COMMISSION
EUROPÉENNE

Bruxelles, le 8.4.2019
COM(2019) 168 final

COMMUNICATION DE LA COMMISSION AU PARLEMENT EUROPÉEN, AU CONSEIL, AU COMITÉ ÉCONOMIQUE ET SOCIAL EUROPÉEN ET AU COMITÉ DES RÉGIONS

Renforcer la confiance dans l'intelligence artificielle axée sur le facteur humain

1. INTRODUCTION – LA STRATÉGIE EUROPÉENNE EN MATIÈRE D'INTELLIGENCE ARTIFICIELLE (IA)

L'intelligence artificielle (IA) est porteuse de transformations positives pour notre monde: elle peut améliorer les soins de santé, réduire la consommation d'énergie, rendre les voitures plus sûres et permettre aux agriculteurs d'utiliser plus efficacement l'eau et les ressources naturelles. Elle peut être utilisée pour prédire les changements climatiques et environnementaux, améliorer la gestion des risques financiers et fabriquer avec moins de déchets des produits adaptés à nos besoins. L'IA peut également contribuer à la détection des fraudes et des menaces informatiques et permet aux services répressifs de lutter plus efficacement contre la criminalité.

L'IA peut profiter à l'ensemble de la société et de l'économie. Le développement et l'utilisation de cette technologie stratégique sont aujourd'hui en plein essor dans le monde entier. Cela étant dit, l'IA crée également de nouveaux défis pour l'avenir du travail et soulève des questions d'ordre juridique et éthique.

Pour relever ces défis et exploiter au mieux les perspectives qu'offre l'IA, la Commission a publié une stratégie européenne en avril 2018[1]**.** Cette stratégie place les citoyens au centre du développement de l'IA — **IA axée sur le facteur humain**. Il s'agit d'une approche en trois volets visant à renforcer la capacité technologique et industrielle de l'UE et à intensifier le recours à l'IA dans tous les secteurs de l'économie, à faire face aux changements socio-économiques et à garantir l'existence d'un cadre éthique et juridique approprié.

Pour atteindre les résultats visés par la stratégie en matière d'IA, **la Commission a élaboré, en association avec les États membres, un plan coordonné dans le domaine de l'intelligence artificielle**[2], qu'elle a présenté en décembre 2018 et qui a pour but de créer des synergies, d'assurer la mise en commun des données (matière première de nombreuses applications d'IA) et d'accroître les investissements conjoints. L'objectif est de favoriser la coopération transfrontière et de mobiliser tous les acteurs pour augmenter les investissements publics et privés afin qu'ils atteignent **au moins 20 milliards d'EUR** par an au cours de la prochaine décennie[3]. La Commission a doublé ses investissements dans l'IA dans le cadre d'Horizon 2020 et prévoit d'investir 1 milliard d'EUR par an au titre du programme «Horizon Europe» et du programme pour une Europe numérique afin de soutenir, notamment, la création d'espaces de données communs dans les domaines de la santé, des transports et de l'industrie manufacturière, les installations expérimentation à grande échelle, telles que les hôpitaux intelligents et les infrastructures pour véhicules automatisés, ainsi que la mise au point d'un programme stratégique de recherche.

Pour mettre en œuvre ce programme stratégique commun de recherche, d'innovation et de déploiement, la Commission a intensifié son **dialogue avec tous les acteurs concernés** du secteur privé, des instituts de recherche et des pouvoirs publics. Le nouveau programme pour une Europe numérique jouera également un rôle déterminant pour offrir aux petites et moyennes entreprises dans tous les États membres la possibilité d'accéder à l'IA au travers de pôles d'innovation numérique, d'installations d'essai et d'expérimentation renforcées, d'espaces de données et de programmes de formation.

---

1 COM(2018) 237.

2 COM(2018) 795.

3 Dans cette perspective, la Commission a proposé que, pendant la prochaine période de programmation 2021-2027, l'Union consacre au moins 1 milliard d'EUR par an au financement d'investissements dans l'IA au titre des programmes «Horizon Europe» et «Europe numérique».

Forte de sa réputation en matière de sécurité et de grande qualité des produits, l'Europe pratique, à l'égard de l'IA, une approche éthique qui renforce la confiance des citoyens dans le développement du numérique et s'efforce de créer un avantage concurrentiel pour les entreprises européennes d'IA. La présente communication a pour objectif de lancer une vaste phase pilote avec la participation d'un maximum de parties prenantes afin de tester la mise en pratique des orientations éthiques à suivre pour le développement et l'utilisation de l'IA.

