



Commission
européenne

*Groupe Européen d'Éthique
des Sciences et des
Nouvelles Technologies*

Déclaration sur
**L'intelligence
artificielle,
la robotique
et les
systèmes
« autonomes »**

Recherche et
Innovation

Groupe Européen d'Éthique des Sciences et des Nouvelles Technologies
Déclaration sur l'intelligence artificielle, la robotique et les systèmes « autonomes »

Commission Européenne
Direction Générale de la Recherche et de l'Innovation
Unité RTD.01 — Mécanisme de Consultation Scientifique

Contact Jim Dratwa, directeur du bureau du GEE
Courriel EC-ETHICS-GROUP@ec.europa.eu
RTD-PUBLICATIONS@ec.europa.eu

Commission Européenne
B-1049 Bruxelles

Imprimé par l'Office de Publications au Luxembourg

Manuscrit achevé en mars 2018

Ni la Commission Européenne ni aucune personne agissant au nom de la Commission n'est responsable de l'usage qui pourrait être fait des informations données ci-après.

Le contenu de cette déclaration relève de la seule responsabilité du Groupe Européen d'Éthique des Sciences et des Nouvelles Technologies (GEE). Bien que des membres du personnel des services de la Commission aient participé à la préparation de cette déclaration, les avis qu'elle exprime reflètent l'opinion collective du GEE et ne peuvent, en aucune circonstance, être considérés comme constituant une prise de position officielle de la Commission Européenne.

Cette déclaration du GEE a été adoptée par les membres du GEE suivants : Emmanuel Agius, Anne Cambon-Thomsen, Ana Sofia Carvalho, Eugenijus Gefenas, Julian Kinderlerer, Andreas Kurtz, Jonathan Montgomery, Herman Nys (vice-président), Siobhán O'Sullivan (vice-présidente), Laura Palazzani, Barbara Prainsack, Carlos Maria Romeo Casabona, Nils-Eric Sahlin, Jeroen van den Hoven, Christiane Woopen (présidente). Rapporteur : Jeroen van den Hoven.

Vous trouverez des informations complémentaires sur l'Union européenne sur Internet (<http://europa.eu>)

Luxembourg : Office des Publications de l'Union Européenne, 2018

Impression	ISBN 978-92-79-88936-3	doi:10.2777/77587	KI-02-18-803-FR-C
PDF	ISBN 978-92-79-88941-7	doi:10.2777/5194	KI-02-18-803-FR-N

© Union Européenne, 2018

Réutilisation autorisée moyennant mention de la source. La politique de réutilisation des documents de la Commission Européenne est régie par la décision 2011/833/UE (JO L 330 du 14.12.2011, p. 39).

Toute utilisation ou reproduction de photos ou d'autres documents dont l'UE n'est pas titulaire des droits d'auteur est interdite sans l'autorisation des titulaires des droits d'auteur.

Image de couverture © EVZ, #141003388, 2018. Source: Fotolia.com

Déclaration sur
**L'intelligence artificielle,
la robotique et les systèmes
« autonomes »**

*Groupe Européen D'Éthique
des Sciences et des Nouvelles Technologies*

Bruxelles, le 9 mars 2018

Table des matières

RÉSUMÉ	5
CONTEXTE	6
RÉFLEXIONS MORALES	9
Questions clés	9
Considérations essentielles	10
Au-delà d'un cadre éthique étroit	12
VERS UN CADRE ÉTHIQUE PARTAGÉ POUR L'INTELLIGENCE ARTIFICIELLE, LA ROBOTIQUE ET LES SYSTEMES « AUTONOMES »	14
PRINCIPES ÉTHIQUES ET CONDITIONS PRÉALABLES DÉMOCRATIQUES	17

Résumé

Les progrès de l'intelligence artificielle (IA), de la robotique et des technologies dites « autonomes »¹ s'accompagnent de nombreuses questions morales, dont l'urgence et la complexité ne cessent de croître. Les efforts actuels pour trouver des réponses aux défis éthiques, sociétaux et juridiques posés par ces progrès, et pour les orienter en faveur de l'intérêt général, se limitent à une mosaïque d'initiatives disparates. Cela souligne la nécessité d'un processus de réflexion et de dialogue collectif, large et inclusif, un dialogue qui se concentre sur les valeurs autour desquelles nous voulons organiser la société et sur le rôle que les technologies doivent jouer.

La présente déclaration appelle au lancement d'un processus qui ouvrirait la voie à un cadre éthique et juridique commun, reconnu au niveau international, concernant la conception, la production, l'utilisation et la gouvernance de l'intelligence artificielle, la robotique et les systèmes « autonomes ». Elle propose aussi un ensemble de principes éthiques fondamentaux, basés sur les valeurs établies par les traités et la charte des droits fondamentaux de l'UE, qui serviraient de guide à l'élaboration de ce processus.

¹ Cette déclaration concerne un ensemble de technologies numériques intelligentes qui convergent rapidement et sont souvent interdépendantes, connectées ou totalement intégrées, par exemple l'intelligence artificielle classique, les algorithmes d'apprentissage automatique, l'apprentissage profond et les réseaux connexionnistes, les réseaux antagonistes génératifs, la mécatronique et la robotique. Les véhicules autonomes et les armes robotisées, les agents conversationnels (« chat bots ») et les systèmes de reconnaissance des images et de la voix sont des exemples bien connus de combinaisons de ces technologies.

