésidé l'Initiative nationale israélienne pour des systèmes intelligents sécurisés (IA) afin de recommander au Premier ministre et au gouvernement un plan national visant à promouvoir Israël en tant que puissance mondiale dans le domaine de l'IA.

²²⁹ Eviatar Matania est professeur à l'École des sciences politiques, du gouvernement et des affaires internationales de l'Université de Tel-Aviv, où il dirige le programme de maîtrise en études de sécurité et le programme de maîtrise en cyberpolitique et gouvernement. Matania est également professeur adjoint à l'école de gouvernement Blavatnik d'Oxford, où il dirige le module sur le cybernétique. Matania a été le chef fondateur et l'ancien directeur général de la Direction nationale israélienne de la cybernétique (INCD) au bureau du Premier ministre israélien, où il relevait directement du Premier ministre, et était responsable de la stratégie cyber globale d'Israël, de la politique et de sa mise en œuvre pour défendre le secteur civil israélien. Au cours des deux dernières années, le professeur Matania a coprésidé l'Initiative nationale israélienne pour des systèmes intelligents sécurisés (IA) afin de recommander au Premier ministre et au gouvernement un plan national visant à promouvoir Israël en tant que puissance mondiale dans le domaine de l'IA.

²³⁰ Leehe Friedman est professeure adjointe et directrice de la filière " *Honors Track in Strategy & Decision Making* " à la *Lauder School of Government, Diplomacy and Strategy* de l'IDC Herzliya. Au cours des deux dernières années, Mme Friedman a été la coordinatrice de l'Initiative nationale israélienne pour les systèmes intelligents sécurisés (IA), chargée de recommander au Premier ministre et au gouvernement un plan national visant à promouvoir Israël en tant que puissance mondiale dans le domaine de l'IA.

réglementaires basées sur le niveau de risque associé à une activité particulière, et une carte de fréquence dynamique qui permet de localiser les domaines complexes en termes d'application des valeurs éthiques au développement d'un système d'IA particulier.

Elle signale également la nécessité de s'engager auprès des pays et des forums internationaux, de tirer des enseignements et de contribuer aux processus internationaux impliquant des questions d'IA, d'éthique, de droit et de gouvernance.

Cet article n'est pas destiné à être un document officiel du gouvernement et ne reflète pas nécessairement la politique du gouvernement israélien. Ses auteurs écrivent à titre personnel, bien qu'ils aient reçu des informations pertinentes de divers responsables gouvernementaux.

Remerciements

Les auteurs tiennent à remercier Cedric Sabbah, directeur de la cybersécurité internationale et du droit des technologies de l'information au bureau du procureur général adjoint (droit international), au sein du Ministère israélien de la justice, le professeur Karin Nahon, chef du groupe de travail sur l'éthique et la réglementation de l'initiative nationale pour les systèmes intelligents sécurisés, le Dr Roy Schöndorf, procureur général adjoint (droit international), et le Dr Ben Moshe, directeur du département des grandes données au Ministère de la santé, pour leur précieuse contribution, leurs conseils professionnels et leurs commentaires éclairés.

II. Introduction

Israël perçoit l'IA comme une technologie émergente de base et comme "une *infrastructure d'infrastructures, une infrastructure qui est essentielle pour l'avenir de l'État d'Israël - pour sa sécurité, son économie et le bien-être de sa population*".²³¹ Les applications de l'IA, en raison de leur potentiel à améliorer la disponibilité, la fiabilité et l'efficacité des infrastructures, des services et des systèmes nationaux, à moindre coût pour l'État et ses citoyens, jouent un rôle clé dans la capacité d'Israël à relever certains de ses défis nationaux au XXI^e siècle.

À l'avenir, l'IA est susceptible de transformer fondamentalement tous les aspects de la vie privée et publique. Afin d'exploiter le potentiel positif des technologies d'IA, Israël s'efforce d'établir un écosystème holistique et durable qui inclut le gouvernement, l'industrie privée et le monde universitaire. Une boucle de rétroaction impliquant ces trois secteurs serait bénéfique pour la société dans son ensemble : (a) en augmentant l'utilisation des applications de l'IA ; (b) en améliorant le travail du gouvernement et les services qu'il fournit ; (c) en encourageant l'économie et l'innovation de nouveaux développements techno-scientifiques qui, à leur tour, augmenteraient à nouveau la demande de nouvelles applications de l'IA.

En conséquence, l'approche israélienne à l'égard de l'IA repose sur deux efforts complémentaires :

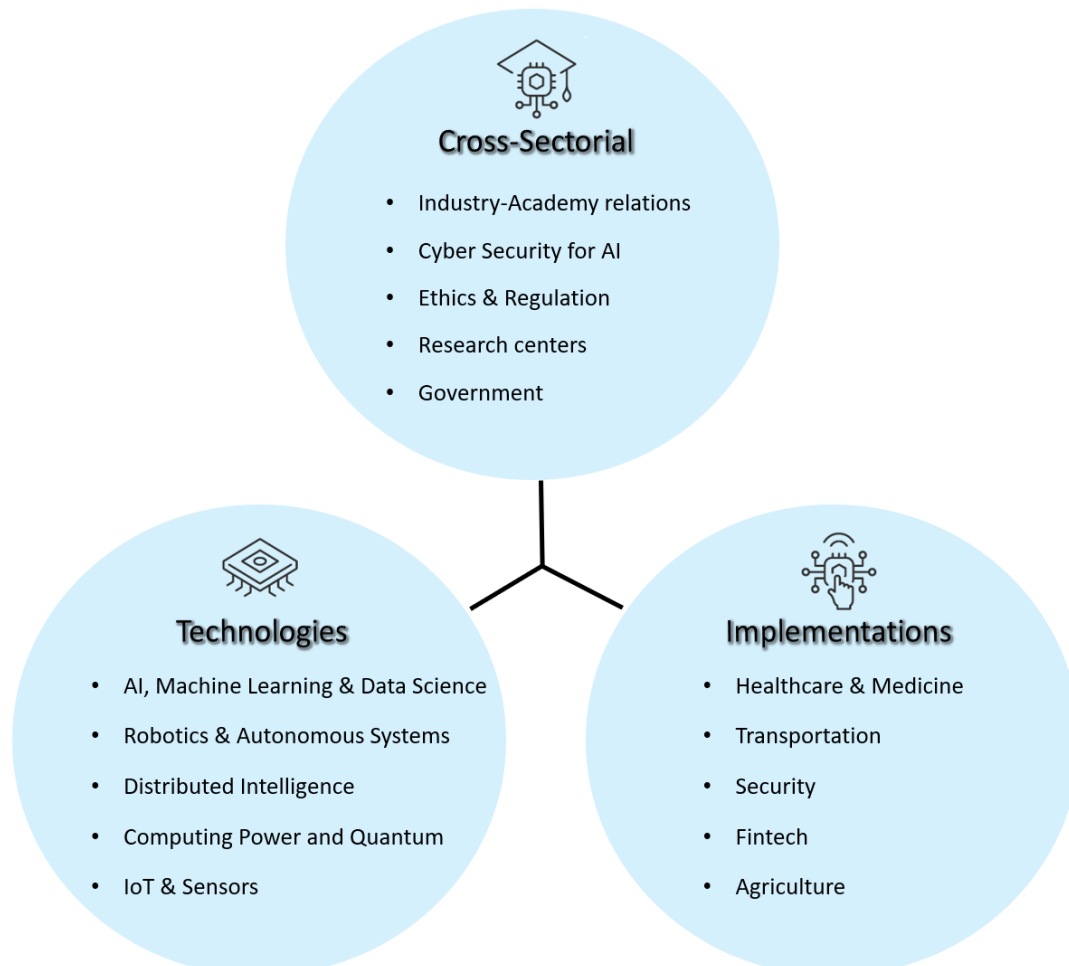
- I. Promouvoir une utilisation large et équitable des applications de l'IA dans les secteurs public et privé.
- II. Encourager une industrie technologique de pointe qui développerait des solutions basées sur l'IA pour les défis émergents en Israël et dans le monde entier.

L'initiative nationale israélienne pour des systèmes intelligents sécurisés

Afin de comprendre le paysage actuel de l'IA en Israël et d'examiner comment ses grands objectifs politiques peuvent être atteints compte tenu des caractéristiques de la société israélienne, le Premier ministre israélien a lancé en 2018 l'Initiative nationale pour des systèmes intelligents sécurisés, et a

²³¹ Ben-Israel, I., Matania, E. & Friedman, L. (Eds.) (sept. 2020). *L'initiative nationale pour des systèmes intelligents sécurisés pour renforcer la sécurité nationale et la résilience techno-scientifique : Une stratégie nationale pour Israël. Rapport spécial au Premier ministre.* (en hébreu) p.3.

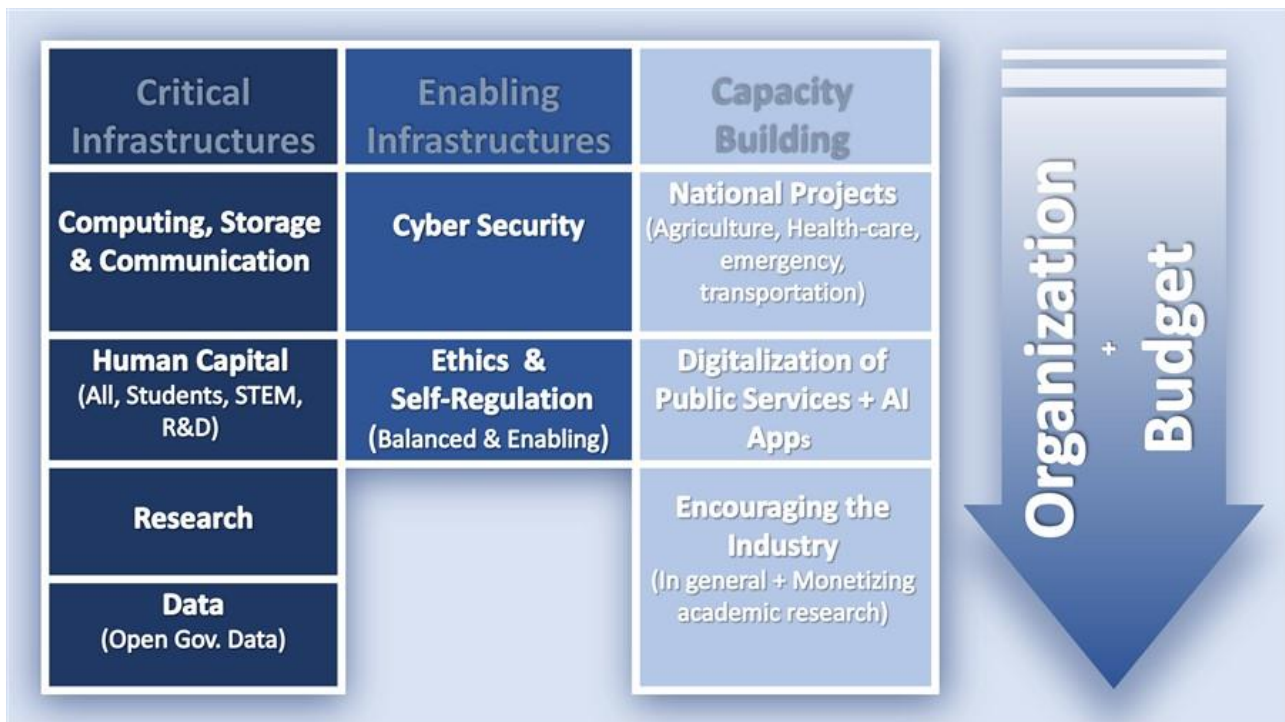
nommé deux des auteurs de cet article, le professeur Isaac Ben-Israel et le professeur Eviatar Matania, pour codiriger l'Initiative. Le mandat de cette dernière était de générer un plan national dans le domaine de l'IA et des technologies intelligentes connexes. L'Initiative a utilisé une approche multipartite : des centaines d'experts israéliens dans divers domaines et issus des secteurs universitaire, industriel et gouvernemental se sont portés volontaires pour participer à cette entreprise. Les experts ont été répartis en quinze groupes de travail traitant de divers aspects technologiques, sectoriels et intersectoriels des systèmes intelligents, selon le modèle suivant²³² :



Chaque groupe de travail a analysé l'environnement de l'IA israélienne selon ses perspectives thématiques. Les groupes de travail ont utilisé comme point de comparaison les travaux de pointe antérieurs qui ont été menés par d'autres juridictions, afin de présenter un plan national adapté aux caractéristiques spécifiques d'Israël. Les conclusions et les recommandations ont été intégrées dans un rapport final qui propose une stratégie nationale pour Israël dans le domaine des systèmes intelligents sécurisés (le "**Rapport d'initiative nationale**"). Ce rapport définit les technologies intelligentes comme une priorité nationale et établit un plan national opérationnel pour l'établissement d'un écosystème durable dans le domaine des systèmes intelligents sécurisés. Le plan national est basé sur trois niveaux : (1) les infrastructures critiques ; (2) les infrastructures habitantes ; (3) le renforcement des capacités ; et se compose des éléments suivants²³³.

²³² Ibid. p.15.

²³³ Ibid. p.23.



Le Rapport d'initiative nationale a été récemment soumis au Premier ministre israélien.

III. Les applications de l'IA en Israël - une opportunité pour les politiques publiques

« Israël est maintenant le numéro trois mondial pour les solutions d'IA. Avec seulement 8,5 millions de citoyens, Israël a une part de marché de 11 % et est à égalité avec la Chine. Israël compte quarante fois plus de sociétés d'IA par habitant que les États-Unis, leader du marché, ce qui fait d'Israël le champion caché de l'intelligence artificielle ». ²³⁴

Israël dispose d'un solide écosystème de haute technologie et d'innovation, ainsi que d'une culture qui adopte et s'adapte aux développements technologiques. L'atmosphère qui règne dans la "Startup Nation" est telle qu'elle encourage les secteurs public et privé à explorer et à utiliser les applications de l'IA dans divers domaines. Toutefois, le paysage des applications de l'IA en Israël est façonné, avant tout, par le marché privé.

i. Le secteur privé

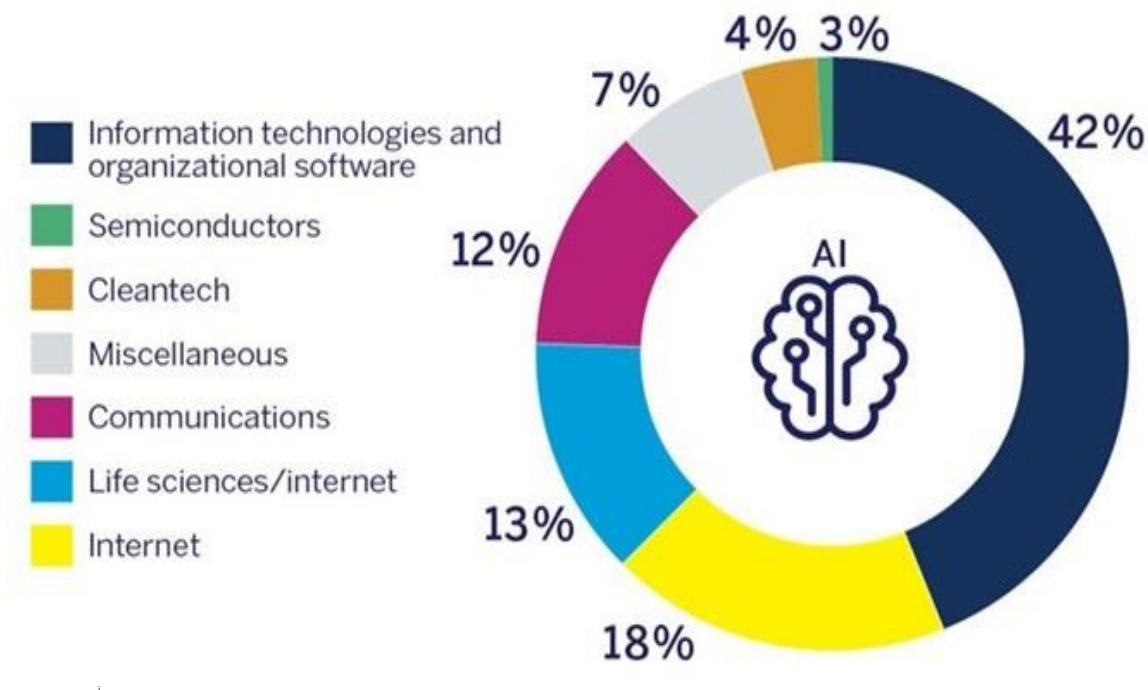
Selon l'Autorité israélienne de l'innovation, les investissements dans les projets israéliens de haute technologie en IA ont été multipliés par 12,5 en 2011-2019, passant de 305 millions de dollars à 4 milliards de dollars. En 2019, 42% de la somme totale investie dans les hautes technologies israéliennes ont été consacrés aux technologies de l'IA.²³⁵ Environ 1400 sociétés d'IA opèrent actuellement en Israël, développant et utilisant des technologies d'IA dans divers domaines tels que l'analyse commerciale, les applications informatiques et de santé, etc. Plus de 40 % des entreprises s'occupent de technologies de l'information et de logiciels d'organisation, tandis que 30 % se concentrent sur les services internet et les communications²³⁶.

²³⁴ ASGARD. Le paysage mondial de l'intelligence artificielle. Récupéré sur <https://asgard.vc/global-ai/7>

²³⁵ Autorité israélienne de l'innovation. (2020). Renforcement de l'intelligence artificielle. Récupéré sur https://innovationisrael.org.il/en/reportchapter/bolstering-artificial-intelligence-0#footnoteref2_bscx727

²³⁶ Ibid ; Israel Innovation Authority. (2020). Innovation in Israel - 2019 Innovation Report. p.62 (hébreu).

Répartition des entreprises d'IA par secteur



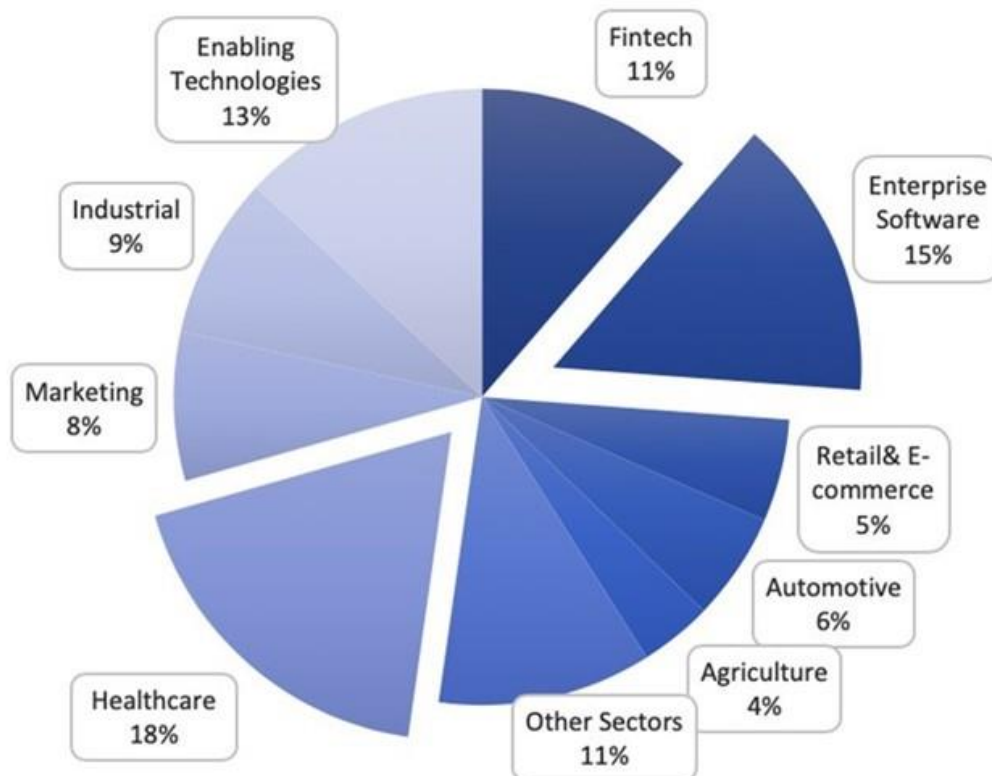
Source : Autorité israélienne de l'innovation (2020). [Renforcer l'intelligence artificielle](#) (non traduit)

1 024 de ces entreprises sont des start-up. Malgré sa petite taille et ses ressources limitées, Israël est le troisième pays au monde en termes de nombre de start-ups d'IA, après les États-Unis et la Chine, et le premier en termes de nombre d'entreprises d'IA par habitant.²³⁷ Au cours des cinq dernières années, une moyenne de 140 nouvelles startups ont vu le jour chaque année, offrant des applications et des produits qui couvrent tous les secteurs et domaines de la vie. Toutefois, le secteur le plus important est celui des soins de santé, avec 188 start-ups (18 %) proposant des solutions d'IA dans des domaines tels que le diagnostic, le suivi, la gestion des maladies, la personnalisation et le flux de travail clinique. Les logiciels d'entreprise suivent de près avec 152 start-ups (15%) développant et utilisant des produits et services d'IA dans les domaines de la vente et de la formation, des RH, des données et des informations, de la clientèle et du support, du développement et de l'informatique, de la gestion et du travail d'équipe, de la sécurité, de la confidentialité et des finances.²³⁸

²³⁷ ROLAND BERGER GMBH & ASGARD. (2018). Intelligence artificielle - Une stratégie pour les start-ups européennes. P. 17

²³⁸ Cardumen Capital (9 juin 2020) Israel's Artificial Intelligence Startups, juin 2020. Medium. extrait de <https://medium.com/@cardumencapital/israels-artificial-intelligence-startups-juin-2020-81e27d9332d8>

Israel's AI startups distribution by sectors



Source : Cardumen Capital. [Israel's Artificial Intelligence Startups, juin 2020](#). Moyen (non traduit)

Israël se classe également en tête du nombre d'entreprises qui développent des technologies d'infrastructure pour l'IA telles que des puces à usage spécifique, des algorithmes d'infrastructure et des systèmes complexes pour l'accélération de l'informatique.²³⁹

Pour une carte complète du paysage de démarrage de l'IA en Israël en 2020, segmentée par secteurs et applications, voir l'annexe I.

ii. Initiatives et politiques du gouvernement

Outre l'activité du secteur privé, le gouvernement israélien joue un rôle important dans la promotion des applications de l'IA, en initiant lui-même des projets ou en créant un environnement encourageant le secteur privé à développer et à utiliser davantage les technologies de l'IA, par exemple par une réglementation favorable, des incitations pour l'industrie, etc. Le potentiel important que le gouvernement attribue aux technologies de l'IA et à leurs applications trouve son origine dans deux caractéristiques de la société israélienne :

- **Un taux de croissance démographique élevé** - La démographie d'Israël est exceptionnelle pour les économies développées et, en tant que telle, elle présente à son gouvernement des défis uniques. Avec 3,1 enfants par femme, Israël a le taux de fécondité le plus élevé parmi les pays de l'OCDE, dont le taux de fécondité moyen est de

²³⁹ Autorité israélienne de l'innovation. (2020). Renforcement de l'intelligence artificielle. Récupéré sur https://innovationisrael.org.il/en/reportchapter/bolstering-artificial-intelligence-0#footnoteref2_bscx727

1,6 enfants.²⁴⁰ En outre, l'espérance de vie moyenne en Israël est de 82,9 ans, ce qui place le pays au cinquième rang des pays de l'OCDE. ²⁴¹Par conséquent, le taux annuel de croissance démographique en Israël - 1,9 % ²⁴²- est presque quatre fois plus élevé que le taux moyen de croissance démographique des pays membres de l'OCDE (0,54 %).²⁴³ La croissance rapide et constante de la population exige de l'État d'Israël qu'il adapte ses services publics et sociaux et qu'il maintienne et augmente en conséquence ses infrastructures dans tous les domaines de la vie (par exemple, les soins de santé, le transport, l'éducation, l'énergie, etc.)

- **Densité de population et surcharge des infrastructures** - la majorité de la population israélienne vit au centre du pays, et plus de 40 % de la population est répartie sur moins de 7 % du territoire du pays.²⁴⁴ Cela entraîne de graves surcharges pour les infrastructures et les services dans les zones très peuplées. Un exemple frappant est la congestion croissante du trafic dans le centre du pays, qui a des ramifications négatives sur la productivité, l'environnement et le nombre d'accidents et de victimes

Le taux de croissance élevé de la population et la densité de la population déterminent l'approche d'Israël en matière d'IA, car ils augmentent considérablement la demande israélienne d'infrastructures et de services nationaux. La motivation du gouvernement pour accroître l'utilisation de l'IA, au-delà des domaines déjà couverts par le secteur privé, réside dans le potentiel de la technologie à répondre au besoin croissant d'amélioration de la disponibilité, la fiabilité et l'efficacité des services publics et des infrastructures nationales, à moindre coût pour l'État et ses citoyens. C'est pourquoi, parmi les premiers secteurs pour lesquels le gouvernement promeut des solutions d'IA, figurent les soins de santé et les transports, où la surcharge des infrastructures actuelles est la plus aiguë.

Cependant, les ministères israéliens ne sont pas tous prêts à adopter les applications d'IA en raison de la différence de maturité numérique et, dans certains cas, des lacunes numériques. L'initiative numérique nationale d'Israël, également appelée "Digital Israel", est l'organisme gouvernemental responsable des services d'e-gouvernement. Digital Israel est le fer de lance des efforts du gouvernement pour la transformation numérique, afin de réduire les écarts socio-économiques, de promouvoir la prospérité économique et de créer un gouvernement plus intelligent et plus accessible. Son champ d'activité englobe un large éventail de services d'e-gouvernement, à tous les échelons (y compris les municipalités). Ainsi, Digital Israel travaille avec d'autres ministères, les aidant à développer et à déployer des plans de numérisation. Digital Israel dirige également les plans d'initiatives gouvernementales pour les villes intelligentes ainsi que le plan national pour l'alphabétisation numérique.

Dans une très large mesure, l'accomplissement du mandat de Digital Israel dépend de la disponibilité et de la transférabilité des données. En effet, afin de maximiser le potentiel de la transformation numérique, les organismes gouvernementaux doivent être en mesure de collecter de grandes quantités d'informations, de les combiner avec des données provenant d'autres sources et de déployer des outils techniques pour analyser les données et tirer des conclusions. Dans de nombreux cas, les données qui doivent être collectées et partagées comprennent des informations personnelles identifiables ("IPI"). En outre, les outils logiciels qui peuvent être utilisés comprennent l'analyse de volumes importants de données, dont une grande partie peut être améliorée par l'apprentissage automatique. Ainsi, les projets que Digital Israel souhaite mettre en œuvre se heurtent à un défi majeur, à savoir comment trouver un équilibre entre, d'une part, les besoins en matière de données et, d'autre part, les considérations juridiques et éthiques.

²⁴⁰ OCDE (2020), Taux de fécondité (indicateur). Consulté sur <https://data.oecd.org/pop/fertility-rates.htm#indicator-chart>

²⁴¹ OCDE (2020), Espérance de vie à la naissance (indicateur). Consulté sur <https://data.oecd.org/healthstat/life-expectancy-at-birth.htm>

²⁴² Bureau central des statistiques. (2020). Israël en chiffres Sélection de données tirées du Statistical Abstract of Israel 2019. Consulté sur https://www.cbs.gov.il/he/publications/DocLib/isr_in_n/isr_in_n19e.pdf p.6.

²⁴³ Banque mondiale, Population Growth for OECD Members [SPPOPGROWOED]. Extrait de FRED, Federal Reserve Bank of St. Louis <https://fred.stlouisfed.org/series/SPPOPGROWOED>

²⁴⁴ Bureau central des statistiques. (2020). Résumé statistique d'Israël 2019 - No.70. Consulté sur https://www.cbs.gov.il/he/publications/DocLib/2019/Shnaton70_mun.pdf (en hébreu). p.21.

À ce jour, il n'existe pas de politique gouvernementale globale pour relever ce défi. Une telle politique est actuellement en cours de finalisation. En attendant son adoption, les activités de Digital Israel s'appuient sur le cadre juridique existant, qui comprend des protections constitutionnelles des droits de l'Homme (vie privée, non-discrimination, liberté d'expression) ainsi que des règles et principes de droit administratif applicables de manière générale à tous les organes gouvernementaux (transparence, responsabilité, équité, respect des procédures et caractère raisonnable).

Vous trouverez ci-dessous une description des principaux projets impliquant l'IA qui sont en cours de développement (dans chaque cas, en collaboration avec le ministère concerné). Dans le cadre du développement de chacun de ces projets, Digital Israel a travaillé avec le Conseil interne et le Conseil constitutionnel du Ministère de la justice, afin de s'assurer que le développement et le déploiement de chaque projet respectent les limites du droit constitutionnel et administratif applicable.

Santé numérique

Le projet a des objectifs ambitieux, notamment : ²⁴⁵

- La personnalisation du traitement : promouvoir la recherche, le développement et la mise en œuvre d'outils qui permettent au patient de recevoir le meilleur traitement et le plus personnalisé possible ;
- La promotion de la santé et la prévention auprès des patients par l'utilisation des outils numériques de manière à d'avantage se concentrer sur la médecine préventive plutôt que les soins aux patient postérieurs ;
- Le développement d'une santé durable : promouvoir le développement et la mise en œuvre de systèmes qui augmentent l'efficacité opérationnelle et gestionnaire du système de santé, de manière à libérer les ressources existantes ;
- Le développement et la mise en œuvre d'outils numériques qui rationalisent la communication entre le Ministère de la santé et les administrés ;
- La prestation de services de traitement d'urgence par le biais d'un système de gestion des rendez-vous et d'une application informant le patient de l'évolution du traitement. Les informations recueillies permettront une meilleure gestion des ressources afin d'éviter l'encombrement des salles d'urgence ;
- Le partage d'informations cliniques entre les plateformes de fournisseurs de services a été étendu, afin de les mettre en relation et leur permettre de consulter les traitements et les diagnostics effectués par d'autres professionnels de santé dans différentes organisations.

Certains de ces objectifs sont déjà en cours de réalisation. Par une résolution du gouvernement, ²⁴⁶le Ministère de la santé a créé une plateforme appelée "TIMNA", qui accorde à des tiers un accès contrôlé aux données de santé afin de promouvoir la recherche appliquée. Ces données comprennent de vastes quantités de dossiers médicaux, recueillis par des hôpitaux et des cliniques de tout le pays, qui constituent une ressource inestimable. L'accès à ces données est soumis à des restrictions strictes en matière de confidentialité et d'éthique. Tout d'abord, les institutions, les chercheurs et les jeunes entreprises qui souhaitent accéder à ces données doivent faire approuver leur projet par le comité d'Helsinki. Ils sont tenus d'identifier les types de données spécifiques dont ils ont besoin, et seules ces données sont fournies. Ils doivent signer des engagements en matière de protection de la vie privée. Toutes les données personnelles sont rendues anonymes. La recherche se déroule entièrement au sein de la plate-forme numérisée - aucune donnée personnelle ne peut être extraite de la plate-forme, ce qui est une garantie supplémentaire de confidentialité. En outre, avant que la recherche ne soit publiée et qu'un algorithme ne soit utilisé, le Ministère de la

²⁴⁵ Page d'accueil des services numériques du ministère de la santé (en hébreu)
<https://www.health.gov.il/About/projects/DigitalHealth/Pages/default.aspx>.

²⁴⁶ Résolution 3709 du gouvernement israélien, "Plan national pour faire progresser la santé numérique comme moyen d'améliorer la santé et de favoriser la croissance" 23 août 2018, https://www.gov.il/he/Departments/policies/des3709_2018 (en hébreu).

santé l'examine pour s'assurer qu'aucune information personnelle n'est utilisée ou divulguée. Jusqu'à présent, les projets TIMNA utilisant l'IA sont les suivants :

- Le Centre israélien de contrôle des maladies applique un algorithme d'IA pour examiner les diagnostics des patients et vérifier leur diagnostic de cancer. Cela représente un gain de temps significatif, car il automatise le processus d'examen de plus de 100 000 formulaires par an. Des audits sont effectués pour s'assurer qu'il n'y a pas de faux négatifs.
- De même, un algorithme est en cours de développement pour aider au suivi médical dans deux domaines : le développement de l'enfant et la grossesse. Dans les deux cas, l'algorithme analyse en temps réel les rapports de situation, les diagnostics et les notes et recommandations des médecins et des infirmières, en les comparant aux protocoles standard. Il alerte ensuite l'hôpital ou la clinique en cas d'erreurs potentielles, de mauvais diagnostics ou de problèmes qui pourraient nécessiter des tests ou des suivis supplémentaires. Le système est destiné à aider les professionnels de la santé à détecter les erreurs et ne comporte pas de risques éthiques importants pour le patient.
- Un autre domaine d'étude est l'utilisation de l'IA pour l'analyse des images médicales (IRM, CT, etc.). Jusqu'à présent, les résultats ont été prometteurs, dans la mesure où les algorithmes ont pu détecter des cas qui n'avaient pas été détectés par les médecins. L'objectif n'est pas de supplanter la prise de décision du médecin, mais plutôt de rationaliser le processus et de l'aider à analyser les images.
- Enfin, le Ministère de la santé a déployé des algorithmes d'IA pour l'aider dans ses efforts visant à ralentir la propagation de COVID-19. Souvent, l'étude épidémiologique basée sur des discussions avec une personne infectée est incomplète, en raison de l'oubli de tous les lieux visités, de l'incapacité de l'enquêteur à saisir correctement toutes les informations, ou d'une autre erreur humaine. Les algorithmes sont utilisés pour dresser un tableau plus complet de l'évolution probable des infections antérieures et prédire les infections futures. Ces informations sont ensuite utilisées pour informer la politique gouvernementale en ce qui concerne les mesures de fermeture à une échelle générale. Il convient de noter que ces informations ne sont pas utilisées pour prendre des décisions concernant des individus ou des communautés spécifiques.

Transport

Le Ministère des transports met en place un projet pilote pour permettre l'essai de véhicules autonomes. Ce projet permettrait aux constructeurs de demander une licence spéciale, avec laquelle ils pourraient tester leur produit dans des conditions réelles, dans des environnements de conduite à faible risque. À cet effet, le Ministère a publié un projet de loi, qui est ouvert aux commentaires du public.²⁴⁷ Le projet de loi ne contient pas encore toutes les règles qui s'appliqueront aux essais ou aux exigences d'un système de conduite autonome (capacités, sécurité et surveillance). Ces questions, ainsi que les questions éthiques qui pourraient se poser, seront abordées ultérieurement. Cette approche reflète une philosophie d'innovation prudente et progressive : afin de comprendre les impacts d'une technologie, et compte tenu des risques pour la vie humaine, les essais sont autorisés dans un environnement qui n'est pas « contrôlé » mais qui présente un risque relativement faible. Cela devrait permettre aux décideurs politiques de procéder à des ajustements avant d'aller de l'avant avec des expérimentations à plus grande échelle.

Des progrès sont également réalisés dans le domaine des transports publics, des solutions innovantes étant mises au point pour réduire les embouteillages²⁴⁸.

²⁴⁷ Le texte du projet de loi peut être consulté ici : https://www.nevo.co.il/law_word/law11/200820-2.doc (en hébreu).

²⁴⁸ Ben Dror, M. et Azaria, M. (24 juillet 2020). Les "déplacements intelligents" en Israël montrent ce que pourraient être les transports publics après COVID-19. Forum économique mondial. Consulté sur <https://www.weforum.org/agenda/2020/07/israel-smart-commuting-after-covid-public-transport-innovation/>

Fiscalité

L'autorité fiscale israélienne a lancé un projet visant à aider les enquêteurs à détecter les fraudes fiscales. Le projet utilise des outils d'IA pour prédire la probabilité de fraude fiscale, sur la base de certains indicateurs. Les préoccupations relatives à la protection de la vie privée ont été au cœur de la conception du projet. En effet, le projet repose sur une approche par couches pour l'accès à l'information : au départ, peu d'indicateurs et peu d'informations sont utilisés pour signaler les risques de fraude ; ce n'est que si l'enquête initiale suggère un risque de fraude plus élevé que des informations personnelles sont requises afin de déterminer avec plus de certitude l'identité du contrevenant potentiel.