2. RENFORCER LA CONFIANCE DANS L'IA AXÉE SUR LE FACTEUR HUMAIN

Il ressort clairement de la stratégie européenne en matière d'IA et du plan coordonné que la **confiance est une condition indispensable pour une approche de l'IA axée sur le facteur humain**: l'IA n'est pas une fin en soi mais un outil qui doit se mettre au service des citoyens pour accroître, en définitive, le bien-être de l'être humain. Pour atteindre cet objectif, **il faut que l'IA soit digne de confiance**. Les valeurs sur lesquelles reposent nos sociétés doivent être pleinement intégrées dans le développement de l'IA.

L'Union est fondée sur **les valeurs de respect de la dignité humaine, de liberté, de démocratie, d'égalité, d'état de droit, ainsi que de respect des droits de l'homme**, y compris des droits des personnes appartenant à des minorités[4]. Ces valeurs sont communes aux sociétés de tous les États membres caractérisées par le pluralisme, la non-discrimination, la tolérance, la justice, la solidarité et l'égalité. En outre la **charte des droits fondamentaux de l'UE** rassemble dans un texte unique l'ensemble des droits individuels, civiques, politiques, économiques et sociaux dont jouissent les citoyens de l'Union.

Elle possède un **cadre réglementaire bien établi**, qui constituera la référence mondiale d'une IA axée sur le facteur humain. Le règlement général sur la protection des données (RGPD) garantit un niveau élevé de protection des données à caractère personnel et impose la mise en œuvre de mesures visant à garantir la protection des données dès la conception et par défaut[5]. Le règlement relatif à la libre circulation des données à caractère non personnel supprime les obstacles à cette libre circulation et garantit que toutes les catégories de données peuvent être traitées partout en Europe. Le règlement sur la cybersécurité, adopté récemment, contribuera à renforcer la confiance à l'égard du monde numérique, objectif que poursuit également la proposition de règlement «vie privée et communications électroniques»[6].

L'IA s'accompagne toutefois de nouveaux défis car elle permet aux machines d'«apprendre» et de prendre et exécuter des décisions sans intervention humaine. D'ici peu, de nombreux types de biens et de services, depuis les téléphones intelligents jusqu'aux voitures automatisées, en passant par les robots et les applications en ligne, seront munis d'office de ce type de fonctionnalité. Or il peut arriver que les décisions prises par des algorithmes soient fondées sur des données incomplètes, donc non fiables, altérées par des cyberattaques ou encore entachées de parti pris ou simplement erronées. Par conséquent, appliquer sans

[4] L'UE est en outre partie à la convention des Nations unies relative aux droits des personnes handicapées.

[5] Règlement (UE) 2016/679. Le règlement général sur la protection des données (RGPD) garantit la libre circulation des données à caractère personnel à l'intérieur de l'Union. Il contient des dispositions sur la prise de décision fondée exclusivement sur le traitement automatisé, dont le profilage. Les personnes concernées ont le droit d'être informées de l'existence d'une prise de décision automatisée et de recevoir des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour elles. Elle ont également le droit, dans de tels cas, d'obtenir une intervention humaine, d'exprimer leur point de vue et de contester la décision.

[6] COM(2017) 10.

discernement la technologie au fur et à mesure de son développement créerait des problèmes et dissuaderait les citoyens de l'accepter ou de l'utiliser.

Il faut, au contraire, développer la technologie de l'IA en plaçant les citoyens en son centre, ce qui lui vaudrait la confiance du public. Cela suppose que les applications de l'IA soient non seulement conformes à la législation, mais respectent également des principes éthiques, et que leur mise en œuvre n'entraîne pas de préjudice involontaire. La diversité au regard du sexe, de la race ou de l'origine ethnique, *de la religion* ou des convictions, *du handicap* et de l'âge devrait être assurée à chaque étape du développement de l'IA. Les applications de l'IA devraient donner plus d'autonomie aux citoyens et respecter leurs droits fondamentaux. Elles devraient augmenter les capacités des personnes, non les remplacer, et s'ouvrir aux personnes handicapées.

Il est, dès lors, nécessaire d'élaborer des **lignes directrices en matière d'éthique** qui s'appuient sur le cadre réglementaire existant et qui devraient être appliquées par les développeurs, les fournisseurs et les utilisateurs d'IA dans le marché intérieur, en créant des conditions de concurrence équitables sur le plan éthique dans tous les États membres. C'est pourquoi la Commission a mis en place un **groupe d'experts de haut niveau sur l'IA**[7], qui représente un large éventail de parties prenantes, et lui a confié la mission d'élaborer des lignes directrices en matière d'éthique dans le contexte de l'IA et de produire un ensemble de recommandations concernant plus largement la politique en matière d'IA. Dans le même temps, l'**Alliance européenne pour l'IA**[8], une plateforme ouverte réunissant des parties prenantes de divers horizons et comptant plus de 2 700 membres, a été créée afin d'élargir les contributions aux travaux du groupe d'experts de haut niveau sur l'IA.

