



Doc. 15150

24 septembre 2020

La nécessité d'une gouvernance démocratique de l'intelligence artificielle

Rapport¹

Commission des questions politiques et de la démocratie

Rapporteure: Mme Deborah BERGAMINI, Italie, Groupe du Parti populaire européen

Résumé

L'intelligence artificielle (IA) fait partie de la nouvelle réalité. Son large usage influencera de plus en plus divers aspects de nos vies et transformera nos sociétés. Cette influence peut être à la fois bénéfique et dommageable.

Le rapport met l'accent sur l'impact de l'IA sur la démocratie. Il donne un aperçu des différentes manières dont l'utilisation de la technologie basée sur l'IA peut, et affecte déjà, le fonctionnement des institutions et processus démocratiques et le comportement social et politique des citoyens. Il conclut que l'utilisation de l'IA et son potentiel d'abus par les États et les organismes publics, ainsi que par les grandes entreprises privées, constituent une menace réelle pour les institutions, les processus et les normes de nos démocraties fondées sur les droits.

Il est nécessaire de mettre en place un cadre garantissant que cette technologie est développée et utilisée dans le plein respect de nos valeurs, de nos droits fondamentaux, de l'État de droit et de la démocratie. Le Conseil de l'Europe devrait jouer un rôle de pionnier dans la conception de méthodes et de formats garantissant que les technologies basées sur l'IA sont utilisées pour améliorer et non pour porter atteinte à la démocratie.

Le projet de résolution se félicite de la mise en place d'un Comité ad hoc sur l'intelligence artificielle (CAHAI) et invite les États membres à travailler ensemble à l'élaboration d'un instrument juridiquement contraignant visant à assurer la gouvernance démocratique de l'IA et, le cas échéant, à le compléter par des instruments juridiques sectoriels. Il énonce une série d'éléments à prendre en compte dans ce processus. Le projet de recommandation invite le Comité des Ministres à exprimer son soutien à l'élaboration d'un tel instrument, éventuellement sous la forme d'une convention.

1. Renvoi en commission : Doc. 14868, renvoi 4445 du 12 avril 2019.



Sommaire	Page
A. Projet de résolution	3
B. Projet de recommandation	6
C. Exposé des motifs, par Mme Bergamini, Rapporteuse	7
1. Introduction	7
2. Intelligence artificielle, définition et principes éthiques	8
3. L'impact de l'intelligence artificielle sur la démocratie	9
4. Initiatives en cours pour l'instauration d'un cadre réglementaire applicable à l'intelligence artificielle	13
5. Les travaux du Conseil de l'Europe	15
6. Intelligence artificielle et grandes entreprises: responsabilité et éthique	16
7. Conclusions	16
Annexe	19

A. Projet de résolution²

1. La technologie a toujours eu une grande incidence sur le cours de l'histoire humaine. Cependant, les progrès technologiques n'ont jamais été aussi rapides, et leurs effets sur les êtres humains aussi directs, palpables et étendus qu'aujourd'hui, à l'aube de la quatrième révolution industrielle. L'intelligence artificielle (IA), qui en est le moteur, est vue au sens large comme un facteur déterminant pour l'avenir de l'humanité, qui transformera en profondeur la vie des individus et retentira fortement sur les communautés humaines.

2. Les appareils dotés de l'IA sont déjà largement présents dans notre quotidien et exécutent de multiples tâches qui étaient autrefois accomplies par des êtres humains, dans la vie privée comme dans le milieu professionnel. Les algorithmes de prédiction, inhérents à l'IA, sont fréquemment employés dans d'importantes décisions comme les admissions à l'université, l'octroi de prêts et la gestion des ressources humaines, mais aussi en ce qui concerne les contrôles aux frontières (y compris dans les aéroports) et la prévention des crimes (par le biais de pratiques policières prédictives et l'utilisation, dans le cadre du système judiciaire pénal, d'instruments d'évaluation des risques quant aux récidives). Alors que toutes nos sociétés s'efforcent de lutter contre la pandémie actuelle de covid-19, l'IA est aussi utilisée pour améliorer la recherche pharmaceutique et aider à analyser les données médicales.

3. Cela étant, les effets à long terme de l'IA sur les êtres humains et sur la société sont encore loin d'être clairs. L'IA pourrait certes permettre de grandes avancées économiques et sociales, mais elle pose aussi toute une série de problèmes complexes. D'un côté, on espère qu'elle apportera des gains importants de productivité, de la croissance économique, des avancées scientifiques, des soins médicaux de meilleure qualité, une augmentation de l'espérance de vie, une sécurité et un confort toujours plus grand. D'un autre côté, certains craignent que l'IA déstabilise gravement les marchés du travail partout dans le monde, conduise à une augmentation des inégalités de revenus, de richesse et des inégalités sociales, et mette en danger la stabilité sociale et politique, ainsi que la sécurité internationale.

4. Les technologies fondées sur l'IA ont une incidence sur le fonctionnement des institutions et des processus démocratiques, ainsi que sur le comportement social et politique des citoyens. Son utilisation peut avoir des effets bénéfiques aussi bien que préjudiciables sur la démocratie. De fait, l'intégration rapide des technologies d'IA dans les outils de communication modernes et les plates-formes des réseaux sociaux crée d'incomparables possibilités d'influencer des individus ou des groupes sociaux de manière ciblée, personnalisée et souvent inaperçue; différents acteurs politiques peuvent être tentés d'exploiter ces possibilités à leur avantage.

5. L'aspect positif est que l'IA permet d'améliorer la mise en œuvre de la responsabilité et de la transparence des gouvernements et de contribuer à la lutte contre la corruption, et qu'elle présente de nombreux avantages pour l'action, la participation et le pluralisme démocratiques, rendant ainsi la démocratie plus directe, plus efficace et plus attentive aux besoins des citoyens. Les technologies fondées sur l'IA peuvent aussi permettre d'élargir la représentation démocratique, en décentralisant les systèmes d'information et les plates-formes de communication. L'IA peut renforcer l'autonomie informationnelle des citoyens, améliorer la manière dont ils s'informent sur les processus politiques et les aider à y participer à distance, en facilitant l'expression politique et en fournissant des canaux de communication avec les acteurs politiques. Elle peut aussi contribuer à l'instauration d'une plus grande confiance entre l'État et la société et entre les citoyens eux-mêmes.

6. Cela étant, l'IA peut être utilisée, et, selon certaines sources, est déjà utilisée pour perturber la démocratie par l'ingérence dans les processus électoraux, le ciblage politique personnalisé, le conditionnement du comportement des électeurs et la manipulation de l'opinion publique. De plus, elle a apparemment été employée pour amplifier la diffusion de fausses informations, de propagande, de «chambres d'écho» et de discours de haine, altérant ainsi la pensée critique et contribuant à la montée du populisme et à la polarisation des sociétés démocratiques.

7. De surcroît, certains États et acteurs privés recourent dans une large mesure à des technologies fondées sur l'IA pour contrôler les personnes, par un filtrage automatisé d'informations s'apparentant à de la censure par exemple, ou encore par une surveillance de masse au moyen de smartphones, le recueil de données à caractère personnel et le suivi de l'activité en ligne et hors ligne; ces pratiques risquant de conduire à une érosion de l'intégrité psychologique des citoyens, des droits civiques et des libertés politiques, et à l'émergence d'un autoritarisme numérique – en d'autres termes, à un nouvel ordre social qui entrerait en concurrence avec la démocratie.

2. Projet de résolution adopté à l'unanimité par la commission le 23 septembre 2020.

8. La concentration de données, d'informations, du pouvoir et des capacités d'influence dans les mains de quelques grands acteurs privés du développement et de la fourniture de technologies et de services fondés sur l'IA, et la dépendance croissante des personnes, des institutions et de la société dans son ensemble vis-à-vis de ces services sont aussi des sources de préoccupation. Ces grandes sociétés ne fonctionnent plus comme de simples canaux de communication entre les individus et les institutions: elles jouent elles-mêmes un rôle de plus en plus important, contrôlent et filtrent les flux d'informations, censurent automatiquement les contenus publiés sur les médias sociaux, définissent l'ordre du jour, façonnent et transforment les modèles sociaux et politiques. Ces acteurs, dont les modèles d'activité privilégient les profits des actionnaires aux dépens du bien commun, pourraient constituer une menace pour l'ordre démocratique et ne devraient pas échapper au contrôle démocratique.

9. L'Assemblée note que, ces dernières années, des gouvernements, la société civile, des institutions internationales et des entreprises ont eu des échanges de vues approfondis dans le but de définir d'un commun accord un ensemble de principes pour répondre aux inquiétudes que suscite l'utilisation de l'IA. Elle se félicite de ce que le Conseil de l'Europe, en tant qu'organisation de premier plan en matière de droits humains, ait activement participé à ces discussions sur l'avenir de l'IA et sa gouvernance, et se réjouit en particulier du fait que le Comité des Ministres, la Commissaire aux droits de l'homme et les organes de coopération intergouvernementale aient contribué à ce processus.

10. L'Assemblée considère que les principes éthiques d'autoréglementation et les politiques mis en place volontairement par des acteurs privés ne sont pas des outils adaptés ni suffisants pour réglementer l'IA, car ils n'entraînent pas nécessairement de contrôle démocratique ni d'obligation de rendre des comptes. L'Europe doit veiller à ce que la puissance de l'IA soit soumise à des règles et soit employée pour le bien commun.

11. Par conséquent, l'Assemblée est fermement convaincue qu'il est nécessaire d'instaurer un cadre réglementaire transversal pour l'IA, définissant des principes spécifiques fondés sur la protection des droits humains, la démocratie et l'État de droit. Toute réflexion dans ce domaine doit associer l'ensemble des parties prenantes, et en particulier les citoyens et les grandes entreprises privées qui prennent part à l'élaboration et à la fourniture de technologies et de services fondés sur l'IA.