Propositions de projets nationaux

Outre les projets en cours susmentionnés, le Rapport d'initiative nationale recommande que le gouvernement lance, en coopération avec les secteurs industriel et universitaire, quatre autres projets nationaux dans les domaines de la Santé, des transports, de la sécurité et de l'agriculture. Ces quatre projets ont été conçus pour répondre à de véritables besoins nationaux découlant de l'évolution de la société israélienne, qui suscite des demandes d'infrastructures et de services améliorés voir nouveaux.

- **Soins de santé** - Réformer le système national de Santé en améliorant la qualité et la disponibilité des services médicaux, et en allégeant la surcharge des hôpitaux en lançant un système national basé sur des technologies intelligentes pour : (i) la gestion des patients à distance ; (ii) un triage et un traitement plus efficaces en médecine d'urgence ; (iii) la création d'indices comparatifs de qualité pour mesurer les résultats cliniques.²⁴⁹
- **Transport** – « Parmi les nombreuses façons dont les systèmes intelligents peuvent résoudre les problèmes critiques dans le domaine du transport, nous avons choisi de recommander, dans un premier temps, l'installation de feux de circulation intelligents dans toute une zone métropolitaine pilote, dans le but de résoudre les problèmes d'embouteillages, qui sont les plus graves et les plus aigus dans le domaine du transport en Israël ». ²⁵⁰
- **Sécurité** – la création d'un double système national qui exploitera le potentiel des technologies intelligentes pour améliorer les capacités de prévision et les processus de décision, pour une meilleure gestion au niveau national. En collectant et en analysant des données pour les applications civiles de commandement et de contrôle en temps normal, il renforcera la capacité nationale à se préparer aux situations d'urgence (catastrophes naturelles, épidémies et menaces à la sécurité de la part d'ennemis) et à prendre des décisions en cas de crise. ²⁵¹
- **Agriculture** – « À une époque où la sécurité alimentaire, la gestion de l'eau et d'autres domaines de l'agriculture deviennent des défis mondiaux critiques, [...] Nous recommandons de promouvoir un projet national visant à développer un système basé sur des technologies intelligentes pour la détection précoce des parasites et des maladies dans les cultures agricoles ; parallèlement à l'intégration de systèmes intelligents dans le secteur agricole pour optimiser l'utilisation des ressources naturelles et des [autres] intrants afin d'assurer une production alimentaire optimale. » ²⁵²

Enfin, reconnaissant la variabilité de la maturité numérique au sein du gouvernement, l'Initiative nationale a inclus un groupe de travail dédié au secteur gouvernemental, afin de définir des lignes directrices pour préparer l'ensemble du gouvernement à l'ère de l'IA. Le groupe de travail a évalué les mesures organisationnelles, technologiques et réglementaires nécessaires pour favoriser la mise

²⁴⁹ Ben-Israel, I., Matania, E. & Friedman, L. (Eds.) (sept. 2020). L'initiative nationale pour des systèmes intelligents sécurisés pour renforcer la sécurité nationale et la résilience techno-scientifique : Une stratégie nationale pour Israël. Rapport spécial au Premier ministre. (en hébreu) p.35.

²⁵⁰ Ibid. P.37

²⁵¹ Ibid.

²⁵² Ibid. P.36

en œuvre des applications d'intelligence artificielle au sein du gouvernement, afin d'améliorer à la fois le travail interministériel et les interactions entre le gouvernement et les citoyens.²⁵³

IV. Risques et défis posés par l'IA dans les domaines des droits de l'Homme, de la démocratie et de l'État de droit

Les thèmes des droits de l'Homme et de l'éthique ont été abordés par l'Initiative nationale, par l'intermédiaire d'un groupe de travail spécialisé qui a étudié la question en profondeur d'un point de vue réglementaire et éthique. Le rapport du groupe de travail ("**Rapport sur l'éthique**")²⁵⁴ identifie les questions nouvelles et uniques à l'IA et les façons dont elle devrait affecter nos vies. À la lumière de ces questions, il présente une série de défis éthiques posés par la technologie en ce qui concerne les droits de l'homme, la démocratie et l'État de droit.

Vous trouverez ci-dessous un aperçu des questions identifiées par le Rapport sur l'éthique, suivi d'un aperçu des risques et des défis éthiques qu'il suggère d'aborder.

i. Quelles sont les nouveautés et les particularités de l'IA ?

a. « Les systèmes d'IA ont tendance à radicaliser les relations sociales existantes. Par exemple, s'il y a une inégalité entre différents groupes sociaux, les systèmes d'IA peuvent la reproduire et même l'exacerber. C'est le cas de la discrimination, des stéréotypes, des violations des droits, de l'extrémisme politique, etc. Pour des raisons de commodité, nous allons démontrer cette affirmation en ce qui concerne l'inégalité. Ce phénomène s'explique par plusieurs raisons principales :

- *Comme les systèmes d'IA dépendent des informations qui leur sont fournies, leur saisie peut refléter une inégalité qui existe déjà. Dans le cas où les données saisies ont été manipulées, le système apprendra cette manipulation.*
- *Les systèmes d'IA sont de plus en plus courants dans un nombre croissant de contextes sociaux. Par conséquent, leur impact - et les biais potentiels - touchent des publics plus larges.*
- *Il existe une tendance erronée à considérer les produits des systèmes d'IA, qui analysent les données rapidement et à grande échelle, comme une vérité scientifique. Par conséquent, il existe un danger que ces systèmes ne soient pas soumis aux contrôles appliqués aux décisions humaines équivalentes, lorsqu'un biais est suspecté.*
- *En raison de la complexité des systèmes, il est difficile d'anticiper et de valider leur comportement à l'avance. Par conséquent, il est souvent difficile de distinguer entre un "vrai" diagnostic basé sur un processus d'examen et de surveillance valable, comme c'est le cas pour les algorithmes classiques ou la prise de décision humaine, et un diagnostic biaisé ».*²⁵⁵

b. Le défi de la procédure : comment « concevoir » les valeurs. Cette question se pose dans les domaines où les systèmes d'IA sont développés pour remplacer les décideurs humains qui sont qualifiés et autorisés à appliquer des considérations normatives. « Lors du

²⁵³ Sharvit, S. et al. (2020). Rapport du groupe de travail gouvernemental. Dans Ben-Israel, I., Matania, E. & Friedman, L. (Eds.). *The National Initiative for Secured Intelligent Systems to Empower the National Security and Techno-Scientific Resilience: A National Strategy for Israel. Rapport spécial au Premier ministre.* pp. 227-214. (en hébreu).

²⁵⁴ Nahon K., Ashkenazi A., Gilad Bachrach R., Ken-Dror Feldman D., Keren A. et Shwartz Altshuler T. (2020). Rapport du groupe de travail sur l'éthique et la réglementation de l'intelligence artificielle. Dans Ben-Israel, I., Matania, E. & Friedman, L. (Eds.). *The National Initiative for Secured Intelligent Systems to Empower the National Security and Techno-Scientific Resilience: A National Strategy for Israel. Rapport spécial au Premier ministre.* pp. 172-119. (en hébreu).

²⁵⁵ Ibid. p. 179.

développement de systèmes d'IA qui remplacent la décision humaine, la responsabilité de ces considérations normatives est transférée de professionnels tels que des médecins et des juristes à des ingénieurs et des informaticiens, ce qui ne se produit pas aussi souvent lorsqu'il s'agit d'algorithmes classiques. »²⁵⁶

- c. La vie privée et l'autonomie sont des risques d'une ampleur et d'une portée sans précédent. Voir les pages 134-136 ci-dessous.
- d. Une complexité qui érode la confiance du public. Le manque de clarté et de compréhension du public quant au fonctionnement des systèmes d'IA et à leur impact sur nos vies entraîne souvent une méfiance, qui peut se traduire par une réticence à adopter la technologie, même dans les domaines où les systèmes d'IA offrent un avantage commercial - et social - évident. Le Rapport mentionne à cet égard l'évaluation du groupe d'experts de haut niveau de l'UE sur l'intelligence artificielle.²⁵⁷
- e. « Des économies d'échelle déloyales. [...] les acteurs puissants disposant des données importantes nécessaires pour développer des systèmes d'IA profitent des économies d'échelle sur Internet pour façonner la manière dont de nouveaux acteurs entrent sur le marché, avec un effet négatif sur la compétitivité. Lorsqu'il s'agit d'acteurs totalement nouveaux, le fait qu'ils ne disposent pas de la quantité de données requises pourrait signifier qu'ils sont en fait exclus du marché de l'IA. »²⁵⁸
- f. « Changements dans les catégories de garantie habituelles. La possibilité de collecter et de traiter des données sur les produits permet aux entreprises de proposer de nouveaux services connexes, mais soulève également de nouvelles questions sur la garantie de ces services et la répartition des responsabilités entre le producteur et les prestataires de services dans la pratique. Les produits intégrés à l'IA, en particulier, comprennent également la combinaison d'un produit physique et d'une calculabilité et d'une opérabilité à distance. Il convient donc de réexaminer la division classique entre produit et service et entre garantie du produit et garantie du service. Les choses deviennent encore plus complexes lorsque ces produits et services sont utilisés par d'autres entités commerciales. Par exemple, lorsqu'une chaîne d'épicerie utilise un drone pour ses livraisons. Le drone est capable de voler, de naviguer et de traiter l'environnement. En outre, il fournit des services de cartographie et de prévisions météorologiques. Tous ces services sont acquis par une chaîne d'épicerie, dans le seul but de livrer des produits alimentaires. »²⁵⁹

ii. Risques et défis éthiques

Le Rapport sur l'éthique aborde les défis éthiques suivants. Il convient de noter qu'il s'appuie largement sur le groupe d'experts de haut niveau de l'UE sur l'intelligence artificielle pour élaborer les différents principes éthiques.

Sécurité

Le droit à la vie et à la sécurité est souvent négligé dans les discussions mondiales sur les droits de l'Homme. Et pourtant, il reste le droit le plus fondamental de tous, sans lequel les autres droits de l'Homme ne peuvent être appliqués. En ce qui concerne l'IA, le Rapport d'éthique souligne la nécessité de sécuriser les applications d'IA et les réseaux et ordinateurs compatibles avec l'IA. Le Rapport fait plusieurs observations importantes à cet égard. Il note que l'information est « l'énergie qui alimente la vague actuelle d'IA », de sorte que la sécurité des applications et des réseaux d'IA est une condition préalable au développement et à la mise en œuvre de la technologie²⁶⁰. Il note en

²⁵⁶ Ibid. p. 180.

²⁵⁷ Ethical Guidelines for Trustworthy AI, The High-Level Expert Group on Artificial Intelligence, EU, 2019, <https://ec.europa.eu/futurium/en/ai-alliance-consultation>. (Ci-après, UE)

²⁵⁸ Nahon K. et al. (2020). p.180.

²⁵⁹ Ibid. p.181.

²⁶⁰ Ibid. p.190.

outre que les informations qui peuvent être collectées par l'IA pour construire et déployer des outils comprennent de vastes quantités d'informations personnelles et commerciales, y compris des informations personnelles, médicales, économiques et autres informations sensibles. Il remarque que même des informations qui semblent « non sensibles » peuvent devenir sensibles lorsqu'elles sont croisées avec d'autres informations.

D'un point de vue politique, cela relie directement la sécurité aux droits de l'Homme. Par exemple, la protection de la vie privée exige, à un niveau fondamental, de protéger les informations privées contre les cyber-opérations malveillantes. De même, au-delà des considérations commerciales et de performance, la protection de l'intégrité des données est également un impératif des droits de l'Homme : afin de prévenir les biais dans une application d'IA particulière, les données ainsi que les algorithmes sur lesquels l'application est basée ne doivent pas être altérés. La liberté d'expression et l'accès à l'information sont également très dépendants de la sécurité des applications et des réseaux liés à l'IA.

L'utilisation de l'IA renforce la dépendance aux ordinateurs, créant ainsi de nouvelles vulnérabilités pour les cyber-attaques. Dans le cadre de cette initiative nationale, un groupe de travail spécialisé dans la cybersécurité à l'ère des systèmes intelligents a identifié les nouvelles cyber-menaces présentées par les technologies de l'IA. Tout d'abord, les attaques contre les systèmes d'IA, qui peuvent entraîner des dommages au système de prise de décision, conduisant ainsi à des décisions fausses, trompeuses ou biaisées, ou à des menaces contre le système IP. Ensuite, l'exploitation malveillante des capacités de l'IA comme cyber-arme pour lancer des « attaques intelligentes » sophistiquées. Ces vulnérabilités des systèmes d'IA soulèvent la question de savoir si et comment un système d'IA peut être authentifié comme étant sécurisé et fiable.²⁶¹ La sécurité de l'IA est donc une couche de base sur laquelle des interactions plus riches peuvent avoir lieu. Son importance est en effet reflétée dans le titre de l'initiative nationale israélienne en matière d'IA - l'initiative nationale pour des systèmes intelligents *sécurisés*.

Vie privée

Il est manifeste que la vie privée soit le premier et le plus important des droits de l'Homme abordés par le Rapport sur l'éthique. Le Rapport souligne que les applications d'IA « reposent en grande partie sur des informations concernant des personnes ou sur le fait de tirer des conclusions à leur sujet à partir d'informations permettant de les identifier ».²⁶² La protection de la vie privée dépend en grande partie d'un cadre législatif solide. En Israël, ce cadre est principalement constitué par la Loi fondamentale d'Israël : Dignité humaine et liberté (1992), de²⁶³ sa loi sur la protection de la vie privée (1981), de plusieurs règlements sur la protection de la vie privée, notamment sur le transfert de données²⁶⁴, et du règlement complet sur la protection de la vie privée (sécurité des données) de 2017²⁶⁵. Le régime juridique est complété par une jurisprudence abondante et un système judiciaire solide. Au fil du temps, un certain nombre de principes de protection de la vie privée sont apparus : la nécessité d'un motif juridique pour la collecte et le traitement des informations (par exemple, le consentement éclairé), les limitations d'utilisation, le droit de revoir et de corriger ses informations personnelles, la transparence vis-à-vis du propriétaire des informations et l'obligation de protéger les informations.

Dans ce contexte, le Rapport sur l'éthique note qu'il subsiste des écarts entre les conceptions traditionnelles de la protection de la vie privée et les défis soulevés par l'IA. En effet, dans la mesure où l'IA s'appuie sur la collecte et le traitement des IPI, l'on peut s'attendre à ce que de nouveaux

²⁶¹ Zack, H. et al. (2020). Rapport du groupe de travail sur les cyber-systèmes et les systèmes intelligents. Dans Ben-Israel, I., Matania, E. & Friedman, L. (Eds.). *The National Initiative for Secured Intelligent Systems to Empower the National Security and Techno-Scientific Resilience: A National Strategy for Israel. Rapport spécial au Premier ministre*. pp. 168-171. (en hébreu). p.168.

²⁶² Nahon K. et al. (2020). p.188.

²⁶³ <http://knesset.gov.il/laws/special/eng/BasicLawLiberty.pdf>. Il s'agit d'une loi quasi-constitutionnelle, dont les principes sous-jacents sont considérés comme constitutionnellement obligatoires, même en l'absence d'une constitution écrite formelle.

²⁶⁴ Règlement sur la protection de la vie privée (transfert de données vers des bases de données à l'étranger) (2001). Voir la traduction non officielle ici : <https://www.gov.il/BlobFolder/legalinfo/legislation/en/PrivacyProtectionTransferofDataabroadRegulationsun.pdf>.

²⁶⁵ Pour la traduction non officielle, voir ici :

https://www.gov.il/BlobFolder/legalinfo/data_security_regulation/en/PROTECTION%20OF%20PRIVACY%20REGULATIONS.pdf

problèmes de protection de la vie privée se posent, ce qui pourrait nécessiter une adaptation plus poussée des lois existantes en la matière.

Le Rapport sur l'éthique observe également que dans certains cas, il peut y avoir un conflit entre la vie privée et l'équité. Si des personnes appartenant à un certain groupe s'abstiennent de partager leurs informations personnelles avec une application d'IA, cette application ne pourra pas tirer parti des données qui tiennent compte de ce groupe, ce qui pourrait conduire à une plus grande discrimination²⁶⁶. Il est donc impératif de permettre la collecte d'IPI tout en garantissant que ces données seront à la fois sécurisées et soumises à de solides protections de la vie privée.

Autonomie

Le Rapport sur l'éthique définit l'autonomie comme « la capacité de l'individu à prendre des décisions intelligentes, y compris la prévention d'une influence injuste ou inconsciente sur le comportement individuel ».²⁶⁷ En termes de droits fondamentaux, cela peut faire référence à des concepts tels que la dignité humaine et le droit d'accès à l'information. Le Rapport sur l'éthique indique ²⁶⁸:

« L'autonomie est fondée non seulement sur la capacité d'un individu à choisir parmi les options, mais aussi sur la disponibilité des informations permettant un choix pertinent et l'évaluation de leur fiabilité. Ces questions ne peuvent être considérées comme allant de soi à l'ère de l'IA. En outre, la capacité d'analyse approfondie des informations sur une personne rendue possible par l'IA permet de concevoir des tentatives de persuasion très intrusives, dont les implications potentielles ne sont pas encore pleinement comprises. »

L'autonomie est également liée à la gamme de décisions humaines impliquées dans l'interaction avec la technologie, que la technologie pourrait réduire. Nous devons donc toujours examiner si, et éventuellement comment, une application donnée affecte l'autonomie. Il convient de noter que dans le cadre de cette discussion, il peut y avoir des cas où l'autonomie est plus étroite au départ (en raison de certaines caractéristiques socio-économiques ou normatives), ou des cas où une autonomie plus étroite est considérée comme plus appropriée sur le plan normatif, il n'est pas nécessaire de prendre des mesures spéciales pour protéger la liberté de choix.

Certaines technologies d'IA, telles que les "deep fakes", sont conçues pour produire des informations peu fiables qu'il est difficile de distinguer des informations réelles. Ces technologies ont le potentiel de réduire la capacité des individus à comprendre la réalité, à prendre des décisions autonomes et éclairées, et d'éroder la confiance entre les personnes et entre celles-ci et leur gouvernement. Par exemple, nous ne sommes pas loin du jour où il serait possible de produire artificiellement un film où un dirigeant déclare la guerre, ce qui entraînerait des résultats catastrophiques. Le Comité estime que l'État d'Israël devrait examiner les moyens de traiter ces technologies dans un rapport séparé.

Un dernier domaine pertinent pour l'autonomie est la pénétration des outils d'IA dans les médias d'information. De nombreux canaux de communication utilisent l'IA pour produire des nouvelles personnalisées. Cet outil présente de nombreux avantages, mais présente également le danger d'une exposition sélective : certains groupes de la population sont exposés à des informations standardisées et ne connaissent pas les preuves et les arguments qui sont incompatibles avec leur vision du monde. Cela priverait cette population de la liberté de choix ou de la liberté d'être exposée à une diversité d'opinions, et la rendrait vulnérable à des campagnes d'influence injustes et très efficaces de la part des parties intéressées. Cela pourrait notamment permettre à des gouvernements étrangers d'intervenir dans les élections. »

Droits civils et politiques

Le Rapport sur l'éthique définit les droits civils et politiques comme incluant « le droit d'élection, la liberté d'expression et la liberté de conscience religieuse ».²⁶⁹ Ces droits sont au cœur des valeurs

²⁶⁶ Nahon K. et al. (2020). p.188.

²⁶⁷ Ibid. p.182.

²⁶⁸ Ibid. p.189.

²⁶⁹ Ibid. p.182.

démocratiques et méritent une protection particulière. À cet égard, le Rapport sur l'éthique note avec inquiétude la manière dont la manipulation automatisée du discours général est manipulée, par exemple en amplifiant à outrance certaines opinions tout en faisant taire d'autres, en polarisant le discours, en donnant une légitimité à des opinions qui pourraient être offensantes pour certains groupes, et en diffusant de fausses informations à grande échelle.²⁷⁰

Tout cela peut nuire au processus démocratique lui-même, en créant des clivages au sein de la société et en sapant la foi dans le processus démocratique.

Équité

Il s'agit d'un principe éthique général, qui fait référence à la nécessité de parvenir à une égalité substantielle, de prévenir les préjugés (dans l'information, dans le processus et dans la production), d'empêcher la discrimination et d'éviter l'élargissement des écarts socio-économiques et éducatifs. Le Rapport sur l'éthique le note : « *La technologie n'est pas neutre, car elle est basée sur une programmation humaine et divers intérêts commerciaux. De plus, les systèmes d'IA sont basés sur des informations liées au comportement humain, qui peuvent refléter et exacerber divers types de préjugés sociaux* ». ²⁷¹ Il fournit les exemples suivants de systèmes d'IA qui soulèvent des questions d'équité :

- Le système décide de l'allocation des ressources telles que les fonds et les traitements médicaux.
- Le système évalue les candidats à un poste de travail ou à l'enseignement supérieur.
- Le système évalue les personnes dans le but de les punir pénalement ou d'en atténuer les effets.
- Le système prend des décisions qui menacent les biens et les intérêts financiers des utilisateurs.

Pour faire face à ces risques, le Rapport sur l'éthique souligne l'importance d'étudier de manière proactive la population cible et d'identifier à l'avance les groupes susceptibles de faire l'objet de fausses déclarations ou de sous-représentation. En outre, il souligne la nécessité de consulter les représentants des utilisateurs cibles eux-mêmes afin de contribuer à produire des systèmes plus équitables.²⁷²

Responsabilité

Le Rapport sur l'éthique sépare la responsabilité en trois catégories : la *transparence*, l'*explicabilité* et la *responsabilité*.

La *transparence* consiste à « *fournir des informations sur le processus et la prise de décision y afférente* »²⁷³ et est considérée comme une « *valeur clé du développement technologique et du développement de produits d'IA en particulier* ». ²⁷⁴ Il s'agit à la fois d'une valeur à part entière et d'un aspect de la responsabilité. Elle permet le suivi et la réalisation d'autres valeurs telles que l'équité. La transparence est un élément essentiel de la confiance du public.

L'*explicabilité* est la capacité d'un système d'IA à expliquer son processus de décision, que ce soit à l'utilisateur final individuel ou à un niveau collectif si la décision affecte le groupe. Elle comprend également la capacité d'un système à fournir des explications significatives aux opérateurs du système eux-mêmes. Le Rapport sur l'éthique adopte la position du rapport du groupe d'experts de l'UE selon laquelle l'explicabilité comprend un principe d'« *information significative* », c'est-à-dire que le niveau d'information fourni doit être suffisant sans être excessivement technique ou détaillé.

²⁷⁰ Ibid. p.190.

²⁷¹ Ibid. p.184.

²⁷² Ibid.

²⁷³ Ibid. p.182.

²⁷⁴ Ibid. p.185.

La *responsabilité*, implique l'élaboration de règles appropriées pour prévenir les risques, en fonction du contexte et de la gravité estimée du risque, la gestion des risques et la nomination d'un employé chargé de la gestion des risques. Le Rapport sur l'éthique note que la diversité des parties prenantes et la complexité des systèmes d'IA rendent cette tâche particulièrement difficile. Cette situation est aggravée par le fait que les systèmes d'IA prennent également leurs propres « décisions ».

Sécurité

Le Rapport sur l'éthique reconnaît la nécessité d'aborder les risques de sécurité qui découlent des systèmes d'IA. En effet, plus un système d'IA est habilité à prendre des décisions ayant un impact direct sur la vie humaine, plus il est risqué de l'utiliser. Le risque survient aussi bien dans le cadre du fonctionnement ordinaire du système d'IA que dans des situations extrêmes. Le Rapport d'éthique fait donc la distinction entre les risques pour la sécurité qui surviennent dans le cadre d'un dysfonctionnement, et ceux qui surviennent lorsque le système fonctionne sans dysfonctionnement mais d'une manière qui cause néanmoins la mort ou un dommage physique.

Les risques pour la sécurité peuvent être atténués par la mise en œuvre d'un certain nombre de mesures. Par exemple, afin d'éviter des décisions incorrectes fondées sur des préjugés erronés, il convient d'utiliser un ensemble de données diversifiées. De même, les considérations de sécurité doivent être prises en compte au stade de la conception. Le Rapport sur l'éthique fournit un exemple intéressant - et peut-être contre-intuitif -. Lors de la conception d'une voiture autonome, il est important que le système ait été formé à diverses conditions, y compris des conditions où un dommage pourrait être imminent. Pour que ce processus de formation ait lieu, il serait nécessaire de placer les individus dans des situations à risque, qui peuvent ensuite faire partie de l'ensemble des données.²⁷⁵ Bien sûr, cela pourrait se faire au moins en partie par des simulations n'impliquant pas d'êtres humains physiquement, mais cela met en évidence les tensions qui existent dans l'ordre entre les différents principes éthiques et droits de l'Homme, et les compromis qui sont parfois nécessaires pour que les systèmes d'IA soient aussi « bons » que possible.

Une concurrence loyale

Parmi les différents principes éthiques qui sont couramment mentionnés dans les divers documents et normes éthiques de l'IA, la concurrence loyale est probablement le moins souvent citée. En revanche, il s'agit d'un principe éthique autonome dans le Rapport sur l'éthique. La concurrence loyale fait référence à la nécessité de permettre aux innovateurs, entrepreneurs, ingénieurs en logiciel et autres parties prenantes de la chaîne d'approvisionnement, de bénéficier d'un accès équitable aux données et aux opportunités de créer et de déployer des systèmes d'IA. Ce principe peut, à première vue, être qualifié d'objectif économique déguisé, mais le Rapport éthique le développe sur la base de critères distincts fondés sur les droits de l'Homme et l'éthique. En effet, le Rapport note : qu'une concurrence équitable est nécessaire « *pour l'innovation et le bien-être social. Ainsi, le maintien d'un marché libre avec une concurrence équitable permettrait à tous les acteurs de la chaîne de valeur, en particulier les petites et moyennes entreprises et les startups, de bénéficier et de tirer profit de l'activité.* »²⁷⁶ L'impératif éthique est sociétal et non individuel : pour que la société puisse récolter les bénéfices de la transformation de l'IA, il faut une plus grande innovation, par une diversité d'acteurs.

Comme le souligne le rapport, les défis posés par la concurrence déloyale sont notamment les suivants

- Le système donne un avantage aux concurrents disposant de grandes données.
- Le système est basé sur une vaste base de données accessible à quelques acteurs du marché seulement.

²⁷⁵ Ibid. p.190.

²⁷⁶ Ibid.

- Au cours de ses opérations, le système produit une base de données importante et unique, inaccessible aux concurrents.
- Accords de non-concurrence et coordination automatique entre les entreprises sur la base de systèmes d'IA.

Les notes du Rapport sur l'éthique : « Les concentrations de pouvoir économique peuvent également conduire à des concentrations de pouvoir politique, permettant aux géants de la technologie de dicter les règles du jeu sur le marché. La crainte est que l'influence de ces méga-acteurs sur le marché ne rende difficile l'entrée de nouvelles technologies ou applications sur le marché, et ne compromette l'innovation si cruciale pour l'IA »²⁷⁷. Elles suggèrent en outre que les lois sur la concurrence, les contrats standardisés et les protections des consommateurs soient mis à jour pour répondre aux défis prévus. « A cela, il faut ajouter le défi international, résultant du fait que certains des acteurs clés sont basés aux États-Unis ».²⁷⁸ En bref, une grande partie du défi consiste à permettre aux PME d'avoir accès à de grandes bases de données.

V. L'approche d'Israël pour relever les défis

Israël est conscient des risques et défis potentiels que présente l'utilisation croissante des applications de l'IA et est déterminé à y faire face. Dans le même temps, comme l'indique le Rapport d'initiative nationale, l'expérience montre qu'une réglementation excessive peut étouffer l'innovation, en particulier lorsqu'il s'agit de technologies émergentes²⁷⁹. L'initiative nationale comprenait deux groupes de travail consacrés respectivement à l'éthique et à la réglementation, et à la cybersécurité des systèmes d'IA. Leur principal défi consistait à établir un modèle qui permettrait d'**équilibrer** la nécessité de (1) assurer un développement et un déploiement éthiques et sécurisés des applications d'IA conformément aux valeurs d'Israël en tant que démocratie ; (2) encourager l'innovation technologique et la recherche et le développement scientifiques qui sont fondamentaux pour l'économie israélienne et la sécurité nationale. Comme nous le verrons plus loin, l'approche du Rapport d'initiative nationale est nouvelle dans sa manière de combiner l'éthique, les droits de l'Homme et l'innovation.

i. Six principes éthiques pour l'IA

Reconnaissant que les droits de l'Homme et les considérations éthiques restent primordiaux, le Rapport sur l'éthique énumère « Six principes éthiques pour l'IA » qui devraient éclairer l'élaboration des politiques publiques :²⁸⁰

1. **L'équité** : Lutte pour une égalité substantielle, prévention des préjugés (dans l'information, dans le processus et dans le produit), prévention de la discrimination, et évitement de l'élargissement des écarts socio-économiques et éducatifs.
2. **Responsabilité** :
 - a. **Transparence** : Fournir des informations sur le processus et la prise de décision correspondante.
 - b. **Explicabilité** : Pouvoir expliquer le processus décisionnel du système (au niveau des utilisateurs individuels, ainsi qu'au niveau collectif si le système affecte un groupe, ainsi que pour les opérateurs du système eux-mêmes).
 - c. **Responsabilité éthique et juridique** - à répartir entre les acteurs concernés de la chaîne de valeur, ainsi que la gestion des risques. Détermination des responsabilités pour la définition des règles relatives aux mesures raisonnables de prévention du risque en

²⁷⁷ Ibid. pp.190-191.

²⁷⁸ Ibid.

²⁷⁹ Ben-Israel, I., Matania, E. & Friedman, L. (Eds.) (Sept. 2020). p.32.

²⁸⁰ Nahon K. et al. (2020). p.182.

fonction du contexte et de la gravité estimée du risque, pour la gestion des risques et pour la désignation d'un employé chargé de la gestion des risques.

3. Protéger les droits de l'Homme :

- a. *L'intégrité corporelle : Prévention de toute atteinte à la vie ou à l'intégrité physique.*
- b. *La vie privée : Prévention des atteintes à la vie privée dues à la collecte, à l'analyse et au traitement, au, partage et à l'utilisation nouvelle et différente des informations.*
- c. *L'autonomie : Maintien de la capacité de l'individu à prendre des décisions intelligentes, y compris la prévention de l'influence injuste ou inconsciente sur le comportement individuel.*
- d. *Droits civils et politiques : Y compris le droit de vote, la liberté d'expression et la liberté de conscience religieuse.*

4. Sécurité de l'information et de la cybernétique : *Maintenir les systèmes en état de fonctionnement, protéger les informations qu'ils utilisent et prévenir les abus par un acteur malveillant.*

5. La sécurité : *Prévention des dangers pour les individus et la société et atténuation des dommages.*

- a. *Sécurité interne : En développant l'outil d'AI.*
- b. *Sécurité externe : Pour les environnements et les clients, dans l'utilisation de l'outil.*

6. Maintenir un marché concurrentiel *et des règles de conduite qui facilitent la concurrence.*

À la lumière de ces principes éthiques directeurs, le Rapport d'initiative nationale suggère un modèle réglementaire équilibré basé d'une part, sur l'application de l'intervention réglementaire minimale requise pour maintenir un environnement éthique adéquat, tout en s'abstenant, d'autre part, de toute restriction inutile à l'innovation et au progrès scientifique. En conséquence, le Rapport d'initiative nationale appelle à « *encourager l'autorégulation par l'utilisation des outils développés dans le cadre de l'Initiative [nationale] pour évaluer les risques et identifier à l'avance les défis éthiques aux stades du développement et de la production. Les limites éthiques doivent être intégrées dans les systèmes intelligents, les conduites interdites doivent être définies et les principes éthiques doivent être mis en œuvre au cours du processus d'apprentissage et de formation de ceux qui s'occupent des systèmes d'IA* ». ²⁸¹

ii. Une réglementation équilibrée pour favoriser l'innovation

L'une des principales caractéristiques du Rapport sur l'éthique est son approche de la réglementation. Plutôt que de dresser une liste d'activités qui doivent être réglementées, il adopte une approche systématique de la question, qui comprend trois étapes : (1) cartographie des différents types d'approches réglementaires, avec leurs avantages et inconvénients respectifs ; (2) identification des principaux domaines et activités d'IA qui pourraient bénéficier d'un certain niveau de réglementation, et des risques associés à chacun d'entre eux ; (3) mise en correspondance des différentes approches réglementaires avec les diverses activités d'IA. Le gouvernement dispose ainsi d'une feuille de route pour élaborer des réglementations adaptées et spécifiques à chaque secteur. Les détails de l'approche sont fournis ci-dessous.

Le Rapport sur l'éthique identifie les grandes approches réglementaires suivantes :

1. Législation ou réglementation
2. Prise de décision judiciaire pour interpréter la législation existante ou combler les lacunes

²⁸¹ Ben-Israel, I., Matania, E. & Friedman, L. (Eds.) (Sept. 2020). p.31.

3. Normes professionnelles (par le gouvernement, l'industrie, le monde universitaire ou la société civile)
4. L'autorégulation par des règles éthiques ou des normes professionnelles généralement élaborées par la communauté professionnelle concernée.

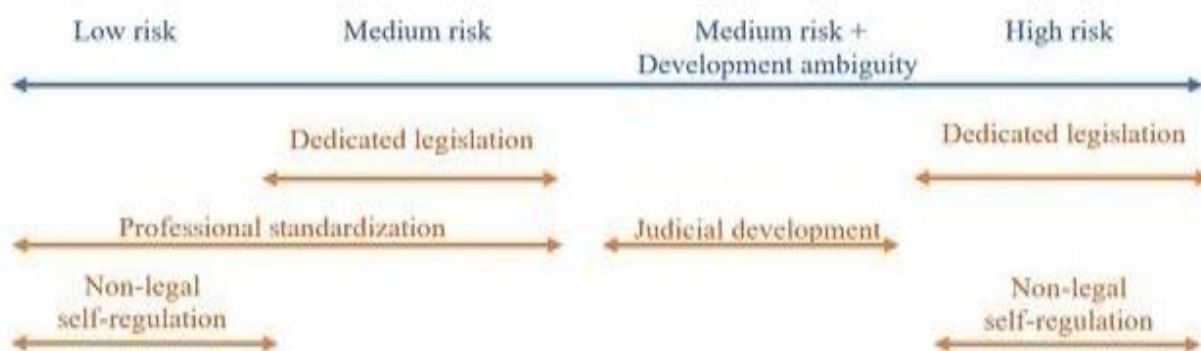
Le Rapport souligne ensuite les avantages et les inconvénients de chaque approche. Le tableau suivant est une traduction de la reproduction tel quel à partir du Rapport sur l'éthique ²⁸²:

| Type de règlement | Caractéristiques | Points forts | Faiblesses | Recommandations du Comité |
|-------------------------------|--|--|--|---|
| Une législation spécifique | Loi ou amendement spécifique appliqué par une autorité publique ou des entités privées | <ul style="list-style-type: none"> • Une plus grande clarté sur les valeurs protégées • Permet un développement judiciaire concret basé sur les directives du législateur • Flexibilité partielle | <ul style="list-style-type: none"> • Manque d'expertise professionnelle dans une seule organisation • Application rétroactive uniquement • Potentiel d'incertitude accrue • Manque d'implication dans les relations de pouvoir actuelles qui peuvent privilégier certains acteurs | Convient principalement aux zones à risque moyen et élevé |
| Développement judiciaire | Pas de loi spécifique | <ul style="list-style-type: none"> • Pas de frictions réglementaires ou juridiques directes • Flexibilité • Permet le développement judiciaire | <ul style="list-style-type: none"> • Généralement applicable à des cas de dommages plus évidents, et peut donc ne pas répondre à toute la gamme des risques de dommages • Manque d'expertise professionnelle dans une seule organisation • Incertitude • Avantage pour les acteurs forts | Convient aux situations à risque moyen présentant une ambiguïté en matière de développement |
| Normalisation professionnelle | Permet l'adoption future par le système juridique | <ul style="list-style-type: none"> • Flexibilité • Une grande légitimité dans la communauté professionnelle • Processus participatif | <ul style="list-style-type: none"> • Risques excluant le droit et ses valeurs • Dépendance à l'égard de la loi pour la validité contraignante, la surveillance et l'application • Avantage pour les acteurs forts | Convient aux situations à risque moyen et faible + comme cadre pour développer et examiner l'application des valeurs éthiques |

²⁸² Nahon K. et al. (2020). p. 200.

| | | | | |
|------------------------------|---|---|---|---|
| Réglementation non juridique | Pas de norme juridique (par exemple, application de principes éthiques) | <ul style="list-style-type: none"> • Flexibilité • Une grande légitimité dans la communauté professionnelle | <ul style="list-style-type: none"> • Risques d'exclusion de la loi et de ses valeurs fondamentales (égalité, équité, droits de l'homme) • Dépendance à l'égard de la communauté professionnelle pour le développement • Absence de mécanisme d'exécution fiable • Avantage pour les acteurs forts | Convient aux situations à faible risque, où une réglementation non juridique est suffisante, et aux situations à haut risque, où l'évolution technologique est relativement rapide pour la voie juridique |
|------------------------------|---|---|---|---|

Le Rapport sur l'éthique propose ensuite le modèle suivant²⁸³, pour faire correspondre différentes approches réglementaires en fonction du niveau de risque associé à une activité particulière :



Ainsi, par exemple, les activités à haut risque sont mieux traitées par la législation et l'autorégulation ex ante que par une intervention judiciaire a posteriori. En revanche, les activités à faible risque ne nécessitent pas nécessairement une législation spécifique et peuvent être traitées par des normes et une autorégulation.

