

# Les lignes directrices en matière d'éthique dans le domaine de l'intelligence artificielle

Une comparaison des lignes directrices de l'UE et de l'OCDE

Alessandro Gasparotti



© shutterstock

En avril 2019, l'Union européenne (UE) a publié des lignes directrices en matière d'éthique pour le développement de l'intelligence artificielle (IA). En mai 2019, l'OCDE a publié un autre ensemble de lignes directrices. Les chefs d'État et de gouvernement du G20 ont adopté les lignes directrices de l'OCDE. Le présent cepInput compare les deux ensembles. Les principales conclusions sont :

- ▶ Il n'existe pas de différence significative entre les lignes directrices de l'UE et celles de l'OCDE. Toutes deux ont pour objectif de créer un cadre pour une IA digne de confiance. Cependant, les lignes directrices publiées par l'UE sont plus claires et plus détaillées.
- ▶ Le fait que le G20 ait approuvé les lignes directrices de l'OCDE ne remet donc pas en cause la pertinence de celles de l'UE.
- ▶ Une importance accrue pourrait même être accordée aux lignes directrices de l'UE car elles comprennent des instructions d'ordre pratique pour leur mise en œuvre, ce que l'OCDE est en train d'élaborer. L'OCDE pourrait utiliser les instructions pratiques de l'UE comme modèle.
- ▶ L'action législative envisagée par l'UE dans le domaine de l'IA devrait s'abstenir de rendre les lignes directrices contraignantes : puisque l'UE accuse un retard concernant le développement et le déploiement de l'IA, des lignes directrices contraignantes augmenteraient encore le désavantage concurrentiel de l'UE par rapport aux États-Unis et à la Chine.

## Table des matières

<b>1</b>	<b>Introduction</b> .....	<b>3</b>
<b>2</b>	<b>L'élaboration des lignes directrices de l'UE et de l'OCDE</b> .....	<b>4</b>
<b>3</b>	<b>Contenu des lignes directrices de l'UE et de l'OCDE</b> .....	<b>7</b>
3.1	Les lignes directrices en matière d'éthique de l'UE .....	7
3.1.1	Trois éléments .....	8
3.1.2	Quatre principes éthiques .....	8
3.1.3	Sept exigences essentielles .....	9
3.2	Les lignes directrices en matière d'éthique de l'OCDE.....	10
3.3	Comparaison des lignes directrices en matière d'éthique de l'UE et de l'OCDE.....	11
<b>4</b>	<b>L'avenir des lignes directrices en matière d'éthique de l'UE</b> .....	<b>12</b>
<b>5</b>	<b>Annexe</b> .....	<b>14</b>

## 1 Introduction

Suite aux rapides progrès de la puissance de calcul et des techniques d'apprentissage automatique, notamment grâce à d'importants réseaux neuronaux, les gouvernements du monde entier ont élaboré des stratégies nationales en matière d'intelligence artificielle (IA). Ces stratégies présentent les conceptions et les politiques portant sur la manière de stimuler la recherche dans le domaine de l'IA et de s'adapter aux changements sociaux qui accompagnent son utilisation croissante. Certaines de ces stratégies traitent également de la question de savoir comment l'IA doit être réglementée.

La nécessité d'une réglementation du développement et de l'utilisation de l'IA résulte des graves conséquences que celle-ci peut avoir pour la démocratie<sup>1</sup>, les droits de l'Homme<sup>2</sup>, la vie privée<sup>3</sup> et la sécurité numérique<sup>4</sup>. Ces conséquences négatives potentielles peuvent être aggravées par le fait que l'IA est une technologie générique, c.-à-d. qu'elle peut théoriquement être appliquée à tous les secteurs de l'économie.

Les deux pays les plus avancés quant au développement et de l'adoption de l'IA sont les États-Unis et la Chine. En 2016, les investissements dans l'IA se sont élevés à environ 15 milliards d'euros en Amérique du Nord et environ 8 milliards d'euros en Asie, principalement en Chine. La même année, les investissements européens ne représentaient que 2,6 milliards d'euros<sup>5</sup>. Bien qu'ils soient conscients des potentielles conséquences négatives de l'IA, les États-Unis et la Chine ont jusqu'à présent décidé de ne pas réglementer le développement et l'utilisation de l'IA. Les deux pays craignent une réduction de leur compétitivité respective dans le développement de l'IA dans le cas où ils imposeraient une réglementation stricte en la matière.

Cependant, étant donné que nombre de parties prenantes considèrent une réglementation nécessaire, plusieurs organisations ont publié des documents sur les normes éthiques minimales à respecter. Il s'agit notamment de centres de recherche<sup>6</sup>, d'entreprises<sup>7</sup> et d'organisations internationales<sup>8</sup>.

L'UE est l'une de ces organisations. Le 8 avril 2019, l'UE a publié des lignes directrices en matière d'éthique dans le domaine de l'IA. Ces lignes directrices font partie de la stratégie européenne en matière d'IA annoncée par la Commission européenne le 25 avril 2018<sup>9</sup>. Elle se compose de trois piliers qui visent à renforcer la capacité technologique de l'UE ainsi que les investissements dans la recherche et le développement<sup>10</sup>, à se préparer aux changements socio-économiques<sup>11</sup> et à garantir l'existence

---

<sup>1</sup> P. ex., les fake news diffusées par les messageries sur les réseaux sociaux peuvent avoir un impact sur les résultats des élections.

<sup>2</sup> P. ex., une prise de décision automatisée par les entreprises pourrait aussi perpétuer les préjugés humains et aboutir à des résultats discriminatoires.

<sup>3</sup> P. ex., la surveillance de masse permise par des technologies de reconnaissance faciale ou des algorithmes capables d'extrapoler l'orientation sexuelle et les préférences politiques à partir des informations de l'utilisateur.

<sup>4</sup> P. ex., l'IA pourrait être utilisée pour créer de nouveaux logiciels malveillants plus efficaces capables d'apprendre et de s'adapter pour lancer de nouvelles attaques.

<sup>5</sup> McKinsey, "10 imperatives for Europe in the age of AI and automation", 2017 [cité par COM(2018) 237, p. 4].

<sup>6</sup> P. ex., la "[Beijing Academy of Artificial Intelligence](#)" ou la "[Montreal Declaration for A Responsible Development of AI](#)".

<sup>7</sup> P. ex., [Google](#), [Deutsche Telekom](#), [SAP](#), [IBM](#), et [Sony](#).

<sup>8</sup> P. ex. l'[UNESCO](#) et le [Conseil de l'Europe](#).

<sup>9</sup> Cf. COM (2018) 237, « L'intelligence artificielle pour l'Europe ».

<sup>10</sup> Cf. [Les analyses du cepFrance n°2019-10](#)

<sup>11</sup> Cf. [Les analyses du cepFrance n°2019-12](#)

d'un cadre éthique et juridique approprié<sup>12</sup>. Dans le cadre de la mise en œuvre de la stratégie, la Commission européenne et les États membres se sont engagés à fournir des ressources pour stimuler la recherche, développer la formation des citoyens de l'UE et évaluer l'adéquation de la législation existante aux défis posés. D'ici la fin 2020, l'investissement total dans l'IA au sein de l'UE devrait atteindre au moins 20 milliards d'euros. En outre, chaque État membre élaborera sa propre stratégie nationale en matière d'IA et l'UE coordonnera les efforts nationaux, facilitant ainsi les projets communs et développant les économies d'échelle.