12. Le Conseil de l'Europe, en tant qu'organisation internationale de normalisation de premier plan dans le domaine de la démocratie, doit faire œuvre de pionnier en concevant des solutions et des formats permettant de s'assurer que les technologies fondées sur l'IA sont employées non pas pour fragiliser la démocratie, mais pour la renforcer.

13. Dans ce contexte, elle se félicite de la création, par le Comité des Ministres, d'un Comité ad hoc sur l'intelligence artificielle (CAHAI), chargé d'examiner la faisabilité et les éléments potentiels, sur la base de vastes consultations multipartites, d'un cadre juridique pour la conception, le développement et l'application de l'IA. Elle invite les États membres du Conseil de l'Europe et les États observateurs participant au CAHAI à travailler ensemble à l'élaboration d'un instrument juridiquement contraignant et destiné à garantir la gouvernance démocratique de l'IA, et, le cas échéant, de le compléter par des instruments juridiques sectoriels.

14. L'Assemblée estime qu'un tel instrument devrait:

14.1. garantir que les technologies fondées sur l'IA sont conçues, développées et exploitées dans le strict respect et en soutien des normes du Conseil de l'Europe en matière de droits humains, de démocratie et d'État de droit;

14.2. promouvoir une compréhension commune et prévoir le respect de principes et concepts éthiques fondamentaux et la mise en œuvre des normes susmentionnées dont:

14.2.1. la transparence, y compris l'accessibilité et l'explicabilité;

14.2.2. la justice et l'équité, y compris la non-discrimination;

14.2.3. la prise de décision par une personne, qui en est responsable, et la mise à disposition de voies de recours;

14.2.4. la sûreté et la sécurité;

14.2.5. le respect de la vie privée et la protection des données.

14.3. s'attacher à ce que l'IA ait le plus d'effets positifs possibles sur le fonctionnement des institutions et processus démocratiques, notamment à ce qu'elle permette:

14.3.1. de renforcer l'obligation des gouvernements à rendre des comptes;

- 14.3.2. de lutter contre la corruption et le crime économique;
- 14.3.3. de faciliter l'action démocratique, la participation et le pluralisme;
- 14.3.4. de rendre la démocratie plus directe, plus efficace et plus attentive aux besoins des citoyens;
- 14.3.5. d'élargir la représentation démocratique, en décentralisant les systèmes d'information et les plates-formes de communication;
- 14.3.6. de renforcer l'autonomie informationnelle des citoyens, d'améliorer la manière dont ils s'informent sur les processus politiques et de les aider à y participer à distance, en facilitant l'expression politique et en fournissant des canaux de communication avec les acteurs politiques;
- 14.3.7. d'améliorer la transparence dans la vie publique et de contribuer à l'instauration d'une plus grande confiance entre l'État et la société et entre les citoyens eux-mêmes;

14.4. comporter des dispositions visant à prévenir et/ou limiter les risques que l'IA soit utilisée pour fragiliser et perturber la démocratie, notamment par:

- 14.4.1. l'ingérence dans les processus électoraux, le ciblage politique personnalisé, le conditionnement du comportement politique des électeurs et la manipulation de l'opinion publique;
- 14.4.2. l'amplification de la diffusion de fausses informations, de la propagande et des «chambres d'écho»;
- 14.4.3. l'altération de la pensée critique des individus et de la société;
- 14.4.4. la contribution à la montée du populisme et à la polarisation des sociétés démocratiques;

14.5. comporter des dispositions visant à limiter les risques d'utilisation des technologies fondées sur l'IA par des États et des acteurs privés pour contrôler les personnes, ce qui pourrait conduire à une érosion de l'intégrité psychologique des citoyens, des droits civiques et des libertés politiques;

14.6. comporter des garde-fous pour prévenir les menaces à l'ordre démocratique résultant de la concentration de données, d'informations, du pouvoir et des capacités d'influence dans les mains de quelques grands acteurs privés du développement et de la fourniture de technologies et de services fondés sur l'IA, et de la dépendance croissante des personnes, des institutions et de la société dans son ensemble vis-à-vis de ces services, et comporter également des dispositions visant à garantir que les activités de ces acteurs sont soumises à un contrôle démocratique.

15. En outre, l'Assemblée estime que, pour garantir le respect de l'obligation de rendre des comptes, le cadre juridique à mettre en place devrait prévoir un mécanisme de contrôle indépendant et préventif auquel participent toutes les parties prenantes, qui garantirait le respect effectif de ses dispositions. Ce mécanisme exigerait la présence d'un organe extrêmement compétent (notamment sur le plan technique, juridique et éthique), capable de suivre les nouvelles évolutions de la technologie numérique et d'évaluer avec précision et autorité ses risques et conséquences.

16. En ce qui concerne les algorithmes et les plates-formes de réseaux sociaux, l'Assemblée estime nécessaire:

- 16.1. de rendre plus transparents les facteurs de décision qui sous-tendent les contenus générés par des algorithmes;
- 16.2. de donner aux utilisateurs plus de souplesse pour décider de la manière dont les algorithmes façonnent leur expérience en ligne;
- 16.3. de demander instamment aux plates-formes de faire preuve de diligence plus systématique en matière de droits humains pour comprendre l'impact social de leurs algorithmes;
- 16.4. d'envisager de créer un organe d'expertise indépendant pour surveiller les plates-formes technologiques et le fonctionnement de leurs algorithmes;
- 16.5. de renforcer les contrôles de confidentialité des données d'utilisateur, pour que, dès leur conception, les algorithmes soient moins en mesure d'exploiter les données.

B. Projet de recommandation³

1. L'Assemblée parlementaire se réfère à sa Résolution ... (2020) «La nécessité d'une gouvernance démocratique de l'intelligence artificielle» et salue les efforts continus menés jusqu'ici aux niveaux national, européen et international, notamment au niveau du Conseil de l'Europe, pour instaurer un cadre réglementaire pour l'intelligence artificielle (IA). En l'absence d'un tel cadre, l'Assemblée note que les entreprises privées qui élaborent et utilisent des technologies fondées sur l'IA ont opté jusqu'ici dans ce domaine, pour une politique d'autoréglementation reposant sur des instruments de droit souple.
2. L'Assemblée salue le travail accompli par le 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 la conception, le développement et l'application de l'IA, fondé sur les normes du Conseil de l'Europe dans le domaine des droits humains, de la démocratie et de l'État de droit.
3. L'Assemblée est convaincue qu'une action forte et rapide est nécessaire de la part du Conseil de l'Europe pour relever les défis que pose l'utilisation des technologies fondées sur l'IA qui peuvent perturber, entre autres, le fonctionnement des institutions et processus démocratiques. Les instruments de droit souple et l'autoréglementation se sont révélés insuffisants jusqu'ici pour relever ces défis et protéger les droits humains, la démocratie et l'État de droit. Le Conseil de l'Europe occupe une position stratégique qui lui permet de fournir les conseils et l'assistance nécessaires, en étroite coopération et coordination avec d'autres institutions et organisations européennes et internationales, en vue de l'instauration d'un cadre réglementaire global de l'IA.
4. Au vu de ce qui précède, l'Assemblée recommande au Comité des Ministres:
 - 4.1. d'exprimer son soutien quant à l'élaboration d'un instrument juridiquement contraignant gouvernant l'intelligence artificielle, éventuellement sous la forme d'une convention, en tenant compte, entre autres, des éléments figurant dans la Résolution ... (2020);
 - 4.2. de garantir qu'un tel instrument juridiquement contraignant soit basé sur une approche globale, se rapporte à l'ensemble des cycles de vie des systèmes fondés sur l'IA, soit destiné à l'ensemble des parties prenantes et comprenne des mécanismes afin de garantir la mise en œuvre de cet instrument.

3. Projet de recommandation adopté à l'unanimité par la commission le 23 septembre 2020.

C. Exposé des motifs, par M^{me} Bergamini, Rapporteure

1. Introduction

1.1. Procédure

1. Le 9 avril 2019, la Commission des questions politiques et de la démocratie a déposé une proposition de résolution sur «La nécessité d'une gouvernance démocratique de l'intelligence artificielle»⁴. Constatant que l'intelligence artificielle (IA) est de plus en plus généralement considérée comme un facteur déterminant de l'avenir de l'humanité, la proposition souligne à quel point l'IA influence d'ores et déjà le fonctionnement de la démocratie (ingérence dans les processus électoraux, ciblage politique personnalisé, conditionnement du comportement des électeurs, propagation de fausses informations pour manipuler l'opinion publique, etc.). Dans sa proposition, la commission note également que la concentration du pouvoir entre les mains de quelques géants du secteur privé qui échappent au contrôle démocratique suscite des inquiétudes. Elle appelle donc à l'adoption de cadres réglementaires à l'échelle nationale et internationale pour garantir la gouvernance démocratique de l'IA et empêcher son utilisation abusive. La proposition a été renvoyée à notre commission pour rapport le 12 avril 2019, et j'ai été nommée rapporteure le 25 juin 2019.

2. Le 2 octobre 2019, la commission a tenu une audition à laquelle ont participé M^{me} Birgit Schippers, Maître de conférences en sciences politiques au St Mary's University College, Belfast, M. Paul Nemitz, Conseiller principal, Direction générale de la justice et des consommateurs, Commission européenne, et M. Yannick Meneceur, Direction de la société de l'information et de la lutte contre la criminalité, Direction Générale Droits de l'homme et État de droit, Conseil de l'Europe. Le 27 janvier 2020, la commission a eu un échange de vues avec M. Dario Fumagalli, juriste spécialisé dans la protection de la vie privée.