Ce modèle, bien sûr, n'est pas censé s'appliquer de manière rigide. Il présente plutôt un cadre qui permet aux décideurs politiques et aux régulateurs d'évaluer les moyens appropriés d'une activité, en tenant compte d'une multitude de variables. Il note en outre que la question de savoir « qui réglemente » n'est pas moins importante : la réglementation par un organisme central d'IA permet d'élaborer des politiques cohérentes ; toutefois, il existe un risque de surréglementation et de freiner l'innovation si une réglementation est adoptée de manière généralisée. À l'inverse, la réglementation pourrait être laissée à différents organismes sectoriels, ce qui permettrait une plus grande expérimentation, au détriment de l'uniformité des règles.

À la lumière de ce qui précède, le groupe de travail sur l'éthique propose onze lignes directrices réglementaires :

1. **Alignement de la réglementation israélienne sur la législation et la normalisation internationales, et promotion de la politique israélienne sur la scène internationale** - il s'agit essentiellement de participer au débat international sur la réglementation de l'IA,

²⁸³ Modèle en version originale non traduit

afin d'être en phase avec les normes internationales émergentes, tout en prenant part à l'élaboration de ces normes à l'avenir.

2. **Cartographier les acteurs pour créer un cadre de responsabilité et d'incitation adapté** - cela nécessite une approche multipartite, pour permettre aux décideurs politiques de comprendre leurs rôles respectifs dans la chaîne de valeur, leurs incitations et leur responsabilité.
3. **Adapter le principe de responsabilité au dynamisme du domaine de l'IA** - il est suggéré ici d'exiger des organisations mettant en œuvre des technologies d'IA qu'elles mettent en place un environnement de test et des périmètres de contrôle avant de mettre en œuvre la technologie, afin de déterminer la meilleure façon d'appliquer le principe de responsabilité dans un cas donné en tenant compte des effets anticipés de la technologie.
4. **Promouvoir la clarté normative dans les étapes critiques de la chaîne de valeur des produits d'IA** - cela souligne l'importance de l'orientation dans les premières étapes du développement de l'IA. Un outil d'évaluation des risques de l'IA, et peut-être dans certains cas une exigence réglementaire pour l'évaluation de l'impact de l'IA, serait utile pour atténuer les risques et permettre aux développeurs de mettre en œuvre les différents principes éthiques et règles juridiques.
5. **Révision constante de la politique réglementaire par le régulateur** - au-delà du contrôle de la mise en œuvre de la réglementation existante et de la mise à jour des textes juridiques, ce principe appelle à l'expérimentation réglementaire. Il exige que les régulateurs adoptent une approche souple et agile, qu'ils encouragent l'innovation tout en tenant compte des risques.
6. **Bacs à sable réglementaires** - le concept de bacs à sable réglementaires est bien connu. Les tests contrôlés sont particulièrement utiles dans un contexte d'IA « en raison de la nécessité de permettre l'innovation, d'une part, et de faire face aux risques imprévisibles pour les intérêts sociaux, d'autre part ».
7. **L'interface entre les principes proposés et les réglementations existantes** - étant donné que les lois et les réglementations s'appliquent déjà dans de nombreux domaines d'activité (santé, transport, finances, éducation, etc.), le paysage juridique existant doit être pris en compte, ainsi que les valeurs, les intérêts et les avantages sociaux potentiels spécifiques de la réglementation, pour déterminer si une nouvelle réglementation est nécessaire et quels devraient être son objectif et sa portée.
8. **Le rôle de l'Autorité de protection de la vie privée** - La protection de la vie privée est une question intersectorielle, de sorte que l'Autorité de protection de la vie privée a un rôle prédominant à jouer dans l'évaluation des implications des systèmes d'IA sur la vie privée, et dans l'élaboration de la réglementation si nécessaire, en coordination avec d'autres organismes gouvernementaux. En outre, il est important que l'Autorité de protection de la vie privée obtienne les ressources nécessaires pour développer un cadre juridique et technologique actualisé dans le domaine de l'anonymisation des informations, car il s'agit d'un élément fondamental du développement de l'IA.
9. **Le rôle de l'Autorité de la concurrence** - comme mentionné ci-dessus, la concurrence et un marché équitable ne sont pas seulement une question d'économie - c'est aussi une question d'éthique. Ainsi, l'Autorité de la concurrence devrait être chargée de « formuler des réglementations visant à maintenir une concurrence loyale dans le domaine de l'IA, à protéger les consommateurs et à garantir l'accessibilité des technologies ; et à empêcher que les risques et les coûts technologiques ne soient répercutés sur les acteurs plus faibles au bas de la chaîne de valeur, d'une manière socialement inefficace ».
10. **La nécessité d'une coordination interministérielle** - pour assurer la cohérence de l'élaboration des politiques et des réglementations, un mécanisme de coordination interministérielle devrait être mis en place

11. Autorités responsables des ressources d'information

Les autorités responsables des ressources d'information de fond utilisées pour les technologies d'IA « ont un rôle clé à jouer pour déterminer si le cadre réglementaire qu'elles appliquent est adapté pour obtenir un avantage sociétal dans ce domaine, tout en maintenant un marché concurrentiel libre et équitable et en protégeant les droits de l'homme. À cet égard, il convient de tenir compte non seulement des risques, mais aussi des espaces d'innovation et [...] de la promotion des intérêts de la société.

*Le Comité recommande donc que les autorités responsables des domaines d'activité touchés par les produits du traitement de l'information soient tenues de se soumettre à une évaluation à la lumière des principes détaillés ci-dessus. Plus précisément, les autorités doivent examiner si, lors du déploiement ou de l'utilisation des technologies de l'IA dans les domaines d'activité qu'elles réglementent, il est nécessaire d'adapter le cadre applicable afin de promouvoir la protection des intérêts réglementés.»*²⁸⁴

iii. Outil original d'évaluation du risque éthique

Comme l'approche israélienne encourage l'autorégulation, le Rapport éthique souligne la responsabilité de tous ceux qui sont impliqués dans l'IA de rester à jour avec les risques de la technologie dynamique. Pour les aider à s'acquitter de cette tâche exigeante, le groupe de travail « Éthique et réglementation » a élaboré un instrument original destiné aux décideurs pour évaluer les défis éthiques. Cet instrument est conçu pour permettre aux professionnels de l'IA d'identifier les risques éthiques tout au long des changements de développement et de production et d'y répondre correctement. Il se compose de deux parties :

Une série de questions préliminaires qui doivent être adressées aux développeurs de produits d'IA afin d'évaluer l'influence du produit :²⁸⁵

1. *Quel est le niveau de préjudice individuel potentiel ?*
2. *Quelle est l'étendue de l'impact perceptuel potentiel ?*
3. *Quel est le degré de préjudice potentiel pour le public ?*
4. *Y a-t-il un impact sur l'allocation des ressources publiques ?*
5. *L'équipe de développement est-elle suffisamment diversifiée ?*
6. *Quelle est l'étendue prévue des dommages dus à une mauvaise utilisation ou à une perte de contrôle du produit ?*
7. *Existe-t-il un moyen rapide d'identifier les défaillances éthiques imprévues ?*

Une carte de fréquence dynamique qui aide à localiser les zones difficiles en termes d'application des valeurs éthiques au développement du système. La carte présente les six principes éthiques juxtaposés aux étapes cardinales du processus de développement. Elle indique la fréquence des problèmes éthiques tout au long de la chaîne de développement du produit en mettant en évidence les domaines où des défaillances ont été constatées dans le passé et en fournissant des informations sur leur taux d'incidence (voir l'annexe II pour un exemple de carte de fréquence). La carte est basée sur l'évaluation de cas réels de systèmes d'IA ayant présenté des défis ou des conflits éthiques dans le passé, ce qui permet de sensibiliser aux domaines dans lesquels les organisations d'IA ont connu des difficultés dans le passé et aux domaines auxquels les décideurs doivent accorder une attention particulière. Il est important de noter que la carte étant façonnée par les cas tests utilisés pour la créer, chaque organisation est censée sélectionner un ensemble de cas pertinents pour le produit qu'elle développe. En outre, pour rester pertinente, la carte doit être fréquemment mise à jour avec de nouveaux cas de test. Le Rapport d'éthique explique en détail

²⁸⁴ Ibid. p.205.

²⁸⁵ Ibid. pp.192-193.

comment une organisation d'IA peut créer et mettre à jour une carte de fréquence pertinente pour son produit. Voir l'annexe II pour de plus amples informations.

iv. **Activité et coopération internationales**

Israël a participé à des forums internationaux traitant de l'éthique et des droits de l'Homme dans le domaine de l'AI. Les représentants israéliens ont participé activement à la rédaction des recommandations et des principes directeurs de l'OCDE en matière d'IA. En outre, Israël est membre des « Nations Digitales » ("DN"), qui regroupent dix des principales économies numériques du monde. En 2018, Israël a accueilli la réunion annuelle des DN, au cours de laquelle une déclaration sur l'IA responsable a été adoptée.²⁸⁶ En 2019, la DN a également adopté une déclaration sur la gouvernance des données.²⁸⁷ Bien que ces déclarations ne soient pas juridiquement contraignantes, elles reflètent l'engagement des Nations Digitales à respecter des normes élevées en matière de droits de l'Homme, d'éthique et de responsabilité dans leur utilisation de la digitalisation. Israël s'est également associé au projet C4IR du Forum économique mondial, en menant des projets de recherche dans les domaines du transport et de la santé.²⁸⁸

²⁸⁶ Shared Approach on the Responsible Use of AI, https://fdfd812d-4234-49d8-8755-ff45ad565157.filesusr.com/ugd/189d02_ef802d92ba5147d2901bde25c6e954a3.pdf, novembre 2018, Rédigé par le groupe de travail sur l'intelligence artificielle, ce cadre a été adopté lors du sommet ministériel du D9 en Israël en 2018.

²⁸⁷ Déclaration de Data 360, https://fdfd812d-4234-49d8-8755-ff45ad565157.filesusr.com/ugd/189d02_abce8f2b8cc140e4baeec7dcab7bee97.pdf, novembre 2019. Rédigée par le groupe de travail Data 360, cette déclaration commune a été présentée lors du sommet ministériel du D9 en Uruguay en 2019.

²⁸⁸ Autorité israélienne de l'innovation. *Création du Centre israélien pour la quatrième révolution industrielle - Forum économique mondial*. Consulté sur <https://innovationisrael.org.il/en/contentpage/establishment-israeli-center-fourth-industrial-revolution-world-economic-forum>.

Annexe I. Le paysage israélien des start-ups d'IA segmenté par secteurs et applications



CARDUMEN CAPITAL

Cardumen Capital updates this map quarterly. If you would like to add your startup or reach out to us regarding our expertise in artificial intelligence, please email us to AI@cardumencapital.com

CARDUMEN CAPITAL

Source : Cardumen Capital (juin 2020). <https://www.cardumencapital.com/ai-israel-landscape>

Annexe II. Carte de fréquence des défis éthiques dans le processus de développement de l'IA

La partie suivante est la traduction de la reproduction telle quelle des pages 194-197 du rapport sur l'éthique

Carte des fréquences

La carte de fréquence indique la fréquence des problèmes éthiques tout au long de la chaîne de développement du produit. Elle indique les domaines où des défaillances ont été constatées dans le passé et fournit des informations sur leur taux d'incidence. Comme la fréquence peut changer avec le temps et les nouveaux événements découverts, nous recommandons de mettre à jour la carte régulièrement, comme démontré ci-dessous.

Pour créer la carte des fréquences, nous avons utilisé dix cas tests sélectionnés parmi des cas réels du passé qui représentent divers défis. La carte illustre tous les principes éthiques énumérés dans la section « Principes éthiques pour l'IA » à la page 138 ci-dessus.

Tableau 1 : Cas types de défis éthiques

| | |
|---|---|
| 1 | Système d'IA pour la sélection des candidats sur le lieu de travail Les entreprises sont contactées par de multiples candidats qui souhaitent travailler pour elles. Afin de sélectionner les meilleurs candidats, plusieurs entreprises ont mis au point des outils basés sur l'IA et formés en fonction des décisions prises par les entreprises dans le passé. Lorsque l'un de ces systèmes mis au point par Amazon a été testé, il s'est avéré discriminatoire à l'égard des femmes candidates à un emploi technique. On suppose que dans le passé, les dirigeants des entreprises avaient l'habitude de pratiquer cette discrimination, et le système a appris à imiter ce comportement ²⁸⁹ . |
| 2 | Utiliser l'IA pour exercer une influence politique Cambridge Analytica a collecté les données personnelles de millions de profils Facebook sans l'accord ou la connaissance des utilisateurs, et les a utilisées pour influencer les utilisateurs à des fins politiques. Il y a probablement eu recours à la technologie de l'IA pour manipuler les esprits. Cette activité s'est poursuivie pendant plusieurs années ²⁹⁰ . |
| 3 | Prévoir le risque de maladie Au cours des années 1990, plusieurs centres de recherche se sont associés pour mettre au point un système permettant d'estimer dans quelle mesure la pneumonie représente un risque pour la vie de certains patients. Ce système a été conçu pour aider les médecins à décider quels patients doivent être hospitalisés et lesquels peuvent être traités dans la communauté. Peu avant le lancement du système, on a constaté que les recommandations qu'il contenait pour les asthmatiques pouvaient mettre leur vie en danger, car les informations utilisées pour construire le système étaient biaisées : les asthmatiques atteints |

²⁸⁹ <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>

²⁹⁰ <https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data>,
<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>

| | |
|---|---|
| | de pneumonie avaient reçu des soins intensifs préliminaires qui leur avaient sauvé la vie, et le système en avait déduit que la pneumonie ne présentait pas de risque pour les asthmatiques. ²⁹¹ |
| 4 | <p>Système d'évaluation de la dangerosité des détenus</p> <p>Lorsqu'on décide de placer un détenu en détention provisoire, l'une des considérations est le danger qu'il représente pour les autres. La décision est basée sur de multiples paramètres, tels que les antécédents criminels. Plusieurs districts des États-Unis ont adopté un système d'IA appelé Compas pour aider les juges à évaluer la dangerosité des suspects. Ce système a été testé et il a été constaté que les détenus blancs étaient moins dangereux que les détenus noirs. ²⁹²</p> |
| 5 | <p>Les joueurs virtuels guidés par l'IA impossibles à battre par des joueurs humains</p> <p>Dans un jeu appelé Elite Dangerous, des joueurs humains s'affrontent à des joueurs guidés par l'IA. Pour rendre le jeu plus intéressant, les restrictions imposées aux joueurs virtuels ont été modifiées dans la version 2.1, afin de leur permettre de mieux voler et de mieux se battre. Les mécanismes d'IA ont trouvé un moyen de tirer profit de ces changements pour accumuler des armes d'une manière qui empêche les utilisateurs humains de pouvoir les égaler. ²⁹³</p> |
| 6 | <p>Le bot raciste</p> <p>Microsoft a lancé un bot afin de lui apprendre à correspondre librement avec les utilisateurs de Twitter. L'idée était que le bot puisse engager la conversation et apprendre à améliorer ses compétences en matière de dialogue au cours du processus. Moins de 24 heures après le lancement, il s'est avéré que puisqu'il imitait les utilisateurs, plusieurs d'entre eux ont choisi d'en faire un bot raciste en utilisant eux-mêmes des commentaires racistes. ²⁹⁴</p> |
| 7 | <p>Le bot imitateur</p> <p>Le Google Duplex permet à un bot de tenir une conversation d'une manière qui rend difficile pour ses interlocuteurs de déterminer si elle est humaine. La construction de cet outil a nécessité l'accès à d'énormes quantités de données qui ne sont accessibles qu'à un très petit nombre d'entreprises à forte intensité de connaissances. ²⁹⁵</p> |
| 8 | <p>Une voiture autonome écrase un piéton</p> <p>Un piéton qui traversait la rue dans une zone sombre a été tué en Arizona par un véhicule autonome Uber. Apparemment, le véhicule a identifié un « obstacle » et aurait pu éviter de le percuter. Néanmoins, comme les ingénieurs avaient auparavant réduit la sensibilité du logiciel aux obstacles, le véhicule ne s'est pas arrêté et la femme a été tuée. Le conducteur humain du véhicule n'était pas assez vigilant pour éviter l'accident. ²⁹⁶</p> |

²⁹¹ <http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf>

²⁹² <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

²⁹³ <https://futurism.com/this-video-games-artificial-intelligence-turned-on-players-using-super-weapons>

²⁹⁴ <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>

²⁹⁵ <https://www.androidcentral.com/google-duplex-will-let-people-know-its-not-human>

²⁹⁶ <https://www.cbc.ca/news/business/uber-arizona-crash-1.4594939>

| | |
|-----------|--|
| 9 | <p>Biais dans la reconnaissance des visages</p> <p>Amazon a développé un outil pour les ingénieurs leur permettant d'ajouter la capacité de reconnaissance des visages au système qu'ils développaient. Le système a été conçu, entre autres, pour être utilisé par les forces de l'ordre, la police des frontières, etc. Un test a révélé que le système commettait des erreurs beaucoup plus fréquemment lorsqu'il était activé sur des personnes à la peau foncée que sur des personnes à la peau claire. ²⁹⁷</p> |
| 10 | <p>Les systèmes de recommandation de contenu montrent des informations différentes à des groupes différents</p> <p>Diverses entreprises utilisent l'IA pour offrir aux utilisateurs des informations plus pertinentes sur le plan personnel. Il a cependant été constaté que le système d'annonces de Google présente des annonces recherchant des informations liées à des actes criminels lorsqu'un utilisateur recherche des informations sous un nom plus courant dans les populations minoritaires. ²⁹⁸</p> |

Les étapes éthiques du processus de développement

Nous présentons ci-dessous des exemples de questions éthiques qui se posent au cours du processus de développement et nous en assurons le suivi au fur et à mesure, afin d'identifier les étapes de développement particulièrement sensibles. Pour ce faire, nous présentons un processus de développement typique de l'IA.

1. Définition du produit

- a. Comprendre le besoin ou le problème commercial que le système tente de résoudre et créer l'organisation de R&D
- b. Collecte de données - identification des sources d'information internes et externes à l'organisation qui seront utilisées pour construire le système et évaluer ses performances

2. Formation sur les produits

- a. Traitement et filtrage des données brutes sous une forme qui permettrait aux algorithmes d'IA de recevoir les données et d'effectuer des calculs avec celles-ci
- b. Modélisation - application d'un algorithme d'IA au traitement de l'information pour tenter d'identifier des modèles pouvant être généralisés

3. Intégration

- a. Évaluer la précision du modèle
- b. Connecter les composants de l'IA avec le reste du système et le distribuer pour une large utilisation

4. Gestion du marché

- a. Suivi des performances pour s'assurer que le système fonctionne comme prévu

²⁹⁷ <https://www.theverge.com/2019/1/25/18197137/amazon-rekognition-facial-recognition-bias-race-gender>

²⁹⁸ <https://www.bostonglobe.com/business/2013/02/06/harvard-professor-spots-web-search-bias/PtOgSh1ivTZMfyEGj00X4l/story.html>

- b. Écosystème - en plus du processus au sein de l'organisation, il faut également tenir compte des considérations éthiques découlant du fait que le processus se déroule dans l'écosystème israélien. L'intégration de l'IA pourrait affecter les systèmes socio-économiques, réglementaires et autres, et cela devrait être surveillé en permanence après le lancement.

Création de la carte des fréquences

Examinez la liste des cas types et les implications et événements signalés en tenant compte de la liste des valeurs éthiques figurant à la page 8 ci-dessus. Remplissez le tableau en fonction des nouveaux défis éthiques. Les chiffres figurant dans les cellules du tableau renvoient au numéro de l'événement. Ensuite, vérifiez le nombre d'événements cumulés. Les cellules ayant une fréquence d'événements faible, moyenne et élevée sont respectivement colorées en beige, jaune et rouge. Notez que ce tableau n'indique pas le degré et l'étendue des dommages potentiels. Un outil plus sophistiqué peut également tenir compte de ces facteurs. La décision concernant ce qui constitue une fréquence faible ou élevée doit être prise lors de la sélection du nombre d'événements auxquels l'organisation se réfère. Dans le tableau 2, nous avons dix événements, et les fréquences ont été déterminées en conséquence.

Le Comité recommande que les décideurs discutent et proposent des solutions aux nouveaux défis selon la carte de fréquences tout au long de leur processus de développement. Étant donné que la carte dépend d'une liste de cas de test, chaque organisation doit choisir un ensemble de cas de test pertinents pour le produit en cours de développement, en supposant que cet ensemble change dans le temps.

Tableau 2 : Carte de fréquence des défis éthiques dans le processus de développement de l'IA

| | Besoin des entreprises | Collecte de données | Organisation des données | Modélisation | Évaluation du modèle | Distribution | Suivi des performances | Ecosystème |
|-------------------|------------------------|---------------------|--------------------------|--------------|----------------------|--------------|------------------------|------------|
| Équité | | 1,3,4 | | 3,4 | 1,3,4 | 1,4,9 | 1,3,4,9 | 1,4,9 |
| Transparence | 4 | | | 3,4 | | | | 4 |
| Explicabilité | 4 | | | 3,4 | | | | 9 |
| Responsabilité | 1,2,3,4 | | | | 1,3,4 | 3,4 | 3,4,5,6 | 2,5,6,9 |
| Vie privée | 2,9 | 1 | 1,2 | 1 | 1 | | | 2,9 |
| Liberté de choix | 7,10 | | | | | | 10 | 6,7,10 |
| Infosécurité | | | 2 | | | | | 2,9 |
| Droits de l'homme | 4,9 | | | 4 | | 4 | 4 | 4,9 |
| Sécurité | 3,4 | 3,4 | | 3,4 | 3,4 | 3,4,5,8 | 3,4,5 | 3,5,8 |
| Marché libre | 5 | | | | | 5,6 | 5 | |

Légende

| | |
|---|---|
| 1 | Présélection des candidats à l'emploi |
| 2 | Influence politique |
| 3 | Prévoir le risque de maladie |
| 4 | Évaluation de la dangerosité des détenus |
| 5 | Les joueurs guidés par l'IA obtiennent des armes de départage |
| 6 | Bot raciste |
| 7 | Bot imitateur |
| 8 | Un véhicule autonome écrase un piéton |
| 9 | Biais dans la reconnaissance des visages |

Faible fréquence des cas problématiques
(cas unique)

Fréquence moyenne des cas problématiques
(deux cas)

Fréquence élevée des cas problématiques
(trois cas ou plus)

CHAPITRE II. La gouvernance de l'IA au Japon

Arisa Ema²⁹⁹ & Hideaki Shiroyama³⁰⁰ (Université de Tokyo)

I. Introduction

Différents acteurs sont impliqués dans le processus de recherche et de développement des nouvelles technologies et de leur utilisation dans la société. À chaque étape de la recherche, du développement et de l'utilisation d'une technologie, chaque acteur prend des décisions et des politiques pour sa mise en œuvre sociale, sur la base d'une évaluation de son impact social. L'ensemble des rôles des différents acteurs dans l'évaluation de l'impact social, la prise de décision et l'élaboration de politiques, et la mise en œuvre mentionnées ci-dessus peut être appelé gouvernance technologique³⁰¹. Actuellement, il existe différentes approches de la gouvernance des technologies de l'intelligence artificielle (IA) au Japon et à l'étranger. Il n'existe pas de réponse unique et appropriée concernant les systèmes et les moyens de mise en œuvre de la technologie dans la société, et en fait, cette question est liée à une combinaison de différents facteurs, notamment la technologie, la culture, la politique, les valeurs des gens, les systèmes juridiques existants, l'environnement et l'économie. En conséquence, les débats dans ces tentatives de gouvernance sont divers.

Bien qu'il soit difficile de couvrir toutes les politiques et activités, ce rapport résume la nature de la gouvernance de l'IA et les caractéristiques des discussions au Japon. Étant donné que ce rapport traite principalement des tentatives et des discussions de gouvernance jusqu'en 2018³⁰², il est facile d'imaginer que les détails qu'il contient changeront avec les développements technologiques futurs et les changements des conditions sociales. Cependant, l'organisation de discussions à un point fixe est utile pour les futures discussions sur la gouvernance de l'IA et pour la recherche comparative avec d'autres gouvernances de technologies de pointe.

Ce rapport est organisé comme suit. Le chapitre II « Gouvernance de l'IA au Japon » organise d'abord les activités et les rapports de chaque acteur de l'industrie, des universités et du gouvernement. Il est suivi d'une analyse des livres blancs et des rapports publiés par les ministères et les agences pour les exercices 2016-2017. Un aperçu et une comparaison de la manière dont les ministères positionnent l'IA et les technologies connexes (par exemple, le Big Data et l'Internet des Objets (IdO)) sont fournis, en plus d'une section sur la manière dont ils perçoivent leurs domaines d'utilisation et leurs défis. Dans le chapitre III, « Comparaison de la gouvernance de l'IA au Japon et à l'étranger », nous résumons les questions et les perspectives qui manquent dans les discussions sur la gouvernance de l'IA au Japon. Enfin, dans le chapitre IV, nous résumons la manière dont les discussions sur la gouvernance de l'IA devraient être traitées à l'avenir.

²⁹⁹ Professeur assistant de projet, *Institute of Future Initiative*, Université de Tokyo.

³⁰⁰ Professeur, École supérieure de politique publique, Université de Tokyo.

³⁰¹ Hideaki Shiroyama, *Science, technologie et politique* (Minerva Shobo, 2018), chapitres 7 et 9 (en japonais).

³⁰² Ce rapport est basé sur une traduction partielle d'un chapitre "AI Governance" écrit par Arisa Ema et Hideaki Shiroyama, dans *Artificial Intelligence, Humanity and Society* (Keisho Shobo, 2020) en japonais (<https://www.keisoshobo.co.jp/book/b498075.html>). Certaines références japonaises sont remplacées par des versions anglaises (traduite en français) et complétées par des informations à partir de 2020 par l'auteur et ses collègues.

II. Gouvernance de l'IA au Japon

i. Le rôle de chaque acteur

Ce chapitre résume les discussions sur la gouvernance de l'IA au Japon par différents acteurs.

Sociétés académiques

En février 2017, le comité d'éthique de la Société japonaise d'intelligence artificielle (JSAI) a publié ses directives éthiques. Il ne s'agit pas d'une éthique pour la technologie de l'IA, mais plutôt d'une éthique pour les chercheurs en IA, car ces directives fournissent des lignes directrices pour le comportement qui doit être adopté en tant que chercheur et comprennent des sections telles que la responsabilité sociale et le respect des lois et règlements³⁰³.

En outre, la Société japonaise de traitement de l'information (IPJS) a créé le groupe d'experts du SC 42 sur l'intelligence artificielle en janvier 2018 pour travailler avec le SC 42, un sous-comité établi par l'ISO/IEC JTC 1 (comité technique conjoint de l'Organisation internationale de normalisation et de la Commission électrotechnique internationale) et la Société a diffusé les avis du Japon sur le sujet³⁰⁴. En outre, le Conseil de promotion académique au sein de l'Association médicale japonaise a publié un rapport intitulé "Intelligence artificielle (IA) et médecine" en juin 2018³⁰⁵.

Universités

Il existe de nombreux instituts de recherche dans les universités japonaises dont le nom des facultés et des départements comporte les mots « IA » et « intelligence ». Par exemple, l'université de Tokyo dispose également d'un centre de recherche sur l'intelligence artificielle de nouvelle génération, qui a accueilli le symposium "*AI and Society*" en 2017³⁰⁶, et a co-organisé le *Beneficial AI Tokyo* avec la CFI au Royaume-Uni, auquel des membres de l'IEEE, de la PAI et d'autres institutions de recherche et entreprises concernées³⁰⁷ ont été invitées.

Organisations industrielles

La Fédération des entreprises du Japon (Keidanren)

La Fédération des entreprises japonaises (Keidanren) a également lancé des discussions sur l'impact de l'IA et d'autres technologies avancées. En mai 2018, la fédération a publié une proposition intitulée « Création de nouvelles valeurs par la mise en œuvre de trois principes commerciaux pour la réalisation des SDG - vers la formulation d'une vision stratégique des DPI »³⁰⁸. La proposition indique que « le Japon a pris du retard dans l'utilisation des données dans les entreprises » et qu'il est important de regagner de la compétitivité internationale en combinant les données avec l'IA et d'autres technologies. La proposition met en avant le concept de société 5.0 dans la résolution des problèmes sociaux et invite à promouvoir les entreprises conformément aux objectifs de développement durable (SDG) des Nations unies. En ce qui concerne l'utilisation des données, la proposition note qu'« il est nécessaire de construire des systèmes qui permettent à

³⁰³ Le comité d'éthique a déclaré qu'il avait élaboré ces lignes directrices pour gagner d'abord la confiance de la communauté des chercheurs avant d'aborder l'impact de l'IA sur la société et les perspectives éthiques. Dans les "Directives éthiques de la Société japonaise d'intelligence artificielle" (<http://ai-elsi.org/archives/514>), le professeur Yutaka Matsuo, alors président du comité d'éthique, a déclaré : "Le professeur Shun Tsuchiya a dit cela : "Le grand public s'inquiète de ce que les chercheurs en intelligence artificielle feraient de cette technologie. Il est donc important de faire savoir d'abord que les chercheurs visent à créer une société meilleure, et qu'ils ne sont pas des scientifiques fous. Je tiens donc à féliciter la JSAI d'avoir publié de telles directives éthiques. "

³⁰⁴ Société de traitement de l'information, Lancement de la normalisation internationale sur l'intelligence artificielle, https://www.ipsj.or.jp/release/20180110_itscjnews.html (en japonais)

³⁰⁵ Intelligence artificielle (IA) et médecine", un rapport de la IX^{ème} réunion du Conseil de promotion académique de l'Association médicale du Japon. http://dl.med.or.jp/dl-med/teireikaiken/20180620_3.pdf (en japonais)

³⁰⁶ IA et société, <http://www.aiandsociety.org/>

³⁰⁷ Par la suite, *Beneficial AI Japan* (<http://bai-japan.org>) a été créé. Notez que le « *Next Generation Intelligence Science Center* » a rejoint le *Partnership on AI* (PAI) en août 2018.

³⁰⁸ Créer de la nouvelle valeur grâce aux trois principes de la société 5.0 *Enabled Business*, 15 mai 2018, http://www.keidanren.or.jp/policy/2018/042_honbun.html (en japonais)

différentes entités d'utiliser les données tout en tenant dûment compte de la vie privée et de la cybersécurité, et de susciter l'innovation dans divers domaines », et ajoute que la participation active à l'élaboration de règles, y compris les réglementations et normes internationales, et au développement des ressources humaines est un défi.

Lors de la première réunion du Conseil des principes sociaux de l'IA centrée sur l'Homme du Cabinet du Premier Ministre, le 8 mai 2018, un document intitulé « Formulation des principes d'utilisation de l'IA pour la mise en œuvre de la société 5.0 » a été fourni par le Keidanren³⁰⁹, qui a soumis des informations sur la formulation d'une version industrielle des lignes directrices de développement de l'IA et du concept stratégique de déploiement³¹⁰.

Instituts nationaux de recherche

Centre de recherche sur l'intelligence artificielle, Institut national des sciences et technologies industrielles avancées

Il a été créé en mai 2015 et, en octobre 2018, 12 équipes de recherche travaillent sur différents projets³¹¹. Des séminaires sont également organisés régulièrement, notamment une discussion sur « l'IA et le droit » en décembre 2017³¹².

Projet RIKEN Center for Advanced Intelligence

Créé en 2016, le projet Riken compte trois groupes de recherche, le groupe de recherche sur la technologie générique, le groupe de recherche sur la technologie orientée vers un but précis et le groupe de recherche sur l'intelligence artificielle dans la société. Le groupe de recherche sur l'intelligence artificielle dans la société se compose de sous-équipes telles que l'équipe chargée de la législation sur l'information et l'équipe chargée de l'éthique et de la société de l'intelligence artificielle, qui effectuent non seulement des recherches techniques, mais aussi des recherches d'enquête dans les domaines³¹³ des sciences humaines et sociales.

Centre de promotion de la recherche et du développement scientifique de l'IA, Institut national des technologies de l'information et des communications

Le Centre a été créé en 2017 en tant qu'organisation de promotion de l'innovation ouverte et de la recherche et développement stratégique pour les technologies de l'IA dans le domaine des sciences du renseignement³¹⁴. Le centre partage une variété de données textuelles et vocales, et construit un banc d'essai de données d'IA de pointe, pour accélérer l'innovation ouverte.

Agence japonaise pour la science et la technologie (JST), Institut de recherche sur la science et la technologie pour la société (RISTEX)

Le RISTEX a créé le domaine d'intervention « Écosystème humain et technologies de l'information » (HITE) en 2016³¹⁵. L'objectif de ce domaine est de promouvoir la coévolution de la technologie et de la société en organisant les questions qui se posent lorsque de nouvelles technologies de l'information, telles que l'IA, sont introduites dans la société et en créant un lieu et une plate-forme où ces questions peuvent être directement renvoyées aux mêmes domaines de recherche et de développement.

³⁰⁹ Keidanren, Stratégie d'utilisation de l'IA pour une société prête pour l'IA, 19 février 2019, https://www.keidanren.or.jp/en/policy/2019/013_outline.pdf.

³¹⁰ Hiroaki Kitano, « Mission Statement Keidanren AI Application Principles Task Force », <http://www8.cao.go.jp/cstp/tyousakai/humanai/1kai/sanko1.pdf> (en japonais)

³¹¹ Centre de recherche sur l'intelligence artificielle : <https://www.airc.aist.go.jp/teams/>

³¹² Centre de recherche sur l'intelligence artificielle [19e séminaire sur l'intelligence artificielle] "IA et droit" https://www.airc.aist.go.jp/seminar_detail/seminar_019.html (en japonais)

³¹³ Centre de recherche sur l'intégration de l'intelligence innovante (AIP) : <https://www.riken.jp/en/research/labs/aip/index.html>

³¹⁴ Centre pour la promotion de la recherche intégrée et le développement de la science intelligente : <http://www2.nict.go.jp/oihq/en/>

³¹⁵ Site web RISTEX-HITE : <https://www.jst.go.jp/ristex/hite/en/>

Organisations gouvernementales

Bureau du cabinet

En avril 2016, le Conseil stratégique pour la stratégie d'IA a été établi en réponse à la directive du Premier ministre dans le "Dialogue entre les secteurs public et privé pour les investissements futurs". Le conseil sert de poste de commandement. En plus des trois ministères (ministère de l'Intérieur et des Communications, ministère de l'Éducation, de la Culture, des Sports, des Sciences et des Technologies, et ministère de l'Économie, du Commerce et de l'Industrie), la recherche et le développement ainsi que la mise en œuvre sociale sont actuellement encouragés par la coopération et la collaboration avec les ministères et les organismes connexes tels que le Bureau du Cabinet, le ministère de la Santé, du Travail et des Affaires sociales, le ministère du Territoire, des Infrastructures, des Transports et du Tourisme, et le ministère de l'Agriculture, des Forêts et de la Pêche. Le concept de Société 5.0 et de "société super-intelligente" a été proposé comme des avantages sociétaux qui devraient être réalisés.