Le groupe d'experts de haut niveau sur l'IA a publié un premier projet de lignes directrices en matière d'éthique en décembre 2018. À la suite d'une **consultation des parties prenantes**[9] et de plusieurs **réunions avec des représentants des États membres**[10], il a soumis un document révisé à la Commission en mars 2019. Dans leurs réactions communiquées à ce jour, les parties prenantes saluent globalement le pragmatisme des lignes directrices et les conseils concrets qu'elles offrent aux développeurs, fournisseurs et utilisateurs d'IA pour gagner la confiance.

2.1. Lignes directrices pour une IA digne de confiance élaborées par le groupe d'experts de haut niveau sur l'IA

Les lignes directrices élaborées par le groupe d'experts de haut niveau sur l'IA, auxquelles se réfère la présente communication[11], s'appuient en particulier sur les travaux du Groupe

---

7 https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

8 https://ec.europa.eu/digital-single-market/en/european-ai-alliance

9 Cette consultation a permis de recueillir les commentaires de 511 organisations, associations, entreprises, instituts de recherche, particuliers et autres. Un résumé des retours d'information est disponible à l'adresse suivante: https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf

10 Les États membres ont fait bon accueil aux travaux du groupe d'experts. Ainsi, dans ses conclusions adoptées le 18 février 2019, le Conseil a notamment pris acte de la publication prochaine des lignes directrices en matière d'éthique et a déclaré soutenir les efforts déployés par la Commission pour porter l'approche éthique de l'UE sur la scène mondiale. https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/fr/pdf

11 https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

européen d'éthique des sciences et des nouvelles technologies et de l'Agence des droits fondamentaux.

Les lignes directrices posent l'hypothèse que, pour parvenir à une «IA digne de confiance», trois éléments sont nécessaires: 1) elle doit respecter la législation, 2) elle doit respecter des principes éthiques et 3) elle doit être robuste.

Sur la base de ces trois éléments et des valeurs européennes rappelées au point 2, les lignes directrices définissent sept exigences essentielles auxquelles les applications de l'IA devraient répondre pour être considérées comme dignes de confiance. Les lignes directrices contiennent également une liste d'évaluation destinée à vérifier si ces exigences sont remplies.

Les sept exigences essentielles sont les suivantes:

- Facteur humain et contrôle humain
- Robustesse technique et sécurité
- Respect de la vie privée et gouvernance des données
- Transparence
- Diversité, non-discrimination et équité
- Bien-être sociétal et environnemental
- Responsabilisation

Bien que ces exigences soient censées concerner tous les systèmes d'IA dans différents contextes et branches d'activité, il faut tenir compte du contexte spécifique dans lequel elles s'appliquent pour en assurer la mise en œuvre concrète et proportionnée, en adoptant une approche fondée sur les incidences. À titre d'illustration, une application de l'IA suggérant à mauvais escient la lecture d'un livre est beaucoup moins dangereuse qu'une application posant un mauvais diagnostic de cancer et pourrait donc faire l'objet d'une surveillance moins stricte.

Les lignes directrices élaborées par le groupe d'experts de haut niveau sur l'IA ne sont pas contraignantes et, à ce titre, ne créent pas de nouvelles obligations juridiques. Cependant, il va de soi que beaucoup de dispositions du droit de l'Union (souvent spécifiques à un usage ou à un domaine donné) intègrent déjà une ou plusieurs de ces exigences essentielles, notamment en matière de sécurité, de protection des données à caractère personnel, de respect de la vie privée ou de protection de l'environnement.

La Commission se félicite des travaux du groupe d'experts de haut niveau sur l'IA et estime qu'ils apportent une contribution très importante à ses travaux d'élaboration des politiques.

2.2. Exigences essentielles pour une IA digne de confiance

La Commission adhère aux exigences essentielles suivantes pour une IA digne de confiance, qui reposent sur des valeurs européennes. Elle encourage les parties prenantes à appliquer les exigences et à expérimenter la liste d'évaluation de leur réalisation afin de créer un climat de confiance propice au développement et à l'utilisation de l'IA. La Commission invite les parties prenantes à faire part de leur expérience afin de déterminer si cette liste d'évaluation proposée dans les lignes directrices demande des ajustements supplémentaires.