3. En ma qualité de rapporteure, j'ai eu l'occasion de représenter l'Assemblée parlementaire à divers événements liés à l'IA, notamment le Réseau parlementaire mondial de l'OCDE (le 11 octobre 2019 à Paris), où je me suis exprimée lors de la session intitulée «Comment les pays abordent leurs stratégies et politiques en matière d'IA». En mai et juin 2020, j'ai eu un échange de vues (par visioconférence) avec M. Steven Feldstein, boursier non-résident, Programme sur la démocratie, les conflits et la gouvernance, Fondation Carnegie pour la paix internationale, et avec trois représentants du siège européen de Facebook à Bruxelles: M^{me} Marisa Jimenez Martin, Directrice adjointe, Affaires et politiques publiques européennes, M. Janne Elvelid, Responsable des politiques, Affaires européennes, et M^{me} Michela Palladino, Cadre, Politiques publiques européennes. Ces réunions ont porté sur l'utilisation éthique de l'IA, le rôle des algorithmes dans les plates-formes de réseaux sociaux et leurs possibles répercussions sur la démocratie, ainsi que sur la nécessité d'un cadre réglementaire instaurant un système fondé sur des valeurs et des principes et d'une coopération plus étroite entre les sociétés privées et les organisations internationales.

1.2. Raison d'être du rapport

4. Les effets de l'IA, principal moteur de la quatrième révolution industrielle, sont visibles partout et dans tous les domaines de la vie. Sous sa forme matérialisée que sont les robots, l'IA conduira bientôt les voitures, approvisionnera les entrepôts et s'occupera des jeunes enfants et des personnes âgées. Les algorithmes de prédiction qui sont inhérents à l'IA sont omniprésents dans notre environnement, que ce soit dans la fonction de lecture automatique de YouTube, dans les recommandations de films de Netflix ou dans la publicité affichée par le moteur de recherche Google. Ces algorithmes sont régulièrement employés dans les décisions de prêt, les admissions à l'université et le recrutement, mais aussi dans le travail de la police, dans les aéroports, aux frontières ou dans les décisions de justice⁵. Alors que nos sociétés s'efforcent toutes de lutter contre la pandémie actuelle de covid-19, l'IA est aussi utilisée pour améliorer la recherche pharmaceutique et aider à analyser les données médicales⁶.

4. [Doc. 14868](#).

5. Deutsche Bank Research, «Digital Politics. AI, big data and the future of democracy», août 2019.

6. Le Président de l'Assemblée, la rapporteure sur «L'intelligence artificielle dans le domaine de la santé: défis médicaux, éthiques et juridiques à venir», M^{me} Selin Sayek Böke, et moi-même avons fait des déclarations pour insister sur le fait que l'utilisation de l'IA dans le contexte de la pandémie de covid-19, notamment pour le traçage et le suivi, doit être responsable, éthique et respectueuse des normes en matière de droits humains défendues par le Conseil de l'Europe. La question du respect de la vie privée et de la protection des données dans le contexte du suivi des patients et du traçage de leurs contacts est traitée dans le rapport intitulé «Les conséquences de la pandémie de covid-19 sur les droits de l'homme et l'État de droit», adopté le 29 juin 2020 par la Commission des questions juridiques et des droits de l'homme. Les questions touchant aux soins médicaux et à l'IA sont examinées dans le rapport de M^{me} Sayek Böke, adopté le 22 septembre 2020 par la Commission des questions sociales, de la santé et du développement durable.

5. L'IA promet de résoudre certains des problèmes les plus urgents de la société, mais elle pose aussi des défis: algorithmes impénétrables à effet «boîte noire»⁷, existence possible de parti pris et d'éléments discriminatoires dans la modélisation et les effets des outils d'analyse de données, maintien des biais avec l'utilisation des données historiques, déplacement potentiel des emplois, quasi-absence des femmes dans les carrières techniques et utilisation non-éthique des données. Si l'IA a porté l'accès à l'information et l'échange d'informations à des niveaux sans précédent, elle a aussi amplifié certaines tendances négatives, contribuant ainsi à la montée du populisme et à la polarisation des sociétés démocratiques.

6. Ces deux ou trois dernières années, des gouvernements, la société civile, des institutions internationales et des entreprises ont eu des échanges de vues approfondis dans le but de définir d'un commun accord un ensemble de principes pour répondre aux défis complexes que pose l'IA. Le Conseil de l'Europe, en tant que principale organisation en matière de droits humains, a pris une part active à ces discussions sur l'avenir de l'IA et sa gouvernance. Dans son dernier rapport en tant que Secrétaire Général, Thorbjørn Jagland a lancé un appel en faveur d'une approche stratégique et transversale de l'IA, conçue et mise en application dans le respect des normes européennes en matière de droits humains, de démocratie et d'État de droit⁸. Le Comité des Ministres a souligné à plusieurs reprises la nécessité éventuelle d'instaurer un cadre réglementaire pour l'IA⁹. De même, l'Assemblée a souligné dans plusieurs documents l'importance d'adopter une approche globale à l'égard de l'IA et de prendre en compte toute la diversité des opportunités et des défis liés à cette technologie¹⁰. Dans ce même esprit, l'Assemblée a initié plusieurs rapports sur l'impact de l'IA dans différents domaines.

7. Le présent rapport s'attache à décrire l'influence et l'impact de l'IA sur le fonctionnement de la démocratie, et à expliquer comment les différentes parties prenantes peuvent participer et contribuer au dialogue sur l'IA. Mais surtout, il présente des arguments en faveur de la définition d'objectifs communs sur la base desquels les institutions et les entreprises pourront, grâce à une coopération ouverte, claire et volontaire, définir un cadre commun pour une gouvernance démocratique de l'IA.

2. Intelligence artificielle, définition et principes éthiques

8. Dans les discussions sur l'IA, certains ont exprimé la crainte que cet outil au service de l'humanité finisse par prendre le pouvoir. Cependant, tous les interlocuteurs ne se basent pas sur la même définition de l'IA; si les éléments de base sont généralement identiques, l'accent est mis sur des aspects qui varient en fonction de l'entité à l'origine de la définition. Par ailleurs, à mesure que les machines intelligentes deviennent omniprésentes, les questions éthiques liées à l'IA se multiplient et retiennent de plus en plus l'attention du public. Par exemple, l'IA reproduit des éléments fondamentaux de la capacité humaine d'agir en fonction de critères moraux, c'est-à-dire la capacité des individus ou des collectivités de porter un jugement moral, de prendre des décisions éthiques fondées sur les notions de bien et de mal, et de répondre de leurs actes. C'est pourquoi il est essentiel d'être conscient des dangers liés au recours à l'IA plutôt qu'à l'intelligence humaine dans les processus décisionnels. En fait, l'apprentissage automatique génère des algorithmes qui semblent être très efficaces pour formuler des prévisions, mais pas pour comprendre pourquoi certaines choses se produisent. L'absence de pensée causale est l'un des principaux problèmes qui devraient nous conduire à soumettre l'IA à de solides principes éthiques¹¹.

9. Plusieurs organisations internationales ont cherché à définir ce concept. Selon le glossaire du Conseil de l'Europe, l'IA désigne un «ensemble de sciences, théories et techniques dont le but est de reproduire par une machine des capacités cognitives d'un être humain»¹². Selon la définition qu'en donne la Commission européenne, l'IA désigne des «systèmes qui font preuve d'un comportement intelligent en analysant leur environnement et en prenant des mesures – avec un certain degré d'autonomie – pour atteindre des objectifs spécifiques»¹³. Pour l'Organisation de coopération et de développement économiques (OCDE), «un système

7. Voir le [glossaire du Conseil de l'Europe](#) sur l'intelligence artificielle.

8. «Relever les défis à venir – Renforcer le Conseil de l'Europe», avril 2019.

9. Entre autres, la 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 Conclusions de la Conférence de haut-niveau «Maîtriser les règles du jeu – Les impacts du développement de l'intelligence artificielle sur les droits de l'homme, la démocratie et l'État de droit» (26-27 février 2019). Pour en savoir plus sur les travaux du Comité des Ministres sur l'IA, voir le chapitre 5 du présent exposé des motifs.

10. [Recommandation 2102\(2017\)](#) «La convergence technologique, l'intelligence artificielle et les droits de l'homme»; [Résolution 2255 \(2019\)](#) «Les médias de service public dans le contexte de la désinformation et de la propagande»; [Résolution 2281 \(2019\)](#) «Médias sociaux : créateurs de liens sociaux ou menaces pour les droits humains?».

11. «What AI still can't do», Brian Bergstein, *MIT Technology Review*, 19 février 2020.

12. Voir le [glossaire du Conseil de l'Europe](#) sur l'intelligence artificielle.

d'intelligence artificielle (ou système d'IA) est un système automatisé qui, pour un ensemble donné d'objectifs définis par l'homme, est en mesure d'établir des prévisions, de formuler des recommandations, ou de prendre des décisions influant sur des environnements réels ou virtuels. Les systèmes d'IA sont conçus pour fonctionner à des degrés d'autonomie divers¹⁴. Enfin, l'UNESCO note que «bien qu'il n'existe pas de définition unique de l'IA, les débats semblent se concentrer sur les machines capables de reproduire certaines fonctionnalités de l'intelligence humaine, notamment sur certaines caractéristiques telles que la perception, l'apprentissage, le raisonnement, la résolution de problème, l'interaction linguistique et même la production d'un travail de création¹⁵.»