En mars 2017, le Une stratégie technologique en matière d'intelligence artificielle a été formulée, et une feuille de route a été créée pour l'industrie. Selon cette feuille de route, les agences nationales sont censées travailler sur (1) la recherche et le développement, (2) le développement des ressources humaines, (3) le développement d'un environnement pour les données et les outils, (4) le soutien aux start-ups, et (5) la promotion d'une compréhension de la technologie de l'intelligence artificielle. Le projet de plan de mise en œuvre de la stratégie technologique en matière d'intelligence artificielle a été élaboré³¹⁶ en août 2018.

Au cours de l'exercice 2016, le Cabinet du Premier Ministre a créé le Conseil consultatif sur l'intelligence artificielle et la société humaine, qui a regroupé plusieurs questions dans un seul rapport³¹⁷, et le Conseil des principes sociaux de l'intelligence artificielle centrée sur l'être humain, qui a été créé au sein du Cabinet du Premier Ministre en mai 2018, ont préparé un projet de résumé des principes sociaux de l'intelligence artificielle centrée sur l'être humain à la fin de l'année 2018³¹⁸. En mars 2019, les principes sociaux de l'IA centrée sur l'être humain ont été publiés par le secrétariat du cabinet en février 2019³¹⁹, et les principes ont été introduits lorsque le Japon a accueilli le G20 en juin 2019.

Ministère des affaires intérieures et des communications (MIC)

L'Institut pour la politique de l'information et de la communication (IICP) du ministère de l'Intérieur et des Communications est un organisme de recherche qui effectue des recherches sur la politique de l'information et de la communication, et qui mène des recherches sur l'IA. En octobre 2016, l'IICP a créé la Conférence vers une société de réseau d'IA dans le but d'examiner les questions sociales, économiques, éthiques et juridiques pour la promotion du réseau d'AI. Cette conférence inclut la participation de parties prenantes de l'industrie et d'autres secteurs, en plus d'experts dans les domaines de la science et de la technologie, des sciences humaines et sociales. Le Comité sur les principes de R&D en IA et le Comité sur l'évaluation de l'impact et des risques ont été créés dans le cadre de la conférence. En juillet 2017, la conférence a publié les lignes directrices pour la R&D en IA. Ces lignes directrices ont été publiées dans le but de contribuer aux discussions internationales et les lignes directrices ont été présentées au G7 et à d'autres conférences. La conférence de l'IICP a également publié le projet de lignes directrices pour l'utilisation de l'IA, qui a été publié sous le nom de lignes directrices pour l'utilisation de l'IA en 2019³²⁰.

³¹⁶ Le Conseil stratégique pour la technologie de l'IA est résumé sur le portail IA du NEDO : http://www.nedo.go.jp/activities/ZZJP2_100064.html (en japonais)

³¹⁷ Cabinet du premier ministre, " Rapport sur l'intelligence artificielle et la société humaine" https://www8.cao.go.jp/cstp/tyousakai/ai/summary/aisociety_en.pdf

³¹⁸ Cabinet du Premier Ministre, Conseil des principes sociaux de l'IA centrée sur l'être humain, <https://www8.cao.go.jp/cstp/stmain/aisocialprinciples.pdf>

³¹⁹ Secrétariat du Cabinet, Principes sociaux de l'IA centrée sur l'être humain, 2019, <https://ai.bsa.org/wp-content/uploads/2019/09/humancentricai.pdf>

³²⁰ La conférence vers une société de réseau d'IA, Lignes directrices sur l'utilisation de l'IA, 2019, https://www.soumu.go.jp/main_content/000658284.pdf

Ministère de l'éducation, de la culture, des sports, des sciences et de la technologie (MEXT)

La réunion des ministres sur le développement des ressources humaines pour la société 5.0 a été créée³²¹ en décembre 2017 et s'est tenue jusqu'en juin 2018. Elle s'est concentrée sur le développement des ressources humaines tandis que des discussions parallèles ont été menées par le groupe de travail du ministère sur la promotion du pouvoir de vivre richement dans la nouvelle ère. Un rapport publié en juin 2018 signale dans sa toute première section les orientations politiques qui doivent être prises en compte dans les sociétés où la technologie de l'IA s'est développée, telles que l'offre de diverses possibilités et lieux d'apprentissage pour réaliser un "apprentissage équitable et individuellement optimisé", l'acquisition de compétences académiques de base et la capacité à utiliser l'information, et à transcender le fossé entre les sciences humaines et les sciences.

Ministère de l'économie, du commerce et de l'industrie (METI)

En janvier 2016, le ministère de l'économie, du commerce et de l'industrie a créé un groupe d'étude sur les approches idéales des politiques de concurrence pour la quatrième révolution industrielle afin d'examiner et de réfléchir aux moyens futurs de faire face à la situation actuelle et aux défis liés aux systèmes transversaux, tels que la politique de concurrence et la politique de propriété intellectuelle, tout en gardant à l'esprit la possibilité de changements rapides des structures industrielles qui surviendront en raison de la quatrième révolution industrielle, notamment le développement de l'IA. L'étude a été réalisée au cours du mois de juillet et un rapport a été rédigé³²².

Le Comité de la nouvelle structure industrielle du Conseil de la structure industrielle du ministère de l'économie, du commerce et de l'industrie³²³ se réunit depuis septembre 2015 et a formulé une vision qui pourrait être partagée par les secteurs public et privé. En outre, le comité a discuté des mesures qui seraient exigées des secteurs public et privé afin de répondre avec précision aux changements causés par l'IdO, le Big Data et l'IA. En mai 2017, un rapport final sur la nouvelle structure industrielle, qui a été rebaptisée "Vision d'avenir vers 2030"³²⁴, a été publié, et quatre domaines stratégiques (mobilité, chaîne d'approvisionnement, soins de santé et vie) ont été définis pour réaliser la société 5.0 comme une vision sociale idéale à viser. En outre, le rapport présente une feuille de route et des projets de percée.

En outre, le La Division de l'économie de l'information du Bureau de la politique commerciale et de l'information a créé un groupe d'étude sur l'IA et les lignes directrices relatives aux contrats de données en décembre 2017, et les "Lignes directrices relatives aux contrats sur l'utilisation de l'IA et des données" ont été publiées en juin 2018³²⁵. Ces lignes directrices comprennent un chapitre sur les données et un chapitre sur l'IA qui, en tant que référence pour les entreprises privées, résumant les principaux points des contrats, des exemples de clauses contractuelles et les facteurs à prendre en compte lors de la rédaction des clauses, de la conclusion de contrats liés à l'utilisation de données ou de contrats liés au développement et à l'utilisation de logiciels utilisant la technologie de l'IA.

Ministère de la santé, du travail et de la protection sociale (MHLW)

En février 2015, le ministère de la santé, du travail et de la protection sociale a entamé des réunions sur la formulation d'une vision de la politique de soins de santé pour les 20 prochaines années (2035) lors de la table ronde Japan Vision ; Health Care 2035³²⁶. Après huit séries de discussions, leurs propositions ont été publiées en juin 2015, et elles énoncent trois valeurs et critères de base : la

³²¹ MEXT, Réunion des ministres sur le développement des ressources humaines pour la société 5.0, https://www.mext.go.jp/b_menu/activity/detail/pdf2018/20180605_001.pdf

³²² Le METI, un groupe d'étude pour des approches idéales des politiques de concurrence pour la quatrième révolution industrielle, rédige un rapport : https://www.meti.go.jp/english/press/2017/0628_001.html

³²³ METI, Conseil de la structure industrielle, https://www.meti.go.jp/english/policy/economy/industrial_council/index.html

³²⁴ METI, un rapport final sur la vision de la nouvelle structure industrielle a été rédigé : https://www.meti.go.jp/english/press/2017/0530_003.html

³²⁵ Ministère de l'économie, du commerce et de l'industrie, version anglaise des directives contractuelles sur l'utilisation de l'IA et des données publiées, 2019, https://www.meti.go.jp/english/press/2019/0404_001.html

³²⁶ la vision du Japon ; les soins de santé 2035 : https://www.mhlw.go.jp/seisakunitsuite/bunya/hokabunya/shakaihoshou/hokeniryoku2035/assets/file/healthcare2035_proposal_150703_slide_en.pdf /

justice et l'équité, la solidarité fondée sur l'autonomie, et la prospérité et la coexistence pour le Japon et le monde³²⁷.

Par la suite, le Conseil sur la promotion de l'utilisation de l'IA dans les assurances et la médecine a commencé à se réunir en janvier 2017³²⁸ et a rédigé un rapport en juin 2017 qui a sélectionné six domaines prioritaires pour la promotion du développement de l'IA. Par la suite, le Consortium pour la promotion du développement de l'IA dans les domaines de l'assurance et de la médecine s'est réuni en juillet 2018³²⁹.

En outre, des réunions sur l'Avenir du travail : 2035 ont été organisées à partir de janvier 2016 concernant les changements de style de travail dus à l'innovation technologique, principalement dans le domaine de l'IA, et un rapport sur "l'Avenir du travail : 2035" a été publié en août 2016³³⁰. En outre, en mars 2017, *Mitsubishi UFJ Research and Consulting* a publié un rapport d'étude, commandé par le ministère de la Santé, du Travail et de la Protection sociale, sur l'impact de l'IdO, du Big Data et de l'IA sur l'emploi et le travail³³¹.

Ministère du territoire, des infrastructures, des transports et du tourisme (MLIT)

Le ministère du territoire, de l'infrastructure, des transports et du tourisme a principalement mené des discussions sur la conduite automatisée, et en septembre 2018, le MLIT a formulé les lignes directrices pour la sécurité dans l'exploitation automatisée des machines agricoles, qui précisent les exigences de sécurité auxquelles doivent satisfaire les véhicules automatisés de niveau 3 et 4³³². En avril 2018, ils ont également annoncé une étude sur le "concept de centre d'AI", qui évaluera et certifiera les données fournies et l'IA nécessaire pour le développement de ponts et de tunnels, etc. afin d'utiliser l'IA pour les inspections des infrastructures³³³.

Ministère de l'agriculture, des forêts et de la pêche (MAFF)

En 2013, le ministère de l'agriculture, des forêts et de la pêche a créé le groupe d'étude sur la promotion de l'agriculture intelligente avec la coopération d'entreprises de robotique et d'informatique, de fabricants de machines agricoles et de ministères et agences connexes. Ce groupe d'étude s'est donné pour mission d'étudier les mesures et les lignes directrices visant à promouvoir l'utilisation de l'IA et des robots dans l'agriculture (agriculture intelligente)³³⁴. En 2017, le groupe de recherche a formulé les lignes directrices pour la sécurité dans les opérations automatisées des machines agricoles, qui définissent les exigences pour les fabricants et les utilisateurs afin de garantir la sécurité de la technologie pour l'opération automatique non habitée des machines agricoles (machines agricoles robotisées). En 2018, un tableau récapitulatif concernant les sources de dangers ainsi que les conditions dangereuses pour l'automatisation des robots de gestion des plantations de thé a été ajouté aux lignes directrices³³⁵.

³²⁷ Assurance soins de santé 2035, Résumé des recommandations :

<https://www.mhlw.go.jp/seisakunitsuite/bunya/hokabunya/shakaihoshou/hokeniryou2035/future/>

³²⁸ MHLW, Conseil sur la promotion de l'utilisation de l'IA dans les assurances et la médecine : https://www.mhlw.go.jp/stf/shingi/other-kousei_408914.html (en japonais)

³²⁹ MHLW, Consortium pour la promotion du développement de l'IA dans les domaines de l'assurance et de la médecine, https://www.mhlw.go.jp/stf/shingi2/0000148680_00002.html (en japonais)

³³⁰ MHLW, *Future of Work : 2035 - For everyone to Shine* : <https://www.mhlw.go.jp/file/06-Seisakujouhou-12600000-Seisakutoukatsukan/0000152705.pdf>

³³¹ *Mitsubishi UFJ Research and Consulting*, Rapport du groupe d'étude sur l'impact de l'IdO, Big Data, AI. sur l'emploi et le travail, mars 2017, <https://www.mhlw.go.jp/file/04-Houdouhappyou-11602000-Shokugyouanteikyoku-Koyouseisakuka/0000166533.pdf> (en japonais)

³³² MLIT, Élaboration de lignes directrices sur les technologies de sécurité pour les véhicules à conduite autonome, 12 septembre 2018, http://www.mlit.go.jp/report/press/jidosha07_hh_000281.html

³³³ Nikkan Kogyo Shimbum, Gestion centralisée des inspections d'infrastructures, Ministère du territoire, des infrastructures et des transports, envisage le "concept de centre d'IA", 26 avril 2018, <https://www.nikkan.co.jp/articles/view/00471269>

³³⁴ MAFF, Groupe d'étude sur la promotion de l'agriculture intelligente, https://www.maff.go.jp/e/policies/tech_res/smaagri/attach/pdf/robot-1.pdf

³³⁵ MAFF, Lignes directrices pour assurer la sécurité du fonctionnement automatique des machines agricoles, mars 2018, http://www.maff.go.jp/j/kanbo/kihyo03/gityo/g_smart_nougyo/attach/pdf/index-6.pdf (en japonais)

ii. Discussions sur l'IA dans les ministères et les agences

Objet

Divers acteurs au Japon discutent des expériences de gouvernance de l'IA. Cette section se concentre donc sur une étude comparative des discussions, telles qu'elles se déroulaient en 2017, qui ont été publiées dans les livres blancs et les rapports des ministères développant des politiques sur l'IA. Ceci afin de comprendre la situation actuelle dans une perspective multidimensionnelle qui n'est pas spécifique à un ministère en particulier. Nous examinerons plus particulièrement : (1) le chevauchement des références aux technologies liées à l'IA dans chaque ministère et agence ; (2) quels sont les domaines d'application et les avantages, les défis et les mesures mentionnés dans chaque ministère et agence en ce qui concerne l'IA ; et (3) quel type de relation existe entre les visions sociales, telles que la société 5.0 et la quatrième révolution industrielle, et les technologies liées à l'IA.

Objectifs de l'enquête

Les ministères et agences japonais ont publié divers livres blancs et rapports³³⁶. Parmi eux, les 19 documents suivants, qui semblaient susceptibles de contenir des discussions sur l'IA, ont été inclus dans l'étude. Seules les versions disponibles en décembre 2017 ont été incluses dans cette étude.

- 1-3 Stratégie de croissance 2017 (Siège de la revitalisation économique du Japon, secrétariat du Cabinet)
- 1-4 Rapport sur les mesures prioritaires pour le renforcement de la compétitivité industrielle 2016 (Siège de la revitalisation économique du Japon, Secrétariat du Cabinet)
- 3-1 Rapport annuel sur l'économie et les finances publiques du Japon (Cabinet du Premier Ministre)
- 5-1 Le Livre blanc sur la police (Agence nationale de police)
- 7-1 Rapport annuel de l'Agence des services financiers (FSA)
- 9-1 Livre blanc sur l'information et les communications (Ministère de l'intérieur et des communications)
- 10-6 Faits marquants du projet de budget pour l'exercice 2008 (Ministère des finances)
- 12-1 Livre blanc sur la science et la technologie (Ministère de l'éducation, de la culture, des sports, de la science et de la technologie)
- 12-2 Livre blanc sur l'éducation, la culture, les sports, la science et la technologie (MEXT)
- 13-1 Rapport annuel sur la santé, le travail et la protection sociale (Ministère de la santé, du travail et de la protection sociale)
- 13-2 Livre blanc sur l'économie du travail (Ministère de la santé, du travail et de la protection sociale)
- 14-1 Rapport annuel sur l'alimentation, l'agriculture et les zones rurales (Ministère de l'agriculture, des forêts et de la pêche)
- 15-1 Livre blanc sur l'économie et le commerce international (Ministère de l'économie, du commerce et de l'industrie)

³³⁶ Le site web du cabinet du Premier ministre présente une liste des livres blancs par année (<https://www.kantei.go.jp/jp/hakusyo/index.html>)

- 16-1 Rapport annuel sur la politique des transports (ministère du territoire, des infrastructures, des transports et du tourisme)
- 16-2 Livre blanc sur le territoire, les infrastructures, les transports et le tourisme (Ministère du territoire, des infrastructures, des transports et du tourisme)
- 17-1 Rapport annuel sur la situation actuelle du service météorologique (Agence météorologique du Japon)
- 18-1 Rapport annuel sur l'environnement (Ministère de l'environnement)
- 19-1 Défense du Japon (Ministère de la défense)
- 20-1 Livre blanc sur la consommation (Agence des consommateurs)

Méthodes de recherche

Sélection de mots

Les 13 mots suivants (en japonais) ont été recherchés et extraits en tant que mots clés possibles dans le débat actuel sur les technologies de l'information.

- Intelligence artificielle, IA (人工知能/AI)
- Données importantes
- IdO (*IoT*)
- Robot
- Drones, véhicules aériens sans pilote et petits véhicules aériens sans pilote (小型無人機/無人航空機/小型無人機)
- Conduite automatique et conduite automatisée (自動運転/自動走行)
- VR, AR et MR (VR/AR/MR)
- Société 5.0 (Society 5.0)
- Quatrième révolution industrielle (第4次産業革命)
- Industries connectées (Connected Industries)
- Intelligence Artificielle Générale (AGI) (汎用人工知能)
- Singularité
- Quantum/technologie quantique (量子)

Nombre de mots

Nous avons recherché les mots dans les 19 documents étudiés et avons évalué leur occurrence.

Structuration

Les descriptions autour des mots ont été extraites et ont été étiquetées sous l'un des quatre points suivants en fonction de leur contenu.

- (a). Positionnement : une section qui décrit la définition et la position du mot sélectionné et une explication de la situation à traiter
- (b). Domaines d'application et avantages : une section où le mot sélectionné décrit un domaine d'application ou un avantage
- (c). Défis et mesures : une section où le mot sélectionné décrit un défi, y compris les politiques et les mesures visant à relever ce défi
- (d). Défis : lorsque le mot sélectionné ne décrit que des défis

iii. Résultats

Nombre d'occurrences par livre blanc et par rapport

Le tableau 1 montre la structure de chaque mot pour la technologie dans les 19 documents étudiés, extraits et structurés par étiquette de technologie. Le même mot apparaissant sous la même étiquette technologique est compté comme un³³⁷. Les mots qui ont été mentionnés à plus de 10 endroits sont indiqués en gras dans le tableau 1.

Le tableau 1 montre qu'il n'y avait aucune mention de RM, de singularité ou d'intelligence artificielle Générale (AGI) dans aucun des documents. Des références à la RV/AR, aux industries connectées et à la technologie quantique ont été trouvées, mais elles sont apparues moins fréquemment.

Par ailleurs, en comparant chaque livre blanc et chaque rapport, « Stratégie de croissance 2017 » (1-3), « Rapport annuel sur l'économie et les finances publiques japonaises » (3-1) et « Livre blanc sur l'information et la communication » (9-1) font référence à de nombreux mots-clés soulevés dans cette enquête.

Certains livres blancs et rapports font un usage intensif de certains mots-clés. Par exemple, « Société 5.0 » est beaucoup utilisé dans le « Livre blanc sur la science et la technologie » (12-1), « robot » dans le « Rapport annuel sur l'alimentation, l'agriculture et les zones rurales » (14-1), « quatrième révolution industrielle » dans le « Livre blanc sur l'économie et le commerce international » (15-1), et « conduite automatique/conduite automatisée », « drones » et « big data » dans le « Livre blanc sur la terre, les infrastructures, les transports et le tourisme » (16-2). En outre, bien que les termes utilisés dans cette enquête ne figurent pas dans le document « Rapport annuel de l'Agence des services financiers » (7-1), la cybersécurité et les technologies financières sont mentionnées dans le chapitre 22 « Autres questions », section 1 « Mesures internationales pour relever les nouveaux défis ».

Comparaison inter-agences pour chaque mot-clé

Les mots recherchés ont été classés dans les quatre groupes suivants : (a) Positionnement, (b) Domaines d'application et avantages, (c) Défis et mesures, et (d) Défis, pour trois domaines technologiques spécifiques, (i) Intelligence artificielle, (ii) Autres mots-clés technologiques, et (iii) Société 5.0/Quatrième révolution industrielle/Industries connectées. Le tableau 2 (Références à l'intelligence artificielle/IA) énumère les principales étiquettes de sous-postes utilisées dans (b) les domaines d'application et les avantages, (c) les défis et les mesures, et (d) les défis. Notez que le chiffre "1" ne fait pas référence au nombre d'occurrences, mais plutôt au nombre de fois où elles sont mentionnées.

Intelligence artificielle/IA

Comme le montre le tableau 2, l'« intelligence artificielle/IA » est largement utilisée dans les documents « Stratégie de croissance 2017 » (1-3), « Rapport annuel sur l'économie et les finances publiques du Japon » (3-1) et « Livre blanc sur l'information et la communication » (9-1). À l'inverse, le « Livre blanc sur la police » (5-1), le « Rapport annuel de l'Agence des services financiers » (7-1), les « Faits marquants du projet de budget pour l'exercice 2018 » (10-6), le « Livre blanc sur l'éducation, la culture, les sports, la science et la technologie » (12-2), le « Rapport annuel sur la politique des transports » (16-1) et le « Livre blanc sur le territoire, les infrastructures, les transports et le tourisme » (16-2) mentionnent à peine le terme IA. Ces livres blancs utilisent d'autres mots clés, à l'exception de 7-1 « Rapport annuel de l'Agence des services financiers » qui n'utilise aucun des

³³⁷ Par exemple, à la page 9 de la "Stratégie d'investissement future 2017", il est dit : "Nous allons promouvoir le développement d'un environnement en nuage pour le développement de l'IA et la mise en place d'un mécanisme de certification, ainsi que l'élaboration de règles pour le mode d'évaluation afin de garantir la qualité et la sécurité des dispositifs médicaux qui utilisent l'IA. Sur la base de ce qui précède, le gouvernement vise à évaluer la qualité des soins médicaux par un soutien précis aux soins médicaux des médecins utilisant l'IA lors de la prochaine révision et des révisions ultérieures des frais de remboursement des soins médicaux". Ici, le mot "IA" est utilisé trois fois. Il est ensuite appelé "c) défis et mesures", puis "évaluation" pour qu'il soit considéré comme un seul et même critère. Notez que certaines notes de bas de page, chiffres et colonnes qui ont été considérés comme triviaux dans la structuration ont été exclus et diffèrent donc du nombre réel de mots recherchés.

mots clés. Par exemple, le « Livre blanc sur le territoire, les infrastructures, les transports et le tourisme » (16-2), utilise des mots tels que « conduite automatique », « drones » et « données importantes ».

Positionnement

L'IA a été abordée comme un outil qui devrait être utilisé pour la réalisation de la Société 5.0 et de la quatrième révolution industrielle. Ce qui est également caractéristique, c'est que ce n'est pas seulement l'IA, mais aussi le Big Data et d'autres technologies qui sont traitées de la même manière.

A travers l'utilisation de l'IA et de l'IdO et la création d'innovation (omis), nous devons travailler à la vision future d'une société ultra-intelligente (Society 5.0). (« Stratégie de croissance 2017 » (1-3) pp. 90-91)

Les nouvelles technologies de la quatrième révolution industrielle, telles que l'IdO et l'IA, peuvent être considérées comme une extension de l'économie numérique (« Rapport annuel sur l'économie et les finances publiques du Japon », (3-1) p. 149)

L'utilisation du Big data est la clé de la croissance économique et de la transformation sociale fondées sur les données. L'Internet des objets (IdO/IoT) est le moyen de collecter du Big data, et l'intelligence artificielle (IA) est le moyen d'analyser et d'utiliser ce Big data. (« Livre blanc sur l'information et les communications », (9-1) p. 52)

La vague de la quatrième révolution industrielle à travers les industries, entraînée par des innovations technologiques telles que l'IdO, le Big data, l'intelligence artificielle et la robotique, apporte des signes de changement dans tous les pans de la société sous la forme de produits et de services innovants qui transcendent les frontières nationales. (« Livre blanc sur l'économie et le commerce international », (15-1) p. 209)

La quatrième révolution industrielle, qui exploite les percées technologiques de l'internet des objets (IdO/IoT), le Big data, l'intelligence artificielle, les robots et les capteurs, devrait créer de nouvelles entreprises qui résoudront les problèmes sociaux et réveilleront les besoins latents des différentes parties prenantes. (« Rapport annuel sur l'état actuel du service météorologique, » (17-1) p. 21)

Pour assurer la croissance de l'économie japonaise, il est important d'améliorer la valeur ajoutée et de promouvoir les efforts visant à résoudre les contraintes de l'offre. À cette fin, il est nécessaire de répondre aux innovations telles que l'IdO et l'IA de la quatrième révolution industrielle, d'augmenter la valeur ajoutée produite par travailleur, en d'autres termes, la productivité du travail, et de répondre aux changements de l'environnement entourant les modes de travail, tels que la participation des femmes et des personnes âgées à la population active. (« Livre blanc sur l'économie du travail », (13-2) p. 71)

Domaines d'application et avantages

Le tableau 2 présente différents domaines d'application de l'IA, notamment la médecine, la santé et les finances. Les soins médicaux, par exemple, sont traités dans le document « Rapport annuel sur la santé, le travail et la protection sociale » (13-1). D'autre part, « Stratégie de croissance 2017 » (1-3), « Rapport annuel sur l'économie et les finances publiques du Japon » (3-1) et « Livre blanc sur l'information et la communication » (9-1) fournissent de nombreux exemples sans s'arrêter à un secteur spécifique.

En outre, les rapports signalent comme avantages les changements qui ont été révélés par des scénarios d'utilisation dans « Affaires/Commercialisation », ainsi que dans « Travail/Changements de style de travail ».

La réalisation de la société 5.0 augmentera les façons de travailler qui ne sont pas limitées par le temps et l'espace. Outre les possibilités offertes aux personnes de développer leurs propres capacités et de mettre en œuvre leurs propres styles de travail, grâce à la collaboration avec l'IA, les robots et autres machines, nous pourrions également voir apparaître des « méthodes de travail plus intelligentes » grâce au télétravail avancé qui utilise la réalité virtuelle, la réalité augmentée et

d'autres formes de TIC. (Rapport annuel sur l'économie et les finances publiques du Japon, (3-1) p. 191)

Bien qu'il y est une crainte que l'IA et les robots ne suppriment des emplois, les résultats de l'analyse du chapitre 3 montrent que, malgré l'impact possible sur certaines professions et certains travailleurs qualifiés, si de nombreux nouveaux biens et services sont créés pour remplacer les biens existants, les gains de productivité de ces innovations de produits peuvent avoir pour effet d'augmenter l'emploi. (3-1 Rapport annuel sur l'économie et les finances publiques du Japon, (3-1) p. 207)

La partie II de ce livre blanc souligne également que l'utilisation des avancées technologiques par la réalisation d'activités d'innovation, y compris l'IA, peut contribuer à la réalisation d'un équilibre entre vie professionnelle et vie privée en créant des styles de travail flexibles tels que le travail non salarié. En outre, il a été souligné que la réalisation d'un équilibre entre vie professionnelle et vie privée peut conduire à l'utilisation efficace des ressources humaines et créer un cercle vertueux d'activités d'innovation. (13-2 Livre blanc sur l'économie du travail (13-2) p.172)

La quatrième révolution industrielle entraîne un changement majeur dans la nature de l'emploi. La quatrième révolution industrielle a permis (1) d'échanger librement toutes sortes de commerces et d'informations dans le monde réel par le biais de données et de réseaux, et (2) d'analyser et d'utiliser de grandes quantités de données de manière à créer une nouvelle valeur. En outre, (3) l'intelligence artificielle a permis aux machines d'apprendre et de prendre des décisions sophistiquées qui dépassent celles des humains, et (4) l'automatisation de tâches diverses et complexes est devenue possible. Ces innovations technologiques ont permis de réaliser une société que l'on pensait auparavant irréalisable et ont donné lieu à la possibilité de changements spectaculaires dans la structure de l'industrie et de l'emploi. (15-1 Livre blanc sur l'économie et le commerce international (15-1) p. 316)

Défis et mesures

Les défis à relever et les mesures à prendre pour apporter les avantages énumérés au point b) comprennent des éléments liés au développement de la technologie et de la recherche, aux ressources humaines et à la propriété intellectuelle. Divers exemples de création de mécanismes pour l'expansion et la collaboration internationales/outre-mer ont également été mentionnés.

Comme au point b), La « Stratégie de croissance 2017 » (1-3) et le « Livre blanc sur l'information et les communications » (9-1) font référence aux défis et aux mesures, mais en outre, le « Livre blanc sur la science et la technologie » (12-1) est unique en ce sens qu'il couvre diverses mesures.

En particulier, le développement des ressources humaines relève de la compétence du ministère de l'éducation, de la culture, des sports, de la science et de la technologie, et est décrit comme suit.

Afin de développer des ressources humaines ayant des connaissances de base en mathématiques et en sciences des données, ce qui est nécessaire pour dynamiser les activités industrielles du Japon à l'avenir, nous renforcerons l'enseignement des mathématiques et des sciences des données, sans distinction pour les sciences humaines ou les sciences dures, dans les universités. L'objectif est de développer des ressources humaines ayant des compétences en mathématiques, capables d'analyser et d'utiliser des données, de résoudre divers problèmes sociaux, de développer de nouvelles idées et de créer de nouvelles technologies. En outre, nous enrichirons l'enseignement pratique, tel que l'apprentissage par projet, en formant un réseau d'enseignement pratique grâce à la coopération entre l'industrie et les universités, et nous encouragerons le développement de programmes d'enseignement systématique pour la rééducation des adultes qui travaillent, afin de renforcer la fonctionnalité des universités en matière de développement des ressources humaines dans le domaine des technologies de l'information. En outre, sur la base des attentes croissantes en matière de formation des ingénieurs, qui joue un rôle central dans le développement des ressources humaines pour la quatrième révolution industrielle et la réalisation de la société 5.0, le ministère de l'éducation, de la culture, des sports, des sciences et de la technologie (MEXT) a mandaté un « Comité d'examen de la formation des ingénieurs dans les universités » depuis janvier 2017. Le comité discute du système éducatif idéal et des programmes d'études des écoles d'ingénieurs de premier et de deuxième cycle, ainsi que de l'approche idéale pour l'enseignement coopératif industrie-université, nécessaire pour réformer le système d'enseignement des ingénieurs

afin qu'il puisse répondre avec souplesse aux changements de la structure industrie-université (« Livre blanc sur la science et la technologie », (12-1) p. 180).

Bien que « Stratégie de croissance 2017 » (1-3) et « Livre blanc sur l'information et la communication » (9-1) et « Rapport annuel sur la santé, le travail et le bien-être » (13-1) fassent référence à des « lois », « institutions » et « règlements » pour introduire des initiatives de mise en œuvre sociale, seul « Livre blanc sur la science et la technologie » (12-1) mentionne des « questions éthiques, sociales et juridiques ».

Le nouveau projet *RIKEN Center for Advanced Intelligence*, qui relève du ministère de l'éducation, de la culture, des sports, des sciences et de la technologie, collabore avec les ministères, les entreprises, les universités et les instituts de recherche concernés pour mettre au point des technologies d'intelligence artificielle de base innovantes pour les dix prochaines années. En outre, ils cherchent à développer davantage les domaines dans lesquels le Japon possède des atouts, tels que les cellules iPS et la fabrication, et à promouvoir le développement des secteurs des soins de santé et de la prévention des catastrophes, ainsi que le développement de nouvelles technologies. Parallèlement à la recherche appliquée visant à résoudre les problèmes sociaux au Japon, tels que les infrastructures, nous menons également des recherches sur les questions éthiques, juridiques et sociales qui se posent avec la diffusion de la technologie de l'intelligence artificielle (« Livre blanc sur la science et la technologie », (12-1) p. 182).

En outre, le « Rapport annuel sur la santé, le travail et la protection sociale » (13-1) se réfère principalement aux conseils et institutions.

En janvier 2017, le ministère de la santé, du travail et de la protection sociale a créé le groupe de travail sur la promotion de la santé des données, dirigé par le ministre de la santé, du travail et de la protection sociale, afin d'étudier la mise en œuvre d'une infrastructure TIC qui relie de manière organique les secteurs de la santé, des soins médicaux et infirmiers, en vue de son fonctionnement à grande échelle au cours de l'exercice 2020. (« Rapport annuel sur la santé, le travail et la protection sociale », (13-1) p. 153).

L'intelligence artificielle (IA) devrait permettre la création de nouvelles méthodes de diagnostic et de traitement, le développement d'un environnement dans lequel les gens peuvent recevoir les soins médicaux les plus avancés partout au Japon, et une réduction de la charge de travail du personnel médical et infirmier. C'est pourquoi le ministère de la santé, du travail et de la protection sociale (MHLW) a organisé le Sommet de la promotion de la réforme de la santé afin de discuter des mesures nécessaires à l'utilisation de l'IA dans les domaines de la santé et de la médecine (« Rapport annuel sur la santé, le travail et la protection sociale », (13-1) p. 163).

Le « Rapport annuel sur l'alimentation, l'agriculture et l'espace rural » (14-1) mentionne également que le projet de création d'une intelligence artificielle pour l'agriculture du futur est mis en œuvre pour accélérer la recherche et le développement de la technologie de l'IA dans le secteur agricole, et pour la mettre en pratique dès que possible.

Défis

Alors que (c) montre les défis et les mesures, (d) est une liste d'éléments qui ne présentent que les défis. Comme le montre le tableau 2, (c) est plus courant dans la « Stratégie de croissance 2017 » (1-3), tandis que (d) est plus courant dans le « Rapport annuel sur l'économie et les finances publiques du Japon » (3-1). Cela peut être attribué à la nature différente des rapports.

Par exemple, la « Stratégie de croissance 2017 » pour les « Ressources humaines » est décrite au point c) comme suit, en indiquant les mesures à prendre pour résoudre le problème :

nous envisagerons des révisions spécifiques pour réformer le système d'enseignement de l'ingénierie afin de développer les ressources humaines nécessaires, en s'appuyant sur les technologies de l'information pour promouvoir les réformes structurelles industrielles telles que l'IA, l'IdO et le Big Data pour créer de nouvelles industries, et en révisant fondamentalement la structure verticale segmentée des départements et des facultés, par exemple en rendant les périodes d'enseignement, comme le système intégré de 6 ans pour les licences et les maîtrises, plus flexibles

et en permettant aux étudiants d'acquérir des sous-spécialités en plus de leur discipline principale, d'ici la fin de cet exercice. L'objectif est de mettre en œuvre les révisions par étapes à partir du prochain exercice fiscal, avec une mise en œuvre complète à partir de l'exercice 2019. (« Stratégie d'investissement future 2017 », (1-3) pp. 92-93)

En revanche, le « Livre blanc sur la politique économique et fiscale » (3-1) se concentre uniquement sur les défis (d) décrits ci-dessous. Bien qu'il énonce la nécessité de mesures, il ne va pas jusqu'à indiquer ce qu'il faut faire spécifiquement.

Aux États-Unis et dans d'autres pays développés, la pratique des salaires relativement élevés pour les personnes très instruites, pour lesquelles la demande augmente, a été signalé en raison de changements dans les structures économiques, et d'autres facteurs tels que le passage aux technologies de l'information. Au Japon, le pourcentage d'étudiants allant dans des établissements d'enseignement supérieur tels que les universités, les collèges et les écoles de formation spéciale (écoles professionnelles) augmente et, jusqu'à présent, l'écart de salaire ne s'est pas creusé en raison d'une demande excessive. Toutefois, à mesure que l'innovation technologique continue de progresser, nous devons nous assurer de disposer de ressources humaines hautement qualifiées qui maîtrisent les technologies les plus avancées et qui peuvent apporter une valeur ajoutée. Ce faisant, le rapport coût-efficacité de l'éducation sera amélioré par la mise en place d'un système qui donne la priorité au développement des ressources humaines nécessaires. Cette démarche s'inscrit dans le cadre des changements structurels économiques et sociaux au Japon, tels que l'augmentation de la demande de soins de santé dans le pays et à l'étranger, la hausse de la demande de tourisme entrant, ainsi que la mise en place et l'approfondissement de l'IA et de l'IdO dans l'économie et la société. L'investissement dans les ressources humaines est important, quel que soit l'âge des employés. Il est important de renforcer le potentiel de croissance de l'économie dans son ensemble en favorisant l'apprentissage des travailleurs (éducation récurrente) et en encourageant la migration de la main-d'œuvre vers des industries à forte valeur ajoutée, tout en améliorant l'employabilité des travailleurs et en augmentant les salaires. (« Livre blanc sur la politique économique et fiscale » (3-1) p. 83)

En ce qui concerne le « travail/emploi », qui était mentionné comme un avantage au point b), le « Livre blanc sur l'économie du travail » (13-2) montre également une certaine reconnaissance des défis.