I. Facteur humain et contrôle humain

Les systèmes d'IA devraient aider les individus à prendre de meilleures décisions et à faire des choix plus éclairés en rapport avec leurs objectifs. Ils devraient être les vecteurs d'une société prospère et équitable en se mettant au service de l'humain et **des droits fondamentaux**, sans diminuer, restreindre ou dévoyer l'autonomie humaine. Le **bien-être** général **de l'utilisateur** doit être au cœur des fonctionnalités du système.

Le contrôle humain permet d'éviter qu'un système d'IA ne mette en péril l'autonomie humaine ou ne provoque d'autres effets néfastes. En fonction du système d'IA concerné et de son domaine d'application, il conviendrait de prévoir des **mesures d'encadrement** d'intensité appropriée, portant notamment sur l'adaptabilité, la précision et l'explicabilité des systèmes fondés sur l'IA[12]. Le **contrôle** peut être assuré en recourant à des mécanismes de gouvernance tels que les approches dites «human-in-the-loop» (l'humain intervient dans le processus), «human-on-the-loop» (l'humain supervise le processus) ou «human-in-command» (l'humain reste aux commandes)[13]. Il convient de veiller à ce que les autorités publiques soient en mesure d'exercer leurs pouvoirs de contrôle conformément à leur mandat. Toutes choses étant égales par ailleurs, moins un être humain peut exercer de contrôle sur un système d'IA, plus il faut approfondir les essais et renforcer la gouvernance.

II. Robustesse technique 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 du système d'IA et pour apporter une réponse adéquate à des résultats erronés. Les systèmes d'IA doivent être **fiables**, suffisamment sûrs pour **résister** à la fois aux attaques directes et aux tentatives plus subtiles de manipulation des données ou des algorithmes proprement dits, et ils doivent offrir des **solutions de secours** en cas de problèmes. Leurs décisions doivent être **précises**, ou au moins correspondre à leur niveau de précision, et leurs résultats doivent être **reproductibles**.

En outre, les systèmes d'IA devraient intégrer des mécanismes de sécurité par conception et de sûreté **permettant d'en vérifier l'innocuité** à chaque étape, en tenant compte de la sécurité physique et mentale de toutes les personnes concernées. Cela comprend la nécessité de réduire au minimum et, si possible, de neutraliser les effets non désirés ou les dysfonctionnements du système. Des processus devraient être mis en place pour définir et évaluer les risques potentiels liés à l'utilisation de systèmes d'IA pour un éventail de domaines d'application.

---

[12] Le règlement général sur la protection des données confère aux personnes le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé lorsque cela produit des effets juridiques sur les utilisateurs ou les affecte de manière significative de façon similaire (article 22 du RGPD) .

[13] L'approche «human-in-the-loop» (HITL) désigne une intervention humaine dans chaque cycle de décision du système, ce qui, dans de nombreux cas, n'est ni possible ni souhaitable. L'approche «human-on-the-loop» (HOTL) désigne une capacité d'intervention humaine dans le cycle de conception du système et la surveillance du fonctionnement du système. L'approche «human-in-command» (HIC) désigne une capacité de contrôle de l'activité globale du système d'IA (y compris de ses incidences économiques, sociétales, juridiques et éthiques au sens large) et la faculté de décider quand et comment utiliser le système dans une situation donnée. Cette faculté peut comprendre la décision de ne pas utiliser le système d'IA dans une situation donnée, de définir des marges d'appréciation pour les interventions humaines lors de l'utilisation du système ou d'ignorer une décision prise par le système.

III. Respect de la vie privée et gouvernance des données

Le respect de la vie privée et la **protection des données** doivent être garantis à **toutes les étapes** du cycle de vie du système d'IA. La numérisation des comportements humains permettrait aux systèmes d'IA de déduire non seulement les préférences, l'âge et le sexe d'une personne, mais aussi son orientation sexuelle, ses convictions religieuses ou ses opinions politiques. Pour que les citoyens aient confiance dans le traitement des données, il faut leur garantir la maîtrise totale de leurs données personnelles et veiller à ce que les données les concernant ne soient pas utilisées contre eux à des fins préjudiciables ou discriminatoires.

Outre la protection de la vie privée et des données à caractère personnel, les systèmes d'AI doivent répondre à des exigences élevées de qualité. La qualité des ensembles de données utilisés est essentielle au bon fonctionnement des systèmes d'IA. Lorsque des données sont recueillies, elles peuvent être entachées de préjugés d'ordre social ou contenir des imprécisions, des fautes et des erreurs. Il faut éliminer ces biais avant d'utiliser un ensemble de données pour entraîner un système d'IA. Par ailleurs, l'**intégrité** des données doit être assurée. Les processus et ensembles de données utilisés doivent être testés et documentés à chaque étape, parmi lesquelles la planification, l'entraînement, les essais et le déploiement. Ce principe devrait s'appliquer également aux systèmes d'IA qui n'ont pas été développés en interne mais qui ont été acquis à l'extérieur. Enfin, l'**accès** aux données doit être géré et contrôlé de manière appropriée.