La publication des lignes directrices en matière d'éthique constitue la dernière étape dans la mise en œuvre de la stratégie de l'UE en matière d'IA. Le 25 avril 2018, la Commission européenne a nommé un groupe indépendant d'experts de haut niveau sur l'intelligence artificielle (GEHNIA), chargé d'élaborer des lignes directrices en matière d'éthique pour le développement et l'utilisation de l'IA. Le 8 avril 2019, le GEHNIA a présenté la version finale de ses « Lignes directrices en matière d'éthique pour une IA digne de confiance »<sup>13</sup>. Bien qu'elles ne soient pas contraignantes, ces lignes directrices constitueront le pilier d'une éventuelle action législative de l'UE à l'avenir.

Un mois après la présentation des lignes directrices de l'UE, le 22 mai 2019, quarante-deux pays<sup>14</sup> ont adopté les premières normes intergouvernementales sur l'IA sous la forme d'une « recommandation de l'organisation de coopération et de développement économiques (OCDE) sur l'intelligence artificielle »<sup>15</sup>. Les ministres des principaux acteurs mondiaux de l'IA, parmi lesquels Israël, le Japon, la Corée et les États-Unis, se sont accordés sur un ensemble commun de principes pour guider le futur développement de l'IA. La majorité des États membres de l'UE, également membres de l'OCDE, se sont mis d'accord sur ces normes. Lors du sommet du G20 de juin 2019, les chefs d'État et de gouvernement ont également approuvé les lignes directrices en matière d'éthique de l'OCDE.

Le présent cepInput évalue les conséquences de cette approbation sur la pertinence des lignes directrices de l'UE. Par conséquent, la section 2 compare le processus d'élaboration des lignes directrices de l'OCDE et de celles de l'UE, tandis que la section 3 compare leur contenu. La section 4 examine enfin la pertinence des lignes directrices de l'UE.

## 2 L'élaboration des lignes directrices de l'UE et de l'OCDE

La Commission européenne a nommé un groupe d'experts (GEHNIA) et lui a demandé d'élaborer des lignes directrices en matière d'éthique pour l'IA. Le GEHNIA comprend 52 membres provenant du milieu universitaire, de l'industrie et de la société civile<sup>16</sup>. La figure 1 ci-dessous illustre la composition du GEHNIA. Le GEHNIA est soutenu par l'Alliance européenne de l'IA, un groupe comprenant plus de 2000 parties prenantes qui nourrissent en permanence le travail du GEHNIA<sup>17</sup>.

<sup>12</sup> Cf. [Les analyses du cepFrance n°2019-13](#)

<sup>13</sup> Cf. COM (2019) 168 et [Les analyses du cepFrance n°2019-13](#).

<sup>14</sup> Outre les 36 États membres de l'OCDE, la recommandation a déjà été signée par six pays non-membres.

<sup>15</sup> <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

<sup>16</sup> <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>.

<sup>17</sup> <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>.

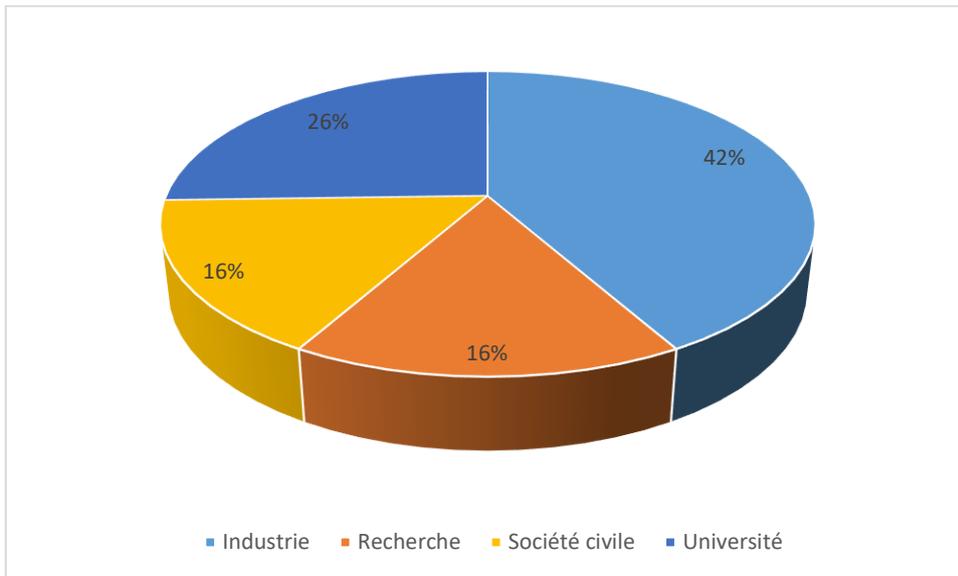
**Fig. 1: membres du GEHNIA**

Diagramme établi par l'auteur d'après la biographie officielle des membres. Chaque fois qu'un membre peut être associé à deux groupes, il est inclus dans les deux.

Sources : <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence> et <https://www.oecd.org/going-digital/ai/oecd-aigo-membership-list.pdf>.

Le GEHNIA a commencé ses travaux sur les lignes directrices en matière d'éthique le 1<sup>er</sup> juin 2018. Six mois plus tard, le 18 décembre 2018, il publiait son projet de lignes directrices en matière d'éthique<sup>18</sup>. Après une révision complète, le GEHNIA a présenté la version finale des lignes directrices en matière d'éthique<sup>19</sup> le 8 avril 2019. Le 26 juin 2019, le GEHNIA a lancé une phase pilote pour évaluer les lignes directrices. Jusqu'à la fin de l'année, les entreprises de l'UE peuvent délibérément adopter ces lignes directrices et faire un compte rendu au GEHNIA concernant leur mise en œuvre. Avant que les lignes directrices ne soient officiellement transmises à la Commission européenne, le GEHNIA examinera la possibilité d'autres amendements. Ces amendements se fonderont sur des consultations supplémentaires avec le public et les informations collectées pendant la phase pilote, p. ex. lors d'entretiens approfondis avec les entreprises retenues. Il est prévu que la Commission européenne disposera de lignes directrices en 2020, après avoir été testées dans le monde réel, ainsi que d'une réserve suffisante d'informations concernant les opinions et les demandes des parties prenantes. Une action supplémentaire pourrait alors suivre.