Autres mots-clés technologiques

Six mots clés liés à l'IA sont abordés dans cette enquête : Big Data, IdO, robotique, drones, conduite automatisée/automatique, et RV/AR/RM. Le tableau 1 montre que toutes les technologies sont principalement mentionnées dans la « Stratégie de croissance 2017 » (1-3), le « Rapport annuel sur l'économie et les finances publiques du Japon » (3-1), « Livre blanc sur l'information et les communications »(9-1), et « Livre blanc sur les terres, les infrastructures, les transports et le tourisme » (16-2).

C'est pourquoi les livres blancs et les rapports sont classés en fonction de sous-étiquettes similaires indiquant les thèmes abordés pour chaque technologie, et non par ministère. Plus précisément, les tableaux 3 à 5 énumèrent respectivement (b) les domaines d'application et les avantages, (c) les défis et les mesures, et (d) les défis.

Comparaison des domaines d'utilisation et des avantages

Le tableau 3 présente la liste des domaines d'application et des avantages. Le tableau 3 montre que l'« intelligence artificielle/AI », le « Big Data », l'« IdO » et la « robotique » sont utilisées dans une variété de domaines relativement grande. Parmi ces six mots, « IdO » apparaît le plus souvent à côté de « IA », ce qui indique un large éventail d'applications, notamment « chantier de construction », « tourisme » et « alimentation ».

À l'exception de VR/AR, qui est apparu moins fréquemment, « logistique » et « remédier à la pénurie de main-d'œuvre » ont été mentionnés dans presque tous les mots-clés. En outre, lorsque les

« drones » ont été exclus, il est apparu clairement que les « soins infirmiers » et le « transport » étaient également des domaines d'intérêt communs.

Défis et mesures

Le tableau 4 présente une liste comparative des c) défis et des mesures. Lorsque les technologies, la promotion du développement technologique et de la recherche et du développement sont mentionnées, ainsi que la nécessité de normes, d'institutions et de réglementations ; à l'exception de la RV/AR, on peut observer que la « mise en œuvre sociale et la démonstration » sont cités, ainsi que l'« application pratique/entrepreneurs » et la « collaboration / mise en réseau ».

Défis

Le tableau 5 présente la liste des défis (d). L'« intelligence artificielle/AI » et l'« IdO » partagent bon nombre des mêmes défis. « La recherche et le développement », les « réglementations et normes », les « ressources humaines », les « entreprises / start-ups » et la « concurrence mondiale » sont également des défis.

Il est intéressant de noter que les rapports sur la gestion de l'aspect psychologique et émotionnel des utilisateurs et du grand public mentionnent des concepts tels que le sentiment de sécurité, qui sont communs aux « robots », à la « conduite automatisée » et aux « drones ».

Il a été souligné que la réticence de certaines personnes aux robots de soins infirmiers est l'une des raisons pour lesquelles les établissements de soins infirmiers n'ont pas été en mesure de les introduire, et il est donc important de faire comprendre au public les nouveaux systèmes de soins infirmiers utilisant des robots à l'avenir. Si les robots de soins infirmiers deviennent plus courants, cela pourrait potentiellement changer l'attitude du public (voir colonne). Dans la « nouvelle stratégie des robots » mentionnée ci-dessus, le gouvernement s'est fixé comme objectif d'augmenter le pourcentage de personnes qui « veulent utiliser » et « veulent que les personnes âgées utilisent » des robots de soins infirmiers à 80% chacune. (Rapport annuel sur la santé, le travail et le bien-être, (13-1) p. 166)

Comme le montrent les enquêtes de sensibilisation du public et les efforts susmentionnés pour mettre en œuvre la conduite automatisée dans la société, pour que les nouvelles technologies et les nouveaux services soient acceptés, il est nécessaire de garantir une sécurité technique, ainsi que de donner aux utilisateurs et à la société un sentiment de sécurité par le biais de règles et d'expériences sociales. Par exemple, l'utilisation de petits drones et d'autres véhicules sans pilote à petite échelle dans le secteur de la logistique, ainsi que l'amélioration des performances et des systèmes, contribuent à améliorer l'environnement de la commercialisation tout en soulageant l'anxiété du public en matière de sécurité. (Livre blanc sur le territoire, les infrastructures, les transports et le tourisme, (16-2) p. 89)

Aucune mention de l'intelligence artificielle générale / de la singularité / de la technologie quantique

L'intelligence artificielle générale et la singularité n'ont été mentionnées dans aucun des livres blancs et rapports. La technologie quantique a été mentionnée dans le rapport « Tendances en matière de sciences et technologies militaires » (19-1) de la « Défense du Japon ».

Ces dernières années, de nouveaux développements ont été réalisés dans le domaine des TIC. Par exemple, en août 2016, la Chine a lancé un satellite appelé Mozi, qui est le premier au monde à tester les communications à codage quantique. En plus des communications codées, de nouvelles technologies telles que l'intelligence artificielle et l'analyse de données importantes pourraient être appliquées au secteur militaire par différents pays dans le futur (« Défense du Japon », (19-1) p. 227).

La « Défense du Japon » (19-1) traite essentiellement de l'utilité des technologies de l'information et de la communication (TIC) dans la défense, et des technologies telles que l'IA et l'analyse du Big

Data sont également positionnées comme « nouvelles technologies », comme les communications quantiques.

Société 5.0 / La quatrième révolution industrielle / Industries connectées

La définition de la société 5.0 est la suivante : « Une nouvelle société qui est la cinquième société dans l'histoire de l'humanité, après (1) une société de chasse, (2) une société agraire, (3) une société industrielle et (4) une société de l'information. De nouvelles valeurs et de nouveaux services seront créés les uns après les autres, apportant la prospérité aux personnes qui sont les principaux acteurs de la société ». La stratégie de croissance pour 2017 (1-3 pp. 1-6) est utilisée comme la vision future du gouvernement japonais pour aller de l'avenir. La différence entre cette stratégie et la quatrième révolution industrielle est également indiquée dans le même rapport « Stratégie de croissance 2017 ».

L'Industrie 4.0 en Allemagne et l'*Industrial Internet* aux États-Unis sont principalement des tentatives d'utiliser l'IdO pour optimiser la production et la gestion des stocks dans les industries manufacturières au-delà du cadre des usines et des entreprises individuelles. À cet égard, le Japon souhaite réaliser des industries connectées qui vont au-delà de la fabrication pour connecter les biens aux biens, les choses aux choses, les personnes aux machines et aux systèmes, les personnes à la technologie, les entreprises aux entreprises dans différentes industries, les personnes aux personnes à travers les générations, et les fabricants aux consommateurs. La société 5.0 que le Japon vise est une tentative de résolution des divers problèmes sociaux en intégrant la technologie de pointe dans toutes les industries et la vie sociale, et en fournissant les biens et services nécessaires aux personnes qui en ont besoin, au moment où elles en ont besoin et en quantité dont elles ont besoin. (Stratégie de croissance 1-3 2017 pp.1-6)

Une des caractéristiques de la société 5.0 est qu'elle est « encore à venir ». Pour cette raison, la description des (b) domaines d'application et des avantages de la Société 5.0 est un processus de discussion pour « le moment où la Société 5.0 deviendra une réalité ».

La réalisation de la société 5.0 augmentera les méthodes de travail qui ne sont pas limitées par le temps et l'espace. En plus des possibilités offertes aux personnes de développer leurs propres capacités et de mettre en œuvre leurs propres styles de travail grâce à la collaboration avec l'IA, les robots et d'autres machines, nous pourrions également voir des « méthodes de travail plus intelligentes » grâce au télétravail avancé qui utilise la réalité virtuelle, la réalité augmentée et d'autres formes de TIC. (Rapport annuel sur l'économie et les finances publiques du Japon, 3-1 p. 191)

Pour cette raison, (c) les défis et mesures et (d) les défis sont présentés dans l'optique d'atteindre la Société 5.0, et il y a peu de discussion concrète sur (b) les domaines d'application et les avantages, ce qui crée une discussion abstraite. Par conséquent, inévitablement, la plupart des discussions sur la Société 5.0 portent sur (c) les défis et les mesures et (d) les défis. Plus précisément, nous pouvons signaler les étiquettes suivantes de « recherche et développement / technologie de base », « innovation », « ressources humaines », « propriété intellectuelle », « expansion à l'étranger », « institutions », « réglementations », « collaboration », « mise en œuvre », « normes », « évaluation » et « démonstration », comme indiqué dans (1) les étiquettes de la sous-section « Intelligence artificielle/AI ».

La quatrième révolution industrielle est souvent utilisée en parallèle avec Society 5.0, mais comme le montre le tableau 1, la quatrième révolution industrielle est plus souvent utilisée dans les livres blancs et les rapports que Society 5.0. En particulier, elle est davantage utilisée dans le « Livre blanc sur l'information et les communications » (9-1) et le « Livre blanc sur l'économie et le commerce international » (15-1) . En effet, de nombreux rapports décrivent la Société 5.0 comme étant postérieure à la quatrième révolution industrielle, et les descriptions de la quatrième révolution industrielle sont donc plus fréquentes³³⁸.

³³⁸ En revanche, le 12-1, "Livre blanc sur la science et la technologie", utilise largement le terme "Société 5.0" uniquement. En outre, le document 10-6, "Faits marquants du projet de budget pour l'exercice 2008", utilise également le terme "Société 5.0" au lieu du terme "Quatrième révolution industrielle". Le terme "Société 5.0" est également souvent utilisé en parallèle avec le terme "société super-intelligente".

Contrairement à la société 5.0, la quatrième révolution industrielle est considérée comme un événement qui a déjà eu lieu.

Ainsi, les entreprises qui utilisent effectivement les nouvelles technologies dans le cadre de la quatrième révolution industrielle se sentent plus à l'aise dans l'innovation de produits et l'augmentation des capacités de vente que dans les aspects d'efficacité de la production. (« Rapport annuel sur l'économie et les finances publiques du Japon », 3-1 p. 171)

Le tableau 6 indique le pourcentage de mentions de Société 5.0 et de la quatrième révolution industrielle par étiquette de sous-poste : alors que Société 5.0 ne compte que 6 % de mentions dans les domaines d'application et les avantages (b), la quatrième révolution industrielle compte 15 % de mentions dans les domaines d'application et les avantages (b).

III. Comparaison de la gouvernance de l'IA au Japon et à l'étranger

i. Comparaison des rôles des différents acteurs

Ce chapitre examine le type de rapports publiés par les acteurs au Japon comparé à d'autres pays.

Le rôle des sociétés académiques et des universités

Au Japon, la Société japonaise d'intelligence artificielle (JSAI) a entamé des discussions sur l'éthique à un stade relativement précoce, en 2014, et a publié ses directives éthiques. Elle a également collaboré avec l'IEEE et TFS³³⁹, mais les activités sont principalement destinées aux membres nationaux et la société ne s'engage pas activement avec le monde extérieur. Par exemple, en décembre 2019, trois sociétés académiques, dont la JSAI, qui étaient engagées dans l'apprentissage machine ont publié la « Déclaration sur l'apprentissage machine et l'équité », et ont déclaré que (1) l'apprentissage machine n'est rien de plus qu'un outil d'aide à la prise de décision humaine, et (2) les chercheurs en apprentissage machine sont engagés à améliorer l'équité dans la société en étudiant les utilisations possibles de l'apprentissage machine³⁴⁰.

En outre, par rapport aux pays étrangers, les initiatives universitaires, les rapports de recherche et les propositions menées principalement par les centres et facultés universitaires, et surtout les discussions, y compris celles qui s'inscrivent dans une perspective de sciences humaines et sociales, sont peu publiées en anglais³⁴¹. La recherche est actuellement laissée à la discrétion des chercheurs individuels.

Le rôle des associations industrielles

La collaboration de l'industrie est essentielle pour créer les meilleures pratiques, et de nombreuses associations industrielles internationales travaillant sur les meilleures pratiques, comme le *Partnership on AI* et le Conseil de l'industrie des technologies de l'information (ITI), qui sont basées aux États-Unis. Si certaines entreprises qui ont particulièrement besoin d'agir viennent des États-Unis, beaucoup d'entre elles créent leurs propres règles de gouvernance. Une enquête Accenture de 2018 a montré que parmi les entreprises qui ont adopté l'IA (72 % des entreprises interrogées), 70 % ont fourni une formation éthique aux techniciens et 63 % ont mis en place un comité d'éthique

³³⁹ Il convient toutefois de noter que c'est la volonté d'examiner cette question qui a lancé le débat, et que la sensibilisation à cette question est légèrement différente de celle des autres pays.

³⁴⁰ Cette déclaration a été publiée en réponse à un discours de haine prononcé par un chercheur de l'Université de Tokyo spécialisé dans l'IA. <https://arisaema0.wordpress.com/2020/01/08/ai-governance-in-2019/>

³⁴¹ "Perspectives on Artificial Intelligence/Robotics and Work/Employment", un "projet de recherche scientifique et technologique" du *Research and Legislative Reference Bureau (RLRB)*, la bibliothèque de la *Diet* nationale, comprend les tendances en matière de recherche et de technologie dans la première partie, la deuxième partie couvre l'impact sur l'emploi dans huit domaines, y compris les soins médicaux, les soins infirmiers, l'éducation et l'agriculture, et la troisième partie couvre les tendances à l'étranger en matière d'IA et d'emploi, ainsi que le développement, l'utilisation et la gestion des ressources humaines. Cette étude a été commandée par l'université et la faculté de l'université en a rédigé une grande partie. Une traduction anglaise de l'étude est disponible sur le site web de l'AIIR (<http://sig-air.org/publications/perspectives-on-ai>).

pour évaluer l'utilisation de l'IA³⁴², ce qui montre les efforts volontaires qui sont exigés des entreprises³⁴³.

À l'heure actuelle, les lignes directrices en matière d'IA sont principalement discutées à l'initiative d'entreprises de Big Tech aux États-Unis et en Chine, et il n'y a pas tant de voix de la part des jeunes entreprises, qui ont moins de ressources pour participer à ce genre de discussions. La structure de l'industrie compte également en termes de types d'entreprises. Il y a beaucoup d'entreprises "Business to Business" (B2B) au Japon, par rapport aux entreprises "Business to Consumer" (B2C). L'ensemble de la structure est comme B2B2B2B2...B2C et la chaîne d'approvisionnement est très longue au Japon. Par conséquent, les entreprises japonaises doivent tenir compte de la responsabilité, de la fiabilité et de l'obligation de rendre compte tout au long de cette très longue chaîne d'approvisionnement industrielle.

Les organisations internationales telles que l'ITI comprennent de nombreuses grandes entreprises japonaises, de sorte que les discussions dans un cadre international devraient s'intensifier à l'avenir. En outre, la *Japan Deep Learning Association*, fondée en 2017, est une organisation qui se consacre à la promotion des applications commerciales de l'apprentissage approfondi, et propose également des examens de certification dans le cadre du développement de ses ressources humaines. Le G-test, qui s'adresse aux ressources humaines à usage professionnel (généralistes), comprend des questions sur les applications industrielles, le droit, l'éthique et les problèmes sociaux actuels. En outre, un groupe d'étude sur la gouvernance de l'IA et son évaluation a commencé à se réunir en juillet 2020. Le groupe d'étude envisage un cadre, et un réseau d'écosystème, impliquant des compagnies d'assurance, des sociétés d'audit ainsi que des systèmes de dénonciation et des comités de tiers pour les enquêtes sur les accidents. Le groupe d'étude étudie également les écosystèmes de réduction des risques liés à l'IA, qui comprennent non seulement les grandes entreprises technologiques, mais aussi les jeunes entreprises.

On espère que les discussions actuelles contribueront à l'élaboration des perspectives éthiques et sociales des sociétés japonaises de capital-risque, notamment en ce qui concerne l'autonomie des industries.

Le rôle des organisations gouvernementales

Dans de nombreux pays, le développement des ressources humaines et de la recherche est une stratégie nationale importante pour l'IA. Le défi consiste donc à créer un centre de collaboration entre l'industrie, les universités et le secteur privé, et à mener les discussions en faisant preuve de leadership.

Il convient de noter que, bien que le Japon encourage la recherche qui ne se limite pas seulement aux sciences et aux technologies, mais inclut également les sciences sociales et humaines, les investissements du gouvernement dans les institutions de recherche et le développement des ressources humaines, par le biais du budget du ministère de l'éducation, de la culture, des sports, des sciences et des technologies, sont inférieurs à ceux des autres pays³⁴⁴. Au lieu de cela, l'Institut pour la politique de l'information et de la communication (IICP), qui multiplie les conférences depuis 2016, a rassemblé un large éventail de parties prenantes pour former un réseau de discussion. Ce réseau est composé de dizaines de personnes issues de différents domaines et secteurs d'activité.

Toutefois, dans le cas du Japon, il existe une forte tendance à lier directement l'évaluation de l'impact social de la technologie à la formulation de stratégies et à la prise de décisions politiques en matière de technologie, ce qui signifie que la distance et l'indépendance entre les deux n'est pas maintenue. La Conférence vers la société du réseau de l'IA, établie par l'IICP du ministère des affaires intérieures et des communications (MIC), a créé deux sous-comités distincts, le comité sur les

³⁴² Accenture constate que les entreprises intensifient leurs efforts pour utiliser l'intelligence artificielle de manière éthique et responsable - SAS, Accenture, Intel et Forbes Insights Dernière enquête, 23 octobre 2018, <https://www.accenture.com/jp-fr/company-news-releases-20181023>

³⁴³ La deuxième édition de l'IEEE "Ethically Aligned Design" propose également la nécessité d'un comité d'éthique ou d'employés qui peuvent tout couvrir en termes de valeur, comme un *Chief Value Officer* (CVO).

³⁴⁴ Le budget global du gouvernement pour l'IA dans le budget global de l'année fiscale 2018 s'élève à 77 milliards de yens, ce qui représenterait moins de 20 % du budget des États-Unis (500 milliards de yens) et de la Chine (450 milliards de yens), selon le Japan Times, le budget du Japon pour l'IA sera inférieur à un cinquième de celui prévu par les États-Unis et la Chine, 25 février 2018, <https://www.japantimes.co.jp/news/2018/02/25/business/tech/japanese-government-spending-ai-less-20-u-s-china/>

principes de la R&D en matière d'IA et le comité sur l'évaluation de l'impact et des risques, mais l'accent a été mis dans les deux comités sur la préparation de principes et de lignes directrices.

Il a également été suggéré que les discussions destinées à un public non japonais et les discussions destinées à un public japonais ne sont pas suffisamment connectées. En termes de présence internationale, le Japon a participé activement aux réunions de l'OCDE sur l'IA, à commencer par la réunion du G7 à Takamatsu en 2016, et le ministère des Affaires intérieures et des Communications a également co-organisé les réunions ultérieures de l'OCDE. Les principes régissant l'IA ont été l'un des points centraux de la réunion des ministres de l'économie numérique du G20 au Japon en 2019.

D'autre part, les discussions sur les principes de l'IA au sein du Cabinet du Premier Ministre et du ministère de l'Intérieur et des Communications, qui ont servi de base aux discussions destinées à un public non japonais, n'étaient pas nécessairement liées aux discussions au sein des ministères et agences gouvernementales qui ont eu lieu en II-1(5), et l'analyse du Livre blanc n'aborde pas beaucoup de questions éthiques ou sociales.

ii. Tendances et problèmes dans les technologies et les domaines ciblés de l'IA

Intensité des discussions sur des technologies d'IA spécifiques dans chaque ministère

Dans la section II de ce rapport, non seulement la technologie de l'IA mais aussi les technologies et concepts périphériques ont été étudiés en même temps. Le tableau 1 montre que l'« intelligence artificielle/AI » est mentionnée dans de nombreux livres blancs et rapports aux côtés de l'« IdO » et des « Big Data ». Parmi eux, ces mots clés sont particulièrement mentionnés dans la « Stratégie de croissance 2017 » (1-3) du secrétariat du Cabinet et dans le « rapport annuel sur l'économie et les finances publiques du Japon » (3-1), ainsi que dans le « livre blanc sur l'information et les communications » (9-1) du ministère de l'intérieur et des communications. L'Institut pour la politique de l'information et des communications (IICP) du ministère de l'intérieur et des communications a organisé la Conférence vers la société de réseau IA, et les résultats et discussions de cette conférence ont également été repris dans le « livre blanc sur l'information et les communications » du MIC.

Au sein du Conseil stratégique pour la technologie de l'intelligence artificielle, outre les trois principaux ministères (ministère de l'intérieur et des communications, ministère de l'éducation, de la culture, des sports, des sciences et des technologies, et ministère de l'économie, du commerce et de l'industrie), la recherche et le développement ainsi que la mise en œuvre sociale sont actuellement encouragés en coopération et en collaboration avec les ministères et organismes concernés, tels que le bureau du cabinet, le ministère de la santé, du travail et des affaires sociales, le ministère du territoire, des infrastructures, des transports et du tourisme, et le ministère de l'agriculture, des forêts et de la pêche. La comparaison des rapports et des livres blancs dans le tableau 2 montre que le Livre blanc sur l'information et la communication aborde les domaines d'utilisation, les avantages, les défis et les mesures de manière générale. En revanche, le Livre blanc sur la science et la technologie du ministère de l'éducation, de la culture, des sports, de la science et de la technologie se concentre sur les questions et les mesures liées à la recherche et au développement technologique, aux ressources humaines et à la collaboration. Au contraire, le Livre blanc sur l'économie et le commerce international du ministère de l'économie, du commerce et de l'industrie fait peu mention de l'IA et n'aborde que les questions et mesures liées à l'emploi, aux styles de travail et à l'innovation.

Le tableau 1 montre à son tour que le Livre blanc sur la science et la technologie se caractérise par ses fréquentes références à la « Société 5.0 » et le Livre blanc sur l'économie et le commerce international par ses fréquentes références à la « Quatrième révolution industrielle ». « Société 5.0 » et « Quatrième révolution industrielle » sont des mots clés pour décrire la vision de la société que nous devrions viser, et la technologie de l'IA n'est qu'une des technologies qui contribuent à la réalisation de cette société. Le Livre blanc sur l'aménagement du territoire, les infrastructures, les transports et le tourisme fait un usage intensif et met donc l'accent sur la « conduite automatique/conduite automatisée », les « drones » et les « données importantes » plutôt que sur

« l'intelligence artificielle/IA ». Le rapport annuel sur la santé, le travail et le bien-être et le Livre blanc sur l'alimentation, l'agriculture et l'espace rural mentionnent bien l'« intelligence artificielle/IA », mais le plus souvent, il est fait référence à des « robots ».

La technologie de l'IA est parfois utilisée dans le logiciel de certains matériels, comme les voitures, les drones et les robots, ce qui ne signifie pas qu'ils n'ont aucun lien entre eux, mais il existe des variations dans les types de technologies sur lesquelles les différents ministères se concentrent.

Technologies périphériques et domaines de l'IA

Le tableau 3 montre que les domaines d'application et les avantages des technologies périphériques de l'IA sont les suivants : soins médicaux/santé/soins infirmiers, transport, logistique, prévention des catastrophes et agriculture. Ce sont des domaines de la société 5.0 qui nécessitent l'intégration de l'espace cybernétique et physique, et on dit aussi qu'ils sont les points forts du Japon.

En revanche, les services financiers (fintech), le logement (maisons intelligentes et assistants) et le tourisme sont moins susceptibles d'être utilisés que dans d'autres pays en termes d'accumulation de données et de développement des infrastructures. En outre, les développements technologiques liés à la défense et à la sécurité nationales sont abordés dans de nombreux débats industrie-université-gouvernement à l'étranger, alors qu'au Japon, ils ne sont abordés que par le ministère de la défense. De même, les systèmes d'armes autonomes létales (LAWS), qui font l'objet de discussions aux Nations Unies, sont rarement discutés en dehors du ministère des affaires étrangères au Japon, bien qu'ils aient été mentionnés à la *Diet* en 2018. Cela peut être dû en partie au fait que le ministère de la défense, le ministère des affaires étrangères et l'agence des services financiers ne sont pas très impliqués dans les organes pertinents du Conseil stratégique pour les technologies d'intelligence artificielle.

Et alors que des mots comme *IdO* et *Big data* sont utilisés avec l'IA, il est peu question de *RV* et de technologie quantique, qui sont censés être des technologies émergentes. Il n'est pas non plus fait mention de mots tels que singularité et intelligence artificielle générale.

Compréhension insuffisante de la situation actuelle

Il est nécessaire de mener des discussions fondées sur des données pour faire progresser la gouvernance de l'IA. C'est pourquoi les gouvernements, les universités et les instituts de recherche de nombreux pays évaluent l'impact social de la technologie et créent des indicateurs à cet effet. Cela dépend du type d'indicateurs à créer, mais au Japon, le problème est qu'il y a un manque de données pour comprendre et évaluer la situation actuelle, bien que des objectifs aient été fixés.

Cependant, comme le montre la section II de ce rapport, alors que presque tous les ministères et agences discutent du développement des ressources humaines, comme par exemple « Développer les ressources humaines avec une connaissance de base des mathématiques et de la science des données », il n'existe pas de données permettant de comprendre la situation actuelle, bien que cette question soit mentionnée comme un objectif important. Bien que l'on estime que les ressources humaines dans le domaine des technologies de l'information risquent de se raréfier à l'avenir³⁴⁵ et que l'indice d'activité entrepreneuriale TEA (2014) indique³⁴⁶ le ratio des ressources humaines entrepreneuriales, il n'existe aucune liste de données permettant de comprendre la situation actuelle.

Le « Projet de stratégie d'IA (Aperçu général) » comprend un graphique sur « l'investissement global dans l'IA, la R&D et les ressources humaines », mais les données relatives au « Nombre de diplômés universitaires formés à l'analyse des données » datent de 2008, il y a plus de 10 ans³⁴⁷. Bien qu'il soit nécessaire d'examiner qui crée quels indicateurs, à un niveau informationnel de base, la création d'indicateurs est nécessaire pour évaluer l'impact de la technologie et identifier les besoins potentiels. Il s'agit certainement d'une question d'avenir. L'élaboration de ces informations

³⁴⁵ 15-1 Livre blanc sur l'économie et le commerce international pp. 235-6

³⁴⁶ 9-1 Livre blanc sur l'information et les communications p.111

³⁴⁷ Le document original du Cabinet du Premier Ministre (<https://www.kantei.go.jp/jp/singi/tougou-innovation/dai2/siryoy1.pdf>) est le Livre blanc sur l'information et la communication 2014 du ministère de l'intérieur et des communications (<http://www.soumu.go.jp/johotsusintokei/whitepaper/en/h26/html/nc134020.html>) (en japonais)

de base servira également de base pour rendre la fonction d'évaluation technologique quelque peu indépendante de l'élaboration des stratégies et des politiques.

iii. Comment créer un forum de discussion et ses enjeux

La place et le rôle de la collaboration de chaque acteur

Notamment, l'industrie, les universités et le secteur privé aux États-Unis sont tous proactifs dans la mise en œuvre des discussions, et en fait, l'industrie et les universités montrent la voie en façonnant les débats plutôt que d'être dirigées par le gouvernement. En outre, étant donné que l'IEEE et l'ISO, qui sont des acteurs internationaux, sont centrés sur l'Occident, il est nécessaire d'examiner comment le Japon participe aux plates-formes internationales, et si le Japon est en mesure de former sa propre plate-forme.

Au Japon, le gouvernement a encouragé la création de principes de sa propre initiative. Cependant, les principes n'ont qu'une existence abstraite, et leur mise en pratique nécessite une collaboration et une coopération entre tous les domaines et toutes les industries. La nature de la gouvernance technologique diffère selon les secteurs³⁴⁸. Au Japon, les ministères et les agences ont entamé des discussions en vue de la mise en œuvre et, par exemple, l'Association médicale japonaise a résumé leurs propres discussions. Toutefois, il est nécessaire de créer un forum où les industries peuvent coopérer pour formuler les structures et les meilleures pratiques de gouvernance technologique, et pour créer et diffuser des normes. Actuellement, les consortiums sont formés par domaine, mais il est nécessaire d'envisager l'avenir de la gouvernance, et de se demander si elle sera comme celle que l'on trouve en Europe ou aux États-Unis, qui impliquent diverses parties prenantes, ou sous d'autres formes.

En outre, la gouvernance par la coopération entre les acteurs est non seulement nécessaire pour chaque domaine, mais elle est également importante pour la réalisation du type de société qui sera construit en utilisant la technologie. Au Japon, la société 5.0 a été mentionnée en termes généraux, mais une vision pour cette société doit encore se concrétiser. Il est également nécessaire d'établir un forum de coopération entre les acteurs comme mécanisme permettant d'envisager avec souplesse le type de société à viser.

Nécessité de discussions impliquant le grand public

Les organisations à but non lucratif et les organisations non gouvernementales sont essentielles pour créer des espaces de discussion qui impliquent un large éventail de parties prenantes. Cependant, il n'y a pas autant d'organisations à but non lucratif au Japon que dans d'autres pays. La Société japonaise d'intelligence artificielle (JSAI) a contribué aux discussions en ligne de la Société du futur. Bien que la conception d'un dialogue en ligne puisse être difficile à certains égards, il s'agit d'une initiative intéressante qui permet d'impliquer diverses parties prenantes. Au Japon, ce type d'activités, qui relie les discussions publiques aux recommandations politiques, est toujours traité comme un processus de consultation publics.

Dans d'autres domaines, des manifestations de dialogue citoyen-participant ont été organisées au Japon dans le passé pour formuler des recommandations politiques sur des sujets tels que les plantes génétiquement modifiées, les nanotechnologies et la biodiversité³⁴⁹. En ce qui concerne la robotique, un événement a été organisé sous la forme d'un « consultation public interactif »³⁵⁰. Il est également nécessaire de réfléchir à la forme idéale de gouvernance qui implique des discussions entre de nombreuses parties prenantes.

³⁴⁸ Shiroyama Hideaki, Science, technologie et politique, chapitre 7

³⁴⁹ Tadashi Kobayashi, "Qui pense à la science et à la technologie ? (2004), Nagoya University Press, Nobuko Kori, et al : "From Participating Observations of WWViews Debate Process on Biodiversity" Science Communication, 13 : 31-46, 2013.

³⁵⁰ Les robots arrivent ! Symposium sur "Robot x Future x Dream Vision" - Un pont entre les "besoins en robots" et la "mise en œuvre dans la société",

http://interactive.pesti.jp/robot/wp/wp-content/uploads/F01_Introduction.pdf

IV. Conclusion

Ce rapport décrit les questions sociales traitées par les différentes parties prenantes en ce qui concerne la gouvernance de l'IA, et examine les caractéristiques et les défis au Japon.

Les leaders en matière de gouvernance, et les défis connexes, changent également selon le pays et le secteur. Dans les domaines où les institutions et les réglementations jouent un rôle majeur, tels que les soins de santé et les transports, la gouvernance nationale tend à montrer la voie. Il est donc nécessaire de développer des essais en utilisant un bac à sable réglementaire. En outre, comme les services fournis par le GAFAM se développent à l'échelle mondiale et que la technologie progresse rapidement, il est nécessaire de collaborer avec diverses entités de bas en haut, et non de haut en bas. Par conséquent, afin de faire avancer les discussions sur la gouvernance de l'IA, il est important d'établir un mécanisme et un forum pour réfléchir en permanence à la mesure dans laquelle les mêmes principes et règles peuvent être appliqués au niveau mondial et local (dans chaque pays et région), au type de parties prenantes qui doivent être impliquées et à qui doit prendre l'initiative.

Remerciements

Ce rapport est basé sur une traduction d'un chapitre de « *AI Governance* » écrit par Arisa Ema et Hideaki Shiroyama dans *Artificial Intelligence, Humanity and Society* (Keisho Shobo, 2020)³⁵¹. Nous remercions Mme Chihiro Saito et M. Hideaki Kojima, M. Yoichi Iida, M. Koichi Takagi, et le Dr. Takashi Egawa pour avoir soutenu notre travail afin de l'intégrer dans le rapport du CAHAI. Nous remercions Mme Haruka Matsumoto, une étudiante diplômée de l'Université de Tokyo (à l'époque), pour son aide dans la réalisation de l'étude de ce rapport. Cette recherche fait partie des résultats du projet de recherche scientifique et technologique de l'exercice 2017 du Bureau de la recherche et de la référence législative (RLRB) : « Perspectives sur l'intelligence artificielle/la robotique et le travail/l'emploi », et la subvention à la recherche scientifique (A) : « Gouvernance internationale des technologies de l'information et des biotechnologies émergentes - Rôle du partage de l'information et des acteurs privés (subvention numéro 18H03620) ».

³⁵¹ Intelligence artificielle, humanité et société (Keisho Shobo, 2020), <https://www.keisoshobo.co.jp/book/b498075.html> (en japonais)

Annexes

Tableau 1

| | Intelligence artificielle/IA | Conduite automatique et conduite automatisée | Drones, véhicules aériens sans pilote et petits véhicules aériens sans pilote | Robot | Données importantes | Id O | Société 5.0 | Quatrième révolution industrielle | Industries connectées | V R, A R | Quantum/technologie quantique | MR | Singularité | AGI |
|-------|------------------------------|--|---|---------------|---------------------|-----------|-------------|-----------------------------------|-----------------------|----------|-------------------------------|----|-------------|-----|
| 1-3 | 45 | 37 | 10 | 34 | 28 | 47 | 7 | 15 | 2 | 4 | 0 | 0 | 0 | 0 |
| **1-4 | Non identifié | Non identifié | 0 | Non identifié | Non identifié | 14 | 1 | Non identifié | Non identifié | 1 | 0 | 0 | 0 | 0 |
| 3-1 | 27 | 2 | 0 | 16 | 9 | 21 | 10 | 18 | 2 | 0 | 0 | 0 | 0 | 0 |
| 4-1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *7-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9-1 | 70 | 8 | 0 | 14 | 37 | 54 | 8 | 32 | 1 | 5 | 0 | 0 | 0 | 0 |
| 10-6 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *12-1 | 7 | 2 | 0 | 3 | 1 | 6 | 32 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| *12-2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13-1 | 7 | 0 | 0 | 9 | 7 | 5 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13-2 | 9 | 0 | 0 | 1 | 2 | 1 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14-1 | 5 | 3 | 2 | 13 | 4 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15-1 | 8 | 1 | 0 | 9 | 5 | 9 | 1 | 28 | 4 | 0 | 0 | 0 | 0 | 0 |
| 16-1 | 0 | 5 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *16-2 | 0 | 13 | 10 | 6 | 12 | 6 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 17-1 | 1 | 0 | 0 | 1 | 3 | 4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18-1 | 1 | 0 | 0 | 0 | 3 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19-1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 20-1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 |

* Cela signifie qu'une recherche de mots n'a pas été effectuée sur l'ensemble du livre blanc, seule une partie de celui-ci a été ciblée.

** Cela signifie que, comme une recherche normale de mots n'était pas possible en raison d'erreurs dans un PDF, seul le nombre qui pouvait être compté à l'œil nu a été comptabilisé.