IV. Transparence

La **traçabilité** des systèmes d'IA doit être assurée; il est important d'enregistrer et de documenter les décisions prises par les systèmes, ainsi que l'ensemble du processus (description de la collecte et de l'étiquetage des données, description de l'algorithme utilisé) qui a abouti aux décisions. Dans le même ordre d'idées, il convient de veiller, dans la mesure du possible, à l'**explicabilité** du processus de prise de décision algorithmique à l'intention des personnes concernées, selon des modalités adaptées à leur cas. Il convient de poursuivre les travaux de recherche en cours sur la mise au point de mécanismes d'explicabilité. Des explications doivent également être fournies sur la manière dont un système d'IA influence et façonne le processus de prise de décision organisationnel, les choix opérés dans la conception du système, ainsi que la justification de son déploiement (de manière à assurer non seulement la transparence des données et du système, mais aussi la transparence du modèle économique).

Enfin, il est important de **communiquer** des informations appropriées sur les capacités et les limites du système d'IA aux différentes parties concernées, selon des modalités adaptées au contexte d'utilisation concerné. En outre, les systèmes d'IA devraient être identifiables en tant que tels, de manière à ce que les utilisateurs sachent qu'ils interagissent avec un système d'IA et puissent identifier les personnes qui en sont responsables.

V. Diversité, non-discrimination et équité

Les ensembles de données utilisés par les systèmes d'IA (tant pour leur entraînement que pour leur exploitation) peuvent être biaisés par des partis pris historiques fortuits, des jeux de données incomplets et de mauvais modèles de gouvernance. La persistance de ces biais pourrait être source de discrimination (in)directe. Des préjudices peuvent également

résulter de l'exploitation intentionnelle de préjugés (des consommateurs) ou d'une concurrence déloyale. En outre, la manière dont les systèmes d'IA sont développés (par exemple, la manière dont le code de programmation d'un algorithme est écrit) peut également être entachée de biais. Il convient de s'attaquer à ces problèmes dès les toutes premières phases du développement du système.

La création d'**équipes de conception diversifiées** et la mise en place de mécanismes de **participation** au développement de l'IA, faisant notamment appel aux citoyens, peuvent également contribuer à la résolution de ces problèmes. Il est souhaitable de consulter les parties prenantes qui peuvent être directement ou indirectement concernées par le système tout au long de son cycle de vie. Les systèmes d'IA devraient prendre en compte tout l'éventail des capacités, aptitudes et besoins humains, et leur accessibilité devrait être garantie par une approche de conception universelle visant l'égalité d'accès pour les personnes handicapées.

VI. Bien-être sociétal et environnemental

Pour que l'IA soit digne de confiance, il faut tenir compte de son incidence sur **l'environnement et les autres êtres sensibles**. Idéalement, tous les êtres humains, y compris les générations futures, devraient pouvoir jouir d'un environnement habitable. La durabilité et la **responsabilité écologique** des systèmes d'IA devraient donc être encouragées. Il en va de même pour les solutions d'IA liées à des problématiques de portée mondiale, telles que, par exemple, les objectifs de développement durable des Nations unies.

Il faut en outre considérer les incidences des systèmes d'IA non seulement du point de vue de l'individu mais aussi du point de vue de la **société dans son ensemble**. L'utilisation des systèmes d'IA réclame une attention toute particulière dans les situations mettant en jeu le processus démocratique, comme la formation de l'opinion publique, la prise de décision politique ou les contextes électoraux. Enfin, l'**impact social** de l'IA devrait être pris en compte. Si les systèmes d'IA peuvent être utilisés pour renforcer les compétences sociales, ils peuvent également contribuer à leur détérioration.

VII. Responsabilisation

Il convient de mettre en place des mécanismes permettant de garantir la responsabilité à l'égard des systèmes d'IA et de leurs résultats, tant avant qu'après leur mise en œuvre, et de les soumettre à une obligation de rendre des comptes. La **vérifiabilité** des systèmes d'IA est essentielle à cet égard, étant donné que l'évaluation des systèmes d'IA par des auditeurs internes et externes et la possibilité de consulter les rapports d'évaluation ainsi produits contribuent fortement à rendre la technologie digne de confiance. La vérifiabilité externe devrait être particulièrement assurée pour les applications mettant eu jeu des droits fondamentaux, notamment les applications critiques pour la sécurité.