10. De même, malgré un consensus apparent sur la nécessité d'une IA éthique, il y a encore débat sur ce qu'il faut entendre par là. Au cours des cinq dernières années, des sociétés privées, des instituts de recherche et des organisations du secteur public ont formulé de nombreuses séries de principes et de lignes directrices pour une IA éthique¹⁶. Ces lignes directrices ont tendance à s'accorder sur certains principes généraux tels que la transparence, la justice, la responsabilité et la vie privée, mais semblent être en profond désaccord sur ce qui devrait être fait, en détail, dans la pratique. De plus, d'autres principes tels que la durabilité, la dignité et la solidarité sont largement sous-représentés, ce qui laisse penser que ces questions passent inaperçues dans le débat général actuel sur l'éthique de l'IA.

11. Aux fins des rapports de l'Assemblée sur l'IA, le concept d'IA et les principes éthiques devant être appliqués aux systèmes d'IA doivent être compris selon la description qui en est faite à l'Annexe I du présent document.

3. L'impact de l'intelligence artificielle sur la démocratie

12. Dans cette section, qui s'inspire largement des études de Kevin Körner (*Deutsche Bank Research*, cité à la note en bas de page n° 5) et de Catelijne Muller¹⁷, ainsi que des contributions apportées par des experts et des membres de la commission, j'aimerais dresser une cartographie des différentes façons dont l'utilisation des technologies basées sur l'IA peut influencer et influence déjà le fonctionnement des institutions et des processus démocratiques ainsi que le comportement social et politique des citoyens.

13. La démocratie, c'est gouverner le peuple par le peuple et pour le peuple. Elle prévoit des contrôles pour s'assurer que le pouvoir n'est pas concentré entre les mains de quelques-uns et ne peut fonctionner correctement que si elle repose sur des institutions solides qui jouissent de la confiance d'une population active, engagée et informée, et qui sont en mesure d'assurer un équilibre dynamique entre les intérêts des parties en présence. La crise des démocraties modernes touche presque tous les aspects de l'ordre démocratique: érosion et perte de la confiance dans les institutions, mésinformation et désinformation du public, rupture de la cohésion sociale et polarisation de la société. Les technologies modernes, notamment les systèmes fondés sur l'IA, peuvent contribuer à résoudre cette crise, mais tout aussi bien l'aggraver.

14. L'utilisation de l'IA par les êtres humains n'est pas neutre. Elle peut être utilisée pour renforcer l'obligation des pouvoirs publics de rendre des comptes et peut présenter de nombreux avantages pour l'action, la participation et le pluralisme démocratiques, et rendre la démocratie plus directe et plus réactive. Cela étant, elle peut aussi servir à renforcer les capacités de répression et à manipuler. De fait, l'intégration rapide des technologies d'IA dans les outils de communication modernes et les plates-formes des réseaux sociaux crée d'incomparables possibilités d'influencer des individus ou des groupes sociaux de manière ciblée, personnalisée et souvent inaperçue; différents acteurs politiques peuvent être tentés d'exploiter ces possibilités à leur avantage.

13. Communication de la Commission au Parlement européen, au Conseil européen, au Conseil, au Comité économique et social européen et au Comité des régions, «Un plan coordonné dans le domaine de l'intelligence artificielle», 7 décembre 2018, COM(2018)795 final.

14. Recommandation du Conseil sur l'intelligence artificielle adoptée le 22 mai 2019, OECD/LEGAL/0449.

15. «Étude préliminaire sur les aspects techniques et juridiques liés à l'opportunité d'un instrument normatif sur l'éthique de l'intelligence artificielle», 206^e session du Conseil exécutif de l'UNESCO, octobre 2019.

16. Selon une étude récente menée par l'Université d'Harvard, il semblerait que toutes les organisations travaillant dans le domaine des politiques technologiques aient rédigé ou adopté un ensemble de principes sur l'IA. «Principled artificial intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI», Berkman Klein Center for Internet Society, Université d'Harvard, 2020.

17. «The impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law», rapport de Catelijne Muller, LLM, CAHAI (2020)06-fin.

15. L'expérience de ces dernières années permet de recenser quelques domaines clés dans lesquels le recours aux technologies fondées sur l'IA risque de saper et de déstabiliser la démocratie, et notamment:

- a. l'accès à l'information (désinformation, «chambres d'écho» et altération de la pensée critique);
- b. la manipulation ciblée des citoyens;
- c. l'ingérence dans les processus électoraux;
- d. l'érosion des droits du citoyen;
- e. la prise de pouvoir financier et politique de l'économie des données.

16. De plus, lorsque des États ont largement recours à des technologies fondées sur l'IA pour contrôler les citoyens, telles que le filtrage automatisé de l'information, qui s'apparente à de la censure, ou la surveillance de masse utilisant les smartphones et des circuits de vidéosurveillance combinés à de gigantesques bases de données intégrées, cela risque d'entraîner l'érosion des libertés politiques et l'émergence d'un autoritarisme numérique – un nouvel ordre social qui ferait concurrence à la démocratie.

3.1. Accès à l'information – désinformation, «chambres d'écho» et altération de la pensée critique

17. Une démocratie en bonne santé suppose que les citoyens soient bien informés et que des personnes ayant des points de vue différents se réunissent pour discuter afin de trouver des solutions communes par le dialogue. En déterminant quelles informations sont montrées et consommées (un algorithme de sites web peut deviner de manière sélective, à partir de renseignements sur un utilisateur, quelles informations celui-ci souhaite voir), les technologies fondées sur l'IA qui sont utilisées dans les médias en ligne peuvent contribuer au développement de la désinformation et du discours de haine, et créer des «chambres d'écho»¹⁸ et des «bulles de filtres», qui conduisent les individus dans un état de cloisonnement intellectuel où le dialogue n'a pas sa place, d'où une altération de la pensée critique et une désorganisation de la démocratie. De plus, en donnant la priorité aux nouvelles et aux informations qui plaisent aux utilisateurs, les algorithmes tendent à renforcer leurs opinions, leurs goûts et leurs habitudes, et à limiter l'accès aux opinions divergentes, ce qui réduit le libre choix des utilisateurs.

18. Pour ce qui est du rôle que jouent les algorithmes dans le développement de la désinformation et du discours de haine, les efforts ont porté avant tout sur la modération des contenus, c'est-à-dire sur la capacité des algorithmes à repérer et à supprimer les messages qui contreviennent aux normes de la collectivité et qui, en diffusant de mauvaises ou de fausses informations, franchissent la ligne rouge. Cela dit, il existe une utilisation tout aussi importante et plus préoccupante des algorithmes par des entreprises telles que Facebook: les algorithmes «d'élaboration de contenu». Plusieurs plateformes fondées sur l'IA exercent une censure automatisée (par des algorithmes définis par les propriétaires) des contenus publiés sur les médias sociaux par des particuliers, des acteurs politiques et même des institutions étatiques, et refusent ou retirent les informations et des opinions qui ne sont pas du goût des propriétaires des plateformes, ce qui limite la liberté d'expression.

19. En fait, les algorithmes d'élaboration de contenu déterminent ce que les utilisateurs, pris individuellement, «voient lorsqu'ils sont en ligne, notamment les messages générés par les utilisateurs, les messages organiques et les publicités payantes. Parmi les exemples les plus visibles d'algorithmes d'élaboration de contenu, on peut citer les fils d'actualité de Facebook et de Twitter ainsi que le moteur de recommandations de YouTube. Les algorithmes déterminent aussi à quels utilisateurs une publicité donnée doit être montrée. En règle générale, l'annonceur fixe les paramètres de ciblage (caractéristiques démographiques, centres d'intérêt supposés, etc.), mais ce sont les systèmes algorithmiques de la plateforme qui choisissent qui verra précisément la publicité concernée et déterminent son insertion au sein de la plate-forme¹⁹.»

20. Les études réalisées en interne par Facebook renforcent ce point de vue. L'équipe de ce réseau social conclut en effet que «dans 64 % des cas, le fait de rejoindre un groupe extrémiste est dû à [ses] outils de recommandation» et, en majorité, aux algorithmes des fonctionnalités «Groupes que vous devriez rejoindre» et «Découvrir» de Facebook, reconnaissant ainsi que ses «systèmes de recommandation ont pour effet

18. Une chambre d'écho peut être définie ainsi: environnement dans lequel les individus ne rencontrent que des convictions ou des opinions qui coïncident avec les leurs, leurs points de vue actuels s'en trouvant renforcés et les autres idées n'étant pas envisagées.

19. «It's not just the content, it's the business model: Democracy's online speech challenge», Nathalie Marechal et Ellery Roberts Biddle, *New America*, 17 mars 2020.

d'amplifier le problème²⁰.» Autrement dit, bien que de nombreuses plates-formes technologiques affirment qu'elles mènent une politique non-interventionniste en matière de contenu en permettant simplement aux utilisateurs d'exprimer leurs préférences, sans empiéter sur leurs droits de liberté d'expression, en réalité, elles trichent, sans faire de bruit, en déterminant quels messages seront vus et lus par des millions d'internautes et deviendront donc viraux. Par conséquent, leurs algorithmes définissent dans une très large mesure ce que voient les utilisateurs et ce à quoi ils réagissent. À ce jour, étant donné que les recettes et le bénéfice sont la motivation primordiale de Facebook et des autres plates-formes, on peut supposer que même si un contenu diffuse de fausses informations, l'algorithme cherchera à en augmenter la visibilité, tant que cela stimule l'implication des utilisateurs sur le site.

3.2. Manipulation ciblée des citoyens et ingérence dans les processus électoraux

21. La propagande et la manipulation de l'information ne sont pas nouveaux dans la boîte à outils politique, mais leur dimension et leur portée ont énormément augmenté sous l'effet des technologies de communication basées sur l'IA. Les médias en ligne et les réseaux sociaux, qui permettent d'utiliser ces technologies pour influencer les individus et promouvoir ou rejeter/refuser des intérêts partisans, jouent un rôle de plus en plus important dans le processus politique. Parmi les méthodes signalées par les experts politiques figurent la désinformation coordonnée à grande échelle, y compris le recours aux «deepfakes»²¹, le microciblage des électeurs, la polarisation du débat public, le travail de sape en ce qui concerne la confiance dans les institutions démocratiques, les partis et responsables politiques et la fiabilité de l'information, ainsi que le contrôle des flux d'information et de l'opinion publique.