Tableau 2

| Point principal | Subdivision | 1-3 | **1-4 | 3-1 | 4-1 | *7-1 | 9-1 | 10-6 | *12-1 | *12-2 | 13-1 | 13-2 | 14-1 | 15-1 | 16-1 | *16-2 | 17-1 | 18-1 | 19-1 | 20-1 |
|--|---|-----|-------|-----|-----|------|-----|------|-------|-------|------|------|------|------|------|-------|------|------|------|------|
| (a) Positionnement. | | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| b) Domaines d'application et avantages | Soins médicaux | 1 | | 1 | | | 1 | | | | 1 | | | | | | | | | |
| | Santé/soins de santé | 1 | | | | | | | | | 1 | | | | | | | | | |
| | Financement | | | 1 | | | 1 | | | | | | | | | | | | | |
| | Univers / Espace | 1 | | | | | 1 | | | | | | | | | | | | | |
| | Soins infirmiers / soutien aux personnes âgées | 1 | | 1 | | | | | | | 1 | | | 1 | | | | | | |
| | Trafic / automobile / conduite automatisée | | | 1 | | | 1 | | | | | | | | | | | | | |
| | Sports | 1 | | | | | | | | | | | | | | | | | | |
| | Fabrication | 1 | | | | | 1 | | | | | | | | | | | | | |
| | Logistique / transport terrestre | 1 | | 1 | | | 1 | | | | | | | | | | | | | |
| | Environnement / développement urbain | | | | | | | | | | | | | | | | | | 1 | |
| | Météo | | | | | | 1 | | | | | | | | | | | | | |
| | Infrastructure / environnement informatique | 1 | | | | | 1 | | | | | | | | | | | | | |
| | Catastrophe | 1 | | | | | | | | | | | | | | | | | | |
| | Administration | 1 | | | | | | | | | | | | | | | | | | |
| | Biotechnologie | 1 | | | | | | | | | | | | | | | | | | |
| | Recherche et développement | 1 | | 1 | | | 1 | | | | | | | | | | | | | |
| | Entreprises / développement des entreprises | 1 | | | | | | | | | | | | | 1 | | | | | |
| | Travail / main-d'œuvre / pénurie de main-d'œuvre / lieu de travail / emploi | | | 1 | | | 1 | | | | | 1 | 1 | 1 | | | | | | |
| Innovation / création | | | 1 | | | 1 | | | | | | | | 1 | | | | | | |

| | | | | | | | | | | | | | | | | | | | |
|---------------------------------------|--|---|--|---|--|--|---|---|---|---|---|---|---|---|--|--|--|--|--|
| | Impact | | | | | | 1 | | | | | | | | | | | | |
| | Données | | | 1 | | | 1 | | | | 1 | | | 1 | | | | | |
| c) Défis et mesures | Développement technologique / recherche et développement | 1 | | | | | 1 | | 1 | | | | 1 | | | | | | |
| | Innovation / création | 1 | | | | | 1 | | | | | | | | | | | | |
| | Ressources humaines | 1 | | | | | 1 | | 1 | | | | 1 | | | | | | |
| | Propriété intellectuelle | 1 | | | | | | | 1 | | | | | | | | | | |
| | Infrastructure | 1 | | | | | 1 | | | | | | | | | | | | |
| | Questions éthiques, juridiques et sociales | | | | | | | | | 1 | | | | | | | | | |
| | Droit | 1 | | | | | 1 | | | | | | | | | | | | |
| | Connectivité / mise en réseau | 1 | | | | | 1 | | 1 | | | | | | | | | | |
| | International / expansion internationale | 1 | | | | | 1 | | | | | | | | | | | | |
| | Soutien / promotion | 1 | | | | | 1 | | | | | | | | | | | | |
| | Utilisation pratique / commercialisation | 1 | | | | | | | | | | | | | | | | | |
| | Évaluation | 1 | | | | | | | | | | | | | | | | | |
| | Système | 1 | | | | | | 1 | | | | 1 | | | | | | | |
| | Normes / lignes directrices / règlements | 1 | | | | | | 1 | | | | | | | | | | | |
| | Conseil | 1 | | | | | | 1 | | | | 1 | | | | | | | |
| | Symposium | | | | | | | 1 | | | | | | | | | | | |
| Mise en œuvre sociale / démonstration | 1 | | | | | | 1 | | 1 | | | | | | | | | | |
| Données | 1 | | | | | | 1 | | | | | | | | | | | | |
| (d) Défis | Recherche et développement | | | 1 | | | | | | | | | | | | | | | |
| | Innovation / croissance économique / transformation / création | | | 1 | | | 1 | | | | | | 1 | | | | | | |
| | Ressources humaines | | | 1 | | | 1 | | | | 1 | 1 | | | | | | | |

| | | | | | | | | | | | | | | | | | | | |
|---|---|--|---|--|--|---|--|--|--|--|---|--|--|--|--|--|--|--|---|
| Amélioration de l'environnement | | | | | | | | | | | 1 | | | | | | | | |
| Investissement | | | 1 | | | 1 | | | | | | | | | | | | | |
| Disparité | | | 1 | | | | | | | | | | | | | | | | |
| International | | | | | | 1 | | | | | | | | | | | | | 1 |
| Mise en réseau | | | 1 | | | 1 | | | | | | | | | | | | | |
| Système | 1 | | 1 | | | 1 | | | | | | | | | | | | | |
| Règlements / lignes directrices | | | 1 | | | 1 | | | | | | | | | | | | | |
| Changement organisationnel / structure organisationnelle / changement d'entreprise | | | 1 | | | 1 | | | | | | | | | | | | | |
| Emploi / travail / main-d'œuvre / marché du travail / pénurie de main-d'œuvre / travail | | | 1 | | | 1 | | | | | 1 | | | | | | | | |
| productivité | | | 1 | | | 1 | | | | | | | | | | | | | |
| Décentralisation (du pouvoir) | | | 1 | | | | | | | | | | | | | | | | |
| Entrepreneuriat / Start-up | | | 1 | | | 1 | | | | | | | | | | | | | |
| Commerces militaires | | | | | | | | | | | | | | | | | | | 1 |

Tableau 3

| | Intelligence artificielle/IA | Big Data | IdO | Robot | Conduite automatique et conduite automatisée | Drone | VR/AR |
|--|------------------------------|----------|-----|-------|--|-------|-------|
| Soins médicaux | 1 | 1 | 1 | 1 | | | |
| Santé / soins de santé | 1 | 1 | 1 | 1 | | | |
| Finance / FinTech | 1 | 1 | | | | | |
| Univers / espace | 1 | 1 | 1 | | 1 | | |
| Soins infirmiers / soutien aux personnes âgées | 1 | 1 | 1 | 1 | 1 | | |
| Trafic / automobile / conduite automatisée | 1 | 1 | 1 | 1 | 1 | | 1 |
| Sports | 1 | 1 | 1 | | | | |
| Fabrication | 1 | | 1 | 1 | | | |

| | | | | | | | |
|--|---|---|---|---|---|---|---|
| Logistique / transport terrestre / navire / avion / fret | 1 | 1 | 1 | 1 | 1 | 1 | |
| Environnement / développement urbain / communauté | 1 | 1 | 1 | | 1 | 1 | |
| Météo / changement climatique | 1 | 1 | 1 | 1 | | | |
| Infrastructure / environnement informatique | 1 | 1 | | 1 | | 1 | |
| Catastrophes / prévention des catastrophes | 1 | 1 | 1 | 1 | | 1 | |
| Administration | 1 | 1 | 1 | | | | |
| Biotechnologie | 1 | | | | | | |
| Recherche et développement | 1 | | | | 1 | | |
| Entreprises / développement des entreprises | 1 | 1 | 1 | | | | |
| Travail / main- d'œuvre / pénurie de main-d'œuvre / lieu de travail / emploi | 1 | 1 | 1 | 1 | 1 | 1 | |
| Innovation / création | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Impact | 1 | | | | | | |
| Données | 1 | 1 | 1 | | | | |
| Ressources humaines | | 1 | | 1 | | | |
| Agriculture, sylviculture et pêche / élevage / produits laitiers | | 1 | 1 | 1 | 1 | 1 | |
| Énergie | | | 1 | | | | |
| Visites guidées | | | 1 | | | | 1 |
| Construction | | | 1 | 1 | | | |
| Résidentiel / Logement | | | 1 | | | | 1 |
| Denrées alimentaires | | | 1 | | | | |
| Gestion des actifs | | | | 1 | | | |
| Expansion internationale | | | | | 1 | | |
| Assurance | | | | | 1 | | |
| Sondage | | | | | | 1 | |

Tableau 4

| | Intelligence artificielle/IA | Big Data | IdO | Robot | Conduite automatique et conduite automatisée | Drone |
|--|------------------------------|----------|-----|-------|--|-------|
| Développement technologique / recherche et développement | 1 | 1 | 1 | 1 | 1 | 1 |
| Innovation / création | 1 | 1 | | | | |
| Ressources humaines | 1 | 1 | 1 | 1 | 1 | |
| Propriété intellectuelle | 1 | 1 | | 1 | | |
| Infrastructure | 1 | | | | | |
| Questions éthiques, juridiques et sociales | 1 | | | | | |
| Droit | 1 | 1 | 1 | | 1 | |
| Connectivité / mise en réseau | 1 | 1 | 1 | 1 | 1 | 1 |
| International / expansion internationale | 1 | 1 | 1 | 1 | 1 | |
| Soutien / promotion | 1 | 1 | 1 | 1 | 1 | |
| Utilisation pratique / commercialisation | 1 | 1 | 1 | 1 | 1 | 1 |
| Évaluation | 1 | | 1 | 1 | 1 | |
| Normes / lignes directrices / règlements / institutions / normes | 1 | 1 | 1 | 1 | 1 | 1 |
| Conseil | 1 | 1 | 1 | | 1 | |
| Symposium / forum / événements | 1 | | 1 | | | 1 |
| Mise en œuvre sociale / démonstration | 1 | 1 | 1 | 1 | 1 | 1 |
| Données | 1 | 1 | | 1 | 1 | |
| Étude de cas avancée | | 1 | 1 | 1 | | |
| Manufacture | | | 1 | | | |
| Sécurité | | | 1 | | 1 | |
| Expérience / banc d'essai | | | 1 | 1 | 1 | |

Tableau 5

| | Intelligence artificielle/IA | Big Data | IdO | Robot | Conduite automatique et conduite automatisée | Drone | VR/AR |
|--|------------------------------|----------|-----|-------|--|-------|-------|
| Recherche et développement | 1 | 1 | 1 | | 1 | | |
| Innovation / croissance économique / transformation / création | 1 | | 1 | | | 1 | |
| Ressources humaines | 1 | 1 | 1 | 1 | | | |
| Amélioration de l'environnement | 1 | | | | | 1 | |
| Investissement | 1 | | 1 | | | | |
| Disparité | 1 | | | | | | |
| International | 1 | | 1 | | 1 | | |
| Réseau | 1 | | 1 | | 1 | | |
| Système | 1 | | | | | | |
| Normes / lignes directrices / règlements / systèmes | 1 | | 1 | 1 | 1 | 1 | |
| Changement organisationnel / structure organisationnelle / Changement d'entreprise | 1 | | | 1 | | | |
| Emploi / main-d'œuvre / population active / marché du travail / pénurie de main-d'œuvre / Travailler | 1 | | 1 | 1 | | | |
| Productivité | 1 | | | | | | |
| Décentralisation (du pouvoir) | 1 | | | | | | |
| Entrepreneuriat/Start-up | 1 | 1 | 1 | 1 | | | |
| Commerces militaires | 1 | | | | | | |
| Concurrence internationale / concurrence | 1 | 1 | 1 | 1 | | | |
| Transactions scripturales/sans espèce | | 1 | | | | | |
| Données | | 1 | 1 | | 1 | | |
| Soins médicaux | | 1 | | | | | |
| Catastrophe | | 1 | | | | | |
| Blockchain | | 1 | 1 | | | | |
| Hygiène | | 1 | | | | | |
| Sécurité | | | 1 | 1 | | | |
| Rassurer / acceptation / anxiété / compréhension du public | | | | 1 | 1 | 1 | |
| Industrialisation | | | | | 1 | | |
| Expérience | | | | | 1 | 1 | |
| Mise en œuvre / démonstration | | | | | 1 | | |

| | | | | | | | |
|--|--|--|--|--|---|--|--|
| Logistique | | | | | 1 | | |
| Vie privée / informations personnelles | | | | | 1 | | |

Tableau 6

| | Société 5.0 | | Quatrième révolution industrielle | |
|--|---------------------|-------|-----------------------------------|-------|
| | Nombre d'étiquettes | % | Nombre d'étiquettes | % |
| (a) Positionnement | 11 | 5%. | 26 | 7%. |
| b) Domaines d'application et avantages | 14 | 6%. | 54 | 15%. |
| c) Défis et mesures | 180 | 76%. | 159 | 43%. |
| (d) Défis | 32 | 14%. | 128 | 35%. |
| Total | 237 | 100%. | 367 | 100%. |

CHAPITRE III. Applications de l'AI au Mexique.

Une vue de l'intérieur

Jorge Cerdio³⁵²

I. Introduction

L'objectif de ce document est de présenter certaines des utilisations les plus répandues et les plus étendues des applications d'intelligence artificielle (IA) au Mexique ainsi que le cadre réglementaire applicable aux applications d'IA. Nous nous efforcerons de représenter avec précision le contexte dans lequel chaque système d'intelligence artificielle fonctionne. Le contexte, ainsi qu'une brève description du système, devraient fournir au lecteur suffisamment d'informations pour permettre des comparaisons avec d'autres juridictions et pour tenter d'éclaircir les complexités de la réglementation potentielle des applications d'IA au Mexique et ailleurs. Le principal problème de la présentation du cas mexicain est l'absence d'une source d'information systématique sur les applications de l'IA. Il existe trois sources d'information principales qui ne sont pas équivalentes en termes de qualité des données contenues ni de pertinence pour les besoins du présent document. La première source d'information est constituée par les travaux universitaires des différents centres de recherche nationaux, les chercheurs dans le domaine de l'IA qui travaillent dans différentes juridictions et institutions locales. La difficulté consiste ici à sélectionner la littérature adéquate pour mettre en évidence les applications déjà matures de l'IA et à écarter les projets de recherche précoces ou les solutions non encore éprouvées. La deuxième source d'information est l'écosystème d'innovateurs et de start-ups qui sont dispersés dans différentes villes du territoire mexicain ; des individus et des petites entreprises qui émergent avec des propositions à valeur commerciale autour de solutions opérationnelles basées sur les applications de l'IA. Enfin, la troisième source d'information est la source officielle, c'est-à-dire la source des gouvernements fédéral et locaux mexicains qui publient des rapports et annoncent les politiques publiques dans différents médias (y compris les sites web officiels). L'information publique qui équivaut à une action publique en faveur des initiatives d'intelligence artificielle. L'état des données publiques au Mexique est loin d'être celui d'un pays à données totalement ouvertes, ce qui fait de la collecte d'informations publiques pertinentes une tâche longue et difficile pour tout chercheur. Au-delà des différents niveaux de qualité des informations provenant de ces sources et donc des exemples d'applications de l'IA, nous nous efforcerons de présenter un puzzle émergent, dont les pièces s'emboîtent selon les deux principaux moteurs de l'utilisation des applications de l'IA au Mexique. En d'autres termes, au lieu de présenter des groupes d'applications par famille de technologies, nous serons guidés par l'idée que la plupart des applications d'IA sont une réponse à un problème social et technique spécifique. Cette approche est appropriée dans un pays comme le Mexique où il est plus difficile d'innover et de développer des idées novatrices que dans d'autres pays où les ressources et les écosystèmes sont mis en réseau pour adopter de nouvelles solutions à des problèmes complexes. Cependant, le Mexique demeure la onzième économie mondiale et le deuxième partenaire commercial (en volume d'exportations) du marché américain. Malgré les profondes inégalités de la société mexicaine, les forces économiques continuent d'encourager et de récompenser la création d'applications d'IA par les secteurs privé et social. Le Mexique a atteint un stade où nous pouvons nous demander comment nous voulons

³⁵² Professeur titulaire au département de droit de l'*Instituto Tecnológico Autónomo de México* (ITAM).

réglementer les applications de l'IA. Le lecteur pourra s'apercevoir que l'intelligence collective est présente au Mexique pour construire un programme national pour l'IA ; un programme qui nécessite la structuration et le leadership des institutions publiques.

D'un point de vue économique, nous pourrions être tenté de dire qu'il vaut mieux laisser de côté les marchés non réglementés des applications de l'IA. Cependant, c'est de ces écosystèmes émergents d'innovateurs au Mexique qu'émerge majoritairement l'opinion selon laquelle nous devons réglementer les applications de l'IA, en gardant toujours à l'esprit les principes de l'État de droit, des droits de l'Homme et de la démocratie. La première partie de ce document traitera de la présentation des principaux exemples des applications de l'IA au Mexique (I), puis nous nous pencherons sur le cadre réglementaire limité et les politiques publiques (II). Nous réserverons une brève dernière partie à quelques conclusions.

II. Observations des applications publiques et privées de l'IA au Mexique

Au Mexique, l'utilisation de l'intelligence artificielle (IA)³⁵³ et l'ensemble des technologies qui lui sont associées progressent dans certains secteurs de l'économie et des programmes publics mexicains. La dynamique générée par l'application de l'intelligence artificielle dans certains secteurs de l'économie mexicaine se produit dans le contexte d'un pays sous-développé en termes de dépenses de recherche et développement. Le Mexique consacre chaque année 0,3 % de son PIB à l'innovation et à la technologie³⁵⁴. Si l'on ajoute à cela le fait que 20 % de la population mexicaine vit avec moins de 5,50 dollars américains par jour aux prix internationaux, l'explication de l'omniprésence des initiatives et des applications de l'intelligence artificielle est à rechercher dans la politique spécifique menée par le gouvernement du pays combinée à la nécessité de rechercher l'innovation dans le secteur privé. À cet égard, il n'y a pas d'adoption normale et généralisée des technologies de l'intelligence artificielle au Mexique. Ce que nous observons, ce sont des secteurs spécifiques qui adoptent rapidement les technologies de l'IA à des fins spécifiques et, en parallèle, des applications prometteuses qui se généraliseront dans les prochaines années. Il existe en effet une nouvelle génération d'innovateurs qui appliquent des technologies d'IA déjà matures, produisant une vague non coordonnée de produits et de services qui contribuent néanmoins à la visibilité de l'IA au Mexique³⁵⁵. Au lieu de regrouper les technologies d'IA par type d'architecture, il est plus utile de décrire les initiatives les plus répandues et les plus importantes en utilisant les critères de la source de l'innovation. D'une part, nous avons vu que le gouvernement fédéral et les autorités locales mexicaines soutiennent l'adoption des technologies d'IA pour des tâches gouvernementales spécifiques de telle sorte qu'elles ont des effets profonds sur la vie quotidienne des citoyens mexicains (A). D'autre part, les acteurs privés, qu'ils soient bien financés (comme le système bancaire) ou qu'ils soient des micro-innovateurs, sont maintenant la force motrice d'une myriade de

³⁵³ Tout au long de ce document, nous nous appuyons sur la définition de l'IA du AI HLEG, groupe d'experts indépendants qui a été mis en place par la Commission européenne en juin 2018. Par conséquent, au lieu de l'IA, nous utilisons indifféremment le concept de système ou d'application d'IA. Voir *Une définition de l'IA : Principales capacités et disciplines scientifiques*, Groupe d'experts de haut niveau sur l'intelligence artificielle, Commission européenne, 8 avril 2019.

³⁵⁴ <https://data.oecd.org/mexico.htm#profile-innovationandtechnology> (consulté le 31 octobre 2020).

³⁵⁵ Bien entendu, nous n'avons pas l'intention d'énumérer dans ce chapitre tous les projets liés à l'IA en cours de développement ou de déploiement. Il existe des innovations passionnantes sur la scène mexicaine de l'IA. Pour une vision plus large des initiatives d'IA, voir les "Cas d'utilisation de l'IA" dans la section "VERS UNE STRATÉGIE D'IA AU MEXIQUE" ci-dessous ; voir "Annexe II" du rapport sur l'IA et la croissance économique <https://news.microsoft.com/uploads/prod/sites/41/2018/11/IA-y-Crecimiento-MEXICO.pdf> (consulté le 31 octobre 2020). Voir également le rapport sur les robots de service préparé par l'Académie mexicaine de calcul <https://www.researchgate.net> (consulté le 31 octobre 2020). Enfin, il existe un registre de projets d'IA documentés dans le *Komputer Sapiens*, une revue éditée par la Société mexicaine d'intelligence artificielle, à l'adresse <http://smia.mx/komputersapiens/> (consulté le 31 octobre 2020).

services *Brotherton* qui disposent d'une technologie basée sur l'IA, des produits et des services qui sont désormais quotidiens pour les Mexicains (B).

i. Les applications d'IA d'initiatives publiques au Mexique

Les applications de l'IA ont été reconnues comme un outil utile au service du bien public et pour atteindre les objectifs politiques dans de nombreux secteurs des gouvernements mexicains, tant au niveau fédéral que local. Certaines applications de l'IA que nous trouvons dans le secteur public mexicain sont comparables aux efforts déployés dans d'autres juridictions, comme l'Agence du service des recettes intérieures, les agences gouvernementales pour la prévention du blanchiment d'argent et des activités de financement du terrorisme, ou les services de santé publique. Il existe une tendance mondiale à utiliser les applications d'IA pour exécuter plus efficacement les tâches qui impliquent le traitement de grandes collectes de données provenant de différentes sources, dont le volume augmente rapidement et dans divers formats (Big Data). Les applications d'IA tentent également de résoudre les problèmes sociaux qui pourraient découler de la pandémie mondiale de la COVID-19. À cet égard, nous pouvons applaudir les efforts du gouvernement mexicain pour maîtriser la pandémie sans aide extérieure. Il existe d'autres applications de l'IA qui répondent aux spécificités du pays mexicain : le vaste flux historique de ressortissants migrant vers les États-Unis, l'industrie naissante de l'électricité éolienne et l'activité sismique destructrice récurrente dans de vastes zones du territoire mexicain.

L'IA pour la prévention de l'évasion fiscale

Depuis sept ans, le Service des impôts du Mexique (SAT, de son sigle en espagnol) suit une application d'IA pour détecter les fraudes fiscales³⁵⁶. Ce système vise à lutter contre une forme de criminalité organisée. Les personnes impliquées dans des sociétés fictives utilisent des transactions légitimes simulées de services et de biens. Ces sociétés fictives produisent des millions de recettes fiscales afin d'éviter le paiement des taxes sur le revenu et sur la valeur ajoutée au gouvernement mexicain. En juin 2020, le Ministère des Finances a fait une déclaration publique annonçant le dépôt d'accusations pénales à l'encontre de 8 212 contribuables (particuliers et entreprises)³⁵⁷. Les opérations frauduleuses se sont élevées à environ 55 083 957 551 086 pesos mexicains (l'équivalent approximatif de 2 223 125 505 255 euros). L'application d'IA du SAT a analysé plus de vingt-deux millions de recettes et d'opérations des contribuables. Cette application très complexe de l'IA a été documentée par certains des plus grands spécialistes du pays dans ce domaine des années auparavant dans un document de recherche³⁵⁸: un livre blanc financé par l'ambassade britannique a révélé par la suite que l'application d'IA du SAT fonctionnait en « identifiant les perturbations de modèle dans les données analysées en utilisant le studio R, le langage Python, et les Redis en mémoire des bases de données ». ³⁵⁹ Ces dernières années, la réglementation fiscale mexicaine est passée à un régime uniquement électronique pour les reçus fiscaux. Chaque reçu fiscal sert de nœud pour identifier les multiples données de l'opération. Avec près de 12 milliards de reçus fiscaux émis juste en 2019, le SAT espère augmenter les recettes internes du PIB de 3 points³⁶⁰.

³⁵⁶ <https://www.gob.mx/innovamx/articulos/inteligencia-artificial-131287> (consulté le 31 octobre 2020).

³⁵⁷ <https://www.gob.mx> (consulté le 31 octobre 2020).

³⁵⁸ http://omawww.sat.gob.mx/gobmxtransparencia/Paginas/documentos/estudio_opiniones/Evasion_en_IVA_Analisis_de_Red.es.pdf (consulté le 31 octobre 2020).

³⁵⁹ TOWARDS AN AI STRATEGY IN MEXICO : Harnessing the AI Revolution, White Paper Report, British Embassy in Mexico through the Prosperity Fund, Oxford Insights and C Minds, juin 2018, p. 23. Consultation électronique à l'adresse <https://www.oxfordinsights.com/mexico> (consulté le 31 octobre 2020).

³⁶⁰ <https://www.eluniversal.com.mx/cartera/sat-va-por-mas-recaudacion-con-inteligencia-artificial> (consulté le 31 octobre 2020).

L'AI contre le blanchiment d'argent

Ce n'est qu'en accord avec la tendance mondiale que la lutte contre le blanchiment d'argent a vu l'émergence d'efforts coordonnés entre les juridictions. Le blanchiment d'argent est un sujet particulièrement sensible au Mexique, compte tenu de la lutte contre les cartels de la drogue. Sur une échelle plus nationale, nous voyons aujourd'hui aussi l'utilisation des applications de l'IA dans la lutte contre la corruption, (c'est-à-dire la corruption au niveau de l'État, organisée à partir des plus hautes sphères du gouvernement, visant à diluer le budget public au moyen de transactions complexes). Au sein du Ministère des Finances, il existe une unité spécialisée dans le renseignement financier (UIF, par son acronyme en espagnol). L'objectif principal de l'UIF est, entre autres, de mettre en œuvre les compromis internationaux du gouvernement mexicain en tant que membre permanent du *Groupe d'action financière sur le blanchiment de capitaux* (GAFI), également connu sous son nom français de *Groupe d'action financière* (GAFI³⁶¹)³⁶². Cette unité spécialisée est remarquable, non seulement par l'utilisation intensive des techniques de *machine learning* et de *Big Data* mais aussi par le fait que le résultat de ses enquêtes soit considéré comme une preuve tangible par le bureau du procureur général³⁶³. En d'autres termes, le Mexique fournit un exemple où une application de l'IA dans les enquêtes criminelles entraîne l'admission comme preuve des conclusions produites par un système d'IA³⁶⁴ : il s'agit jusqu'à présent, d'un précédent notoire en Amérique latine.

L'IA pour le développement des services de santé publique

Fin 2019, les résultats d'une enquête nationale sur la santé et la nutrition ont révélé qu'environ 75,2 % des adultes mexicains étaient en surpoids ou obèses, tandis que 10,3 % d'entre eux souffraient également de diabète³⁶⁵. Ces données révèlent un grave problème de santé publique qui touche non seulement les adultes mais aussi les enfants. Les résultats de l'enquête ont montré que les enfants de femmes qui étaient obèses pendant leur grossesse ont 1,4 fois plus de chances d'être en surpoids à l'âge de 3 ans que les enfants de femmes qui avaient un poids adéquat. Les femmes obèses avant la grossesse ont 4,5 chances de plus d'avoir un enfant en surpoids à l'âge de 7 ans que les enfants de mères ayant un poids adéquat avant la grossesse³⁶⁶. Cette information est essentielle pour prévenir et inverser les effets de l'obésité ; on sait que l'accent mis par les politiques publiques sur l'information a un effet positif sur les choix de santé que les gens font au fil du temps ; et cela a été particulièrement vrai dans le cas du Mexique³⁶⁷. Le coût de la diffusion d'informations pertinentes en temps réel à l'ensemble de la population est élevé mais les applications de l'IA peuvent contribuer à diminuer ce coût et aider la population à prendre de meilleures décisions en matière de santé. *MiSalud* est une application d'IA qui délivre des messages interactifs en utilisant

³⁶¹ <https://www.uif.gob.mx/> (consulté le 31 octobre 2020).

³⁶² https://www.gob.mx/cms/uploads/attachment/file/520516/Comunicado_UIF_011.pdf (consulté le 31 octobre 2020).

³⁶³ Dans un autre précédent, un juge a décidé que l'UIF sera admise comme demandeur avec le bureau du procureur général pour poursuivre une affaire de corruption politique. Voir le numéro de dossier FED/FECC/UNAI-CDMX/0000002/2020 sur <https://datos.cdmx.gob.mx> (consulté le 31 octobre 2020).

³⁶⁴ À cet égard, il existe des points de comparaison entre le cas mexicain et le cas australien. *Australian Transaction Reports and Analysis Centre* (Austrac), tel que rapporté par *l'Association of Certified Financial Crime Specialist* à l'adresse <https://www.acfcs.org/austrac-using-machine-learning-to-better-uncover-interconnected-criminal-groups-improve-aml-alerts/> (consulté le 31 octobre 2020).

³⁶⁵ https://ensanut.insp.mx/encuestas/ensanut2018/doctos/informes/ensanut_2018_presentacion_resultados.pdf (consulté le 31 octobre 2020).

³⁶⁶ Yu Z, Han S, Zhu J, Sun X, Ji C, Guo X. Pre-pregnancy body mass index in relation to infant birth weight and offspring overweight/obesity : a systematic review and meta-analysis. *PloS one*. 2013;8:e61627. DOI. <https://doi.org/10.1371/journal.pone.0061627> et Weng SF, Redsell SA, Swift JA, Yang M, Glazebrook CP. Revue systématique et méta-analyses des facteurs de risque de surpoids chez les enfants identifiables pendant la petite enfance. *Arch Dis Child*. 2012;97:1019-26. DOI. <https://doi.org/10.1136/archdischild-2012-302263>

³⁶⁷ Voir l'étude sur l'obésité au Mexique réalisée par l'Institut national de santé publique, disponible sur https://www.insp.mx/resources/imagenes/stories/2019/Docs/190213_LaObesidadenMexico.pdf, p. 28 (consulté le 31 octobre 2020).

les systèmes de messagerie Facebook et Twitter. Elle fournit non seulement des conseils de santé, mais interagit également avec les utilisateurs pour leur rappeler leurs rendez-vous et contrôles médicaux. Elle fournit également des informations en temps réel sur les pandémies ou les risques pour la santé, et a la capacité de donner des conseils de santé. Le programme pilote a débuté en 2017, avec un accent sur la santé maternelle. *MiSalud* a envoyé des messages à cinq mille femmes, avec des conseils pour les aider à améliorer leur santé et celle de leurs bébés. On a constaté une amélioration de la fréquence des consultations médicales et du poids des nouveau-nés, ainsi qu'une diminution du taux de surpoids des mères et des bébés. Le programme s'est maintenant étendu au diabète, à l'obésité, à la vaccination des enfants et à l'évaluation des risques de toxicomanie³⁶⁸.

Il existe plusieurs possibilités d'utilisation de l'IA au Mexique et ailleurs. Face à la pandémie mondiale de la COVID-19, différents pans de la société ont accéléré la mise en œuvre de plusieurs technologies afin de contenir l'épidémie et de faire face à la propagation du virus.

Systèmes d'IA de réponse à la pandémie COVID-19

Dans le scénario COVID-19, la détection précoce des groupes de personnes potentiellement infectées est essentielle pour contenir le virus. La mégapole de Mexico est habitée par près de neuf millions de personnes. Au cours des premiers stades de la pandémie COVID-19, le gouvernement local de Mexico a mis en place deux systèmes d'aide utilisant les techniques d'IA. Le premier système est un système interactif de diagnostic par SMS. Il fonctionne de la manière suivante : en envoyant un message SMS avec le mot "COVID-19", l'utilisateur reçoit une série de questions. Ces questions permettent d'évaluer les risques et les probabilités pour que l'individu soit infecté par le virus. Lorsque le risque d'infection est élevé, le système déclenche une alarme que les services de santé utilisent pour localiser la personne ou pour lui fournir des informations complémentaires afin qu'elle puisse se rendre dans un centre de santé pour y être traitée. Lorsque la contagion est certaine, ce même système sert à fournir un kit de test de diagnostic rapide à l'usage du groupe d'individus qui ont interagi avec le porteur du virus diagnostiqué, ainsi qu'à recevoir une aide économique et sociale supplémentaire pour l'individu et sa famille. Le second système fonctionne en parallèle du premier et fournit une mise à jour en temps réel sur la disponibilité des places et l'encombrement des hôpitaux en fonction des informations GPS de l'utilisateur. En estimant l'hôpital ou l'établissement de santé le plus proche et le moins encombré, ce système maximise le temps de réponse pour les personnes potentiellement infectées par le virus et minimise le risque de contagion ultérieure. Jusqu'à présent, le premier système - le système de diagnostic rapide - dessert environ vingt millions d'utilisateurs et a été développé par l'Agence numérique pour l'innovation publique de Mexico (DAPI). Le second système est disponible sous la forme d'une application pour les systèmes Android et iOS.³⁶⁹ Les informations exploitées par les deux systèmes sont traitées dans un *Big Data Center* afin de créer une stratégie au jour le jour pour les équipes d'épidémiologistes qui travaillent sur les clusters de la ville, ainsi que pour réévaluer le comportement de l'épidémie sur les 1495 km² de la mégapole. Parallèlement aux systèmes DAPI, il existe COVIDBot, un agent d'IA disponible sur WhatsApp formé pour répondre aux questions sur le contenu multimédia interactif produit par les organisations de santé. Le COVIDBot est une création philanthropique de la société mexicaine *Intevolution*. L'objectif principal de COVIDBot est de lutter contre les fausses nouvelles (fake news) liées à la pandémie qui circulent sur l'application WhatsApp. COVIDBot est un service gratuit qui absorbe en temps réel les registres et les données de l'Organisation mondiale de la santé, du Centre de contrôle des maladies (*Center for Disease*

³⁶⁸ <https://www.gob.mx/misalud/> (consulté le 31 octobre 2020).

³⁶⁹ <https://adip.cdmx.gob.mx/proyectos> (consulté le 31 octobre 2020).

Control) aux États-Unis et du ministère mexicain de la Santé. L'une des principales caractéristiques de COVIDBot est la localisation du laboratoire certifié le plus proche pour effectuer des tests COVID-19³⁷⁰.

Outre la pandémie mondiale, l'utilisation de l'IA dans les services de santé publique pourrait être appliquée à une politique plus large au Mexique : celle qui vise à cibler les besoins des personnes défavorisées. La prise en charge des groupes défavorisés dans un pays où les inégalités sont très marquées représente un défi de taille pour tout gouvernement au pouvoir, et un terrain inconnu pour l'exploitation des applications de l'IA afin de venir en aide à ceux qui en ont le plus besoin.

L'aide aux immigrants mexicains aux États-Unis

Historiquement, les Mexicains émigrent du Mexique vers les États-Unis à la recherche de meilleures opportunités, qu'il s'agisse d'une activité saisonnière liée à l'agriculture ou d'une immigration permanente visant un séjour plus long. En 2018, il y avait environ 11,17 millions d'immigrants nés au Mexique aux États-Unis³⁷¹. Les immigrants mexicains aux États-Unis sont confrontés à un certain nombre de défis et de besoins. Répondre aux besoins de la population mexicaine vivant aux États-Unis a également été une priorité pour l'État mexicain. Un consulat mexicain n'a pas de grandes marges de manœuvre pour aider pour les groupes dispersés (et souvent cachés) de ressortissants qui ont besoin d'accéder à des informations vitales concernant leurs droits et l'aide juridique. L'enregistrement d'un nouveau-né comme mexicain, le retour d'un défunt relatif en sol étranger et le respect des accords judiciaires sur les pensions alimentaires sont des exemples de situations qui nécessitent l'aide du service extérieur mexicain. La demande la plus fréquente est le renouvellement d'un passeport, puisqu'il s'agit du principal moyen d'identification des Mexicains aux États-Unis. Si une personne est détenue, elle a le droit de bénéficier de l'aide juridique du consulat mexicain, et l'exercice du droit à une identification par passeport est essentiel. L'utilisation de robots pour interagir avec les immigrants mexicains est actuellement une réponse prometteuse pour répondre aux besoins de millions de Mexicains aux États-Unis. L'application AI appelée "*Asistente Virtual SRE-UNAM*" [Assistant *virtuel* SRE-UNAM] reconnaît et identifie le type de demande administrative qu'un immigrant peut avoir. Cette application génère un rendez-vous automatisé en fonction des réponses fournies lors de l'interaction avec l'utilisateur. L'assistant virtuel est un développement conjoint entre le Ministère des Affaires étrangères et l'Université nationale autonome du Mexique (UNAM, par son acronyme en espagnol)³⁷².

L'innovation motivée par l'intérêt public peut parfois découler d'un contexte particulier qui offre l'occasion de s'attaquer à un problème complexe, et c'est le cas de la migration humaine. Dans d'autres cas, la complexité va de pair avec la nécessité de faire face aux besoins de créer de nouveaux moyens pour promouvoir un meilleur environnement. C'est le cas de l'IA appliquée à la production d'électricité d'Eolic au Mexique.

³⁷⁰ <https://www.infobae.com/tecnologia/2020/04/01/asi-es-el-chatbot-mexicano-que-lucha-contra-la-desinformacion-sobre-coronavirus/> (consulté le 31 octobre 2020).

³⁷¹ Duffin Erin, *Number of Mexican Immigrants in the United States 1850-2018*, <https://www.statista.com/statistics/673350/mexican-immigrants-in-the-united-states/> (consulté le 31 octobre 2020).

³⁷² https://www.youtube.com/watch?reload=9&v=1dv-gVKBABA&ab_channel=UNAMGlobal (consulté le 31 octobre 2020). Un rapport technique sur l'application se trouve ici : <https://cscwvicttechnologies.files.wordpress.com/2020/10/asistente-virtual-saul-esparza.pdf> (consulté le 31 octobre 2020). Derrière l'équipe d'innovateurs qui a développé l'application se trouve le professeur Saip Savage, qui a été nommé l'un des Innovateurs de moins de 35 ans Amérique latine 2018 par le MIT Technology Review : <https://www.innovatorsunder35.com/the-list/saip-savage/> (consulté le 31 octobre 2020).