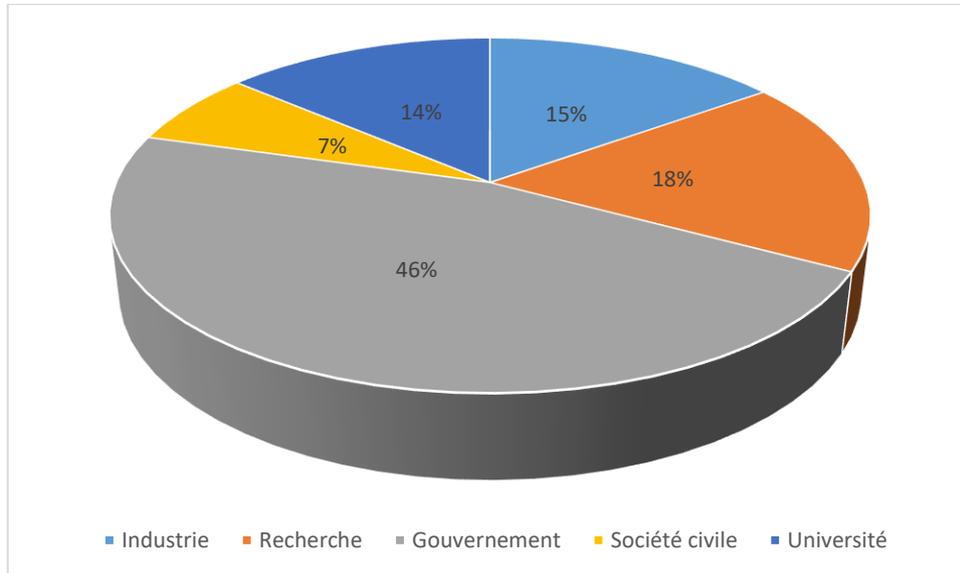
En mai 2018, un mois après l'annonce par la Commission européenne de son projet d'élaborer des lignes directrices en matière d'éthique, le Comité de la politique de l'économie numérique (CPEN) de l'OCDE a décidé de constituer un groupe d'experts chargé de rédiger des principes éthiques pour le développement de l'IA. Le groupe d'experts de l'OCDE (AIGO) est composé de 56 membres, assistés de 17 experts externes. Les membres et les experts sont issus du milieu universitaire, de l'industrie, de la société civile, et principalement des représentants de gouvernements nationaux par ailleurs. Il est intéressant de noter que des membres de la Commission européenne, de l'Institut des ingénieurs en

<sup>18</sup> <https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai>.

<sup>19</sup> <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

électricité et en électronique (IEEE)<sup>20</sup> et de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO) font également partie de l'AIGO. Les trois organisations ont élaboré leurs propres lignes directrices, ce qui suggère que l'OCDE souhaite travailler en étroite collaboration avec ses partenaires afin d'harmoniser ses principes sur l'IA. La figure 2 illustre la composition de l'AIGO.

**Fig. 2 : les membres de l'AIGO**



Source : <https://www.ieee.org/about/index.html>

Calculs de l'auteur fondés sur la biographie officielle des membres. Chaque fois qu'un membre peut être associé à deux groupes, il est inclus dans les deux.

Les 13 et 14 mars 2019, le CPEN a approuvé un projet de recommandation pour des lignes directrices en matière d'éthique pour l'IA et a accepté de les transmettre au Conseil de l'OCDE. Les 22 et 23 mai 2019, le Conseil de l'OCDE a adopté<sup>21</sup> la recommandation lors de son assemblée au niveau ministériel. Une telle recommandation est le premier acte législatif au monde sur une réglementation en matière d'éthique de l'IA. Bien qu'il s'agisse d'un instrument législatif, les lignes directrices de l'OCDE ne sont pas contraignantes pour les pays signataires. Toutefois, compte tenu de la portée mondiale des membres de l'OCDE – allant des pays de l'UE au Japon en passant par la Corée, Israël, le Canada et les États-Unis, le consensus sur les normes éthiques pour l'IA semble être une remarquable réussite.

Afin de renforcer la mise en œuvre des lignes directrices de l'OCDE, le Conseil de l'OCDE a chargé le CPEN de développer des instructions pratiques pour leur mise en œuvre. Il n'y a actuellement pas de délai pour la publication de ce document pratique.

Le tableau 1 compare le développement des deux ensembles de lignes directrices. Il montre que l'élaboration de ces deux ensembles était relativement similaire. Ils ont tous deux été élaborés par un groupe d'acteurs variés et se sont conclus sous la forme d'un document non contraignant. La différence la plus frappante entre les deux ensembles de lignes directrices réside dans le fait que, pour

<sup>20</sup> « L'IEEE invite la communauté mondiale à innover pour un avenir meilleur à travers plus de 423 000 membres dans 160 pays. L'IEEE est la « voix » de la confiance en matière d'ingénierie, d'informatique et d'information technologique dans le monde entier. » <https://www.ieee.org/about/index.html>

<sup>21</sup> <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

l'élaboration des lignes directrices de l'OCDE, près de 50% des membres étaient des représentants gouvernementaux, alors qu'aucun d'entre eux n'était directement impliqué dans l'élaboration des lignes directrices de l'UE. Plus précisément, les lignes directrices de l'UE ont été élaborées par les utilisateurs des lignes directrices et par les parties directement concernées par celles-ci. La section suivante évaluera si cela implique des différences dans le contenu des deux ensembles de lignes directrices.

**Tab. 1 : Informations sur les lignes directrices de l'UE et de l'OCDE**

		GEHNIA (Commission européenne)	AIGO (OCDE)
Composition du groupe d'experts	Recherche	16 %	18 %
	Université	26 %	14 %
	Industrie	42 %	15 %
	Société civile	16 %	7 %
	Gouvernement	0 %	46 %
Nature supranationale de l'organisme demandeur		+	+
Caractère contraignant		-	-
Autres étapes en cours		+	+

### 3 Contenu des lignes directrices de l'UE et de l'OCDE

Les deux ensembles de lignes directrices visent à fournir un cadre général, qui pourra éventuellement être complété par une réglementation d'ordre sectoriel<sup>22</sup> ou contractuel<sup>23</sup>. L'OCDE et l'UE estiment cela opportun compte tenu des développements imprévisibles de l'IA et de ses diverses applications. Selon les deux organisations, tout cadre réglementaire devrait être général, c.-à-d. englober la totalité du cycle de vie de la technologie, tout en étant flexible, c.-à-d. pouvant être adapté à de nouveaux scénarios sans entraver involontairement les progrès technologiques.

La section suivante décrit les deux ensembles de lignes directrices, en débutant par les lignes directrices de l'UE. À la fin de cette section seront présentées les similitudes et les différences entre les lignes directrices.

#### 3.1 Les lignes directrices en matière d'éthique de l'UE <sup>24</sup>

En ce qui concerne l'UE, la confiance est un élément essentiel à l'acceptation et à l'adoption des technologies de l'IA<sup>25</sup>. Les lignes directrices en matière d'éthique sont proposées en tant qu'instrument permettant d'établir la confiance entre les producteurs, les fournisseurs et les utilisateurs finaux de l'IA. C'est la raison pour laquelle le GEHNIA apporte une définition détaillée de ce qu'est une « IA digne

<sup>22</sup> L'IA peut être utilisée dans différents produits, p. ex. un logiciel de reconnaissance vocale, des drones de surveillance, des voitures autonomes et des dispositifs médicaux. Étant donné que ces applications interviennent dans différents domaines, il pourrait s'avérer impossible de leur appliquer des règles identiques.