Contexte

Les deux premières décennies du XXI^e siècle nous offrent des exemples éclatants de ce que l'on qualifie communément de « technologie autonome » et d'« intelligence artificielle ». Véhicules autonomes, drones, robots d'exploration des fonds marins et de l'espace, systèmes d'armes, agents logiciels comme les bots de transactions financières, ou encore apprentissage profond dans le domaine des diagnostics médicaux, sont parmi les exemples les plus marquants, mais ils sont loin d'être les seuls. L'intelligence artificielle (IA), notamment sous forme d'apprentissage automatique, et la mise à disposition croissante d'énormes ensembles de données dans des domaines divers, sont des moteurs importants pour ces évolutions. La convergence de ces technologies numériques en accroît rapidement la puissance; elles sont appliquées dans un nombre grandissant de nouveaux produits et services, dans le secteur privé comme public, et peuvent avoir des applications civiles comme militaires. L'IA intégrée dans ces systèmes peut redéfinir les tâches ou améliorer les conditions du travail humain, et réduire la nécessité pour l'homme de contribuer, d'intervenir et d'interférer au cours des opérations. Elle peut aider à assister ou à remplacer les hommes par des technologies intelligentes dans des situations difficiles, pour des travaux salissants, ennuyeux, dangereux ou autres.

Sans intervention humaine directe ni contrôle extérieur, les systèmes intelligents sont aujourd'hui en mesure d'entretenir un dialogue avec des clients dans des centres d'appels en ligne, de piloter des bras robotisés afin de ramasser et de manipuler des objets de façon précise et ininterrompue, d'acheter et vendre de grandes quantités d'actions en quelques millisecondes, de commander à un véhicule de dévier ou de freiner pour éviter une collision, de

classer des personnes en fonction de leur comportement, ou d'infliger des amendes.

Malheureusement, certains de ces outils cognitifs les plus puissants sont aussi les plus opaques. Leurs actions ne sont plus programmées par l'homme de manière linéaire. Google Brain développe une IA prétendument capable d'élaborer d'autres IA plus rapidement et plus efficacement que l'homme ne pourrait le faire. AlphaZero peut apprendre seul les règles du jeu d'échecs en quatre heures, en partant de zéro et jusqu'à atteindre le niveau d'un champion du monde. Il est impossible de comprendre exactement comment AlphaGo a réussi à battre le champion du monde de jeu de Go. L'apprentissage profond et les « approches de réseaux antagonistes génératifs » permettent aux machines de s'enseigner mutuellement de nouvelles stratégies et de rechercher de nouveaux éléments à analyser. En ce sens, leurs actions échappent souvent à la compréhension et ne sont plus à la portée du contrôle par l'homme. C'est le cas d'abord parce qu'il est impossible d'établir comment elles obtiennent leurs résultats, au-delà des algorithmes initiaux. Ensuite, leurs performances sont basées sur des données utilisées pendant le processus d'apprentissage, mais qui ne sont plus nécessairement disponibles ou accessibles. Ainsi, les distorsions et erreurs qui leur ont été présentées par le passé s'ancrent dans le système.

Lorsque des systèmes peuvent apprendre à réaliser ces tâches sans instruction ni supervision humaine, on les qualifie désormais d'« autonomes ». Ces systèmes « autonomes » peuvent prendre la forme de systèmes robotiques high-tech ou d'agents logiciels intelligents tels que les bots. Beaucoup d'entre eux sont déployés dans le monde sans supervision et accomplissent des tâches qui n'étaient pas prévues par leurs concepteurs ou propriétaires humains.

Nous assistons par conséquent aux évolutions technologiques suivantes :

- (1) L'intelligence artificielle sous forme d'apprentissage automatique (notamment l'apprentissage profond), alimenté par les mégadonnées, gagne rapidement en puissance. Elle est appliquée dans un nombre croissant de nouveaux produits et services numériques, dans les secteurs privé et public, et peut avoir des applications militaires ou civiles. Comme mentionné précédemment, les mécanismes internes de l'IA peuvent être extrêmement difficiles (si ce n'est impossibles) à tracer, expliquer et évaluer de façon critique. Ces capacités avancées s'accumulent en grande partie entre des mains privées et sont, pour une large part, exclusives.

- (2) La mécatronique avancée (combinaison d'IA et d'apprentissage profond, de science des données, de technologie des capteurs, d'Internet des objets et d'ingénierie mécanique et électrique) offre une large gamme de systèmes robotiques et high-tech de plus en plus sophistiqués, pour des applications pratiques dans l'industrie des services et de la production, la santé, la distribution, la logistique, la domotique (automatisation des bâtiments) ainsi que dans le domaine de la sûreté et de la sécurité. Les deux domaines d'application qui prédominent dans les débats publics sont les systèmes d'armes robotisés et les véhicules « autonomes ».
- (3) Des systèmes toujours plus intelligents sont produits et affichent un degré élevé de ce qui est souvent qualifié d'« autonomie », ce qui signifie qu'ils se développent et peuvent accomplir des tâches indépendamment des opérateurs humains et sans contrôle de leur part.
- (4) Il semble y avoir un mouvement en faveur de niveaux d'automatisation et d'« autonomie » toujours plus élevés dans le domaine de la robotique, de l'IA et de la mécatronique. Les investissements des États et des grandes compagnies dans ce domaine sont énormes et l'un des objectifs majeurs des superpuissances du monde est de parvenir à occuper une position de leader dans la recherche en IA.
- (5) On assiste aussi à une évolution vers une interaction toujours plus étroite entre l'homme et la machine (cobots, cyber-équipages, jumeaux numériques et même intégration de machines intelligentes dans le corps humain, sous forme d'interfaces ordinateur-cerveau ou cyborgs). Des développements similaires sont constatés dans l'univers de l'IA. Des équipes bien préparées de systèmes d'IA et de professionnels obtiennent de meilleures performances dans certains domaines que les hommes et les machines séparément.