L'IA pour la production d'énergie éolienne

Fin 2019, les parcs éoliens produisaient 15 000 mégawatts d'énergie pour un prix de 826 millions de dollars. La mise en œuvre d'Eolic Electricity pose de nombreux défis, parmi lesquels la prévision de la force du vent pour estimer l'énergie qui sera fournie au réseau électrique public. Le défi de la variabilité est généralement relevé à l'aide de modèles de simulation et de simulations prédictives qui nécessitent des ressources informatiques importantes (en raison de la nécessité de nombreuses itérations). L'utilisation de techniques d'IA pour prévoir la quantité d'énergie qui sera disponible pour être injectée dans le réseau à partir d'un parc éolien a permis de réduire considérablement ce problème. La technique d'IA pour la prévision de l'énergie éolienne utilise des informations météorologiques collectées pendant plusieurs années à l'aide de réseaux bayésiens dynamiques (DBN). Les résultats de cette application d'IA ont été comparés de manière satisfaisante aux résultats des prévisions des techniques de séries chronologiques précédentes, ce qui indique que les DBN sont un outil prometteur pour la prévision³⁷³ de l'énergie éolienne. La tendance que nous observons ici est l'utilisation d'une approche de *machine learning* qui peut être utilisée pour remplacer le modèle de simulation rigoureux par un modèle de substitution (par exemple en utilisant des algorithmes de régression vectorielle)³⁷⁴. Le modèle de substitution peut être obtenu en peu de temps et avec beaucoup moins de ressources de calcul³⁷⁵. La recherche et les applications de l'IA sont en phase avec la croissance rapide des parcs éoliens sur le territoire mexicain. L'une des questions clés pour promouvoir la compétitivité de la récolte d'énergie éolienne est l'atténuation des changements de la demande et la variabilité des éoliennes.

La croissance économique associée aux conditions naturelles est un facteur clé pour l'industrie des énergies vertes. D'autres conditions naturelles qui représentent un défi en raison de leur pouvoir perturbateur sont des phénomènes qui génèrent la destruction et le chaos, tels que les éruptions volcaniques ou les tremblements de terre, qui sont pourtant imprévisibles. Le Mexique a une longue histoire de tremblements de terre. Une grande partie de son territoire est touchée par des mouvements sismiques. Au cours des 365 derniers jours, il y a eu 2 215 tremblements de terre d'une magnitude de 1,5 ou plus³⁷⁶. L'IA et l'IdO, combinées à des techniques d'apprentissage machine dans le cadre de l'informatique dématérialisée, font partie des outils dont disposent les Mexicains pour gérer les tremblements de terre et les victimes.

L'IA pour la prévention des dommages causés par les tremblements de terre

Le Mexique dispose depuis 1991 d'un des rares systèmes d'alerte précoce aux tremblements de terre (EEW) au monde, appelé SASMEX ; mais une phase de développement intense dans ce même domaine a commencé plus tôt, plus précisément après le tremblement de terre de 1985³⁷⁷. SASMEX travaille avec des stations de capteurs le long de la côte sud du Pacifique pour détecter l'activité sismique. Lorsqu'un événement sismique est détecté, un système d'alerte radio à ondes

³⁷³ Ibarguengoytia-González P H, Borunda-Pacheco M, Reyes-Ballesteros A, García-López Uriel Alejandro, Wind Power Forecasting using Artificial Intelligence Tools, *Ingeniería, Investigación y Tecnología*, 2018, 19-4 : 1-11. DOI. <http://dx.doi.org/10.22201/ii.25940732e.2018.19n4.033>

³⁷⁴ Santamaría-Bonfil G, Reyes-Ballesteros A, Gershenson C, *Wind speed forecasting for wind farms : A method based on support vector regression*, *Renewable Energy*, 85, 2016 : 790-809, ISSN 0960-1481, DOI. <https://doi.org/10.1016/j.renene.2015.07.004>.

³⁷⁵ Rosado-Tamariz E, Zuniga-Garcia M A, Santamaria-Bonfil G, Batres R, *A machine-learning approach to speed-up simulation towards the design of optimum operating profiles of power plants*. Dans les Actes de la 8e Conférence internationale sur l'informatique, l'environnement, l'énergie et les applications (IEEA '19). Association for Computing Machinery, New York, NY, USA : 194-199. DOI. <https://doi.org/10.1145/3323716.3323735>. Le projet DeepMind d'Alphabet a également mené des recherches pour résoudre les problèmes de stabilité et de variance des parcs éoliens, de sorte que le Mexique a la ressource qui pourraient permettre de trouver une solution au problème énergétique mondial. Voir <https://deepmind.com/blog/article/machine-learning-can-boost-value-wind-energy> (consulté le 31 octobre 2020).

³⁷⁶ https://earthquaketrack.com/p/mexico/recent?mag_filter=7 (consulté le 31 octobre 2020).

³⁷⁷ http://www.cires.org.mx/sasmex_es.php (consulté le 31 octobre 2020).

longues est activé. À Mexico, chaque école publique est équipée d'un récepteur et d'un système sonore permettant d'émettre une alerte. Entre les écoles publiques et les autres sites, on compte environ 90 000 récepteurs à Mexico et dans la zone métropolitaine. En moyenne, les systèmes d'alarme donnent l'alerte cent secondes avant l'arrivée de l'onde sismique. Ces cent secondes, accompagnées d'une formation civique intensive et de répétitions pour mener des évacuations en toute sécurité des bâtiments, sont la clé pour prévenir les décès en cas de séisme massif. SASMEX dispose également d'une application pour Android et iOS, ainsi que d'un compte Twitter pour diffuser les avis des stations de capteurs. La demande de systèmes d'alerte précoce précis et généralisés a donné naissance à des applications d'IA sur le marché mexicain. Il existe deux principales entreprises dans ce domaine : Skyalert³⁷⁸ et Grillo³⁷⁹. Les deux entreprises utilisent l'Internet des objets et *cloud computing* pour générer des alertes sismiques par des applications sur des appareils mobiles ; et, dans le cas de Skyalert, par des systèmes d'alarme vocale. Skyalert est davantage orientée vers le marché des entreprises en proposant des formations et des actions de sensibilisation ainsi que des systèmes d'alarme sur site. Cette société fournit également une alerte de cent vingt secondes avant l'arrivée de la vague destructrice, avec un système d'alarme personnalisé à focalisation GPS qui filtre les événements sismiques pour l'utilisateur. Grillo a un record de zéro faux positifs, et a régulièrement réussi les analyses comparatives avec Skyalert et SASMEX. Les deux entreprises utilisent le *cloud computing* pour traiter les données depuis leurs propres capteurs. Grillo travaille avec les services de *cloud computing* d'AWS, qui ont des délais de livraison de 100 ms du capteur au Cloud, tandis que Skyalert traite ses données avec la plateforme Azure de Microsoft. Grillo dispose d'une gamme de capteurs destinés à la détection des dommages structurels sur les bâtiments afin d'éviter d'autres pertes humaines après qu'un tremblement de terre ait touché la structure. Au total, entre Grillo et Skyalert, ces entreprises détiennent 97 % des parts de marché des services d'alerte sismique précoce. Une initiative très particulière de Grillo est un partenariat avec IBM et la Linux Foundation. Grillo a ouvert ses données et son code dans le cadre du projet OpenEEW de collaboration mondiale renforcée autour de la technologie de Grillo, dans l'espoir d'encourager des initiatives similaires dans le monde³⁸⁰ entier.

L'objectif des applications de l'IA dans le secteur public au Mexique est le bien commun, la portée et la protection de l'intérêt public au profit des citoyens en général. De nombreux projets d'IA, tant au niveau de la recherche que des applications, s'efforcent de s'attaquer à un problème social, même s'il s'agit d'une initiative privée. Parfois, la frontière entre l'intérêt privé et l'intérêt public n'est pas assez claire pour classer une recherche ou une application spécifique d'IA. Il existe cependant des cas bien précis où l'application de l'IA apparaît comme une réponse à une opportunité commerciale, à un créneau de marché ou à l'innovation délibérée d'entreprises.

ii. Les applications de l'IA d'initiatives privées au Mexique

En l'absence d'un registre général ouvert des demandes d'aide humanitaire au Mexique, nous ne pouvons recenser que quelques-unes des entreprises privées les plus remarquables fondées sur des demandes d'aide humanitaire. Presque toutes les grandes industries utilisent déjà des applications de l'IA, de l'industrie automobile qui utilise des robots et des outils mécatroniques pour assembler les voitures à l'armée de bots que les partis politiques ont utilisée dans le passé pour désinformer ou pour exploiter la division du public. Dans le même temps, la proximité de l'économie mexicaine avec le marché américain a stimulé l'offre de services basés aux États-Unis (ou avec une

³⁷⁸ <https://skyalert.mx/> (consulté le 31 octobre 2020).

³⁷⁹ <https://blog.grillo.io/the-grillo-journey-e60b2dcee224> (consulté le 31 octobre 2020).

³⁸⁰ <https://grillo.io/linux-foundation-hosting-openeew/> et <https://grillo.io/open-source/> (consulté le 31 octobre 2020).

matrice d'entreprises aux États-Unis) sur le territoire mexicain. À cet égard, les Mexicains interagissent avec l'offre désormais mondiale de services autour des produits numériques : Netflix, Uber, DiDi, Facebook, Twitter, Instagram, Amazon et Google, par exemple. Dans ce contexte général, il existe certains domaines d'innovation locale qui sont nés de jeunes entreprises désormais consolidées ou de sociétés plus importantes qui ont investi dans l'utilisation de l'IA dans le cadre de leurs programmes de recherche et développement pour conserver leur avantage concurrentiel. Du côté des start-ups, nous avons un large éventail d'innovateurs qui fournissent aujourd'hui des applications d'IA au marché mexicain des services numériques. Cette gamme d'innovateurs comprend FinTechs, CRM-ChatBots, les analyses de media sociaux, l'immobilier et les robots médicaux en tant que service. Étant donné que les exemples que nous examinons ne concernent que quelques participants du marché américain et mexicain, il est inévitable de constater que, dans la plupart des cas, les initiatives sont la réponse (ou la version) locale d'autres solutions existantes pour les consommateurs et les services anglophones aux États-Unis.

Applications de l'IA dans les FinTechs

Selon l'Agence mexicaine de régulation financière (CNBV, en espagnol), 500 entreprises FinTech opéreraient au Mexique en 2019³⁸¹. Cependant, une autre entreprise indépendante en décompte 640³⁸². Selon les données de Statista, les entreprises FinTech mexicaines exerceront d'ici 2022 des activités d'une valeur d'environ 69 000 millions de dollars, avec un taux de croissance annuel de 17,3 %³⁸³. Avec plus de 158 start-ups, le Mexique possède le plus grand marché FinTech d'Amérique latine, surpassant les marchés brésiliens et colombiens³⁸⁴. Par ailleurs, seulement 4,5 % des entreprises FinTech du Mexique ont cessé leurs activités entre juin 2019 et juin 2020³⁸⁵. Les principales activités de FinTech sont le *machine learning* pour l'évaluation des risques, l'utilisation de *Big Data* et l'IA. La quasi-totalité du cycle économique des entreprises FinTech tourne autour de l'utilisation des technologies liées à l'IA : des paiements et des transferts de fonds, de la gestion des finances personnelles, aux financements participatifs et aux prêts. Il existe également un vaste univers de FinTech et de « technologies » connexes telles qu'InsurTech. L'accent mis sur l'application de l'IA que chaque entreprise FinTech utilise est étroitement lié au domaine dans lequel l'utilisation de l'IA apportera le plus de compétitivité.

Pour Conekta³⁸⁶, une FinTech qui propose ses services en tant qu'agrégateur certifié de paiements (qu'il s'agisse de paiements en espèces, de point à point ou de commerce électronique), la sécurité est primordiale. Cette société a développé Conekta Shield, qui utilise le *machine learning* et les réseaux bayésiens pour modéliser les comportements des consommateurs et des utilisateurs afin de créer des modèles dynamiques pour minimiser la fraude et les risques transactionnels pour les clients³⁸⁷ de Conekta. Comme Conekta, Dapp a une philosophie commerciale similaire, qui consiste à automatiser les paiements dans différents écosystèmes. Contrairement à Conekta, Dapp fournit aux banques la technologie nécessaire pour générer des codes QR. Les utilisateurs scannent le code QR à partir de leur téléphone portable et payent avec la *Wallet App* (application porte-monnaie)

³⁸¹ <https://www.eleconomista.com.mx/sectorfinanciero/Tenemos-mapeadas-mas-de-500-fintech-en-Mexico-CNBV-20190526-0072.html> (consulté le 31 octobre 2020).

³⁸² <https://elceo.com/tecnologia/cnbv-alista-autorizaciones-para-80-empresas-fintech-entre-octubre-y-noviembre/> (consulté le 31 octobre 2020).

³⁸³ <https://www.statista.com/study/45600/statista-report-fintech/>, (consulté le 31 octobre 2020).

³⁸⁴ <https://panamericanworld.com/en/magazine/startups/meet-the-five-most-innovative-mexican-fintechs/> (consulté le 31 octobre 2020).

³⁸⁵ <https://www.finnovista.com/radar/el-numero-de-startups-fintech-en-mexico-crecio-un-14-en-un-ano-hasta-las-441/> (consulté le 31 octobre 2020).

³⁸⁶ <https://conekta.com> (consulté le 31 octobre 2020).

³⁸⁷ <http://www.ebizlatam.com/seguridad-e-inclusion-soluciones-conekta/> (consulté le 31 octobre 2020).

de leur choix. Une fois que le *Wallet* transfère les fonds pour la transaction, le commerce reçoit le paiement³⁸⁸.

D'autres FinTechs se concentrent sur l'utilisation du *Big Data* et du *machine learning* pour établir le profil du client et « l'encourager » tout au long du processus de remboursement.

C'est le cas de Kubo Financiero³⁸⁹, la première communauté de prêt entre pairs en ligne au Mexique, et la première en Amérique latine autorisée par une autorité financière. Kubo Financiero propose des prêts en ligne à partir d'un téléphone portable, allant de 400 à 4 100 dollars américains mais offre également un taux d'intérêt attractif aux micro-investisseurs, qui commence à seulement 22 dollars américains. Le processus de profilage comprend un formulaire de demande, le téléchargement de quelques documents et une vidéoconférence. Les demandeurs de crédit ayant de bons antécédents obtiennent un meilleur taux d'intérêt et de meilleures conditions, tandis que les investisseurs qui prêtent aux emprunteurs obtiennent un meilleur rendement de leurs investissements³⁹⁰. Kubo utilise des modèles de supervision issus de la *data science* pour enrichir ses processus de collecte de données auprès des clients, ainsi que du *machine learning* pour calibrer son mécanisme de notation de crédit. Sur un segment de marché similaire, Kueski³⁹¹ utilise des applications d'IA pour déterminer un profil d'évaluation des risques pour prêter de l'argent en utilisant des informations provenant de points de données non traditionnels, y compris la présence dans les réseaux sociaux³⁹². Les demandes de prêt sont acceptées et payées en quelques minutes. Les emprunteurs peuvent progressivement demander des montants plus élevés. La spécialité de Kueski sont les microcrédits, qui vont de 100 à 200 dollars US pour une durée maximale de trente jours. L'évaluation des risques à l'aide de l'IA peut être appliquée pour améliorer la prédiction du succès d'une petite entreprise. Dans ce domaine, l'IA peut être utilisée pour prédire la probabilité qu'une idée ou un entrepreneur réussira à devenir une entreprise à fort impact dans l'économie mexicaine. Konfío³⁹³ utilise des algorithmes exclusifs et l'analyse de données pour développer des opérations de crédit abordables et pour accélérer le processus de financement des petites et moyennes entreprises (PME). Les PME au Mexique représentent 95 % du marché. Le succès de Konfío repose sur l'utilisation intensive de points de données dans le bilan de ses clients potentiels afin de prédire leurs chances de réussite et d'adapter ensuite un taux de prêt optimal pour produire ce succès dans la PME. En utilisant des algorithmes dans les cycles d'amélioration, Konfío a pu réduire le taux de gravité à la valeur du prêt 10 fois de 2016 à 2017³⁹⁴.

Si le profilage des risques est particulièrement sensible pour les plateformes de prêt d'argent, il est au cœur des préoccupations des entreprises utilisant des devises numériques telles que le Bitcoin³⁹⁵. Au Mexique, Bitso est une société qui pense que la clé de l'inclusion financière mondiale sera obtenue par l'utilisation de Bitcoin et d'autres Crypto-monnaies (la société utilise actuellement 10 types différents de Crypto-monnaies). Elle a beaucoup investi dans les technologies d'IA pour automatiser la vérification préalable des clients afin d'éviter le blanchiment d'argent et le financement du terrorisme. Dans le même temps, elle a simplifié l'interface de son application pour investir, utiliser et effectuer des transactions avec les Crypto-monnaies qui reposent sur la technologie BlockChain.

³⁸⁸ <https://dapp.mx> (consulté le 31 octobre 2020).

³⁸⁹ <https://www.kubofinanciero.com/Kubo/Portal/productos/productos.xhtml>

³⁹⁰ <https://publications.iadb.org/publications/english/document/FINTECH--Innovations-You-May-Not-Know-were-from-Latin-America-and-the-Caribbean.pdf> (consulté le 31 octobre 2020).

³⁹¹ <https://kueski.com> (consulté le 31 octobre 2020).

³⁹² <https://kueski.com/blog/tecnologia/machine-learning-aprendizaje-supervisado/> (consulté le 31 octobre 2020).

³⁹³ <https://konfio.mx/> (consulté le 31 octobre 2020).

³⁹⁴ <https://expansion.mx/tecnologia/2017/08/15/konfio-un-ejemplo-de-como-las-fintech-dejan-atras-a-los-bancos> (consulté le 31 octobre 2020).

³⁹⁵ <https://bitso.com> (consulté le 31 octobre 2020).

Enfin, il existe un segment de marché qui s'appuie également sur des applications intensives de *machine learning* et de *Big Data* pour interagir avec les clients afin de leur offrir des services et de garantir des transactions compétitives. Les entreprises FinTech, appelées « néobanques », sont un exemple florissant de l'utilisation d'applications liées à l'IA. Il serait hors-sujet, dans le cadre de cette enquête, de détailler tous les participants du marché mexicain³⁹⁶. Il suffit de reconnaître que toutes les néobanques utilisent une forme ou une autre de technologie liée à l'IA : qu'il s'agisse d'une interaction autonome avec les clients, de l'extraction de données pour évaluer les risques, de la prévention de la cybercriminalité ou de services personnalisés et micro-ciblés à partir de modèles de structure de Big Data.

Applications de l'IA dans l'immobilier

L'évaluation des risques à l'aide de points non traditionnels pour créer un profil est une technique d'IA assez courante dans le secteur FinTech. Une utilisation moins courante des techniques de modélisation prédictive des risques est utilisée dans le domaine du crédit-bail des ménages. Le marché mexicain de la location de biens urbains pour l'usage domestique des ménages est un marché dominant. Certaines enquêtes révèlent qu'environ 42 % de la population mexicaine vit dans une maison en location³⁹⁷. Homie³⁹⁸ utilise un modèle prédictif comportant des centaines de variables pour estimer la probabilité qu'un locataire manque à ses obligations de paiement du loyer (taux de défaillance des locataires). Le modèle d'apprentissage automatique n'analyse pas seulement les revenus et les antécédents de crédit du demandeur, mais aussi le degré de respect des engagements financiers. Homie sélectionne le meilleur candidat pour les propriétaires. Après avoir sélectionné le meilleur candidat, Homie s'occupe de la collecte des loyers et a même une police d'assurance pour payer le propriétaire en cas de défaut de paiement et récupérer la propriété en quatre mois³⁹⁹. Le service de réponse aux clients de Homie est soutenu par des algorithmes prédictifs très précis, qui sont l'un des arguments de vente qui incitent les propriétaires à confier leurs biens à Homie. À cet égard, le premier répondant entre les clients et les services est essentiel pour maintenir une clientèle solide et engagée.

Applications de l'IA pour l'assistance virtuels

Les applications d'IA sous forme de ChatBots avec différents niveaux de complexité, sont maintenant une tendance mondiale pour répondre aux questions, commentaires et réclamations des clients, ainsi que pour en attirer des nouveaux. Selon un rapport d'Oracle sur les technologies émergentes, les principales entreprises investiront dans les interactions basées sur les chatbots, y compris les investissements dans la « voix intelligente » (intelligence voice). 51% des répondants ont déclaré que les avantages de l'investissement dans "*l'Intelligent Voice*" sont « une résolution plus rapide des problèmes des clients », tandis que 50% ont répondu que son principal avantage est « une efficacité opérationnelle accrue »⁴⁰⁰. Juniper Research prévoit que d'ici 2022, les entreprises économiseront 8 milliards de dollars US grâce à l'utilisation des ChatBots pour les activités liées aux services à la clientèle⁴⁰¹. De même que le rapport d'Oracle, Gartner estime que

³⁹⁶ Il existe une liste assez complète des néo-Banques sur le marché mexicain ici : <https://www.legalparadox.com/post/neo-banks-who-will-control-the-mexican-market> (consulté le 31 octobre 2020).

³⁹⁷ <https://www.eleconomista.com.mx/finanzaspersonales/En-el-2019-70-de-las-nuevas-familias-compraria-una-casa-20181216-0053.html> (consulté le 31 octobre 2020).

³⁹⁸ <https://homie.mx/h/> (consulté le 31 octobre 2020).

³⁹⁹ <https://inmobiliare.com/homie-la-plataforma-que-facilita-la-renta-de-departamento> (consulté le 31 octobre 2020).

⁴⁰⁰ <https://www.oracle.com> (consulté le 31 octobre 2020).

⁴⁰¹ <https://www.juniperresearch.com/new-trending/analytsexpress/july-2017/chatbot-conversations-to-deliver-8bn-cost-saving> (consulté le 31 octobre 2020).

d'ici 2022, 70 % des employés interagiront quotidiennement avec des plates-formes de conversation ⁴⁰².

Au Mexique, il y a un petit nombre d'entreprises basées sur l'IA qui offrent des services ChatBot avec différents degrés d'autonomie et de gamme d'interaction. Selon M. Nanalyze, YaloChat⁴⁰³ a rejoint sept autres sociétés mondiales qui facilitent l'adoption de la technologie ChatBot, c'est-à-dire la gamme de sociétés qui peuvent « nous fournir un ChatBot rapidement et avec un minimum d'efforts »⁴⁰⁴. YaloChat travaille avec les entreprises pour comprendre les relations avec leurs clients afin de concevoir un ChatBot sur mesure et améliorer la qualité du service. YaloChat envoie et gère les notifications aux clients, y compris les démonstrations de produits. L'intelligence artificielle de YaloChat automatise les questions fréquemment posées, avec une estimation de la prévision de 90 % des conversations potentielles entre le client et le service. Enfin, YaloChat détaille les profils des clients pour offrir une expérience interactive distinctive afin d'accroître l'engagement du service et du produit. YaloChat a produit des Bots orientés sur l'IA pour de nombreuses grandes entreprises opérant au Mexique, notamment Amazon, Pepsi, Volkswagen, Aeromexico et Walmart. Gus Chat⁴⁰⁵, une autre entreprise mexicaine spécialisée dans la conception et le développement de robots AI pour l'automatisation du service à la clientèle, est également à l'origine de YaloChat, avec une spécialisation dans le commerce électronique, les technologies financières et les assurances⁴⁰⁶. Gus Chat conçoit un algorithme différent pour chaque tâche différente assignée à un Bot : génération de pistes de vente, Bots transactionnels, Bots de service de réponse aux clients, Bots B2B et Bots de marketing et de publicité. Cette grande variété d'algorithmes fait de Gus Chat une entreprise innovante dans le domaine du traitement du langage naturel dans la région.

L'utilisation d'algorithmes de langage naturel pour soutenir les clients et les services sous la forme de robots interactifs est l'une des nombreuses utilisations des technologies d'IA. La reconnaissance de caractères et la reconnaissance vocale de texte sont également deux technologies clés qui améliorent le processus commercial au Mexique.

Nowports⁴⁰⁷ est une plate-forme qui met en relation les clients, les fournisseurs, les services douaniers, les ports et les transporteurs. Le processus utilise une *Blockchain* pour rendre chaque transaction sûre et transparente pour chaque participant impliqué. Nowports traite automatiquement chaque demande d'expédition par courrier électronique, quel que soit le format ou la documentation. Le système d'IA traite la demande, puis établit un budget avec un prix fixe sans variation. L'agent IA connecte tous les autres points terminaux pour compléter la transaction. Une fois la transaction approuvée par le client, il passe la commande pour expédier le produit. Le client peut suivre en temps réel le cheminement de la marchandise jusqu'au point de livraison. Grâce à sa technologie d'intelligence artificielle, Nowports peut traiter les demandes logistiques jusqu'à 70 % plus rapidement que les autres entreprises sur le marché concurrent.

Lors de l'utilisation des AI-Bots, l'association de l'informatique cognitive, du *cloud computing* et d'une architecture d'IA d'entraînement général constituent une puissante combinaison. Nearshore Delivery Solutions (NDS)⁴⁰⁸ est une entreprise mexicaine qui travaille avec IBM pour former et créer des bots

⁴⁰² <https://www.gartner.com/smarterwithgartner/chatbots-will-appeal-to-modern-workers/> (consulté le 31 octobre 2020).

⁴⁰³ <https://www.yalochat.com> (consulté le 31 octobre 2020).

⁴⁰⁴ <https://www.nanalyze.com/2017/07/17-chatbot-platforms-chatbots-easy/> (consulté le 31 octobre 2020).

⁴⁰⁵ N.B. Le nom de la société est peut-être en hommage du système GUS qui est le pionnier de l'architecture des systèmes de dialogue. Voir page 11, chapitre 24 du *Traitement de la parole et du langage*. Daniel Jurafsky & James H. Martin à l'adresse <https://web.stanford.edu/~jurafsky/slp3/24.pdf> (consulté le 31 octobre 2020).

⁴⁰⁶ <https://gus.chat> (consulté le 31 octobre 2020).

⁴⁰⁷ <https://nowports.com> (consulté le 31 octobre 2020).

⁴⁰⁸ https://nearshoremx.com/cognitive_computing (consulté le 31 octobre 2020).

d'informatique cognitif. NDS pourrait créer un assistant virtuel pour une grande banque mexicaine en seulement 12 semaines. NDS a formé l'assistant Watson d'IBM à l'analyse des e-mails, des transcriptions d'appels téléphoniques et des messages de chat échangés pendant un an entre les clients et la zone de support de la banque. Grâce aux connaissances acquises lors de cette analyse approfondie et détaillée, l'assistant IBM Watson a pu répondre à près de 1 000 questions. Le Bot a également détecté quand une question nécessitait une intervention humaine pour rediriger l'appel vers un expert humain. En outre, la NDS a utilisé l'analyseur de tonalité Watson pour détecter l'anxiété ou la colère d'un client en vue d'une intervention humaine rapide afin de prévenir toute nouvelle frustration du client. 75% des appels des clients sont maintenant traités par le robot cognitif qui résout les problèmes en deux minutes en moyenne, soit 80% plus rapidement qu'un assistant de centre d'appel typique⁴⁰⁹.

L'utilisation d'une architecture générale d'IA comme celle de Watson fournit un cadre solide pour innover en matière d'informatique cognitif. En particulier, l'analyse du ton et les réponses émotionnelles sont des informations cruciales pour produire des réponses précises aux clients et aux consommateurs.

Deux entreprises mexicaines développent et appliquent des technologies d'IA pour gérer les messages. Metric est une entreprise mexicaine qui a commencé à créer du contenu automatisé pour les réseaux sociaux. Elle a ensuite développé le programme Bots pour détecter les émotions et les attitudes, et donc, pour répondre avec un contenu spécifique afin d'influencer ces émotions. À l'heure actuelle, leur activité consiste à modéliser et à prévoir le comportement des réseaux sociaux. Leurs produits génèrent des prévisions estimatives de l'impact des campagnes, des positions des parties prenantes et des tendances avant qu'elles ne se produisent⁴¹⁰. Ils peuvent évaluer des scénarios de risque (et d'opportunité) pour la valeur des marques, y compris la détection des Bots, des trolls et des fausses informations. De la même manière, mais orienté vers les annonceurs et les agences de publicité, l'Adext mexicain⁴¹¹ exploite la puissance de l'IA pour les campagnes de réseaux sociaux. Adext a créé des modèles d'apprentissage automatique qui effectuent des simulations afin de déterminer le moment opportun, le public cible et l'espace numérique (placement) pour les publicités dans les réseaux sociaux. La simulation détermine l'impact marketing optimal compte tenu des contraintes budgétaires à partir de milliers de scénarios potentiels pour produire une granularité incroyable pour les publics. Adext assure fournir une impulsion dans les campagnes. Cette société propose également une démonstration pour voir en temps réel le nombre de conversions qu'une campagne engendre avec l'utilisation de leurs modèles⁴¹².

Les Bots (comme leurs homologues physiques - les robots) fonctionnent en utilisant des modèles et des informations pour accomplir des tâches spécifiques. La notion d'espace numérique rend les Bots et leur présence moins évidente que celle des robots humanoïdes. Cependant, le domaine de l'IA au Mexique a connu des progrès dans la conception et le développement de Robots en tant que service (la plupart d'entre eux étant fabriqués à l'étranger)⁴¹³.

⁴⁰⁹ <https://www.ibm.com/downloads/cas/JOEXYB08> (consulté le 31 octobre 2020).

⁴¹⁰ <https://buy.metricser.com/revelio/> (consulté le 31 octobre 2020).

⁴¹¹ <https://www.adext.ai> (consulté le 31 octobre 2020).

⁴¹² <https://www.latamdigitalmarketing.com/blog/software-campanas-digitales/> (consulté le 31 octobre 2020).

⁴¹³ Il existe de nombreux cas de robots déjà déployés comme service, voir par exemple <https://integritas.mx/robots-de-servicio/>, <http://robotsmexico.mx> ainsi que le document de recherche de Sucar pour une analyse de la maturité du domaine au Mexique dans <https://ccc.inaoep.mx/~emorales/Papers/2009/eduardo.pdf> (tous les sites ont été consultés le 31 octobre 2020).

Applications de l'IA pour les services de santé

La détection et l'isolement rapide des groupes infectieux est l'un des nombreux aspects de la lutte contre la pandémie de la COVID-19. Un autre défi consiste à prendre en charge les patients infectés par le virus pendant leur hospitalisation. Il existe aujourd'hui une grande quantité d'informations décrivant les risques et la contagion des professionnels de santé qui soignent les patients infectés. Une entreprise mexicaine, GESEDIG, propose trois modèles de *Robots as Service* qui peuvent aider le personnel de santé à soigner les patients atteints de la COVID-19⁴¹⁴. Les robots humanoïdes peuvent interroger les patients lorsqu'ils arrivent au centre de santé. Ils peuvent faire des rondes dans les zones d'entrée des hôpitaux et les salles d'attente en montrant des vidéos d'information et en fournissant des mesures préventives de soins de santé. Ils peuvent également orienter les patients vers différentes zones de l'hôpital à l'aide de capteurs de mouvement, ainsi que livrer de la nourriture et administrer des médicaments. Les robots humanoïdes constituent un moyen de prévenir la propagation du virus parmi les professionnels de la santé. L'IA du robot a été conçue et programmée par des ingénieurs mexicains. Le noyau de l'IA comprend un module de traitement du langage naturel, la reconnaissance des visages, des lecteurs de codes-barres, des processus de gestion des relations avec les clients et des fonctions "*Click to Call*"⁴¹⁵.

Nous avons vu les nombreuses utilisations des applications de l'IA au Mexique dans différents domaines. De l'intérêt public au secteur privé, on peut voir que les applications de l'IA ont un impact sur les industries, les services et les tâches gouvernementales. Le cycle de vie d'une application d'IA implique des ressources financières, technologiques et humaines. Dans une économie émergente telle que l'économie mexicaine, les variables qui auront un impact négatif sur l'essor de la recherche et des applications de l'IA sont vastes, mais pas infinies. De même, il existe un degré inconnu d'interaction entre la mentalité individuelle, la conscience collective, l'infrastructure institutionnelle, l'élaboration de politiques et le cadre juridique pour susciter et encourager le développement de l'IA. Un compte rendu des politiques publiques, du cadre réglementaire et des initiatives de coordination privées peut donner une idée partielle de l'élan de l'IA au Mexique.

III. Comptabilisation de la (dé)réglementation de l'IA au Mexique

Au Mexique, il n'existe pas de cadre juridique et législatif qui réglemente directement l'IA. L'absence d'un outil législatif pour l'IA se constate à la fois au niveau fédéral et au niveau de l'État local (c'est-à-dire l'organe législatif local). Il existe cependant une législation et un ensemble de règles de différentes hiérarchies et sources dans les domaines spécifiques où des applications de l'IA apparaissent. Il est utile de reconnaître, dans une perspective ciblée, la réglementation indirecte des systèmes d'IA au Mexique. L'absence d'un instrument législatif pour réglementer les systèmes d'IA est compensée par des politiques publiques qui ont été mises en œuvre au moins depuis 2019, bien qu'il existe des politiques publiques mises en œuvre dans la transformation numérique et l'inclusion numérique qui remontent au moins à une décennie. Le Mexique est une république fédérale, ce qui signifie qu'il y a deux décideurs politiques en matière d'IA. Nous constatons que le gouvernement

⁴¹⁴ N.B. Il existe des procédures chirurgicales utilisant des robots ou des instruments robotisés au Mexique depuis 1993 selon Miller FHS, *Cirugía robótica en México. Los sistemas inteligentes, perspectivas actuales y a futuro en el ámbito mundial*, *Revista Mexicana de Cirugía Endoscópica*, 2003, 4(1) : 45-50. Avec les progrès de la réalité augmentée attachée aux robots dans la salle d'opération, il y a certainement un nouveau champ d'applications pour l'IA en médecine. Au Mexique, cependant, ce champ n'est pas très étendu. La première chirurgie en réalité augmentée assistée par l'IA a eu lieu à l'hôpital La Conchita dans l'État de Nuevo Leon en utilisant la plateforme *BedsideXR* en février 2020, voir <https://www.christusmuguerza.com.mx/sala-de-prensa/es-hospital-conchita-sede-de-la-primera-cirugia-con-realidad-aumentada-en-mexico/> (consulté le 31 octobre 2020). Nous avons laissé de côté ce domaine des applications de l'IA au Mexique car nous les considérons comme étant à leurs débuts par rapport aux robots comme service de soins de santé.

⁴¹⁵ <https://gesedig.com> (consulté le 31 octobre 2020).

fédéral a fait des efforts pour encourager l'innovation et le développement des technologies d'IA. De plus, certains gouvernements locaux ont également mis en place différentes actions gouvernementales en faveur de l'adoption des technologies d'IA. En parallèle des deux niveaux de réglementation et d'élaboration des politiques, fédéral et local, nous constatons qu'il existe des agences d'État dotées de pouvoirs constitutionnels qui formulent également des réglementations relatives aux systèmes⁴¹⁶ d'IA. Au Mexique, il existe des universités publiques et privées, ainsi que des centres de recherche publics et privés, qui ont, de manière directe ou indirecte, un intérêt dans la politique publique menée par les gouvernements fédéral et local. Nous devons avoir un intérêt particulier sur une agence du gouvernement fédéral en charge de la technologie et du développement pour apprécier les mécanismes spécifiques de la façon dont la politique publique en matière d'IA a été menée. Si nous examinons les cas que nous avons présentés dans la première partie de ce chapitre, nous pouvons constater que la plupart des initiatives datent d'avant l'entrée en vigueur d'une politique publique fédérale officielle au Mexique. Dans le même temps, les applications d'IA motivées par l'intérêt privé que nous avons analysées ne sont pas le résultat direct d'une politique spécifique pour le développement des initiatives d'IA au Mexique. La plupart des applications de l'IA dans le secteur privé sont le résultat d'une combinaison de forces du marché et d'une volonté individuelle d'innovation. L'absence d'un cadre juridique solide et d'un environnement réglementaire favorable a peut-être produit l'élan de l'IA au Mexique que nous avons constaté ces dernières années. En outre, des collectifs organisés dont les membres proviennent de divers domaines (tels que l'industrie, l'université, l'entrepreneuriat et l'activisme) proposent un programme national pour le développement de l'IA au Mexique. L'exercice démocratique et spontané du collectif en faveur de l'IA n'a pas de précédent dans le pays. Nous présenterons et analyserons le cadre juridique qui est indirectement lié aux applications de l'IA et à la politique publique tant au niveau fédéral que local qui a été proposé et mis en œuvre au Mexique (A) avant d'analyser le mouvement collectif en faveur de l'IA qui met en avant un programme de thèmes à réglementer par l'État mexicain (B).

i. Cadre juridique et politiques publiques concernant l'IA

Pour présenter le cadre réglementaire qui touche indirectement les applications de l'IA, nous distinguerons les réglementations nationales et les réglementations locales. Au Mexique, certains domaines juridiques peuvent uniquement et exclusivement être réglementés par le Congrès fédéral, ce qui exclut le pouvoir des législateurs locaux. Un domaine juridique qui relève de la compétence exclusive du Congrès fédéral est appelé réglementation nationale. Il existe d'autres domaines juridiques dans lesquels les législateurs locaux et fédéraux peuvent produire des normes applicables en même temps. Dans le même ordre d'idées, nous distinguerons la politique du gouvernement fédéral et la politique du gouvernement local en matière d'IA.