<sup>23</sup> Un cadre flexible garantissant la liberté contractuelle devrait exister en particulier pour les relations entre les entreprises. Les relations entre entreprises et consommateurs peuvent nécessiter des règles plus structurées en raison des asymétries potentielles de pouvoir et d'information entre les parties.

<sup>24</sup> Pour une évaluation détaillée, voir [Les analyses du cepFrance no. 2019-13](#). Un cepPolicyBrief évaluant le contenu des lignes directrices de l'UE est en cours de publication.

<sup>25</sup> Groupe d'experts de haut niveau sur l'intelligence artificielle, Les lignes directrices en matière d'éthique pour une IA digne de confiance p. 4. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

de confiance », en exposant trois éléments généraux que l'IA doit respecter pour être considérée comme tel. L'un de ces éléments comprend quatre principes éthiques. Puisque ces principes sont très généraux et donc difficiles à mettre en œuvre dans la pratique, les principes éthiques sont traduits en sept exigences essentielles qui fournissent des indications – notamment aux développeurs – sur la mise en pratique de ces principes. Les lignes directrices comprennent en outre un questionnaire permettant d'aider – notamment les développeurs – à évaluer la conformité d'un produit avec les lignes directrices en matière d'éthique. La figure 3 ci-dessous décrit la logique des lignes directrices en matière d'éthique de l'UE.

**Fig. 3 : Logique des lignes directrices en matière d'éthique de l'UE**

**3 éléments → 4 principes éthiques → 7 exigences essentielles → questionnaire**

Le reste de cette section décrit les trois éléments, les principes éthiques et les exigences essentielles.

### 3.1.1 Trois éléments

Une IA digne de confiance comporte trois éléments qui doivent être respectés tout au long du cycle de vie du système d'IA. Cela signifie qu'une IA digne de confiance ne concerne pas seulement la fiabilité du système en lui-même. Elle englobe également la fiabilité de tous les processus et des acteurs qui font partie du cycle de vie du système. Les trois éléments d'une IA digne de confiance sont<sup>26</sup> :

- **Une IA licite** : L'IA doit respecter toutes les lois et la réglementation applicables. Étant donné que la législation prévoit à la fois des obligations positives et négatives, cet élément d'une IA digne de confiance renvoie aussi bien à ce qui peut être fait qu'à ce qui ne doit pas être fait.
- **Une IA éthique** : L'IA doit faire preuve de respect et assurer son adhésion aux principes et valeurs éthiques.
- **Une IA robuste** : L'IA devrait garantir que son utilisation ne cause aucun préjudice involontaire.

### 3.1.2 Quatre principes éthiques

Les lignes directrices de l'UE comprennent quatre principes éthiques ancrés dans les droits fondamentaux. Les lignes directrices soulignent que les praticiens de l'IA doivent toujours s'efforcer de s'y conformer. Ces quatre principes sont<sup>27</sup> :

- **Le respect de l'autonomie humaine** : les êtres humains devraient garder une capacité totale d'autodétermination et un contrôle sur l'IA. Par ailleurs, l'IA ne devrait pas induire en erreur, contraindre ou manipuler les êtres humains.
- **La prévention de toute atteinte** : l'IA ne devrait ni porter atteinte, ni accentuer toute atteinte ou nuire de quelque façon que ce soit aux êtres humains, d'autant plus qu'elle peut provoquer des asymétries de pouvoir ou d'informations<sup>28</sup>.

<sup>26</sup> Lignes directrices en matière d'éthique pour une IA digne de confiance, p. 6-7, p. 38.

<sup>27</sup> Ibid., p. 11-13.

<sup>28</sup> P. ex., une entreprise peut imposer des primes d'assurance plus élevées aux personnes davantage susceptibles de tomber malade.

- **L'équité** : il conviendrait de partager équitablement les coûts et les avantages liés à l'IA, d'éviter les discriminations et les biais injustes et de mettre en place des mécanismes de recours efficaces contre les décisions prises par l'IA.
- **L'explicabilité** : il conviendrait de communiquer l'objectif des systèmes d'IA, les procédés devraient être transparents et les décisions explicables, dans la mesure du possible en fonction du contexte et de la gravité des conséquences d'un résultat erroné.

Ces principes sont déjà largement intégrés dans la législation existante, p. ex. sur la protection des données. L'IA doit donc déjà respecter la plupart de ces principes. En cas d'arbitrage entre ces quatre principes, il conviendrait d'y procéder par une réflexion fondée sur des preuves<sup>29</sup>. Il conviendrait de communiquer l'existence d'arbitrages et la manière dont ils sont effectués, et d'assurer que cela soit compréhensible pour toutes les parties concernées afin de garantir la confiance.

### 3.1.3 Sept exigences essentielles

La nature générale des principes éthiques rend ces derniers difficiles à mettre en œuvre. Ils sont donc traduits en sept exigences essentielles qui fournissent une orientation – notamment aux développeurs – sur la manière de les mettre en œuvre. Ces sept exigences essentielles doivent être respectées pour parvenir à une IA digne de confiance<sup>30</sup>. Le respect de ces exigences est nécessaire mais pas suffisant pour garantir que l'IA soit digne de confiance.

Les sept exigences sont les suivantes :

- **Action et contrôle humains** :
  - les effets négatifs sur les droits fondamentaux devraient être évalués avant le développement de l'IA et réduits ou justifiés ;
  - les systèmes d'IA ne devraient pas restreindre l'autonomie humaine, mais aider les individus à prendre de meilleures décisions et des choix plus éclairés ;
  - un contrôle humain approprié devrait être assuré sur l'IA par des mécanismes de gouvernance, p. ex. en permettant de décider de l'utilisation de l'IA dans des situations données, en contrôlant l'activité de l'IA et en veillant à ce que l'être humain garde une certaine marge d'appréciation et la capacité d'ignorer des décisions prises par l'IA ; moins il est possible de contrôler l'IA, plus il est nécessaire de la tester ;
  - les autorités publiques de contrôle devraient avoir le mandat et la capacité d'exercer un contrôle sur les utilisateurs et les résultats de l'IA.
- **Robustesse technique et sécurité** :
  - L'IA devrait être fiable, sécurisée et résistante aux attaques, p. ex. au piratage et à la manipulation.
  - L'IA devrait disposer de mesures de protection assurant une solution de secours en cas de problème, p. ex. avec un opérateur humain.
  - L'IA devrait avoir un niveau de sécurité proportionné à la magnitude du risque que son résultat présente.
  - L'IA devrait être précise et informer les utilisateurs de la probabilité d'erreurs éventuelles.
  - L'IA devrait créer les mêmes résultats, c.-à-d. des résultats « reproductibles », dans les mêmes conditions, de manière à pouvoir décrire son comportement.

<sup>29</sup> P. ex., les technologies de reconnaissance faciale peuvent réduire la criminalité (prévenir les préjudices) tout en limitant la vie privée et la liberté individuelle (c.-à-d. l'autonomie humaine).