Réflexions morales

Questions clés

L'avancée des systèmes et logiciels high-tech capables de fonctionner de plus en plus indépendamment de l'homme, et d'exécuter des tâches qui requièrent de l'intelligence lorsqu'elles sont accomplies par l'homme, mérite réflexion. Ces systèmes inspirent de nombreuses questions morales importantes et difficiles.

Tout d'abord, des questions sur la sécurité, la sûreté, la prévention des préjudices et l'atténuation des risques. Comment pouvons-nous assurer la sécurité et la sûreté d'un monde faisant appel à l'IA interconnectée et aux dispositifs « autonomes », et comment évaluer les risques associés?

Ensuite, se posent des questions de responsabilité morale de l'homme. Où se situe l'entité moralement pertinente dans les systèmes sociotechniques dynamiques et complexes de l'IA et des composants robotiques avancés? Comment attribuer et répartir la responsabilité morale, et qui est responsable (et en quels termes) en cas de résultats indésirables? Peut-on parler de « contrôle partagé » et de « responsabilité conjointe » entre des humains et des machines intelligentes? L'homme fera-t-il partie des écosystèmes de dispositifs « autonomes » en tant que « zone tampon », servant juste à absorber la responsabilité, ou sera-t-il en mesure d'assumer réellement la responsabilité des actes posés par ces systèmes?

Troisièmement, cette évolution soulève des questions en matière de gouvernance, de réglementation, de conception, de développement, d'inspection, de surveillance, de test et de certification. Comment faut-il modifier nos institutions et nos lois afin qu'elles assurent le bien-être des personnes et de la société et fassent de cette société un endroit sûr face à ces technologies?

Quatrièmement, cela pose des questions quant à la prise de décision démocratique, y compris au sujet des institutions, des politiques et des valeurs qui sous-tendent toutes les interrogations précédentes. Des études sont menées partout dans le monde pour établir la mesure dans laquelle les citoyens se laissent abuser par le recours à des techniques de coup de pouce basées sur une combinaison d'apprentissage automatique, de mégadonnées et de science du comportement, qui rendent possibles le profilage discret, le microciblage, la personnalisation et la manipulation de structurations de choix servant des objectifs commerciaux ou politiques.

Enfin, les capacités d'explication et la transparence de l'IA et des systèmes « autonomes » posent aussi question. Quelles valeurs ces systèmes servent-ils visiblement et réellement? Sur quelles valeurs s'appuie la conception de nos politiques et de nos machines? Autour de quelles valeurs voulons-nous organiser nos sociétés? Quelles valeurs, enfin, sommes-nous prêts à laisser ébranler (ouvertement ou en silence) par le progrès technologique et les compromis au nom de l'utilité? L'« optimisation » des processus sociaux entraînée par l'IA et basée sur des systèmes d'évaluation sociale que certains pays expérimentent, viole l'idée fondamentale d'égalité et de liberté de la même manière que les systèmes de castes, parce qu'elle crée «différentes catégories d'individus» là où il n'y a en réalité que différentes «caractéristiques» des individus. Comment empêcher ces attaques contre les systèmes démocratiques et cette utilisation de systèmes d'évaluation comme fondement de la domination de ceux qui ont accès à ces puissantes technologies?

Considérations essentielles

D'un point de vue éthique, il est important de garder à l'esprit que:

Le terme « autonomie » est un terme philosophique qui fait référence à la capacité de l'homme à légiférer pour lui-même, à formuler, penser et choisir des normes, des règles et des lois à respecter. Cela comprend le droit de définir librement ses propres normes et de choisir ses propres buts et objectifs de vie. Les processus cognitifs qui accompagnent et facilitent cela sont parmi ceux qui sont le plus étroitement associés à la dignité d'être humain, à la nature et à l'activité humaines par excellence. Ils impliquent généralement les aspects de connaissance de soi, de conscience de soi et de conscience de sa propre capacité à créer, conformément à des raisons et des valeurs. Le terme d'autonomie dans le sens éthique du terme peut donc uniquement être appliqué aux êtres humains. Il est donc quelque peu inapproprié pour désigner des objets, aussi avancés, complexes, adaptatifs ou même « intelligents » soient-ils. La terminologie de systèmes « autonomes » a cependant gagné du terrain dans la littérature scientifique et dans le débat public pour désigner le degré le plus élevé d'automatisation et d'indépendance vis-à-vis de l'homme, en matière d'« autonomie » opérationnelle et décisionnelle. Mais l'autonomie au sens premier est un aspect important de la dignité humaine qui ne devrait pas être relativisé.