22. Lors des élections, l'IA peut être utilisée efficacement pour s'adresser aux électeurs de manière individuelle tout au long du processus électoral. Les chatbots et les forums de discussion sur les réseaux sociaux qui encouragent les utilisateurs à envoyer des commentaires, des retours ou des critiques sont autant de moyens de sonder l'opinion publique. De plus, l'IA peut être utilisée pour collecter toutes ces données en temps réel, permettant ainsi aux partis politiques de modifier leurs campagnes en fonction du ressenti du public. L'IA peut aussi être employée pour manipuler individuellement les électeurs. Elle est utilisée pour analyser les profils comportementaux et psychographiques individuels des électeurs dans le but de les convaincre de voter pour un candidat donné, pour discréditer des opposants ou encore pour conforter les électeurs dans leur choix.

23. Le microciblage employé dans le cadre des campagnes politiques peut être considéré comme une publicité ordinaire, mais lorsqu'il s'appuie sur la collecte et la manipulation des données des utilisateurs (analyse des mégadonnées) afin d'anticiper et d'influencer les opinions politiques et les résultats électoraux (propagande numérique), cela fait peser une grave menace sur la démocratie, le débat public et le libre arbitre des électeurs.

24. Les cas les plus connus d'ingérence présumée dans un processus démocratique à l'aide de l'IA concernent l'élection présidentielle de 2016 aux États-Unis. La société de conseil politique Cambridge Analytica (qui a fermé entre-temps) est accusée d'avoir aidé Donald Trump à gagner l'élection en diffusant des contenus hostiles à Hillary Clinton, tandis que certains grands agrégateurs d'informations et médias «traditionnels» auraient privilégié les informations et les vidéos offrant une image positive de Clinton et une image négative de Trump.

3.3. Érosion des droits du citoyen

25. La disponibilité des données et les progrès rapides des systèmes d'IA entraîneront un recours accru à l'analyse prédictive, non seulement par les entreprises, les banques et les recruteurs, mais aussi par les institutions gouvernementales et les autorités publiques. Si les insuffisances et les risques qui en découlent ne sont pas traités de manière appropriée, l'amplification technologique des orientations tendancieuses et des préjugés ainsi que les inexactitudes et les erreurs statistiques pourraient conforter les inégalités préexistantes. Cela pourrait compromettre la protection contre la discrimination et les garanties d'égalité de traitement consacrées par les constitutions des sociétés démocratiques modernes et par la Convention européenne des droits de l'homme (STE n° 005, Article 14) et d'autres instruments du Conseil de l'Europe²².

20. «Facebook executives shut down efforts to make the site less divisive», Jeff Horwitz et Deepa Seetharaman, *The Wall Street Journal*, 26 mai 2020.

21. Le terme *deepfake* est fréquemment employé pour désigner des vidéos qui ont été modifiées à l'aide d'un algorithme en remplaçant une personne dans l'original par une autre personne (en particulier une personnalité publique) de façon à faire croire que la vidéo est authentique.

26. L'utilisation de systèmes d'IA pour établir le profil des individus, les suivre et les identifier, ainsi que pour recenser, classer et même orienter leur comportement peut avoir un effet dissuasif sur la liberté d'expression et la liberté de réunion et d'association (garanties par les articles 10 et 11 de la Convention européenne des droits de l'homme). Appliquée dans l'espace public, la reconnaissance faciale peut porter atteinte à la liberté individuelle d'opinion et d'expression, car elle supprime la protection qu'offre l'«anonymat de la foule». Cela pourrait décourager la participation à des manifestations et à des rassemblements pacifiques, qui comptent parmi les éléments les plus importants d'une société démocratique. De plus, certaines personnes pourront préférer s'abstenir d'exprimer leur point de vue et d'accéder à certaines sources d'information si elles craignent que les données sur leurs activités ainsi collectées ne soient utilisées par des outils basés sur l'IA servant à prendre des décisions à leur sujet (par exemple, le recrutement ou la promotion à un nouveau poste).

26.1. *Concentration du pouvoir entre les mains d'entreprises du numérique*

27. Une inquiétude plus générale, en ce qui concerne l'impact des technologies d'IA sur la démocratie, tient au risque d'une concentration sans précédent et hors de tout contrôle des données, de l'information et du pouvoir entre les mains d'un petit groupe de grandes entreprises du numérique qui élaborent et détiennent les algorithmes, ainsi qu'à la centralisation d'internet en tant que tel. Ces grandes entreprises ne fonctionnent plus comme de simples canaux de communication entre les individus et les institutions: elles jouent elles-mêmes un rôle de plus en plus important, définissent l'ordre du jour, façonnent et transforment les modèles sociaux et politiques. 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. Il apparaît donc clairement nécessaire de réduire l'influence des grandes entreprises sur la prise de décision démocratique. En outre, les collaborations public-privé dans le domaine de l'IA et de son utilisation dans des domaines sensibles tels que l'ordre public, la sécurité et le renseignement et le contrôle aux frontières, etc., mais aussi dans la recherche-développement brouillent les délimitations entre, d'un côté, les responsabilités, les processus et les institutions des États démocratiques et, de l'autre, les intérêts des sociétés privées.²³

3.4. *Surveillance de masse et renforcement de l'autoritarisme*

28. L'IA pourrait faciliter les abus de pouvoir par les États et les organismes publics. En effet, cette technologie à double usage peut être déployée pour porter atteinte à des droits fondamentaux importants qui sont constitutifs du bon fonctionnement des démocraties.²⁴ Les technologies évoluées de surveillance fondées sur l'IA, telles que la reconnaissance faciale, vocale et de mouvements, associées aux réseaux de caméras de surveillance dans les lieux publics, permettent d'observer les individus dans le monde réel. Ces capacités de l'IA étaient au premier plan lors de la pandémie de covid-19 (voir note en bas de page n° 6). De même qu'avec les progrès accomplis par d'autres technologies, les outils de surveillance associés à l'analyse prédictive peuvent être utilisés aussi bien pour améliorer la sûreté, la sécurité ou le contrôle de la circulation que pour permettre aux gouvernements de contrôler de grands rassemblements et d'anticiper la formation de manifestations et d'émeutes²⁵. Ainsi, les mesures de surveillance généralisée par le biais de l'IA menacent notre droit à la vie privée et à la liberté d'expression.

3.5. *IA et prise de décision politique*

29. Ces dernières décennies, nous avons assisté à un certain degré de dépolitisation de la prise de décision. Une enquête de 2019 sur les attitudes des Européens envers la technologie a montré qu'un quart des personnes interrogées préféreraient que les décisions politiques soient prises par l'IA plutôt que par les responsables politiques – sans tenir compte du fait que les décisions de l'IA reposent sur une analyse statistique des données disponibles et non sur une relation de cause à effet entre un événement et une décision²⁶. Cet état d'esprit témoigne probablement de la défiance croissante des citoyens à l'égard des gouvernements et des responsables politiques, et démontre une remise en cause du modèle occidental de démocratie représentative.

22. «Principled artificial intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI», Berkman Klein Center for Internet Society, Université d'Harvard, 2020.

23. Contribution de M^{me} Birgit Schippers, Maître de conférences en politique, Collège universitaire St Mary, Belfast, à l'échange de vues en Commission le 2 octobre 2019, [AS/Pol/Inf \(2019\) 10](#).

24. Idem.

25. «Algorithms and human rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications», Conseil de l'Europe, 2017.

26. Center for the Governance of Change, European Tech Insights 2019. «Mapping European attitudes to technological change and its governance».

30. Une telle attitude peut engendrer la passivité des électeurs plutôt que les encourager à remettre en question les motifs des décisions prises et à comprendre que les décisions sont motivées par des intérêts (au sens noble du terme) ou des valeurs qui ne doivent pas nécessairement être incontestables, absolus, ni «scientifiques» pour pouvoir être considérés comme valables. Ce phénomène, souvent poussé à l'extrême, incite à exploiter une forme contemporaine *d'ipse dixit* – le maître l'a dit. Il est difficile, voire impossible, pour un être humain de remonter à l'origine d'une décision prise par une machine ou de reproduire le raisonnement qui a mené à cette décision. Lorsque des algorithmes impénétrables de type «boîte noire» prennent des décisions qui ont des répercussions sur la vie des personnes, en particulier dans des domaines sensibles, les valeurs démocratiques que sont la transparence, la responsabilité et l'égalité, et le principe de légitimité démocratique sont gravement menacés. Le biais de l'automatisation, qui désigne l'acceptation des décisions générées par la machine, sans contrôle par l'homme ou avec un contrôle limité, menace la transparence et l'obligation de rendre des comptes. Les systèmes fondés sur l'IA peuvent produire des résultats profondément injustes, inéquitables et même discriminatoires qui mettent en danger les processus et les institutions démocratiques, et qui ont des effets préjudiciables sur les personnes, en particulier sur celles qui appartiennent à des communautés vulnérables.²⁷

31. Amener la société à accepter des choix non sur la base d'une réflexion critique, mais selon les diktats de l'autorité est profondément injuste, et donc néfaste, compte tenu de l'impossibilité d'établir de manière incontestable quelles sont les sources qui devraient faire autorité aux yeux de l'opinion publique. Les technologies assistées par l'IA peuvent faire croire aux gens qu'ils font leurs propres choix alors qu'en réalité ils ne font que suivre des schémas. L'IA peut ainsi être utilisée comme un instrument aux fins d'abuser de la démocratie directe. Plus généralement, la prise de décision politique assistée par l'IA peut conduire à terme à établir une forme de démocratie automatisée et à priver l'homme de son autonomie dans les processus politiques. La définition des objectifs politiques ne doit pas être laissée aux algorithmes; elle doit rester du ressort des hommes et des femmes qui sont investis de la légitimité démocratique et assument des responsabilités politiques et juridiques.