<sup>30</sup> Lignes directrices en matière d'éthique pour une IA digne de confiance, op. cit., p. 4-6 et p. 15-20.

- **Respect de la vie privée et gouvernance des données :**
  - Les êtres humains doivent avoir la maîtrise totale de leurs données collectées pour ou par l'IA ; les données ne doivent pas être utilisées illégalement.
  - La qualité, l'intégrité et la pertinence des données utilisées par un système d'IA doivent être garanties pour éviter tout biais ou toute erreur.
- La **transparence** des données, des systèmes d'IA et des modèles économiques doit être garantie. Cela comprend :
  - la traçabilité des systèmes d'IA, notamment en documentant leurs décisions et le processus sous-jacent (y compris les données utilisées dans l'analyse et l'« entraînement » de la machine d'IA) ;
  - dans la mesure du possible, l'explicabilité des processus de prise de décision algorithmiques, en tenant compte d'une possible réduction de la précision et à condition que l'IA ait « un impact significatif sur la vie des gens » ;
  - une notification appropriée des individus concernant leur interaction avec un système d'IA, les capacités et les limitations de celui-ci (p. ex. sa précision limitée) et la possibilité de s'opposer à cette interaction au profit d'une interaction humaine lorsque cela est nécessaire pour assurer le respect des droits fondamentaux.
- **Diversité, non-discrimination et équité :** L'IA devrait être « centrée sur l'utilisateur » afin que tout le monde puisse utiliser le produit ou le service. Il faut éviter les biais injustes, p. ex. dans les ensembles de données, car ils pourraient conduire à des discriminations.
- **Bien-être sociétal et environnemental :** L'IA doit être durable et respectueuse de l'environnement et de la société.
- **Responsabilité :** L'IA devrait être conçue de manière à pouvoir être vérifiée – sans avoir besoin de publier des informations relatives à la propriété intellectuelle ou toute autre propriété – en particulier lorsque les applications de l'IA ont une incidence sur les droits fondamentaux. Les effets négatifs devraient être signalés et minimisés, et des mécanismes de réparation adéquats devraient être prévus.

Parallèlement aux lignes directrices en matière d'éthique, l'UE a publié un questionnaire. Celui-ci comprend une série de questions sur chacune des sept exigences, ce qui devrait laisser une chance à toutes les parties prenantes à l'IA d'évaluer si un produit de l'IA respecte les sept exigences et donc les quatre principes éthiques.

### 3.2 Les lignes directrices en matière d'éthique de l'OCDE

Pour l'OCDE, la confiance dans l'IA est également une exigence-clé. Les lignes directrices en matière d'éthique de l'OCDE ont donc pour objectif de développer une IA digne de confiance. Au lieu de fournir une définition exhaustive à l'instar de l'UE, l'OCDE définit une IA digne de confiance simplement comme une IA qui respecte les lignes directrices. Par rapport aux lignes directrices de l'UE, l'OCDE a élaboré un document beaucoup plus concis. Alors que les premières comportent 38 pages, celles de l'OCDE tiennent en une page où sont énumérés cinq principes éthiques que l'IA doit respecter pour être considérée comme digne de confiance. Ces cinq principes sont<sup>31</sup> :

- **L'IA devrait soutenir la croissance inclusive, le développement durable et le bien-être**, c.-à-d. tendre vers des résultats bénéfiques pour les individus et la planète. Ces résultats comprennent le renforcement des capacités humaines, l'inclusion des populations sous-représentées, la réduction des inégalités économiques, sociales, entre les sexes et autres, et la protection de l'environnement.

<sup>31</sup> Cf. OECD/LEGAL/0449, section 1. <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449>

- **L'IA devrait soutenir les valeurs centrées sur l'humain et l'équité**, c.-à-d. que :
  - les acteurs de l'IA devraient respecter l'Etat de droit, les Droits de l'homme et les valeurs démocratiques tout au long du cycle de vie du système d'IA. Ces droits et ces valeurs comprennent la liberté, la dignité et l'autonomie, la protection de la vie privée et des données, la non-discrimination et l'égalité, la diversité, l'équité, la justice sociale et les droits des travailleurs reconnus à l'échelle internationale ;
  - les acteurs de l'IA doivent ensuite établir des garanties et des mécanismes de protection, tels que l'attribution de la capacité de décision finale à l'être humain, qui soient adaptés au contexte et conformes à l'état des connaissances.
- **L'IA devrait être transparente et explicable**, c.-à-d. que les acteurs de l'IA devraient fournir des informations pertinentes, adaptées au contexte et à l'état des connaissances :
  - pour favoriser une compréhension générale des systèmes d'IA,
  - pour faire prendre conscience aux parties prenantes de leurs interactions avec les systèmes d'IA,
  - pour permettre aux personnes concernées de comprendre le résultat,
  - pour permettre aux personnes qui en subissent les effets néfastes de contester les résultats sur la base d'informations sur les facteurs et la logique qui ont servi à la formulation de la décision.
- **L'IA devrait être robuste, sûre et sécurisée** :
  - Dans des conditions d'utilisation normale ou prévisible, ou d'utilisation abusive, elle devrait fonctionner convenablement et ne pas présenter de risque de sécurité démesuré.
  - Les acteurs de l'IA devraient assurer la traçabilité, p. ex. des ensembles de données, des processus et des décisions prises au cours du cycle de vie du système d'IA, en fonction du contexte.
  - Les acteurs de l'IA devraient appliquer une approche systématique de gestion des risques à chaque phase du cycle de vie d'un système d'IA afin de gérer les risques liés aux systèmes d'IA, p. ex. le respect de la vie privée et la sécurité numérique.
- **L'IA devrait être responsable**, c.-à-d. que les acteurs de l'IA devraient être responsables du bon fonctionnement des systèmes et du respect des principes susmentionnés.

Bien que chaque principe soit explicité, les lignes directrices de l'OCDE sont très courtes et rédigées de manière très générale. Contrairement aux sept exigences de l'UE, aucun détail supplémentaire sur la mise en œuvre de ces cinq principes n'est fourni et aucun questionnaire n'est compris dans les lignes directrices. L'évaluation du respect de ces principes est donc difficile pour les praticiens de l'IA, p. ex. les développeurs.

### 3.3 Comparaison des lignes directrices en matière d'éthique de l'UE et de l'OCDE

Cette section compare les lignes directrices en matière d'éthique de l'UE et celles de l'OCDE. Comme nous l'avons indiqué dans la section 3.2, les deux ensembles de lignes directrices ont pour objectif de créer un cadre pour le développement d'une IA digne de confiance. Pour atteindre cet objectif, les deux ensembles de lignes directrices comprennent une liste de principes éthiques qui, dans le cas de l'UE, sont traduits par sept exigences essentielles.