Protection des données personnelles

Les droits en matière de protection des données personnelles sont traités au niveau constitutionnel. La Constitution mexicaine établit une agence nationale, dotée d'une pleine autonomie, pour superviser les droits relatifs à la protection des données⁴¹⁷. Mais, en même temps, elle reproduit la conception d'un contrôleur autonome des droits de protection des données dans chaque État local

⁴¹⁶ Un exemple est l'Institut fédéral mexicain des télécommunications qui a un programme explicite pour l'utilisation des applications de l'IA sur le secteur des télécommunications mexicain. Voir <http://www.ift.org.mx/sites/default/files/contenidogeneral/transparencia/1vision19-23.pdf> (consulté le 31 octobre 2020).

⁴¹⁷ Article 4 du Statut général pour la transparence et l'accès à l'information publique sur http://www.diputados.gob.mx/LeyesBiblio/pdf/LGTAIP_130820.pdf (consulté le 31 octobre 2020).

de la Fédération mexicaine. L'Agence nationale de contrôle de la protection des droits relatifs aux données à caractère personnel (appelée INAI en espagnol⁴¹⁸) dispose de pouvoirs suffisants pour contrôler, imposer des sanctions et même interdire des opérations en cas de violation des principes et des normes relatifs aux données à caractère personnel⁴¹⁹. Il est intéressant de noter que le Mexique est partie à la Convention 108 ainsi qu'au Protocole additionnel à la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, concernant les autorités de contrôle et les flux transfrontaliers de données. Pour compléter les instruments juridiques nationaux et internationaux, l'INAI fait partie du Réseau ibéro-américain pour la protection des données⁴²⁰ (NDP). Le NDP a publié des lignes directrices spécifiques pour le respect des principes et des droits qui régissent la protection des données personnelles dans les projets⁴²¹ d'intelligence artificielle ainsi que des recommandations générales pour le traitement des données personnelles dans l'intelligence artificielle⁴²². Ces deux instruments sont considérés comme des instruments juridiques non contraignants en raison de leur force institutionnelle dans le système juridique mexicain, mais ils fournissent néanmoins un cadre directeur très détaillé de principes et de normes pour l'INAI.

Toute personne peut saisir l'INAI d'une plainte contre tout organisme gouvernemental, toute entité privée ou tout individu qui aurait violé les obligations constitutionnelles et légales en matière de protection des données personnelles. Il n'y a pas eu une seule affaire portée devant l'autorité nationale sur les droits des données personnelles qui implique la collecte, le traitement ou l'utilisation de données dans une demande d'IA au Mexique.

Réglementation financière

Contrairement à la protection des données, la réglementation financière est exclusive à la réglementation fédérale, à l'échelle nationale, au Mexique. Si une entreprise ou une startup FinTech au Mexique a accès à de l'argent du grand public, ou si elle fournit des services bancaires, elle est tenue de se conformer au statut fédéral de FinTech⁴²³ ainsi qu'à tous les règlements détaillés émis par l'Autorité fédérale de réglementation financière⁴²⁴ et ceux applicables de la Banque centrale mexicaine⁴²⁵. Du point de vue d'un système normatif, les applications de l'IA ne sont pas le thème principal du système normatif régissant FinTech. Il n'y a pas une seule référence qui indique une quelconque norme pour le développement ou l'utilisation de systèmes d'IA lors de l'acquisition de services financiers par le biais de la technologie. L'évaluation des risques financiers fondée sur des preuves, les différentes méthodes de profilage et les informations que toute entité au sein du marché des services financiers est tenue de fournir à l'autorité de régulation sont toutes les mêmes sur le plan de la justification de cette dernière. Il est justifié de chercher à obtenir une assurance lorsqu'on gère de l'argent et une gestion adéquate des risques lorsqu'on prête de l'argent, quels que soient les moyens techniques et technologiques utilisés pour décider de la clientèle. Il est largement reconnu au sein de la communauté FinTech au Mexique que la réglementation fédérale favorise les

⁴¹⁸ <https://home.inai.org.mx> (consulté le 31 octobre 2020).

⁴¹⁹ Article 6 de la Constitution mexicaine sur http://www.diputados.gob.mx/LeyesBiblio/pdf_mov/Constitucion_Politica.pdf (consulté le 31 octobre 2020).

⁴²⁰ <https://www.redipd.org/es/la-red/historia-de-la-red-iberoamericana-de-proteccion-de-datos-ripd> (consulté le 31 octobre 2020).

⁴²¹ <http://inicio.inai.org.mx/nuevo/SPECIFICGUIDELINESARTIFICIALINTELLIGENCEPROJECTS2019.pdf> (consulté le 31 octobre 2020).

⁴²² <http://inicio.inai.org.mx/nuevo/GeneralRecommendationsfortheProcessingofPersonalDatainArtificialIntelligence.pdf> (consulté le 31 octobre 2020).

⁴²³ http://www.diputados.gob.mx/LeyesBiblio/pdf/LRITF_090318.pdf (consulté le 31 octobre 2020).

⁴²⁴ <https://www.cnbv.gob.mx/SECTORES-SUPERVISADOS/Fintech/Paginas/NORMATIVIDAD-FINTECH.aspx> (consulté le 31 octobre 2020).

⁴²⁵ <https://www.banxico.org.mx/marco-normativo/normativa-agrupada-por-sujeto.html> (consulté le 31 octobre 2020).

affaires et l'innovation. À cet égard, le statut FinTech adopte indirectement une approche légère de l'innovation technologique sur le marché des services financiers. Grâce à une politique législative qui n'est pas trop intrusive en matière d'innovation, on s'attend à ce que les systèmes et applications d'IA sur le marché FinTech prospèrent également.

Bien que la réglementation des services financiers ne soit pas spécifiquement axée sur les applications de l'IA, les gouvernements fédéral et locaux ont mis en place une politique publique active pour promouvoir et soutenir les applications de l'IA au Mexique.

Le 21 mars 2018, une étude commandée par l'ambassade du Royaume-Uni au Mexique, avec le soutien du bureau du président mexicain, a été présentée à Oxford Insights. L'étude était un projet de stratégie nationale d'IA pour le gouvernement mexicain ⁴²⁶. À la fin de la présentation, le gouvernement mexicain a dévoilé la stratégie 2018 pour l'intelligence artificielle au Mexique (stratégie MX-AI2018). Cette stratégie comprend cinq actions de politique publique ⁴²⁷: 1) Développer un cadre de gouvernance adéquat pour susciter un dialogue ouvert entre les différents acteurs. Le gouvernement créerait une commission de ministères pour le développement d'un gouvernement électronique. 2) Identifier les besoins de l'industrie et les meilleures pratiques gouvernementales en matière d'IA. 3) Soutenir un effort international sur l'IA en mettant l'accent sur le rôle du Mexique auprès de l'OCDE et du G-7. 4) Ouvrir la consultation publique concernant les recommandations de l'étude d'Oxford Insights. 5) Travailler avec des experts et des citoyens au sein du sous-comité pour l'intelligence artificielle et l'apprentissage approfondi créé au ministère de la fonction publique⁴²⁸. À l'époque, le Mexique était ainsi devenu l'un des dix premiers pays au monde à lancer des actions concrètes de politique publique en faveur du développement, de l'adoption et de l'utilisation de l'intelligence artificielle.

En 2019, le Mexique a adhéré aux Principes de l'OCDE sur l'intelligence artificielle « pour promouvoir une intelligence artificielle innovante et digne de confiance, et qui respecte les droits de l'homme et les valeurs démocratiques »⁴²⁹. En adhérant aux Principes de l'OCDE sur l'intelligence artificielle, le Mexique s'est engagé à suivre les cinq principes complémentaires fondés sur des valeurs pour une gestion responsable d'une intelligence artificielle digne de confiance ainsi que les cinq recommandations de l'OCDE aux gouvernements. Le gouvernement fédéral a alors créé une commission de ministères pour le développement d'un gouvernement électronique afin de coordonner les différents efforts dans le déploiement de la stratégie MX-AI2018.

Un résultat important de la politique publique nationale a été une initiative organisée par le Conseil national pour la science et la technologie (CONACYT en espagnol). Le CONACYT est chargé de superviser et de financer la recherche publique par l'intermédiaire des centres nationaux de recherche publique (CNRP) dans tout le pays. Chaque centre a un programme de recherche indépendant et autonome, en partie parce qu'il est créé dans des universités publiques et ensuite accrédité par le CONACYT en tant que CNR. Sous la direction de CONACYT, huit CNR ont uni leurs forces pour se conformer à l'initiative sur l'intelligence artificielle. Cette initiative permet la recherche conjointe dans des domaines interdisciplinaires entre des chercheurs et des étudiants diplômés de différentes origines et disciplines. L'initiative a concentré ses efforts sur la modélisation des phénomènes naturels et sociaux dans les domaines de la médecine, de la sécurité publique, de la mobilité humaine et des transports, de la prévention des catastrophes naturelles et de la formation

⁴²⁶ L'étude a finalement été publiée en juin 2018. Voir <https://www.oxfordinsights.com/mexico> (consulté le 31 octobre 2020).

⁴²⁷ <https://datos.gob.mx/blog/estrategia-de-inteligencia-artificial-mx-2018> (consulté le 31 octobre 2020).

⁴²⁸ <https://www.gob.mx/cidge/articulos/crea-sfp-subcomision-de-inteligencia-artificial-y-deep-learning-de-la-cidge-161421?tab=> (consulté le 31 octobre 2020).

⁴²⁹ <http://www.oecd.org/going-digital/ai/principles/> (consulté le 31 octobre 2020).

des ressources humaines en IA⁴³⁰. L'initiative a depuis lors produit un grand nombre de recherches, exposées dans différentes publications sur des articles dans des revues prestigieuses dans de multiples disciplines. L'initiative a également créé des groupes de recherche entre universitaires et étudiants de troisième cycle de différents CNR et, à ce jour, sept au total, mais avec une révision permanente des thèmes pour rester en phase avec la recherche de pointe⁴³¹. L'alliance s'est avérée être la plus riche en matière de ressources humaines pour le travail multidisciplinaire sur les applications de l'IA au Mexique

Parallèlement à la politique publique du gouvernement fédéral, nous constatons que les autorités locales ont également examiné les applications de l'IA de manière proactive. Dans un contexte plus large, il faut toujours garder à l'esprit le vaste territoire que possède le Mexique, ainsi que la diversité entre les différents États locaux. À cet égard, certains États sont plus développés que d'autres en termes de ressources humaines, d'industrie et de revenu par habitant. Il n'est donc pas étrange que certaines des initiatives en matière de politique publique prises soit par le gouvernement local, soit en collaboration avec le secteur privé, proviennent de régions spécifiques du pays.

Les États de Puebla⁴³² et de Querétaro⁴³³ ont lancé des programmes locaux pour transformer leurs capitales en « villes intelligentes » grâce à l'utilisation d'applications de l'IA. Les gouvernements du Yucatan⁴³⁴ et de Guanajuato⁴³⁵ ont également lancé des plans de politique publique agressifs pour intégrer l'industrie technologique, ils ont créé des alliances avec des entreprises technologiques ou des centres de recherche pertinents, et ils ont soutenu leurs universités publiques locales pour qu'elles jouent un rôle majeur dans leur politique publique. Dans la capitale du Yucatan, Mérida, un projet de création d'un centre de recherche sur l'IA est en cours. Veracruz, un État du sud, dispose d'un centre de recherche sur l'IA qui fonctionne depuis 1994⁴³⁶. Les autres États locaux disposant d'un centre de recherche sur l'IA sont Chihuahua⁴³⁷, Nuevo León⁴³⁸ et Jalisco.

Jalisco est peut-être le plus important des états mentionnés ci-dessus, le Mexique. Jalisco a été nommé la Silicon Valley mexicaine⁴³⁹ en raison du nombre de start-ups, de grands acteurs technologiques qui ont installé leurs laboratoires de recherche et de développement dans l'État, et de la forte densité de professionnels des compétences technologiques⁴⁴⁰. L'année dernière, le gouvernement local de Jalisco a conclu un accord de collaboration avec Tec de Monterrey (une des plus prestigieuses universités technologiques privées du Mexique) et la Banque interaméricaine de développement pour créer un centre d'intelligence artificielle, le premier au Mexique. Ce centre implique un investissement de 900 000 USD de la part de Tec de Monterrey, la participation d'Intel, IBM et Amdocs, Sparkcognition, Tata Consultancy Services et Wizeline, l'Université de Berkeley en Californie, l'Institut de technologie de Pékin et l'Institut de recherche en informatique et en automatisation de France. L'une des nouveautés de ce consortium est que l'accent sera mis sur la résolution de problèmes complexes de politique publique à orientation sociale à Jalisco sur la santé, l'éducation et la sécurité publique. Le centre travaillera à la production d'applications d'IA pour le

⁴³⁰ <https://www.consorcioia.mx/nosotros> (consulté le 31 octobre 2020).

⁴³¹ <https://www.consorcioia.mx/grupos-investigacion> (consulté le 31 octobre 2020).

⁴³² <https://www.eluniversal.com.mx/estados/proyectan-puebla-como-ciudad-de-inteligencia-artificial> (consulté le 31 octobre 2020).

⁴³³ <https://amqueretaro.com/queretaro/2020/10/22/convertiran-a-queretaro-en-ciudad-inteligente/> (consulté le 31 octobre 2020).

⁴³⁴ <https://yucatanahora.mx/yucatan-busca-desarrollar-proyectos-de-inteligencia-artificial/> (consulté le 31 octobre 2020).

⁴³⁵ <https://www.eleconomista.com.mx/estados/Guanajuato-epicentro-de-la-cuarta-Revolucion-Industrial-20201029-0104.html> (consulté le 31 octobre 2020).

⁴³⁶ <https://www.uv.mx/ciia/> (consulté le 31 octobre 2020).

⁴³⁷ <https://www.facebook.com/InteligenciaArtificial.Center/> (consulté le 31 octobre 2020).

⁴³⁸ <https://www.eleconomista.com.mx/estados/Centro-de-Inteligencia-Artificial-en-NL-alista-inicio-de-operaciones-20180820-0020.html> (consulté le 31 octobre 2020).

⁴³⁹ https://english.elpais.com/elpais/2017/03/13/inenglish/1489403756_441981.html (consulté le 31 octobre 2020).

⁴⁴⁰ <https://inmobiliare.com/por-que-guadalajara-es-el-silicon-valley-mexicano/> (consulté le 31 octobre 2020).

traitement et la détection de la rétinopathie diabétique, pour prévenir le décrochage scolaire et pour éviter la délinquance juvénile⁴⁴¹.

Le secteur privé a été très actif en termes de création d'applications de la GA et d'encouragement des initiatives dans tout le pays. L'exercice le plus marquant en faveur de la configuration d'une feuille de route complète pour un programme national sur l'intelligence artificielle est peut-être l'IA2030 collective.

ii. Un cadre collectif concernant la réglementation de l'IA

IA2030 est un collectif d'organisations et d'acteurs clés d'organisations de la société civile, du monde universitaire, du secteur privé, de l'industrie en général, de consultants indépendants, de fonctionnaires et de centres de recherche publics⁴⁴². Le collectif a été dirigé par une start-up appelée C Minds. Deux produits sont le résultat d'un effort de coordination très intensif : une enquête nationale⁴⁴³ et un programme national pour l'intelligence artificielle au Mexique⁴⁴⁴.

L'enquête a été élaborée par près de cinquante organisations et professionnels indépendants dans le cadre d'un engagement volontaire. L'enquête a été disponible en ligne pendant un mois du 15 août au 18 septembre 2018. Bien que les résultats de l'enquête ne puissent pas être généralisés (c'est-à-dire qu'ils manquent de validité statistique), elle sert de thermomètre social autour des parties intéressées dans le domaine de l'IA. Sur les 1 585 personnes interrogées, près de 90 % étaient originaires de Mexico, de l'État de Mexico et de Jalisco. Ces trois sites concentrent une grande partie des efforts technologiques et des ressources humaines liés à l'IA au Mexique. 43 % des personnes interrogées travaillent pour le gouvernement, 31 % pour le secteur privé et 18 % pour le monde universitaire. Parmi les principales conclusions, l'enquête montre que 80 % des personnes interrogées pensent que l'IA aura un effet positif sur leur vie. Cependant, 53 % pensent que le chômage augmentera avec l'IA et 45 % sont préoccupés par le fait que leur vie privée et leurs données personnelles seront compromises par les applications de l'IA. Près de 45 % des personnes interrogées ont déclaré que l'utilisation généralisée de l'IA sans réglementation adéquate posait des problèmes d'éthique et d'inégalité. Sur le plan des politiques publiques, les opinions se sont rejointes sur le rôle actif de l'État mexicain pour encourager la recherche, le développement et l'adoption des technologies d'IA, pour améliorer les services publics et pour générer davantage de ressources humaines spécialisées dans les applications d'IA.

L'Agenda national sur l'IA est le résultat d'un effort de collaboration de quatre cents personnes de différents secteurs et milieux, réparties en six groupes de travail selon l'expertise de chaque participant : Données, infrastructure numérique et cybersécurité, éthique, gouvernance, gouvernement et services publics, recherche et développement, éducation, capacités et compétences, et sensibilisation des immigrants mexicains. L'exercice qui en résulte est la présentation de problèmes spécifiques à chaque thème du groupe de travail, la proposition de lignes d'action et d'indicateurs de points clés pour évaluer les lignes d'action proposées. Bien que le contenu spécifique nécessiterait une analyse plus détaillée que celle fournie dans ce chapitre, certains thèmes communs émergeant des groupes de travail méritent d'être mentionnés. D'une part, l'idée

⁴⁴¹ <https://tec.mx/es/noticias/guadalajara/investigacion/asi-sera-el-innovador-hub-de-inteligencia-artificial-del-tec> (consulté le 31 octobre 2020).

⁴⁴² <https://www.ia2030.mx/> (consulté le 31 octobre 2020).

⁴⁴³ https://36dc704c-0d61-4da0-87fa-917581cbce16.filesusr.com/ugd/7be025_9e91bfff6ea647a0a663630ea716aa8f.pdf (consulté le 31 octobre 2020).

⁴⁴⁴ https://36dc704c-0d61-4da0-87fa-917581cbce16.filesusr.com/ugd/7be025_6f45f669e2fa4910b32671a001074987.pdf (consulté le 31 octobre 2020).

d'objectifs stratégiques dans chaque thème. Ensuite, l'utilisation de principes généraux. Enfin, l'analyse intersectionnelle entre les thèmes tout au long du document. D'autre part, le rapport présente des indicateurs de points clés pour chaque ligne d'action suggérée avec l'indication de la personne responsable de la mise en œuvre de la ligne d'action (c'est-à-dire le législateur, une agence de régulation ou la société civile). De ce point de vue, le rapport n'est pas seulement une feuille de route pour la réglementation de l'IA au Mexique, mais un GPS de la manière dont les cadres réglementaires possibles s'entrecroisent, tant en termes de contenu, de portée que de normes.

IV. Conclusions

Regarder une image en mouvement n'est pas la même chose que regarder un film. Les applications de l'IA au Mexique sont une image en mouvement, une composition d'acteurs et d'initiatives, d'entreprises et de technologies en phase de développement rapide. Même si la disposition des projets dans ce chapitre peut sembler avoir un ordre séquentiel, il existe en réalité une interaction complexe entre l'écosystème de l'IA au Mexique qu'il est difficile de systématiser. Cette complexité provient d'une vision fragmentaire, fruit d'initiatives épisodiques qui fleurissent avec un fondement spécifique sur un programme institutionnel. La plupart des cas d'utilisation présentés ici sont davantage le produit de mentalités et de talents individuels que le résultat d'un environnement de type incubateur pour l'IA. Cependant, le mouvement des initiatives d'IA présenté et retenue lors des recherches effectuées pour ce document laissent entendre que le marché mexicain des applications de l'IA est dynamique et prometteur. Dans un pays où les inégalités sont profondes et où l'État de droit et les droits de l'Homme sont encore en suspens, il est toujours possible d'investir dans les institutions, et davantage encore si elles favorisent le bien-être et l'inclusion grâce aux progrès technologiques. Il y a suffisamment de parties intéressées, tant dans le secteur privé que dans le secteur public - et dans le monde universitaire - pour déclencher une institutionnalisation fructueuse de l'IA au Mexique. Toutefois, l'inertie d'un marché largement non réglementé pourrait devenir le trait principal du cas mexicain.

Références

Une définition de l'IA : Principales capacités et disciplines scientifiques Groupe d'experts de haut niveau sur l'intelligence artificielle, Commission européenne, 8 avril 2019.

Article 4 du Statut général pour la transparence et l'accès à l'information publique sur http://www.diputados.gob.mx/LeyesBiblio/pdf/LGTAIP_130820.pdf (consulté le 31 octobre 2020).

Article 6 de la Constitution mexicaine sur http://www.diputados.gob.mx/LeyesBiblio/pdf_mov/Constitucion_Politica.pdf (consulté le 31 octobre 2020).

Duffin Erin, *Number of Mexican Immigrants in the United States 1850-2018*, <https://www.statista.com/statistics/673350/mexican-immigrants-in-the-united-states/> (consulté le 31 octobre 2020).

<http://inicio.inai.org.mx/nuevo/GeneralRecommendationsfortheProcessingofPersonalDatainArti%E F%AC%81cialIntelligence.pdf> (consulté le 31 octobre 2020).

<http://inicio.inai.org.mx/nuevo/SPECIFICGUIDELINESARTIFICIALINTELLIGENCEPROJECTS2019.pdf> (consulté le 31 octobre 2020).

[http://omawww.sat.gob.mx/gobmxtransparencia/Paginas/documentos/estudio_opiniones/Evasion en IVA Analisis de Redes.pdf](http://omawww.sat.gob.mx/gobmxtransparencia/Paginas/documentos/estudio_opiniones/Evasion_en_IVA_Analisis_de_Red.es.pdf) (consulté le 31 octobre 2020).

<http://smia.mx/komputersapiens/> (consulté le 31 octobre 2020).

http://www.cires.org.mx/sasmex_es.php (consulté le 31 octobre 2020).

http://www.diputados.gob.mx/LeyesBiblio/pdf/LRITF_090318.pdf (consulté le 31 octobre 2020).

<http://www.ebizlatam.com/seguridad-e-inclusion-soluciones-conekta/> (consulté le 31 octobre 2020).

<http://www.ift.org.mx/sites/default/files/contenidogeneral/transparencia/1vision19-23.pdf> (consulté le 31 octobre 2020).

<http://www.oecd.org/going-digital/ai/principles/> (consulté le 31 octobre 2020).

<https://adip.cdmx.gob.mx/proyectos> (consulté le 31 octobre 2020).

<https://amqueretaro.com/queretaro/2020/10/22/convertiran-a-queretaro-en-ciudad-inteligente/>

<https://bitso.com> (consulté le 31 octobre 2020).

<https://blog.grillo.io/the-grillo-journey-e60b2dcee224> (consulté le 31 octobre 2020).

<https://buy.metricser.com/revelio/> (consulté le 31 octobre 2020).

<https://conekta.com> (consulté le 31 octobre 2020).

<https://dapp.mx> (consulté le 31 octobre 2020).

<https://data.oecd.org/mexico.htm#profile-innovationandtechnology> (consulté le 31 octobre 2020).

<https://datos.cdmx.gob.mx/explore/dataset/carpetas-de-investigacion-pgi-cdmx/table/?q=FED%2FFECC%2FUNAI-CDMX%2F0000002%2F2020> (consulté le 31 octobre 2020).

<https://datos.gob.mx/blog/estrategia-de-inteligencia-artificial-mx-2018> (consulté le 31 octobre 2020).

https://earthquaketrack.com/p/mexico/recent?mag_filter=7 (consulté le 31 octobre 2020).

<https://elceo.com/tecnologia/cnbv-alista-autorizaciones-para-80-empresas-fintech-entre-octubre-y-noviembre/> (consulté le 31 octobre 2020).

https://english.elpais.com/elpais/2017/03/13/inenglish/1489403756_441981.html (October 31, 2020).

https://ensanut.insp.mx/encuestas/ensanut2018/doctos/informes/ensanut_2018_presentacion_res ultados.pdf (consulté le 31 octobre 2020).

<https://expansion.mx/tecnologia/2017/08/15/konfio-un-ejemplo-de-como-las-fintech-dejan-atras-a-los-bancos> (consulté le 31 octobre 2020).

<https://gesediq.com> (consulté le 31 octobre 2020).

<https://grillo.io/linux-foundation-hosting-openeew/> et <https://grillo.io/open-source/> (consulté le 31 octobre 2020).

<https://gus.chat> (consulté le 31 octobre 2020).

<https://home.inai.org.mx> (consulté le 31 octobre 2020).

<https://homie.mx/h/> (consulté le 31 octobre 2020).

<https://inmobiliare.com/homie-la-plataforma-que-facilita-la-renta-de-departamento> (consulté le 31 octobre 2020).

<https://inmobiliare.com/por-que-guadalajara-es-el-silicon-valley-mexicano/> (consulté le 31 octobre 2020).

<https://konfio.mx/> (consulté le 31 octobre 2020).

<https://kueski.com> (consulté le 31 octobre 2020).

<https://kueski.com/blog/tecnologia/machine-learning-aprendizaje-supervisado/> (consulté le 31 octobre 2020).

https://nearshoremx.com/cognitive_computing (consulté le 31 octobre 2020).

<https://news.microsoft.com/uploads/prod/sites/41/2018/11/IA-y-Crecimiento-MEXICO.pdf> (consulté le 31 octobre 2020).

<https://nowports.com> (consulté le 31 octobre 2020).

<https://panamericanworld.com/en/magazine/startups/meet-the-five-most-innovative-mexican-fintechs/> (consulté le 31 octobre 2020).

<https://publications.iadb.org/publications/english/document/FINTECH--Innovations-You-May-Not-Know-were-from-Latin-America-and-the-Caribbean.pdf> (consulté le 31 octobre 2020).

<https://skyalert.mx/> (consulté le 31 octobre 2020).

<https://tec.mx/es/noticias/guadalajara/investigacion/asi-sera-el-innovador-hub-de-inteligencia-artificial-del-tec> (consulté le 31 octobre 2020).

<https://www.adext.ai> (consulté le 31 octobre 2020).

<https://www.banxico.org.mx/marco-normativo/normativa-agrupada-por-sujeto.html> (consulté le 31 octobre 2020).

<https://www.cnbv.gob.mx/SECTORES-SUPERVISADOS/Fintech/Paginas/NORMATIVIDAD-FINTECH.aspx> (consulté le 31 octobre 2020).

<https://www.consorcioia.mx/grupos-investigacion> (consulté le 31 octobre 2020).

<https://www.eleconomista.com.mx/estados/Centro-de-Inteligencia-Artificial-en-NL-alista-inicio-de-operaciones-20180820-0020.html> (consulté le 31 octobre 2020).

<https://www.eleconomista.com.mx/estados/Guanajuato-epicentro-de-la-cuarta-Revolucion-Industrial-20201029-0104.html> (consulté le 31 octobre 2020).

<https://www.eleconomista.com.mx/finanzaspersonales/En-el-2019-70-de-las-nuevas-familias-compraria-una-casa-20181216-0053.html> (consulté le 31 octobre 2020).

<https://www.eleconomista.com.mx/sectorfinanciero/Tenemos-mapeadas-mas-de-500-fintech-en-Mexico-CNBV-20190526-0072.html> (consulté le 31 octobre 2020).

<https://www.eluniversal.com.mx/cartera/sat-va-por-mas-recaudacion-con-inteligencia-artificial> (consulté le 31 octobre 2020).

<https://www.eluniversal.com.mx/estados/proyectan-puebla-como-ciudad-de-inteligencia-artificial> (consulté le 31 octobre 2020).

<https://www.facebook.com/InteligenciaArtificial.Center/> (consulté le 31 octobre 2020).

<https://www.finnovista.com/radar/el-numero-de-startups-fintech-en-mexico-crecio-un-14-en-un-ano-hasta-las-441/> (consulté le 31 octobre 2020).

<https://www.gartner.com/smarterwithgartner/chatbots-will-appeal-to-modern-workers/> (consulté le 31 octobre 2020).

<https://www.gob.mx/cidge/articulos/crea-sfp-subcomision-de-inteligencia-artificial-y-deep-learning-de-la-cidge-161421?tab=> (consulté le 31 octobre 2020).

https://www.gob.mx/cms/uploads/attachment/file/520516/Comunicado_UIF_011.pdf (consulté le 31 octobre 2020).

<https://www.gob.mx/innovamx/articulos/inteligencia-artificial-131287> (consulté le 31 octobre 2020).

<https://www.gob.mx/misalud/> (consulté le 31 octobre 2020).

<https://www.gob.mx/shcp/prensa/comunicado-no-054-shcp-mediante-sat-y-pff-anuncia-acciones-en-contra-de-defraudacion-fiscal-a-traves-de-empresas-factoreras> (consulté le 31 octobre 2020).

<https://www.ibm.com/downloads/cas/JOEXYB08> (consulté le 31 octobre 2020).

<https://www.infobae.com/tecnologia/2020/04/01/asi-es-el-chatbot-mexicano-que-lucha-contra-la-desinformacion-sobre-coronavirus/> (consulté le 31 octobre 2020).

<https://www.innovatorsunder35.com/the-list/saiph-savage/> (consulté le 31 octobre 2020).

https://www.insp.mx/resources/images/stories/2019/Docs/190213_LaObesidadenMexico.pdf, p. 28 (consulté le 31 octobre 2020).

<https://www.juniperresearch.com/new-trending/analytixpress/july-2017/chatbot-conversations-to-deliver-8bn-cost-saving> (consulté le 31 octobre 2020).

<https://www.kubofinanciero.com/Kubo/Portal/productos/productos.xhtml> (consulté le 31 octobre 2020).

<https://www.latamdigitalmarketing.com/blog/software-campanas-digitales/> (consulté le 31 octobre 2020).

<https://www.legalparadox.com/post/neo-banks-who-will-control-the-mexican-market> (consulté le 31 octobre 2020).

<https://www.nanalyze.com/2017/07/7-chatbot-platforms-chatbots-easy/> (consulté le 31 octobre 2020).

<https://www.oracle.com/a/ocom/docs/dc/em/lpd100807811-impact-of-emerging-technology-on-cx-excellence.pdf?elqTrackId=d368a11a304041c8a7bf8e7a2f2a71e2&elqaid=82669&elqat=2> (consulté le 31 octobre 2020).

<https://www.oxfordinsights.com/mexico> consulté le 31 octobre 2020).

<https://www.redipd.org/es/la-red/historia-de-la-red-iberoamericana-de-proteccion-de-datos-ripd> (consulté le 31 octobre 2020).

https://www.researchgate.net/profile/Yasmin_Hernandez2/publication/340634786_Robotica_de_Servicio/links/5e96334a4585150839de623a/Robotica-de-Servicio.pdf (consulté le 31 octobre 2020).

<https://www.statista.com/study/45600/statista-report-fintech/> (consulté le 31 octobre 2020).

<https://www.uif.gob.mx/> (consulté le 31 octobre 2020).

<https://www.uv.mx/ciia/> (consulté le 31 octobre 2020).

<https://www.yalochat.com> (consulté le 31 octobre 2020).

https://www.youtube.com/watch?reload=9&v=1dv-gVKBRBA&ab_channel=UNAMGlobal (consulté le 31 octobre 2020). Un rapport technique sur l'application se trouve ici :

<https://cscwcivictكنولوجies.files.wordpress.com/2020/10/asistente-virtual-saul-esparza.pdf> (consulté le 31 octobre 2020).

<https://yucatanahora.mx/yucatan-busca-desarrollar-proyectos-de-inteligencia-artificial/> (consulté le 31 octobre 2020).

Rosado-Tamariz E, Zuniga-Garcia M A, Santamaria-Bonfil G, Batres R, *A machine-learning approach to speed-up simulation towards the design of optimum operating profiles of power plants*. Dans les Actes de la 8e Conférence internationale sur l'informatique, l'environnement, l'énergie et les applications (IEEA '19). Association for Computing Machinery, New York, NY, USA : 194-199. DOI. <https://doi.org/10.1145/3323716.3323735>.

Santamaria-Bonfil G, Reyes-Ballesteros A, Gershenson C, *Wind speed forecasting for wind farms : A method based on support vector regression*, *Renewable Energy*, 85, 2016 : 790-809, ISSN 0960-1481, DOI. <https://doi.org/10.1016/j.renene.2015.07.004>.

TOWARDS AN AI STRATEGY IN MEXICO : Harnessing the AI Revolution, White Paper Report, British Embassy in Mexico through the Prosperity Fund, Oxford Insights and C Minds, juin 2018, p. 23. Consultation électronique à l'adresse <https://www.oxfordinsights.com/mexico> (consulté le 31 octobre 2020).

Wind Power Forecasting using Artificial Intelligence Tools, *Ingeniería, Investigación y Tecnología*, 2018, 19-4 : 1-11. DOI. <http://dx.doi.org/10.22201/fi.25940732e.2018.19n4.033>

Yu Z, Han S, Zhu J, Sun X, Ji C, Guo X. *Pre-pregnancy body mass index in relation to infant birth weight and offspring overweight/obesity : a systematic review and meta-analysis*. *PloS one*. 2013;8:e61627. DOI. <https://doi.org/10.1371/journal.pone.0061627> et Weng SF, Redsell SA, Swift JA, Yang M, Glazebrook CP. *Revue systématique et méta-analyses des facteurs de risque de surpoids chez les enfants identifiables pendant la petite enfance*. *Arch Dis Child*. 2012;97:1019-26. DOI. <https://doi.org/10.1136/archdischild-2012-302263>.

Les systèmes d'intelligence artificielle (IA) sont de plus en plus utilisés dans notre vie quotidienne et dans presque tous les types d'activité humaine, par exemple dans des domaines tels que l'éducation et la protection sociale, la société de l'information, les systèmes judiciaires et répressifs et, récemment, pour lutter contre la pandémie de Covid-19. Très souvent cités en tant qu'outils révolutionnaires, les systèmes d'IA peuvent apporter de nombreux avantages, mais ils peuvent également soulever des questions complexes et importantes sur les plans juridique, éthique, politique et économique.

Cette publication vise à alimenter les réflexions en cours au sein du CAHAI sur l'analyse des défis posés par les systèmes d'IA et les réponses réglementaires possibles.

Tout d'abord, elle vise à informer le lecteur de l'avancement des travaux du Comité ad hoc sur l'intelligence artificielle (CAHAI) et elle présente plusieurs études réalisées dans le cadre du CAHAI.

Deuxièmement, elle présente les perspectives nationales de différents États observateurs, l'Israël, le Japon et le Mexique, afin de soutenir le développement d'un cadre juridique international de l'intelligence artificielle basé sur les normes établies par le Conseil de l'Europe.

www.coe.int/cahai

www.coe.int

Le Conseil de l'Europe est la principale organisation de défense des droits de l'homme du continent. Il comprend 47 États membres, dont l'ensemble des membres de l'Union européenne. Tous les États membres du Conseil de l'Europe ont signé la Convention européenne des droits de l'homme, un traité visant à protéger les droits de l'homme, la démocratie et l'État de droit. La Cour européenne des droits de l'homme contrôle la mise en œuvre de la Convention dans les États membres.