Puisqu'aucun objet ou système (aussi avancé et sophistiqué soit-il) ne peut en lui-même être qualifié d'« autonome » au sens éthique original, on ne peut pas lui accorder le statut moral de la personne humaine et lui octroyer

la dignité humaine. La dignité humaine, en tant que fondement des droits de l'homme, implique qu'une intervention et une participation humaines significatives doivent être possibles pour ce qui concerne les hommes et leur environnement. Par conséquent, contrairement à l'automatisation de la production, il n'est pas approprié de gérer le sort des hommes et d'en décider de la même manière que nous gérons et décidons de ce qu'il advient des objets ou des données, même si c'est techniquement concevable. Une telle gestion « autonome » des êtres humains serait contraire à l'éthique et nuirait aux valeurs européennes fondamentales et profondément enracinées. L'homme doit pouvoir déterminer quelles sont les valeurs que sert la technologie, ce qui est moralement pertinent et quels sont les objectifs et conceptions du bien qui valent la peine d'être poursuivis. Cela ne peut pas être laissé à des machines, quelle que soit leur puissance.

La capacité et la volonté d'assumer et d'attribuer une responsabilité morale font partie intégrante de la conception de la personne sur laquelle reposent toutes nos institutions morales, sociales et juridiques. La responsabilité morale est ici comprise au sens large, qui peut faire référence à divers aspects de la nature humaine: causalité, responsabilité (obligation de rendre des comptes et obligation de compenser les préjudices), attitudes réactives telles que l'éloge et le blâme (pertinence d'une variété d'émotions morales) et devoirs associés aux fonctions sociales. La responsabilité morale, quel que soit son sens, ne peut pas être attribuée ou transférée à une technologie « autonome ».

Lors des récents débats au sujet des systèmes d'armes létales autonomes (SALA) et des véhicules autonomes, il semble que la nécessité d'un *contrôle important de la part de l'être humain* pour assurer la responsabilité morale ait fait consensus. Le principe de contrôle important de la part de l'être humain a été suggéré au départ pour limiter le développement et l'utilisation des futurs systèmes d'armes. Il implique que les êtres humains (et non les ordinateurs et leurs algorithmes) doivent fondamentalement rester aux commandes et, de ce fait, demeurer moralement responsables².

² ONG Article 36, 2015

Au-delà d'un cadre éthique étroit

Les deux domaines dans lesquels le développement de systèmes « autonomes » a déjà donné lieu à des débats éthiques intenses sont les véhicules autonomes et les systèmes d'armes létales autonomes. Bien que les voitures capables de rouler entièrement sans conducteur ne soient pas encore commercialisées, plusieurs pays se préparent à la possibilité d'autoriser les véhicules « autonomes » sur la voie publique. En 2016, une controverse morale a éclaté lors du décès de la première victime d'un accident impliquant une voiture conduite en mode « autonome ». Les débats moraux se limitent désormais souvent à des cas d'utilisation exceptionnels concernant les expériences de pensée qualifiées de « dilemme du tramway ». Ces cas s'intéressent au dilemme des accidents inévitables, dans lesquels le seul choix possible est lié à la perte de vies humaines. Cette conception restrictive des problèmes éthiques incite à une approche par le calcul et implique souvent des facteurs exagérément simplistes dans le domaine humain. Les questions centrales dans ce cadre semblent principalement concerner la responsabilité des systèmes « autonomes », leurs effets et la façon dont ils doivent être programmés pour que leur mise en œuvre conduise à des résultats moralement acceptables du point de vue des pertes en vies humaines par rapport aux vies sauvées. Cela omet des questions plus larges comme «quelles décisions de conception ont été prises par le passé pour aboutir à cette impasse morale?», «quelles valeurs doivent guider la conception?», «quel poids doivent avoir les valeurs associées à la conception en cas de conflit et à qui appartient-il de l'évaluer?», «quel est le statut des observations empiriques massives accumulées concernant la façon dont les personnes tranchent les cas de dilemme du tramway, qui sont transposées pour la configuration des véhicules automatisés?»

Le deuxième domaine de contestation et de controverse est celui des systèmes d'armes « autonomes ». Ces systèmes militaires peuvent être équipés d'armes létales, mais en ce qui concerne la dimension logicielle, ils ne sont pas très différents des systèmes « autonomes » que l'on peut trouver dans divers domaines civils de proximité. Une grande partie du débat se déroule lors de la conférence d'examen de la Convention sur certaines armes classiques, à Genève, en ce qui concerne l'acceptabilité morale des armes « autonomes » et la responsabilité juridique et morale du déploiement de ces systèmes. Mais l'attention doit se porter sur des questions relatives à la nature et à la signification du « contrôle important de la part de l'être humain » sur ces systèmes et sur la façon d'instituer des formes de contrôle moralement souhaitables.

Un troisième domaine d'application important est le logiciel « autonome », qui inclut les bots. Les marchés commerciaux, financiers et boursiers sont largement gérés par des algorithmes et des logiciels. Sans intervention humaine ni contrôle extérieur, les systèmes intelligents mènent désormais des dialogues avec les clients dans des centres d'appels en ligne; les interfaces de reconnaissance vocale et les systèmes de recommandation des plateformes en ligne, comme Siri, Alexa et Cortana, émettent des suggestions à l'intention des utilisateurs. Au-delà des questions directes de protection des données et de la vie privée, on peut se demander si les personnes concernées ont le droit de savoir si elles traitent avec un être humain ou avec un dispositif d'intelligence artificielle. De plus, la question se pose de savoir s'il doit y avoir des limites à ce que les systèmes d'IA peuvent suggérer à une personne, en se basant sur une interprétation de la propre conception de son identité par cette personne.