Les **effets négatifs potentiels** des systèmes d'IA devraient être définis, analysés, documentés et réduits au minimum. Le recours aux analyses d'impact facilite ce processus. Ces analyses devraient être proportionnées à l'ampleur des risques associés aux systèmes d'IA. Les **arbitrages** entre les exigences — qui sont souvent inévitables — devraient être effectués avec raison et méthode et devraient être dûment justifiés. Enfin, lorsqu'un système d'AI produit une injustice, il convient de prévoir des mécanismes accessibles afin d'en permettre une **juste réparation**.

2.3. Prochaines étapes: une phase pilote associant un maximum de parties prenantes

La recherche d'un consensus sur ces exigences essentielles auxquelles devraient répondre les systèmes d'IA constitue une première étape importante dans l'élaboration de lignes directrices pour une IA éthique. Dans un deuxième temps, la Commission veillera à ce que ces orientations puissent être testées et mises en pratique.

À cette fin, la Commission lancera une phase pilote ciblée afin d'obtenir un retour d'information structuré de la part des parties prenantes. Cet exercice se concentrera en particulier sur la liste d'évaluation que le groupe d'experts de haut niveau a établie pour chacune des exigences essentielles.

Ces travaux comprendront deux volets: i) une phase pilote pour les lignes directrices, à laquelle participeront les parties prenantes qui développent ou utilisent l'IA, y compris les administrations publiques, et ii) un processus continu de consultation et de sensibilisation des parties prenantes dans l'ensemble des États membres, s'adressant à différents groupes de parties prenantes, dont les entreprises et le secteur des services.

(i) À partir de juin 2019, toutes les parties prenantes et tous les citoyens seront invités à tester la liste d'évaluation et à fournir un retour d'information sur les possibilités de l'améliorer. En outre, le groupe d'experts de haut niveau sur l'IA procédera à un large tour d'horizon avec les parties prenantes du secteur privé et du secteur public afin de recueillir des informations plus détaillées sur les modalités de mise en œuvre des lignes directrices dans un large éventail de domaines d'application. Tous les retours d'information sur l'exploitabilité et la faisabilité des lignes directrices seront évalués d'ici à la fin de 2019.

(ii) Parallèlement, la Commission organisera de nouvelles activités de communication, en donnant aux représentants du groupe d'experts de haut niveau en matière d'IA la possibilité de présenter les lignes directrices aux parties concernées dans les États membres, notamment aux entreprises privées et au secteur des services, et en offrant à ces parties concernées une occasion supplémentaire de formuler des observations et de contribuer aux lignes directrices concernant l'IA.

La Commission tiendra compte des travaux du groupe d'experts sur l'éthique dans le domaine de la conduite connectée et automatisée[14] et collaborera avec des projets de recherche sur l'IA financés par l'UE et avec les partenariats public-privé concernés en vue de la mise en œuvre des exigences essentielles[15]. Elle soutiendra, par exemple, en coordination avec les États membres, le développement d'une base de données commune d'images médicales, qui sera initialement consacrée aux formes de cancer les plus courantes, afin que des algorithmes puissent être entraînés à diagnostiquer les symptômes avec une très grande précision. De même, la coopération entre la Commission et les États membres permet de multiplier les corridors transfrontières pour tester les véhicules connectés et automatisés. Les lignes directrices devraient être appliquées dans le cadre de ces projets et testées, et les résultats seront pris en compte dans le processus d'évaluation.

La phase pilote et la consultation des parties prenantes bénéficieront de la contribution de l'Alliance européenne pour l'IA et d'AI4EU, la plateforme d'IA à la demande. Le projet

[14] Voir la communication de la Commission sur la mobilité connectée et automatisée, COM(2018) 283.

[15] Dans le cadre du Fonds européen de la défense, la Commission élaborera également des orientations éthiques spécifiques pour l'évaluation des propositions de projets dans le domaine de l'IA pour la défense.

AI4EU[16], lancé en janvier 2019, rassemble des algorithmes, des outils, des jeux de données et des services pour aider les organisations, en particulier les petites et moyennes entreprises, à mettre en œuvre des solutions d'IA. L'Alliance européenne pour l'IA continuera, en collaboration avec AI4EU, à mobiliser l'écosystème de l'IA dans toute l'Europe, notamment en vue de piloter les lignes directrices en matière d'éthique de l'IA et de promouvoir le respect de l'approche axée sur le facteur humain en matière d'IA.

Au début de 2020, sur la base de l'évaluation des retours d'information reçus au cours de la phase pilote, **le groupe d'experts de haut niveau sur l'IA réexaminera et actualisera les lignes directrices**. Sur la base de ce réexamen et de l'expérience acquise, **la Commission évaluera les résultats et proposera, le cas échéant, de nouvelles mesures**.