Les deux ensembles de lignes directrices sont très similaires. Les cinq principes de l'OCDE coïncident avec les quatre principes de l'UE. Comme le montre le tableau 2 en annexe, les sept exigences décrites dans les lignes directrices de l'UE peuvent être intégrées aux cinq principes définis par l'OCDE. Aucune disposition majeure ne diffère dans les deux ensembles de lignes directrices. Cependant, les lignes directrices de l'UE sont plus précises car elles fournissent deux outils – les exigences et le questionnaire

– qui peuvent faciliter leur mise en œuvre. Même si les lignes directrices de l'UE sont plus précises que celles de l'OCDE, toutes deux restent toutefois générales et manquent d'exemples pratiques. Une large marge d'appréciation est donc à prévoir dans la mise en œuvre des deux ensembles de lignes directrices. La présence d'un questionnaire facilite le respect et éventuellement la future exécution<sup>32</sup> des lignes directrices de l'UE. La principale différence entre les deux ensembles de lignes directrices réside dans la tentative de traduire des principes éthiques vagues en exigences plus précises – mais encore vagues – et de donner les étapes à suivre.

#### 4 L'avenir des lignes directrices en matière d'éthique de l'UE

Lors de la réunion ministérielle du G20 sur le commerce et l'économie numérique tenue à Tsukuba (Japon) les 8 et 9 juin 2019, les ministres des puissances économiques mondiales, parmi lesquelles notamment les États-Unis et la Chine, ont approuvé des principes éthiques pour une approche de l'IA centrée sur l'être humain. Ces principes sont annexés à la déclaration ministérielle du G20 et intégrés à la déclaration finale du G20 publiée le 29 juin 2019<sup>33</sup>. Ils sont de nature non contraignante<sup>34</sup>. Ces principes sont « tirés » des lignes directrices de l'OCDE<sup>35</sup>, ce qui signifie qu'ils sont en parfaite adéquation avec la recommandation de l'OCDE. En conséquence, il semble que les États-Unis, la Chine et l'UE<sup>36</sup> approuvent les lignes directrices de l'OCDE.

Un consensus est cependant encore loin d'être atteint. Le même paragraphe de la déclaration finale du G20 insiste sur l'importance d'approches réglementaires souples et flexibles, « notamment par l'utilisation de sas réglementaires »<sup>37</sup>. Les sas sont par nature des solutions temporaires nationales pour contourner les normes existantes. À l'avenir, il est donc possible que les législations nationales, p. ex. sur la protection des données, soient supprimées pour favoriser le développement de l'IA au lieu de définir des normes éthiques communes et robustes.

L'adoption de normes éthiques non contraignantes, définies au sens large, difficiles à mettre en œuvre et pouvant être contournées au moyen de sas, ne constitue pas un progrès décisif dans l'établissement de la réglementation mondiale en matière d'IA.

Dans le contexte mondial, l'UE a donc élaboré des lignes directrices plus précises et plus pratiques. La présidente élue de la Commission européenne, Ursula von der Leyen, a annoncé que l'IA serait l'objet d'une action législative. À ce stade, les mesures législatives que la Commission européenne a l'intention de présenter sont incertaines. Si les lignes directrices devaient devenir contraignantes, il faudrait déployer beaucoup d'efforts pour réduire les imprécisions et éclaircir les ambiguïtés. Une telle tâche prendrait du temps et comporterait un certain nombre d'incertitudes, compte tenu du rythme rapide auquel l'IA se développe. En outre, il est difficilement concevable que la Chine et les États-Unis acceptent une législation définissant des normes en matière d'IA et qui émanerait de l'UE.

---

<sup>32</sup> Alors que les lignes directrices de l'UE et de l'OCDE sont par nature non contraignantes, la Commission européenne veut proposer [une législation pour l'IA](#). Il reste à déterminer s'il est question de prendre des mesures pour réglementer l'IA ou promouvoir la recherche et l'innovation.

<sup>33</sup> [https://g20.org/pdf/documents/en/FINAL\\_G20\\_Osaka\\_Leaders\\_Declaration.pdf](https://g20.org/pdf/documents/en/FINAL_G20_Osaka_Leaders_Declaration.pdf).

<sup>34</sup> Ibid., paragraphe 12 de la déclaration finale.

<sup>35</sup> Ibid.

<sup>36</sup> Jean-Claude Juncker représente la Commission européenne au sommet du G20.

<sup>37</sup> Paragraphe 12 de la déclaration finale, op. cit., note 33.

Par conséquent, l'UE étant actuellement en retard dans le développement et le déploiement de l'IA, des lignes directrices contraignantes dans le domaine de l'IA augmenteraient davantage le désavantage concurrentiel de l'UE.

Compte tenu de la similitude des lignes directrices de l'UE et de l'OCDE, ainsi que du projet de l'OCDE de publier un guide pratique sur la mise en œuvre de leurs principes éthiques, l'UE peut jouer un rôle important pour façonner à l'avenir les normes mondiales en matière d'IA. Tel serait le cas si le guide pratique de l'OCDE était fondé sur les exigences essentielles de l'UE et sur son questionnaire qui ne doit pas en être exclu, étant donné que certaines grandes sociétés<sup>38</sup> se sont déjà engagées à mettre en œuvre les lignes directrices en matière d'éthique de l'UE.

---

<sup>38</sup> <https://www.ibm.com/blogs/policy/ai-ethics-eu/>, IBM par exemple les applique déjà .

## 5 Annexe

Tab. 2 : Comparaison des principes éthiques de l'OCDE et des exigences essentielles de l'UE

Principes de l'OCDE	Exigences essentielles de l'UE correspondantes
<p><b>Croissance et bien-être inclusifs et durables</b> Une IA digne de confiance tendant vers des résultats bénéfiques pour les individus et la planète, réduisant les inégalités, protégeant les ressources naturelles, favorisant la croissance inclusive et le développement durable.</p>	<p><b>Bien-être sociétal et environnemental</b> La durabilité et la responsabilité écologique des systèmes d'IA doivent être encouragées et la recherche de solutions à des problématiques de portée mondiale, telles que les objectifs de développement durable, favorisée.</p>
<p><b>Valeurs centrées sur l'humain et équité</b> Les acteurs de l'IA doivent respecter l'Etat de droit, les Droits de l'Homme et les valeurs démocratiques. Ceux-ci comprennent la liberté, la dignité et l'autonomie, le respect de la vie privée et la protection des données, la non-discrimination et l'égalité, l'équité, la justice sociale et les droits des travailleurs.</p>	<p><b>Action et contrôle humains</b> Compte tenu de la portée et de la capacité des systèmes d'IA, ils peuvent aussi avoir des effets négatifs sur les droits fondamentaux. Dans les cas où de tels risques existent, une évaluation de l'impact sur les droits fondamentaux devrait être réalisée. Cela devrait être effectué avant le développement du système, en évaluant notamment si ces risques peuvent être réduits ou justifiés comme étant nécessaires dans une société démocratique pour respecter les droits et libertés d'autrui.</p>
	<p><b>Respect de la vie privée et gouvernance des données</b> Les systèmes d'IA doivent garantir le respect de la vie privée et la protection des données tout au long du cycle de vie d'un système. Cela comprend les informations fournies initialement par l'utilisateur, ainsi que les informations générées sur l'utilisateur au cours de son interaction avec le système (p. ex. les résultats que le système d'IA a générés pour des utilisateurs spécifiques ou comment les utilisateurs ont répondu à des recommandations particulières). La numérisation des comportements humains peut permettre aux systèmes d'IA de déduire non seulement les préférences des individus, mais aussi leur orientation sexuelle, leur âge, leur sexe, leurs opinions religieuses ou politiques. Pour permettre aux individus d'avoir confiance dans les processus de collecte de données, il faut s'assurer que les données collectées sur ces individus ne seront pas utilisées pour les discriminer de manière illégale et injuste.</p>
	<p><b>Diversité, non-discrimination et équité</b> Dans la mesure du possible, les biais identifiables et discriminatoires doivent être supprimés dans la phase de collecte. La manière dont les systèmes d'IA sont développés (p. ex. la programmation des algorithmes) peut également souffrir de biais injustes. On peut contrecarrer cela par la mise en place de processus de contrôle permettant d'analyser et de répondre à l'objectif, aux contraintes, aux exigences et aux décisions du système de manière claire et transparente.</p>
<p><b>Bien-être sociétal et environnemental</b> Outre l'évaluation des incidences sur les individus du développement, du déploiement et de l'utilisation d'un système d'IA, ces incidences doivent être évaluées du point de vue de la société, en tenant compte des effets sur les institutions, la démocratie et la société au sens large. L'utilisation de systèmes d'IA devrait également être examinée attentivement, particulièrement dans les situations ayant trait au processus démocratique, que ce</p>	