Alors que la prise de conscience du besoin de répondre à ces questions est grandissante, l'IA et la robotique progressent plus rapidement que le processus de réponse à ces épineuses questions éthiques, juridiques et sociétales. Les efforts actuels s'apparentent à une mosaïque d'initiatives disparates. Il est à l'évidence nécessaire d'établir un processus collectif, large et inclusif, qui ouvrirait la voie à un cadre éthique et juridique commun, reconnu au niveau international, concernant la conception, la production, l'utilisation et la gouvernance des systèmes robotiques, « autonomes » et d'intelligence artificielle.

Cette déclaration appelle au lancement d'un tel processus et propose un ensemble de principes éthiques fondamentaux et de conditions démocratiques préalables qui pourraient aussi guider la réflexion sur un droit contraignant. Le GEE est d'avis que l'Europe devrait jouer un rôle actif et de premier plan en la matière. Dans le cadre de la supervision des débats sur la responsabilité morale de l'IA et des technologies dites « autonomes », le GEE appelle à une réflexion et à des études plus systématiques sur les aspects éthiques, juridiques et de gouvernance de ces systèmes high-tech qui agissent sur le monde sans contrôle direct par leurs utilisateurs humains, au bénéfice ou au détriment de l'homme. Il y a urgence.

Vers un cadre éthique partagé pour l'intelligence artificielle, la robotique et les systèmes « autonomes »

Certaines des initiatives les plus importantes en faveur de la formulation de principes éthiques relatifs à l'IA et aux systèmes « autonomes » sont issues de l'industrie, des professionnels et de leurs associations, telles que le document d'orientation de l'IEE (Institute of Electrical and Electronics Engineers) sur la « conception alignée sur l'éthique »,³ le sommet mondial de l'UIT (Union internationale des télécommunications) consacré à « L'intelligence artificielle au service du bien »⁴ à l'été 2017, et les travaux de l'ACM (Association for Computing Machinery) sur le sujet, dont une conférence majeure AAAI/ACM sur l'intelligence artificielle, l'éthique et la société,⁵ en février 2018. Dans le secteur privé, des sociétés comme IBM, Microsoft et DeepMind de Google ont établi leurs propres codes d'éthique sur l'IA et ont uni leurs forces pour créer d'importantes initiatives comme le « Partenariat sur l'IA »⁶ ou l'« OpenAI »⁷, qui réunissent industriels, organisations à but non lucratif et organismes universitaires.

L'une des initiatives majeures appelant à un développement responsable de l'IA a été lancée par le « Future of Life Institute » et a donné lieu à la création des « principes d'Asilomar encadrant l'IA ». Cette liste de 23 principes fondamentaux d'orientation de la recherche et des applications de l'IA a été signée par des centaines de parties prenantes,⁸ issues en majeure partie du monde scientifique, de la recherche et de l'industrie. Un processus participatif similaire a été lancé à l'initiative du Forum sur le développement socialement responsable de l'intelligence artificielle, organisé par l'université de Montréal en novembre 2017, à la suite duquel une « Déclaration pour un développement responsable de l'intelligence artificielle » a été élaborée. Elle est désormais accessible au public sur une plateforme en ligne où tous les secteurs de la société sont invités à en commenter le texte⁹.

Un débat mondial sur l'utilisation militaire de l'IA a été lancé par les Nations unies et les réunions de la Convention sur l'emploi de certaines armes

³ http://standards.ieee.org/news/2016/ethically_aligned_design.html

⁴ <https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx>

⁵ <http://www.aies-conference.com/>

⁶ <https://www.partnershiponai.org/>

⁷ <https://openai.com/>

⁸ <https://futureoflife.org/ai-principles/>

⁹ <http://nouvelles.umontreal.ca/article/2017/11/03/declaration-de-montreal-pour-un-developpement-responsable-de-l-intelligence-artificielle/>

classiques (CCAC, Genève), où plusieurs hautes parties contractantes ont approuvé le principe dit du « contrôle humain effectif » pour les SALA, affirmant que « les systèmes d'armes autonomes qui ne requièrent aucun contrôle humain effectif devraient être interdits » (Assemblée générale des Nations unies, 2016). Les Nations unies ont aussi établi un institut de recherche spécial, à La Haye, chargé d'étudier la gouvernance de la robotique et de l'IA (UNICRI)¹⁰. Plusieurs initiatives et ONG qui s'intéressent à l'IA et aux systèmes « autonomes » « au service du bien » font quant à elles campagne pour une interdiction des armes « autonomes ». C'est le cas par exemple de la « Foundation for Responsible Robotics ».

Cependant, au niveau international, les initiatives sont variables, certains pays donnant la priorité au développement de règles pour la robotique et l'intelligence artificielle, allant jusqu'à adopter une législation (par exemple, pour réglementer les voitures autonomes sur la voie publique), tandis que d'autres n'ont pas encore abordé le sujet. Cette absence d'approche européenne harmonisée a poussé le Parlement européen à appeler à une série de mesures visant à préparer la réglementation de la robotique avancée,¹¹ y compris par l'élaboration d'un cadre d'orientation éthique pour la conception, la production et l'utilisation des robots.