Une IA respectueuse de l'éthique est bénéfique pour toutes les parties. Garantir le respect des valeurs et des droits fondamentaux est non seulement essentiel en soi, mais cela favorise également l'acceptation par le public et accroît l'avantage concurrentiel des entreprises européennes dans le domaine de l'IA en leur permettant de se démarquer par une approche de l'IA centrée sur l'humain et digne de confiance, associée à l'image de produits éthiques et sûrs. D'une manière plus générale, ce processus profitera de la solide réputation dont jouissent les entreprises européennes en matière de produits sûrs et de haute qualité. La phase pilote contribuera à garantir que les produits de l'IA respectent cette promesse.

2.4. Vers des lignes directrices internationales en matière d'éthique de l'IA

Les discussions internationales sur l'éthique en matière d'IA se sont intensifiées lorsque la présidence japonaise du G7 en a fait une grande priorité en 2016. Compte tenu de l'imbrication des intérêts internationaux dans le développement de l'IA sur le plan de la circulation des données, du développement des algorithmes et des investissements dans la recherche, **la Commission poursuivra ses efforts pour porter l'approche de l'Union sur la scène mondiale et parvenir à un consensus sur une approche de l'IA axée sur le facteur humain[17]**.

Les travaux réalisés par le groupe d'experts de haut niveau sur l'IA, et plus précisément la liste des exigences et le processus de dialogue avec les parties prenantes, fournissent à la Commission d'autres éléments particulièrement utiles pour contribuer aux discussions internationales. L'Union européenne peut jouer un rôle de chef de file dans l'élaboration de lignes directrices internationales en matière d'IA et, si possible, d'un mécanisme d'évaluation connexe.

Par conséquent, les intentions de la Commission sont les suivantes.

Renforcer la coopération avec les partenaires partageant les mêmes idées:

- étudier les possibilités de convergence avec les projets de lignes directrices en matière d'éthique de pays tiers (par exemple, le Japon, le Canada, Singapour) et, à partir de ce groupe de pays partageant les mêmes idées, préparer une discussion plus large,

[16] https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-ai4eu-project-launches-1-january-2019

[17] La haute représentante de l'Union pour les affaires étrangères et la politique de sécurité, avec l'aide de la Commission, se basera sur les consultations au sein des Nations unies, le Global Tech Panel, et d'autres enceintes multilatérales et coordonnera notamment des propositions visant à résoudre les problèmes complexes de sécurité qui se posent.

accompagnée d'actions de mise en œuvre de l'instrument de partenariat pour la coopération avec les pays tiers[18]; et

- étudier la manière dont les entreprises de pays tiers et les organisations internationales peuvent contribuer à la phase pilote des lignes directrices en procédant à des essais et des validations.

Continuer à jouer un rôle actif dans les discussions et les initiatives internationales:

- contribuer aux travaux menés dans les enceintes multilatérales telles que le G7 et le G20;
- engager des dialogues avec les pays tiers et organiser des réunions bilatérales et multilatérales pour parvenir à un consensus autour de l'IA axée sur le facteur humain;
- contribuer aux activités de normalisation pertinentes au sein des organisations internationales de normalisation afin de promouvoir cette vision; et
- renforcer la collecte et la diffusion d'analyses sur les politiques publiques, en collaboration avec les organisations internationales concernées.

3. CONCLUSIONS

L'UE est fondée sur un ensemble de valeurs fondamentales, socle sur lequel elle a bâti un cadre réglementaire solide et équilibré. Le caractère radicalement nouveau de l'IA et les défis spécifiques qu'elle implique imposent d'élaborer, au départ de cet acquis réglementaire, des lignes directrices en matière d'éthique pour encadrer le développement et l'utilisation de cette technologie. Pour emporter la confiance, l'IA doit être développée et utilisée dans le respect de valeurs éthiques largement partagées.

Dans cette optique, la Commission se félicite de la contribution apportée par le groupe d'experts de haut niveau sur l'IA. Sur la base des exigences essentielles à respecter pour que l'IA soit considérée comme digne de confiance, la Commission lancera une phase pilote ciblée destinée à vérifier que les lignes directrices en matière d'éthique qui en découlent pour le développement et l'utilisation de l'IA peuvent être mises en pratique. La Commission s'emploiera également à forger un large consensus sociétal autour de l'IA axée sur le facteur humain, notamment avec toutes les parties concernées et nos partenaires internationaux.