32. Pour résumer, la technologie fondée sur l'IA offre les moyens d'interférer sur les procédures et sur les processus des démocraties et de fragiliser les institutions démocratiques. L'utilisation de l'IA par les États, les organismes publics et les sociétés privées, et son risque potentiel d'abus font peser une vraie menace sur les institutions, les processus et les normes de nos démocraties fondées sur des droits.²⁸ En vue de contrer cette menace, «nous avons besoin d'un cadre qui garantisse que cette technologie soit développée et déployée dans le plein respect non seulement de nos valeurs, mais aussi de notre droit écrit, de nos droits fondamentaux, de l'État de droit, de la démocratie et de l'ensemble du droit dérivé [...] Le principe doit être que rien de ce qui est réalisé de façon automatique à l'aide de l'intelligence artificielle ne peut être légal s'il est illégal lorsque c'est une personne qui s'en charge.»²⁹

4. Initiatives en cours pour l'instauration d'un cadre réglementaire applicable à l'intelligence artificielle

33. Comme indiqué ci-dessus, des organisations nationales et internationales tentent de répondre aux inquiétudes qui entourent l'utilisation de l'IA. Cette section présente un aperçu des mesures prises par les principales organisations internationales ainsi que quelques initiatives nationales, pour établir un cadre réglementaire applicable à l'IA.

4.1. Organisation des Nations Unies

34. L'UNESCO³⁰ élabore en ce moment le premier instrument normatif mondial sur l'éthique de l'intelligence artificielle, suite à la décision prise par la Conférence générale de l'UNESCO lors de sa 40^e session en novembre 2019. Ce processus inclusif et pluridisciplinaire devrait comporter la consultation d'un large éventail de parties prenantes, notamment la communauté scientifique, des personnes d'origines culturelles et de perspectives éthiques différentes, des groupes minoritaires, la société civile, le gouvernement et le secteur privé. La première version du projet de texte de la recommandation a été publiée en ligne et ouverte à la consultation. L'inclusion, la fiabilité, la protection de l'environnement et de la vie privée font partie des principes énoncés par cette recommandation.

27. Contribution par M^{me} Birgit Schippers, op. cit.

28. *Ibidem*.

29. Contribution par M. Paul Nemitz, Conseiller principal, «Stratégies pour des politiques de justice transversales et les actions en justice», Direction générale de la justice et des consommateurs, Commission européenne, à l'échange de vues en Commission, 2 octobre 2019, AS/Pol/Inf (2019) 10.

30. <https://en.unesco.org/artificial-intelligence>.

4.2. Union européenne

35. Comme cela est décrit dans la stratégie de l'Union européenne en matière d'IA, la Commission européenne suit une démarche en trois étapes: définir les exigences clés pour une IA digne de confiance, lancer une phase pilote à grande échelle pour recueillir les réactions et retours d'informations des parties prenantes, et établir un consensus international pour une IA centrée sur l'humain. La Commission a présenté sept éléments essentiels pour parvenir à une IA digne de confiance:

- Facteur humain et contrôle humain: les systèmes d'IA devraient être les vecteurs de sociétés équitables en se mettant au service de l'humain et des droits fondamentaux, sans restreindre ou dévoyer l'autonomie humaine;
- Robustesse et sécurité: une IA digne de confiance nécessite des algorithmes suffisamment sûrs, fiables et robustes pour gérer les erreurs ou les incohérences dans toutes les phases du cycle de vie des systèmes d'IA;
- Respect de la vie privée et gouvernance des données: il faut que les citoyens aient la maîtrise totale de leurs données personnelles et que les données les concernant ne soient pas utilisées contre eux à des fins préjudiciables ou discriminatoires;
- Transparence: la traçabilité des systèmes d'IA doit être assurée;
- Diversité, non-discrimination et équité: les systèmes d'IA devraient prendre en compte tout l'éventail des capacités, aptitudes et besoins humains, et leur accessibilité devrait être garantie;
- Bien-être sociétal et environnemental: les systèmes d'IA devraient être utilisés pour soutenir des évolutions sociales positives et renforcer la durabilité et la responsabilité écologique;
- Responsabilisation: il convient de mettre en place des mécanismes pour garantir la responsabilité à l'égard des systèmes d'IA et de leurs résultats, et de soumettre les responsables à une obligation de rendre des comptes.

36. Enfin, dans son Livre blanc³¹ présenté le 19 février 2020, la Commission européenne envisage la création d'un cadre pour une IA digne de confiance, fondée sur l'excellence et la confiance. En partenariat avec les secteurs privé et public, le cadre vise à mobiliser des ressources tout au long de la chaîne de valeur et à créer les incitations appropriées pour accélérer l'adoption de solutions fondées sur l'IA, notamment par les petites et moyennes entreprises. Cela inclut une collaboration avec les États membres et le milieu de la recherche pour attirer et garder les talents. Étant donné que les systèmes d'IA peuvent être complexes et comporter des risques importants dans certains contextes, il est essentiel d'instaurer la confiance. Des règles claires doivent s'appliquer aux systèmes d'IA présentant un risque élevé, sans exercer une pression trop importante sur les systèmes moins critiques.

4.3. OCDE

37. En mai 2019, l'OCDE a adopté la Recommandation sur l'intelligence artificielle³². La Recommandation énonce cinq principes complémentaires fondés sur des valeurs pour une approche responsable en appui d'une IA digne de confiance:

- l'IA devrait servir les intérêts des individus et de la planète en favorisant la croissance inclusive, le développement durable et le bien-être;
- les systèmes d'IA devraient être conçus de manière à respecter l'État de droit, les droits humains, les valeurs démocratiques et la diversité, et être assortis de garanties appropriées – permettant par exemple l'intervention humaine en cas de besoin –, afin de tendre vers une société juste et équitable;
- il convient d'assurer la transparence et la divulgation responsable des informations liées aux systèmes d'IA, pour que les individus sachent lorsqu'ils interagissent avec de tels systèmes et puissent en contester les résultats;
- les systèmes d'IA devraient être robustes, sûrs et sécurisés tout au long de leur cycle de vie; les risques connexes éventuels devraient être évalués et gérés en permanence;

31. «Livre blanc sur l'intelligence artificielle – une approche européenne axée sur l'excellence et la confiance». Un vaste processus de consultation sur le Livre blanc a été lancé par la Commission européenne après sa publication, et un examen des différentes contributions est en cours afin d'affiner les options politiques et législatives qui seront retenues par la Commission à l'avenir.

32. OECD/LEGAL/0449, adopté le 22 mai 2019.

- les organisations et les individus chargés de développer, de déployer ou d'exploiter des systèmes d'IA devraient être responsables de leur bon fonctionnement, dans le respect des principes susmentionnés.

4.4. Actions menées au niveau national

38. Les autorités nationales de nombreux pays ont élaboré des stratégies et des politiques pour promouvoir et réglementer l'IA. En voici quelques exemples parmi d'autres:

- Finlande: un groupe de pilotage spécialisé mis en place par le gouvernement a publié en 2017 une stratégie nationale en matière d'IA³³, dont un chapitre est consacré aux questions d'éthique; la Finlande a été le premier pays de l'Union Européenne à se doter d'une telle stratégie;
- Allemagne: l'Allemagne a publié une stratégie sur l'IA en 2018³⁴, qui est axée sur la nécessité de promouvoir la recherche-développement tout en veillant à ce que l'IA se développe de manière socialement responsable;
- Fédération de Russie: la Douma a récemment adopté la stratégie nationale pour le développement de l'IA à l'horizon 2030. Cette stratégie mentionne notamment les principes sur lesquels reposent le développement et l'utilisation de l'IA (protection des droits humains et des libertés fondamentales, sécurité et transparence). Elle insiste également sur la nécessité de sensibiliser le public et de créer un système intégré de réglementation des relations sociales dans le contexte du développement et de l'utilisation des technologies d'IA³⁵;
- Royaume-Uni: l'accord sectoriel sur l'IA (*AI Sector Deal*)³⁶ et le rapport de la Chambre des Lords sur l'IA³⁷ (publiés tous deux en avril 2018) soulignent l'importance d'une réflexion approfondie et d'une politique solide en ce qui concerne l'éthique de l'IA.

5. Les travaux du Conseil de l'Europe

39. Le Conseil de l'Europe, en tant que l'une des principales organisations en matière de droits humains, joue un rôle important dans la promotion d'une IA qui respecte les droits humains. Le 13 février 2019, le Comité des Ministres a adopté une importante Déclaration sur les capacités de manipulation des processus algorithmiques³⁸. Il a invité les États membres à combattre le risque que les individus ne soient plus en mesure de se forger leur propre opinion et de prendre des décisions indépendamment des systèmes automatisés, voire qu'ils fassent l'objet de manipulations liées à l'utilisation des technologies numériques de pointe et des techniques de microciblage. Notant que les outils d'apprentissage automatique sont de plus en plus capables non seulement de prédire les choix, mais aussi d'influencer les émotions et les pensées, parfois de façon subliminale, le Comité des Ministres a encouragé les États à assumer leur responsabilité pour combattre cette menace grandissante en prenant des mesures législatives appropriées et proportionnées contre les ingérences illégitimes, et en permettant aux utilisateurs de développer des compétences clés dans la culture numérique.