Dans ce contexte, le GEE attire l'attention sur les risques inhérents à des approches non coordonnées et déséquilibrées en matière de réglementation de l'IA et des technologies « autonomes ». Les mosaïques réglementaires peuvent donner lieu à un « shopping éthique » qui entraînerait la délocalisation des évolutions et de l'utilisation de l'IA vers des régions dans lesquelles les normes éthiques seraient moins rigoureuses. Permettre au débat de se concentrer autour de certaines régions, disciplines, éléments démographiques ou acteurs industriels risque d'exclure un ensemble plus vaste de points de vue et d'intérêts sociétaux. Les discussions actuelles manquent aussi parfois d'une vue d'ensemble des technologies « autonomes » susceptibles d'être étudiées, développées et mises en œuvre dans les dix ans à venir, créant un angle mort en matière d'anticipation réglementaire.

¹⁰ Il convient aussi de noter, à cet égard, sous l'égide de l'UNESCO: le rapport de la COMEST sur l'éthique de la robotique et le rapport du CIB sur les mégadonnées et la santé, tous deux adoptés en septembre 2017.

¹¹ Parlement européen, commission des affaires juridiques, rapport contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103[INL]), rapporteur Mady Delvaux.

Le GEE appelle à un engagement public large et systématique et à une délibération sur l'éthique de l'IA, de la robotique et de la technologie « autonome », ainsi que sur les valeurs que les sociétés choisissent d'intégrer dans le développement et la gouvernance de ces technologies. Ce processus, auquel le GEE est prêt à prendre part, devrait fournir une plateforme permettant de réunir les différentes initiatives internationales mentionnées ci-avant. Il devrait intégrer un débat sociétal large, inclusif et ambitieux, s'appuyant sur la contribution de divers points de vue et permettant à tous ceux qui disposent d'une expertise et de valeurs différentes de se faire entendre. Le GEE appelle l'Union européenne à se placer à l'avant-garde d'un tel processus et invite la Commission européenne à lancer et soutenir son élaboration.

Comme première étape vers la formulation d'un ensemble de lignes directrices éthiques pouvant servir de base à l'établissement d'une norme internationale et d'une action législative, **le GEE propose un ensemble de principes de base et de conditions préalables démocratiques, reposant sur les valeurs fondamentales édictées par les traités et la charte des droits fondamentaux de l'UE.**

Principes éthiques et conditions préalables démocratiques

(a) Dignité humaine

Le principe de la dignité humaine, compris comme la reconnaissance de l'état inhérent à l'homme d'être digne de respect, ne doit pas être violé par les technologies « autonomes ». Cela signifie par exemple qu'il y a des limites aux déterminations et classifications concernant des personnes, effectuées sur la base d'algorithmes et de systèmes « autonomes », notamment lorsque les personnes concernées n'en sont pas informées. Cela implique aussi qu'il doit y avoir des limites (légales) à la façon dont les gens peuvent être amenés à croire qu'ils traitent avec des êtres humains quand, en réalité, ils ont affaire à des algorithmes et des machines intelligentes. La conception relationnelle de la dignité humaine, caractérisée par nos relations sociales, exige que nous sachions si et quand nous interagissons avec une machine ou avec un autre être humain, et que nous nous réservions le droit de confier certaines tâches à l'homme ou à la machine.

(b) Autonomie

Le principe d'autonomie implique la liberté de l'être humain. Celle-ci se traduit en responsabilité humaine et, par conséquent, en contrôle et en connaissance sur les systèmes « autonomes », sans nuire à la liberté des êtres humains d'établir leurs propres normes et de pouvoir vivre en les appliquant. Toutes les technologies « autonomes » doivent donc respecter la capacité humaine à choisir si, quand et comment leur déléguer des décisions et des actions. Cela implique aussi la transparence et la prévisibilité des systèmes « autonomes », sans lesquelles les utilisateurs ne seraient pas en mesure d'intervenir ou d'y mettre fin s'ils l'estimaient nécessaire d'un point de vue moral.

(c) Responsabilité

Le principe de responsabilité doit occuper une place fondamentale dans la recherche et les applications de l'IA. Les systèmes « autonomes » doivent être développés et utilisés uniquement de manière à servir le bien social et environnemental général, tel que déterminé par les résultats de processus démocratiques délibératifs. Cela implique qu'ils doivent être conçus de façon

à ce que leurs effets soient cohérents avec une pluralité de valeurs et de droits humains. La possibilité d'abus d'utilisation des technologies « autonomes » constitue un défi majeur qui justifie qu'une importance cruciale soit accordée à la sensibilisation au risque et à une approche prudente. Les applications de l'IA et de la robotique ne devraient pas faire peser de risques inacceptables de nuisance sur les êtres humains ni compromettre la liberté et l'autonomie des hommes en réduisant subrepticement et de façon illégitime les choix et les connaissances des citoyens. Leur développement et leur utilisation devraient au contraire être adaptés de façon à accroître l'accès à la connaissance et à des opportunités pour les individus.

La recherche, la conception et le développement de l'IA, de la robotique et des systèmes « autonomes » devraient être orientés par une authentique préoccupation éthique de la recherche, une responsabilité sociale des développeurs et une coopération académique internationale visant à protéger les valeurs et droits fondamentaux en ayant pour objectif de concevoir des technologies au service de ces derniers et non à leurs dépens.