	<p>soit le processus de prise de décision politique ou les contextes électoraux.</p>
<p><b>Valeurs centrées sur l'humain et équité</b>                  Les acteurs de l'IA doivent mettre en place des garanties et des mécanismes, tels que l'attribution de la capacité de décision finale, qui soient adaptés au contexte et conformes à l'état des connaissances.</p>	<p><b>Responsabilité</b>                  La capacité tant à signaler des actions ou des décisions contribuant au résultat d'un système donné qu'à remédier aux conséquences d'un tel résultat doit être garantie. Il est essentiel pour les personnes (in)directement concernées que les effets négatifs potentiels des systèmes d'IA soient identifiés, analysés, documentés et réduits au minimum.</p> <p><b>Robustesse technique et sécurité</b>                  Les systèmes d'IA devraient disposer de mécanismes assurant une solution de secours en cas de problème. Cela peut impliquer que les systèmes d'IA passent d'une procédure statistique à une procédure fondée sur des règles ou qu'ils fassent appel à un opérateur humain avant de poursuivre leur action.</p> <p><b>Action et contrôle humains</b>                  Il doit être possible pour les utilisateurs de prendre des décisions de manière autonome et en connaissance de cause concernant les systèmes d'IA. Les connaissances et les outils pour comprendre et interagir avec des systèmes d'IA de manière satisfaisante devraient leur être fournis et, si possible, il devrait leur être possible d'auto-évaluer ou de contester le système d'une manière appropriée. (...) 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. (...) 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. En outre, il doit être garanti que les autorités publiques aient la capacité d'exercer un contrôle conformément à leur mandat.</p>
<p><b>Transparence et explicabilité</b>                  Les acteurs de l'IA doivent s'engager à divulguer de manière responsable des informations sur les systèmes d'IA. Ils devraient fournir des informations qui favorisent la compréhension générale des systèmes d'IA.</p>	<p><b>Transparence</b>                  L'explicabilité d'un système d'IA porte sur la capacité à expliquer à la fois les processus techniques de celui-ci et les décisions humaines associées (p. ex. les domaines d'application d'un système). L'explicabilité technique nécessite que les décisions prises par un système d'IA puissent être comprises et tracées par des êtres humains. En outre, des arbitrages peuvent être faits entre la nécessité d'améliorer l'explicabilité du système (qui peut réduire sa précision) et d'accroître sa précision (au détriment de son explicabilité). Chaque fois qu'un système d'IA a des incidences significatives sur la vie des personnes, il doit être possible de demander une explication convenable sur le processus de prise de décision du système d'IA. Une telle explication devrait être donnée en temps utile et adaptée au niveau d'expertise de la partie prenante concernée (p. ex. un citoyen, un régulateur ou un chercheur). En outre, des explications sur la mesure dans laquelle le système d'IA influence et façonne le processus de prise de décision organisationnel, les choix de conception du système et la logique de son déploiement devraient être disponibles (assurant ainsi la transparence du modèle économique).</p>

<p><b>Transparence et explicabilité</b> Fournir des informations utiles pour que les parties prenantes aient conscience de leurs interactions avec des systèmes d'IA.</p>	<p><b>Transparence</b> Les systèmes d'IA ne doivent pas être présentés aux utilisateurs comme des êtres humains ; les êtres humains ont le droit d'être informés qu'ils interagissent avec un système d'IA. Cela implique que les systèmes d'IA doivent être identifiables en tant que tels. En outre, la possibilité de se prononcer contre cette interaction et en faveur d'une interaction humaine devrait être fournie si nécessaire pour garantir le respect des droits fondamentaux.</p>
<p><b>Transparence et explicabilité</b> Fournir des informations utiles aux personnes concernées par l'IA pour qu'elles en comprennent les résultats.</p>	<p><b>Transparence</b> Les capacités et les limites d'un système d'IA doivent être communiquées aux praticiens de l'IA ou aux utilisateurs finaux selon des modalités adaptées au contexte d'utilisation concerné. Cela pourrait comprendre la communication du niveau de précision du système d'IA, ainsi que de ses limites.</p>
<p><b>Transparence et explicabilité</b> Fournir des informations utiles pour permettre aux personnes qui subissent les effets néfastes de l'IA de contester ses résultats.</p>	<p><b>Action et contrôle humains</b> Les utilisateurs doivent être en mesure de prendre des décisions autonomes et éclairées concernant les systèmes d'IA. Les connaissances et les outils pour comprendre et interagir avec des systèmes d'IA de manière satisfaisante devraient leur être fournis et ils devraient pouvoir, dans la mesure du possible, raisonnablement auto-évaluer et contester le système de manière appropriée. Les systèmes d'IA devraient aider les individus à prendre de meilleures décisions et à faire des choix davantage en connaissance de cause et en rapport avec leurs objectifs. Les systèmes d'IA peuvent parfois être déployés pour façonner et influencer les comportements humains au moyen de mécanismes qui peuvent être difficiles à détecter puisqu'ils peuvent exploiter des processus inconscients, y compris diverses formes de manipulation déloyale, de tromperie, d'asservissement et de conditionnement, toutes pouvant menacer l'autonomie individuelle. Le principe général d'autonomie de l'utilisateur doit être au cœur de la fonctionnalité du système.</p> <p><b>Responsabilité</b> Lorsque des effets négatifs injustes se produisent, des mécanismes accessibles doivent être prévus pour permettre une juste réparation. Il est essentiel de savoir qu'une réparation est possible en cas de problème pour garantir la confiance. Une attention particulière devrait être accordée aux personnes et aux groupes vulnérables.</p>
<p><b>Robustesse et sécurité</b> Les systèmes d'IA doivent être robustes, sûrs et sécurisés tout au long du cycle de vie, de manière à ne pas présenter de risque de sécurité excessif dans des conditions d'utilisation normales et prévisibles, ou en cas d'utilisation abusive.</p>	<p><b>Robustesse technique et sécurité</b> Il doit être garanti que le système fera ce qu'il est censé faire sans porter préjudice aux êtres humains ou à l'environnement. Cela comprend la réduction au minimum des conséquences et des erreurs inattendues. En outre, des processus permettant de clarifier et d'évaluer les risques potentiels associés à l'utilisation des systèmes d'IA dans divers domaines d'application devraient être mis en place. Le niveau des mesures de sécurité dépend de l'ampleur du risque posé par le système d'IA qui, à son tour, dépend des capacités du système. Lorsqu'il est prévisible que le processus de développement ou le système lui-même présentera des risques particulièrement élevés, il est essentiel que des mesures de sécurité soient développées et testées de manière proactive.</p>