40. Les 26 et 27 février 2019, la Conférence d'Helsinki intitulée «Maîtriser les règles du jeu – L'impact du développement de l'intelligence artificielle sur les droits humains, la démocratie et l'État de droit», organisée par le Conseil de l'Europe et la Présidence finlandaise du Comité des Ministres, a souligné que «des mécanismes de surveillance effectifs et des structures de contrôle démocratiques doivent être mis en place pour tout ce qui touche à la conception, au développement et au déploiement des systèmes d'IA», et qu'«un processus démocratique fonctionne si la population est informée de manière indépendante et si l'on encourage des débats ouverts et inclusifs. Il convient de la sensibiliser davantage aux risques et aux bénéfices potentiels de l'IA et de développer les nouvelles compétences nécessaires. Il faut en outre stimuler la confiance du public dans l'environnement de l'information et les applications de l'IA³⁹.»

33. «Finland's Age of Artificial Intelligence. Turning Finland into a leading country in the application of artificial intelligence. Objective and recommendations for measures,» Publications of the Ministry of Economic Affairs and Employment 47/2017.

34. «Strategie Künstliche Intelligenz der Bundesregierung», novembre 2018.

35. Informations fournies par M. Piotr Tolstoi, membre de la Délégation de la Fédération de Russie, sous la forme d'une note d'information adressée à la rapporteure et datée du 10 juin 2020.

36. Document d'orientation, «AI Sector Deal», mis à jour le 21 mai 2019.

37. «AI in the UK: ready, willing and able?», House of Lords, Select Committee on Artificial Intelligence, Report of Session 2017-19.

38. Decl (13/02/2019)1.

39. Conclusions de la Conférence de haut-niveau «Maîtriser les règles du jeu - L'impact du développement de l'intelligence artificielle sur les droits de l'homme, la démocratie et l'état de droit» (26-27 février 2019).

41. Le 11 septembre 2019, le Comité des Ministres a institué un Comité ad hoc sur l'intelligence artificielle (CAHAI) afin 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, fondé sur les normes du Conseil de l'Europe dans le domaine des droits humains, de la démocratie et de l'État de droit.

42. Dans sa Recommandation CM/Rec(2020)1 aux États membres sur les impacts des systèmes algorithmiques sur les droits humains, adoptée le 8 avril 2020 dans le contexte de la pandémie de covid-19, le Comité des Ministres a publié une série de lignes directrices appelant les gouvernements à veiller à ne pas violer les droits humains par leur propre utilisation, développement ou acquisition de systèmes algorithmiques. De plus, leur pouvoir réglementaire donne à ces derniers l'obligation d'établir des cadres législatifs, réglementaires et de supervision efficaces et prévisibles, capables de prévenir, de détecter, d'interdire et de réparer les violations des droits humains, qu'elles soient imputables à des acteurs publics ou privés.

6. Intelligence artificielle et grandes entreprises: responsabilité et éthique

43. Les démarches éthiques contribuent à créer un langage commun pour discuter et débattre de préoccupations sociales et politiques. Elles apportent aux développeurs, aux employés des entreprises et aux autres parties prenantes un ensemble d'énoncés de valeurs et d'objectifs de haut-niveau sur la base desquels les actions pourront ensuite être évaluées. Elles jouent aussi un rôle éducatif, contribuant souvent à attirer l'attention sur les risques particuliers liés à l'IA, aussi bien au sein d'une institution donnée qu'à l'extérieur, pour le public concerné. Mais ce ne sont pas des outils appropriés pour garantir la mise en œuvre des responsabilités.

44. Comme le souligne un rapport publié par l'institut AI Now⁴⁰ de l'Université de New York en 2019, les entreprises qui créent des solutions basées sur l'IA pour toutes sortes d'activités, depuis le calcul des notes dans l'enseignement jusqu'à l'évaluation du potentiel criminel des migrants, ne sont guère liées que par quelques déclarations éthiques auxquelles elles choisissent elles-mêmes d'adhérer. Tout en reconnaissant que les entreprises font des efforts, les auteurs considèrent que «les cadres qui régissent actuellement l'IA ne permettent pas de définir clairement les responsabilités». «À mesure que s'accroissent l'omniprésence, la complexité et la puissance de ces systèmes, l'absence de responsabilisation réelle et de contrôle – notamment des garanties de base de responsabilité et de régularité des procédures – est une préoccupation de plus en plus pressante.»

45. En 2018, Sundar Pichai, le PDG de Google, a rendu public un ensemble de sept principes directeurs élaborés dans le but de veiller à ce que les travaux de sa société sur l'IA soient socialement responsables. Ces principes éthiques comprennent l'engagement d'«être utile à la société» et d'«éviter de créer ou d'accroître une partialité injuste». Dans un article du *Financial Times*, Sundar Pichai s'est déclaré favorable à une réglementation de l'IA tout en plaidant pour une approche de bon sens, soutenant que certains secteurs du développement de l'IA, tels que les voitures autonomes et les technologies médicales, nécessitaient des règles sur mesure⁴¹. Il convient également de mentionner que Google a lancé son propre Comité d'éthique indépendant en 2019, mais l'a dissolu moins de deux semaines plus tard à la suite d'une polémique concernant les membres qui y avaient été nommés.

46. D'autres sociétés, dont Microsoft, Facebook et Axon, le fabricant de caméras mobiles pour la police, ont également mis sur pied des comités d'éthique, des conseillers et des équipes. De telles avancées sont encourageantes; il faut se réjouir que ceux qui sont au cœur du développement de l'IA aient déclaré prendre la question de l'éthique au sérieux. Cela dit, comme mentionné plus haut, ces avancées ne constituent pas une base solide qui engagerait la responsabilité des entreprises.

7. Conclusions

47. La démocratie suppose que des personnes ayant des points de vue différents se réunissent pour trouver des solutions communes par le dialogue. Plutôt que créer un espace public et un ordre du jour communs, les plates-formes de communication basées sur l'IA semblent favoriser l'individualisme et la polarisation et conduire à la formation de communautés en ligne fermées, partageant les mêmes points de vue; ce phénomène, qui sape la cohésion sociale et le débat démocratique, contribue à la prolifération du

40. https://ainowinstitute.org/AI_Now_2019_Report.pdf.

41. «Why Google thinks we need to regulate AI», Sundar Pichai, 20 janvier 2020.

discours de haine, au cloisonnement et à la segmentation de la société. Le fait que des segments entiers de la population n'utilisent pas les plates-formes, ceci étant imputable à diverses raisons se rapportant à l'utilisation des nouvelles technologies de l'information et de la communication (c'est à dire genre, âge et origine sociale) doit également être pris en compte dans cette réflexion. Dans l'Union européenne, par exemple, il y a une différence imputable au genre de 11 % en ce qui concerne les compétences numériques, la différence étant encore plus importante en ce qui concerne les compétences numériques de base et plus spécifiquement pour les personnes âgées de plus de 55 ans⁴². Les sociétés privées qui appliquent les règles du marché et non celles de la démocratie n'assument pas la responsabilité d'autoriser l'alimentation du discours de haine et la diffusion des contenus violents.

48. Les technologies fondées sur l'IA perturbent le fonctionnement des institutions et processus démocratiques, et ont une incidence sur le comportement social et politique des citoyens. Tous les algorithmes d'apprentissage automatique s'appuient sur une structure de classification dans laquelle la machine apprend à formuler un ensemble d'hypothèses au sujet de différents lots de données. Comme tous les processus d'apprentissage itératifs, l'apprentissage automatique aussi peut produire des résultats faussement positifs ou faussement négatifs. Si ces erreurs sont courantes dans certains types d'études, la marge d'erreur peut, une fois reportée dans des décisions politiques, conduire à la répression systématique de certains groupes sociaux ou ethniques, à la mise en accusation à tort de suspects ou au profilage systématique non nécessaire de citoyens⁴³.

49. Manifestement, le cadre réglementaire est en retard par rapport aux progrès technologiques. Les principes et politiques d'autoréglementation ne sauraient être les seuls outils d'encadrement de l'IA, car ils n'entraînent pas d'obligation de rendre des comptes. L'Europe doit veiller à ce que la puissance de l'IA soit réglementée et employée au service du bien commun. Il est donc nécessaire de créer un cadre réglementaire pour l'IA, assorti de principes spécifiques fondés sur la protection des droits humains, la démocratie et l'État de droit. Toute réflexion dans ce domaine doit associer l'ensemble des parties prenantes, et en particulier les citoyens et les entreprises privées. Les travaux du CAHAI, qui doivent conduire à terme à l'instauration d'un cadre juridique pour la gouvernance démocratique de l'intelligence artificielle fondé sur les normes du Conseil de l'Europe, doivent être pleinement soutenus et encouragés.

50. Afin d'assurer le respect de l'obligation de rendre des comptes, le cadre juridique à mettre en place devrait prévoir un mécanisme de contrôle indépendant qui garantirait le respect effectif de ses dispositions. Sans un tel mécanisme, les grandes entreprises des technologies de l'information et de la communication continueront simplement à fonctionner comme par le passé et, face à l'énorme pouvoir et à la nature transnationale de ces entreprises, la plupart des États continueraient à fermer les yeux sur les comportements non-conformes, considérés comme un prix à payer nécessaire et acceptable pour la poursuite d'un intérêt vital.

51. Toutefois, ce mécanisme de contrôle ne peut être efficace que s'il est proactif et déclenché au préalable. De fait, alors qu'il serait important de mettre en place des sanctions pour les comportements non-conformes, un mécanisme qui se limiterait à des pénalités et des amendes après coup – qui sont généralement facilement abordables par les grandes entreprises privées, quel que soit leur montant – n'atteindrait pas le résultat escompté. Il est, en effet, souvent très difficile, voire impossible, de rétablir la situation antérieure ou de réparer les dégâts causés par la mise en place et l'utilisation d'une technologie d'IA donnée, aussi contraire à l'éthique et/ou aux droits humains, à la démocratie et à l'État de droit qu'elle puisse être.