(d) Justice, équité et solidarité

L'IA devrait contribuer à la justice mondiale et à l'égalité d'accès aux bénéfices et avantages que l'IA, la robotique et les systèmes « autonomes » peuvent apporter. Les partis pris discriminatoires dans les ensembles de données utilisés pour former et faire fonctionner les systèmes d'IA devraient être empêchés ou détectés, signalés et neutralisés au stade le plus précoce possible.

Nous avons besoin d'un effort mondial concerté pour assurer un accès égal aux technologies « autonomes » et à une juste distribution de leurs bénéfices, ainsi qu'à une égalité des chances au sein de chaque société et entre les sociétés elles-mêmes. Cela inclut la formulation de nouveaux modèles de distribution équitable et de partage des bénéfices aptes à répondre aux transformations économiques causées par l'automatisation, la numérisation et l'IA, en assurant l'accessibilité des technologies d'IA essentielles et en facilitant la formation aux disciplines des sciences, de la technologie, de l'ingénierie et des mathématiques (STIM) et du numérique, notamment au niveau des régions et des groupes sociaux défavorisés. La vigilance est de rigueur face aux inconvénients de l'accumulation de données

massives et détaillées concernant les individus et qui vont peser sur le principe de solidarité, par exemple dans les systèmes d'assistance mutuelle comme l'assurance sociale et les systèmes de santé. Ces processus peuvent nuire à la cohésion sociale et donner naissance à un individualisme radical.

(e) Démocratie

Les décisions clés sur la réglementation du développement et de l'application de l'IA doivent être le résultat d'un débat démocratique et d'un engagement public. Un esprit de coopération international et un dialogue public sur le sujet assureront que ces décisions sont prises de façon inclusive, éclairée et clairvoyante. Le droit à recevoir une éducation ou l'accès aux informations sur les nouvelles technologies et leurs implications éthiques permettra à chacun de comprendre les risques et les opportunités et donnera la possibilité de participer aux processus décisionnels qui détermineront notre avenir.

Les principes de dignité humaine et d'autonomie impliquent fondamentalement le droit de l'homme à l'autodétermination au moyen de la démocratie. Les valeurs de pluralisme, de diversité et d'acceptation d'une variété de conceptions d'une vie heureuse pour les citoyens sont indispensables à nos systèmes politiques démocratiques. Elles ne doivent pas être mises en danger, perverties ou compensées par de nouvelles technologies qui inhibent ou influent sur la prise de décision politique et portent atteinte à la liberté d'expression et au droit de recevoir et de partager des informations sans ingérence. Les technologies numériques doivent au contraire servir à exploiter l'intelligence collective et à soutenir et améliorer les processus civiques sur lesquels reposent nos sociétés démocratiques.

(f) État de droit et responsabilité

L'état de droit, l'accès à la justice et le droit à la réparation des préjudices et à un procès équitable constituent un cadre nécessaire pour assurer le respect des normes des droits de l'homme et d'une éventuelle réglementation spécifique de l'IA. Ce cadre inclut des protections contre les risques associés aux systèmes « autonomes » susceptibles de porter atteinte à des droits humains comme la sécurité et la vie privée.

L'ensemble des difficultés juridiques apparaissant sur le terrain devraient être abordées en investissant de manière opportune dans l'élaboration de solutions robustes attribuant de manière claire et juste les responsabilités, et de mécanismes efficaces de droit contraignant.

À cet égard, les gouvernements et les organisations internationales devraient accroître leurs efforts de clarification en matière de détermination des responsabilités en cas de dommages causés par le comportement indésirable des systèmes « autonomes ». De plus, des systèmes efficaces d'atténuation des préjudices devraient être mis en place.

(g) Sécurité, sûreté, intégrité physique et mentale

La sécurité et la sûreté des systèmes « autonomes » se concrétisent de trois manières: (1) sécurité externe pour leur environnement et leurs utilisateurs, (2) fiabilité et robustesse interne, par exemple contre le piratage, et (3) sécurité émotionnelle en ce qui concerne les interactions homme-machine. Toutes ces dimensions de la sécurité doivent être prises en compte par les développeurs d'IA et faire l'objet de tests rigoureux avant leur mise sur le marché, afin d'assurer que les systèmes « autonomes » ne portent pas atteinte aux droits humains, à l'intégrité physique et mentale et à la sécurité de l'environnement. Une attention particulière doit ainsi être accordée aux personnes qui se trouvent en situation de vulnérabilité. Une vigilance particulière doit aussi être portée à la possibilité d'un double usage et de la militarisation de l'IA, par exemple dans le cadre de la cybersécurité, de la finance, des infrastructures et des conflits armés.

(h) Protection des données et de la vie privée

À une époque où le recueil massif de données par l'intermédiaire des technologies de communication numérique est omniprésent, le droit à la protection des informations personnelles et au respect de la vie privée est profondément mis en cause. Les robots physiques utilisant l'IA dans le cadre de l'Internet des objets, comme les agents logiciels qui fonctionnent sur le World Wide Web, doivent respecter les réglementations relatives à la protection des données et ne pas recueillir et diffuser de données ou être utilisés sur des ensembles de données pour lesquelles aucun consentement à l'utilisation et à la diffusion n'a été donné.