<p><b>Robustesse et sécurité</b>                  Les acteurs de l'IA doivent garantir la traçabilité, y compris en ce qui concerne les ensembles de données, les processus et les décisions prises, afin de permettre l'analyse des résultats et le traitement des demandes d'informations, compte tenu du contexte et de l'état des connaissances.</p>	<p><b>Robustesse technique et sécurité</b>                  Il est essentiel que les résultats des systèmes d'IA soient reproductibles et fiables. Un système d'IA fiable est un système qui fonctionne correctement avec un éventail d'informations et dans diverses situations. Cela est nécessaire pour examiner un système d'IA et éviter les préjudices involontaires. La reproductibilité décrit si une expérimentation de l'IA présente le même comportement quand elle est répétée dans les mêmes conditions. Cela permet aux scientifiques et décideurs politiques de décrire de façon précise ce que les systèmes d'IA font. Des fichiers de reproduction peuvent faciliter le processus de test et de reproduction des comportements.</p> <p><b>Transparence</b>                  Les ensembles de données et les processus qui conduisent à la décision du système d'IA, y compris les processus de collecte et d'étiquetage des données ainsi que des algorithmes utilisés, devraient être documentés dans les meilleures conditions possibles pour assurer la traçabilité et une transparence accrue. Cela s'applique également aux décisions prises par le système d'IA. Cela permet de déterminer les raisons pour lesquelles une décision réalisée par l'IA était fautive, ce qui peut à son tour aider à éviter de futures erreurs. La traçabilité facilite la vérifiabilité et l'explicabilité.</p> <p><b>Responsabilité</b>                  La vérifiabilité implique la possibilité d'évaluer les algorithmes, les données et les processus de conception. Cela ne signifie pas nécessairement que les informations sur les modèles économiques et sur la propriété intellectuelle liés au système d'IA doivent toujours être librement disponibles. Une évaluation par des auditeurs internes et externes et la disponibilité des rapports d'évaluation peuvent contribuer à rendre la technologie digne de confiance. Pour les applications qui ont un effet sur les droits fondamentaux, y compris les applications présentant un risque critique pour la sécurité, les systèmes d'IA devraient pouvoir être vérifiés de manière indépendante.</p>
<p><b>Robustesse et sécurité</b>                  Les acteurs de l'IA devraient, en fonction du contexte, de leur rôle et de leurs capacités, adopter une approche systématique de gestion des risques à chaque phase du cycle de vie d'un système d'IA afin de gérer les risques liés à l'IA.</p>	<p><b>Action et contrôle humains</b>                  Compte tenu de la portée et de la capacité des systèmes d'IA, ceux-ci peuvent avoir des effets négatifs sur les droits fondamentaux. Dans les cas où de tels risques existent, une analyse d'impact sur les droits fondamentaux devrait être réalisée. Cela devrait être fait en amont du développement du système et inclure une évaluation concernant une possible réduction de ces risques ou leur justification éventuelle dans une société démocratique afin de respecter les droits et les libertés d'autrui. En outre, des mécanismes devraient être mis en place pour recevoir les commentaires extérieurs.</p> <p><b>Responsabilité</b>                  La vérifiabilité implique la possibilité d'évaluer les algorithmes, les données et les processus de conception. Cela ne signifie pas nécessairement que les informations sur les modèles économiques et sur la propriété intellectuelle liés au système d'IA doivent toujours être librement disponibles. Une évaluation par des auditeurs internes et externes et la disponibilité des rapports d'évaluation peuvent contribuer à rendre la technologie</p>

	digne de confiance. Pour les applications qui ont un effet sur les droits fondamentaux, y compris les applications présentant un risque critique pour la sécurité, les systèmes d'IA devraient pouvoir être vérifiés de manière indépendante.
<p><b>Responsabilité</b> Les acteurs de l'IA devraient être responsables du bon fonctionnement des systèmes d'IA et du respect des principes susmentionnés.</p>	<p><b>Responsabilité</b> L'exigence de responsabilité complète les exigences susmentionnées. Elle est étroitement liée au <i>principe d'équité</i>. Cela nécessite que des mécanismes soient mis en place pour garantir la responsabilité à l'égard des systèmes d'IA et de leurs résultats, tant avant qu'après leur développement, leur déploiement et leur utilisation.</p> <p><b>Action et contrôle humains</b> Il doit être garanti que les autorités publiques ont la capacité d'exercer un contrôle conformément à leur mandat. Les mécanismes de contrôle peuvent être requis à différents degrés pour soutenir d'autres mesures de sécurité et de contrôle, en fonction du domaine d'application du système d'IA et du risque potentiel. Toutes choses étant égales par ailleurs, moins un être humain pourra exercer un contrôle sur son système d'IA, plus il sera nécessaire que soient réalisés des tests approfondis et que soit mise en place une gouvernance stricte.</p>

**Récemment publiés dans cette série (en anglais):**

- 06/2019 : Gas Supply in the EU (juillet 2019)
- 05/2019 : Road Safety Management (juillet 2019)
- 04/2019 : Internal Electricity Market (juin 2019)
- 03/2019 : The EU Green Bond Standard (juin 2019)
- 02/2019 : Energy Union Governance (mai 2019)
- 01/2019 : Renewable Energy in the EU (avril 2019)
- 05/2018 : Energy Efficiency Policy (décembre 2018)
- 04/2018 : Climate Protection outside the EU ETS (août 2018)
- 03/2018 : Climate Protection by way of the EU ETS (juillet 2018)
- 02/2018 : French Vocational Training (février 2018)

**L'auteur :**

Alessandro Gasparotti est économiste au département *Marché intérieur* du cep | Centre de politique européenne.

Texte original : Ethics Guidelines on Artificial Intelligence

Texte traduit et publié par :

**cepFrance | Centre de politique européenne**

350, rue Lecourbe | F-75015 Paris

Téléphone +33 1 45 54 91 50 | [france.cep.eu](http://france.cep.eu)

Le cepFrance est un think tank spécialisé sur les questions relatives à l'intégration européenne, partenaire du cep | Centrum für europäische Politik de Fribourg-en-Brigau (Allemagne). C'est un centre d'excellence indépendant qui examine, analyse et évalue les projets politiques de l'Union européenne.