52. Un mécanisme de contrôle proactif exige l'existence d'un organe extrêmement compétent (notamment sur le plan technique, juridique et éthique), capable de suivre les nouvelles évolutions de la technologie numérique et d'évaluer avec précision et autorité ses risques et conséquences. Il va sans dire que toutes les parties prenantes concernées devraient prendre part à un tel organe.

53. Autre point encore plus essentiel, le rôle de l'IA dans la modification de l'équilibre des pouvoirs entre les institutions, les acteurs politiques concernés et les organes exécutifs doit faire l'objet de travaux de recherche plus structurés. Compte tenu de l'ampleur des problèmes de légitimité et de souveraineté liés à la délégation de décisions politiques à des algorithmes, le rôle des constitutions, des parlements et des élites politiques en lien avec l'IA doit être étudié en profondeur, en mettant notamment l'accent sur la place que devrait occuper l'autorité politique à l'ère des décisions automatisées⁴⁴.

42. Commission européenne, «Women in digital scoreboard» (2019).

43. Hamid Akin Ünver, «Artificial Intelligence, Authoritarianism and the Future of Political Systems» (The Centre for Economics and Foreign Policy Studies (EDAM), juillet 2018).

44. Hamid Akin Ünver, *op. cit.*

54. Cela ne signifie pas que l'IA ne saurait être un vecteur de bien, ou qu'elle ne peut pas améliorer l'efficacité des politiques ou leur capacité à répondre aux besoins des citoyens. Bien employée, l'IA peut permettre d'élargir la représentation démocratique en décentralisant les systèmes d'information et les plateformes de communication. Elle peut encourager l'autonomie informationnelle des citoyens, améliorer la manière dont ils s'informent sur les processus politiques et les aider à y participer à distance. De même qu'elle peut être utilisée pour accentuer l'opacité et dispenser les responsables d'avoir à rendre des comptes, elle peut aussi améliorer la transparence et contribuer à l'instauration d'une plus grande confiance entre l'État et la société et entre les citoyens eux-mêmes⁴⁵.

55. Pour sa part, le Conseil de l'Europe, en tant qu'organisation internationale de normalisation de premier plan dans le domaine de la démocratie, doit faire œuvre de pionnier en concevant des solutions et des formats permettant de s'assurer que les technologies fondées sur l'IA sont employées pour renforcer la démocratie, au moyen d'assemblées citoyennes, d'agoras électroniques et d'autres formes délibératives et participatives d'engagement des citoyens dans les processus démocratiques.

45. Idem.

Annexe

Intelligence artificielle – Description et principes éthiques

On a tenté à plusieurs reprises de définir le terme «intelligence artificielle» depuis sa première utilisation en 1955. Ces initiatives s'intensifient aujourd'hui, car les organes normatifs, notamment le Conseil de l'Europe, réagissent aux capacités et à l'omniprésence croissantes de l'intelligence artificielle en œuvrant en faveur de son encadrement juridique. Il n'existe cependant toujours pas de définition «technique» ou «juridique» unique qui soit universellement admise⁴⁶. Aux fins du présent rapport, il est toutefois indispensable de décrire cette notion.

À l'heure actuelle, le terme «intelligence artificielle» (IA) désigne en général les systèmes informatiques capables de percevoir et d'extraire des données de leur environnement, puis d'utiliser des algorithmes statistiques pour traiter ces données, afin d'obtenir des résultats qui correspondent à des objectifs prédéterminés. Les algorithmes se composent de règles définies par l'homme ou par l'ordinateur lui-même, qui «forme» l'algorithme en analysant des ensembles de données considérables et continue à affiner ces règles à mesure qu'il reçoit de nouvelles données. Cette méthode, connue sous le nom «d'apprentissage automatique» ou «d'apprentissage statistique», est actuellement la technique la plus utilisée pour les applications complexes; elle est uniquement devenue possible ces dernières années grâce à l'augmentation de la puissance de traitement des ordinateurs et à la disponibilité de données suffisantes. «L'apprentissage en profondeur» représente une forme particulièrement avancée d'apprentissage automatique, qui utilise plusieurs couches de «réseaux neuronaux artificiels» pour traiter les données. L'analyse ou la compréhension intégrale par l'homme des algorithmes développés par ces systèmes n'est pas toujours possible; aussi sont-ils parfois qualifiés de «boîtes noires» (il arrive que ce terme désigne également, mais pour une raison différente, les systèmes d'IA propriétaires protégés par les droits de propriété intellectuelle).

Les formes actuelles d'IA sont toutes «restreintes», c'est-à-dire affectées à une tâche unique définie. L'IA «restreinte» est également qualifiée parfois de «faible», même si les systèmes modernes de reconnaissance faciale, de traitement du langage naturel, de conduite autonome et de diagnostic médical, par exemple, sont incroyablement sophistiqués et effectuent certaines tâches complexes avec une rapidité et une précision étonnantes. «L'intelligence artificielle générale», parfois qualifiée d'IA «forte», qui est capable d'exécuter toutes les fonctions du cerveau humain, est encore à réaliser. La «super-intelligence artificielle» désigne un système dont les capacités dépassent celles du cerveau humain.

Comme le nombre de domaines dans lesquels les systèmes d'intelligence artificielle sont appliqués est en augmentation, puisqu'ils se propagent dans des domaines qui peuvent avoir un impact important sur les droits individuels et les libertés, ainsi que sur les systèmes démocratiques et l'État de droit, la dimension éthique de ce phénomène a fait l'objet d'une attention croissante et de plus en plus urgente.

Un large éventail d'acteurs a formulé de nombreuses propositions d'ensemble de principes éthiques qui devraient être appliqués aux systèmes d'IA. Ces propositions sont rarement identiques et diffèrent à la fois dans les principes qu'elles énoncent et par la manière dont elles définissent ces principes. Les études montrent que la teneur essentielle des principes éthiques qui devraient être appliqués aux systèmes d'IA fait néanmoins l'objet d'un large consensus; c'est notamment le cas des principes suivants⁴⁷:

- **Transparence.** Le principe de transparence peut faire l'objet d'une interprétation élargie, de manière à englober l'accessibilité et l'explicabilité d'un système d'IA, en d'autres termes la possibilité donnée à un individu de comprendre le fonctionnement de ce système et le mode de production de ses résultats.
- **Justice et équité.** Ce principe englobe la non-discrimination, l'impartialité, la cohérence et le respect de la diversité et du pluralisme. Il suppose également que la personne à laquelle est appliqué un système d'IA puisse en contester les résultats, disposer d'une voie de recours et obtenir réparation.
- **Responsabilité.** Ce principe englobe le fait d'exiger qu'un être humain soit responsable de toute décision qui a des conséquences sur les droits et libertés individuels, et que l'obligation de rendre des comptes et la responsabilité juridique de ces décisions soient définies. Il est donc étroitement lié au principe de justice et d'équité.

46. Pour une vue d'ensemble élargie des tentatives de définition de «l'intelligence artificielle», voir *AI Watch: Defining Artificial Intelligence – Towards an operational definition and taxonomy of artificial intelligence*, Samoili S., López Cobo M., Gómez E., De Prato G., Martínez-Plumed F. et Delipetrev B., European Commission Joint Research Centre, 2020.

47. Voir *Lignes directrices sur l'éthique en matière d'IA: situation en Europe et dans le monde*, rapport provisoire commandé par le Comité ad hoc sur l'intelligence artificielle (CAHAI) du Conseil de l'Europe, Ienca M. et Vayena E., mars 2020.

- *Sûreté et sécurité.* Ce principe suppose que les systèmes d'IA fassent preuve de solidité, de sécurité contre toute ingérence extérieure et de sûreté contre la commission d'actes involontaires, conformément au principe de précaution.
- *Respect de la vie privée.* Si le respect des droits de l'homme en général peut être considéré comme inhérent aux principes de justice et d'équité, de sûreté et de sécurité, le droit au respect de la vie privée est particulièrement important chaque fois qu'un système d'IA procède au traitement de données à caractère personnel ou privé. Les systèmes d'IA doivent par conséquent respecter les normes contraignantes du Règlement général sur la protection des données (RGPD) de l'Union Européenne et de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel du Conseil de l'Europe (STE n° 108, et sa version actualisée, la Convention 108+, STCE n° 223), le cas échéant.

La mise en œuvre effective des principes éthiques applicables aux systèmes d'IA exige une approche intégrée de l'éthique, notamment une évaluation de leur impact sur les droits de l'homme, de manière à garantir le respect des normes établies. Il ne suffit pas que ces systèmes soient conçus uniquement sur la base de normes techniques et que des éléments soient ajoutés à un stade ultérieur pour tenter de faire respecter les principes éthiques.

Dans quelle mesure le respect de ces principes doit-il être intégré dans des systèmes particuliers d'IA? Cela dépend des utilisations prévues et prévisibles de ces systèmes: plus leur impact sur l'intérêt général et les droits et libertés individuels est important, plus les garanties doivent être strictes. La réglementation éthique peut donc être mise en œuvre de différentes manières, depuis les chartes internes volontaires pour les domaines les moins sensibles jusqu'aux normes juridiques contraignantes pour les plus délicats. Dans tous les cas, il importe qu'elle prévoie des mécanismes de contrôle indépendants selon le niveau de réglementation.

Ces principes essentiels portent sur les systèmes d'IA et leur environnement immédiat. Ils n'ont pas vocation à être exhaustifs ni à exclure des préoccupations éthiques plus générales, telles que la démocratie (participation pluraliste des citoyens à l'élaboration de normes éthiques et réglementaires), la solidarité (qui admet les différences de point de vue des divers groupes) ou la durabilité (préservation de l'environnement de la planète).