Les systèmes « autonomes » ne doivent pas porter atteinte au droit à la vie privée, qui comprend le droit de s'affranchir des technologies qui influent sur le développement et les opinions personnels, le droit à établir et à développer des relations avec d'autres êtres humains, et le droit de ne pas être sous surveillance. En la matière, des critères exacts devraient aussi être définis et des mécanismes établis pour assurer le développement et l'application éthiques des systèmes « autonomes ».

À la lumière des inquiétudes concernant les conséquences des systèmes « autonomes » sur la vie privée et la confidentialité, il pourrait être utile de s'intéresser au débat relatif à l'introduction de deux nouveaux droits: le droit à un contact humain significatif et le droit à ne pas être profilé, mesuré, analysé, encadré ou encouragé.

(i) Durabilité

L'IA doit être cohérente avec la responsabilité humaine qui consiste à assurer les conditions préalables indispensables à la vie sur notre planète, à la prospérité continue de l'humanité et à la préservation d'un environnement convenable pour les générations futures. Les stratégies visant à empêcher les futures technologies de porter préjudice à la vie humaine et à la nature doivent être basées sur des politiques qui donnent la priorité à la protection de l'environnement et à la durabilité.

L'intelligence artificielle, la robotique et les systèmes « autonomes » peuvent apporter la prospérité, contribuer au bien-être et participer à l'atteinte des idéaux moraux et des objectifs socio-économiques européens s'ils sont conçus et mis en œuvre avec sagesse. Les considérations éthiques et les valeurs morales partagées peuvent être utilisées pour former le monde de demain et doivent être envisagées comme un stimulus et une opportunité d'innovation, et non comme un obstacle ou une entrave.

Le GEE appelle la Commission européenne à étudier quels sont les instruments juridiques existants à sa disposition pour traiter efficacement les problèmes abordés dans cette déclaration et à déterminer si de nouveaux instruments réglementaires et de gouvernance sont nécessaires.

Le GEE appelle au lancement d'un processus qui ouvre la voie à un cadre éthique et juridique commun, reconnu au niveau international, concernant la conception, la production, l'utilisation et la gouvernance de l'intelligence artificielle, la robotique et les systèmes « autonomes ».

Comment prendre contact avec l'Union européenne ?

EN PERSONNE

Dans toute l'Union européenne, des centaines de centres d'information Europe Direct sont à votre disposition. Pour connaître l'adresse du centre le plus proche, visitez la page suivante :

<http://europa.eu/contact>

PAR TÉLÉPHONE OU COURRIER ÉLECTRONIQUE

Europe Direct est un service qui répond à vos questions sur l'Union européenne. Vous pouvez prendre contact avec ce service :

- par téléphone : via un numéro gratuit : **00 800 6 7 8 9 10 11** (certains opérateurs facturent cependant ces appels),
- au numéro de standard suivant : **+32 22999696**;
- par courrier électronique via la page <http://europa.eu/contact>

Comment trouver des informations sur l'Union européenne?

EN LIGNE

Des informations sur l'Union européenne sont disponibles, dans toutes les langues officielles de l'UE, sur le site internet Europa à l'adresse <http://europa.eu>

PUBLICATIONS DE L'UNION EUROPÉENNE

Vous pouvez télécharger ou commander des publications gratuites et payantes sur le site EU Bookshop à l'adresse suivante : <http://publications.europa.eu/eubookshop>. Vous pouvez obtenir plusieurs exemplaires de publications gratuites en contactant Europe Direct ou votre centre d'information local (<http://europa.eu/contact>).

DROIT DE L'UNION EUROPÉENNE ET DOCUMENTS CONNEXES

Pour accéder aux informations juridiques de l'Union, y compris à l'ensemble du droit de l'UE depuis 1951 dans toutes les versions linguistiques officielles, consultez EUR-Lex à l'adresse suivante :

<http://eur-lex.europa.eu>

DONNÉES OUVERTES DE L'UNION EUROPÉENNE

Le portail des données ouvertes de l'Union européenne (<http://data.europa.eu/euodp>) donne accès à des ensembles de données provenant de l'UE. Les données peuvent être téléchargées et réutilisées gratuitement, à des fins commerciales ou non commerciales.

Les progrès de l'intelligence artificielle (IA), de la robotique et des technologies dites « autonomes » s'accompagnent de nombreuses questions morales, dont l'urgence et la complexité ne cessent de croître. Les efforts actuels pour trouver des réponses aux défis éthiques, sociétaux et juridiques posés par ces progrès, et pour les orienter en faveur de l'intérêt général, se limitent à une mosaïque d'initiatives disparates. Cela souligne la nécessité d'un processus de réflexion et de dialogue collectif, large et inclusif, un dialogue qui se concentre sur les valeurs autour desquelles nous voulons organiser la société et sur le rôle que les technologies doivent jouer.

La présente déclaration appelle au lancement d'un processus qui ouvrirait la voie à un cadre éthique et juridique commun, reconnu au niveau international, concernant la conception, la production, l'utilisation et la gouvernance de l'intelligence artificielle, la robotique et les systèmes « autonomes ». Elle propose aussi un ensemble de principes éthiques fondamentaux, basés sur les valeurs établies par les traités et la charte des droits fondamentaux de l'UE, qui serviraient de guide à l'élaboration de ce processus.

Politique en matière de Recherche et d'Innovation