La dimension éthique de l'IA n'est pas un luxe superflu ou un élément accessoire: elle doit faire partie intégrante du développement de l'IA. En nous efforçant d'orienter l'IA vers une approche axée sur le facteur humain et fondée sur la confiance, nous préservons nos valeurs de société fondamentales et nous permettons à l'Europe et à ses entreprises de se distinguer en tant qu'acteurs de premier plan dans le domaine de l'IA de pointe, en jouissant d'un capital de confiance dans le monde entier.

[18] Règlement (UE) nº 234/2014 du Parlement européen et du Conseil du 11 mars 2014 instituant un instrument de partenariat pour la coopération avec les pays tiers (JO L 77 du 15.3.2014, p. 77). Par exemple, le projet intitulé «Alliance internationale pour une approche de l'intelligence artificielle axée sur le facteur humain» facilitera la réalisation d'initiatives conjointes avec des partenaires partageant les mêmes idées afin de promouvoir les lignes directrices en matière d'éthique et d'adopter des conclusions opérationnelles et des principes communs. Il permettra à l'UE et aux pays partageant les mêmes idées d'analyser les conclusions opérationnelles résultant des lignes directrices en matière d'éthique dans le domaine de l'IA proposées par le groupe d'experts afin de parvenir à une approche commune. Il permettra en outre de suivre la pénétration des technologies de l'IA au niveau mondial. Enfin, le projet prévoit d'organiser des activités de diplomatie publique en marge d'événements internationaux, par exemple au sein du G7, du G20 et de l'Organisation de coopération et de développement économiques.

Afin de garantir le développement éthique de l'IA en Europe dans un contexte élargi, la Commission poursuit une approche globale comprenant notamment les lignes d'action suivantes, à mettre en œuvre d'ici au troisième trimestre de 2019.

- Elle commencera à lancer un ensemble de **réseaux de centres d'excellence en matière d'IA** dans le cadre d'Horizon 2020. Elle sélectionnera un maximum de quatre réseaux en mettant l'accent sur de grands défis scientifiques ou technologiques, tels que l'explicabilité et l'interaction avancée entre l'homme et la machine, qui sont des éléments essentiels pour une IA digne de confiance.
- Elle entamera la création de **réseaux de pôles d'innovation numérique**[19] centrés sur l'IA dans l'industrie manufacturière et sur les mégadonnées.
- En collaboration avec les États membres et les parties prenantes, la Commission entamera des discussions préparatoires afin d'élaborer et de mettre en œuvre **un modèle de partage des données et d'utilisation optimale des espaces de données communs**, en mettant particulièrement l'accent sur les transports, les soins de santé et la fabrication industrielle[20].

En outre, la Commission élabore actuellement un rapport sur les défis liés à l'IA pour les cadres applicables en matière de sécurité et de responsabilité ainsi qu'un document d'orientation sur la mise en œuvre de la directive sur la responsabilité du fait des produits[21]. En parallèle, l'entreprise commune européenne pour le calcul à haute performance européen (EuroHPC)[22] développera la prochaine génération de supercalculateurs, la capacité de calcul étant essentielle pour le traitement des données et l'entraînement de l'IA, et l'Europe se devant de maîtriser l'ensemble de la chaîne de valeur numérique. Le partenariat en cours avec les États membres et l'industrie sur les composants et les systèmes microélectroniques (ECSEL)[23] ainsi que l'initiative relative à un processeur européen[24] contribueront au développement d'une technologie de microprocesseurs de faible puissance pour des systèmes de calcul à haute performance et de traitement des données à la périphérie («edge computing») dignes de confiance et sûrs.

À l'instar des travaux sur les lignes directrices en matière d'éthique pour l'IA, toutes ces initiatives s'appuient sur une **coopération étroite entre tous les acteurs concernés**, les États membres, les entreprises, les acteurs de la société et les citoyens. D'une manière générale, l'approche de l'Europe dans le domaine de l'intelligence artificielle montre dans quelle mesure la compétitivité économique et la confiance au sein de la société doivent s'appuyer sur les mêmes valeurs fondamentales et se renforcer mutuellement.

---

19 http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs

20 Les ressources nécessaires seront puisées dans le budget d'Horizon 2020 (au titre duquel près de 1,5 milliard d'EUR est réservé à l'IA pour la période 2018-2020) et du programme qui doit lui succéder, Horizon Europe, du volet «numérique» du mécanisme pour l'interconnexion en Europe et surtout du futur programme pour une Europe numérique. Les projets bénéficieront également de ressources provenant du secteur privé et des programmes des États membres.

21 Voir la communication de la Commission intitulée «L'intelligence artificielle pour l'Europe», COM(2018) 237.

22 https://eurohpc-ju.europa.eu

23 www.ecsel.eu

24 www.european-processor-initiative.eu