

Les 10 grands principes pour une intelligence artificielle éthique



UNI Global Union

Nyon, Suisse

Les 10 grands principes pour une intelligence artificielle éthique

Introduction

Les syndicats se doivent être d'impliqués à une époque où l'intelligence artificielle (IA), la robotique, les données et l'apprentissage automatique font irruption sur les lieux de travail dans le monde entier et déplacent et perturbent les travailleurs et les emplois. Le présent document fournit aux syndicats, aux délégués syndicaux et aux travailleurs un ensemble de revendications concrètes portant sur la transparence et l'application de l'IA. Il souligne pour les concepteurs de l'IA et les dirigeants l'importance de l'inclusion des travailleurs. Il est urgent d'agir sans tarder afin de protéger les intérêts des travailleurs et de préserver un équilibre sain entre les forces en présence sur les lieux de travail. Les 10 principes du présent document ont été élaborés par UNI Global Union dans ce but.

L'IA est présente dans un grand nombre d'appareils ménagers et de lieux de travail : dans les agents conversationnels (chatbots), les robots, l'analyse de systèmes et les bases de données qui génèrent des informations et des réactions telles que des mouvements ou la parole. Comme le dit judicieusement Arvind Narayanan, de l'Université de Princeton, ce phénomène se rencontre « dès lors que le comportement n'émane pas exclusivement du programmeur, mais d'autres moyens tels que des bases de connaissance. »

Les données sont les pièces maîtresses de l'intelligence artificielle : il peut parfois s'agir de simples algorithmes, mais aussi et de plus en plus souvent de fils complexes de jeux de données multiples combinés en un code de plus en plus long.

L'intelligence artificielle n'est pas un phénomène nouveau. Elle existe depuis 40, voire 50 ans. Mais l'extension des technologies numériques et l'énorme quantité de données produites jour après jour par vous comme par moi ont donné à l'IA une nouvelle importance et une dimension totalement nouvelle : l'apprentissage automatique. L'apprentissage automatique est une application de l'intelligence artificielle (IA) qui offre aux systèmes la faculté d'apprendre automatiquement et de s'améliorer en tirant parti de l'expérience sans avoir été explicitement programmés pour cela. L'apprentissage automatique se concentre sur la mise au point de programmes informatiques capables d'accéder à des données et de les utiliser pour apprendre par eux-mêmes.

Nous avons donc désormais des formes d'intelligence augmentée capable d'apprendre par elles-mêmes. Dans un processus d'apprentissage en spirale sans fin s'appuyant sur des données émanant initialement de nous tous, où l'être humain dit à l'ordinateur qu'une image représente un panneau de signalisation routière, une cellule cancéreuse, une personne ou une voiture, les machines, sur la base des informations précédentes, deviennent capables de le déterminer par elles-mêmes. Elles aussi peuvent découvrir des corrélations complexes entre des jeux de données. Citons à titre d'exemple le fait que des chercheurs utilisant l'IA en tant qu'outil ont désormais découvert les sept critères qui doivent être remplis pour qu'une personne souffre d'une dépression *plus tard dans sa vie*.

L'intelligence artificielle et le monde du travail

L'intelligence artificielle et ses applications entraînent déjà des déplacements de travailleurs, et vu les progrès rapides de ses capacités, on peut s'attendre à ce qu'un bien plus grand nombre de tâches réalisées aujourd'hui par l'être humain le seront à l'avenir par l'IA et les robots. Au sein des entreprises, les tâches typiques relevant des ressources humaines sont complétées, voire

remplacées par l'IA. Cela s'observe dans le recours à l'IA pour procéder à des recrutements ou à des promotions, ainsi que dans la surveillance sur le lieu de travail et dans les tests d'efficacité/de productivité. C'est précisément à cause de cela que les syndicats doivent être impliqués dans la compréhension de l'IA, de ses potentialités et de ses problèmes pour le monde du travail, et se battre pour exercer une influence sur son application.

Certains travailleurs sont déjà en train de perdre leur emploi à cause de l'IA ; des recherches donnent même à penser que plus de 50% du travail actuellement fourni par l'être humain peut être réalisé plus rapidement et plus efficacement par des systèmes automatisés.

L'IA, l'apprentissage automatique, la robotique et l'automatisation peuvent également profiter aux travailleurs. Dans le secteur des soins de santé, les robots vont pouvoir aider à soulever les patients ou surveiller leur état de santé. Dans de nombreux emplois du tertiaire, les systèmes d'IA peuvent améliorer le service fourni aux clients car des bases de données de plus en plus riches en informations peuvent aider le travailleur à offrir un service approprié ou une information correcte.

Pour que l'IA et toutes ses applications soient mises en œuvre de manière durable et éthique, les syndicats doivent demander à disposer d'idées, d'influences et de droits se rapportant aux décisions de management basées en tout ou en partie sur l'IA. De par le monde, rares sont les conventions d'entreprises qui incluent ces droits des travailleurs.

Les experts reconnaissent que le moment est venu de discuter pour déterminer l'utilisation appropriée qu'il convient de faire de l'IA. UNI Global Union réclame une convention mondiale sur l'IA éthique, capable de veiller à prévenir et à gérer les conséquences négatives indésirables de l'IA tout en accentuant ses avantages pour les travailleurs et pour la société. Nous soulignons que les agents responsables doivent être les êtres humains et les entreprises.

Le présent document concrétise une revendication majeure d'UNI Global Union : **l'intelligence artificielle doit faire passer la planète et l'être humain au premier plan. C'est la raison pour laquelle il est essentiel de discuter à l'échelle mondiale de l'IA éthique. Une convention mondiale sur l'IA éthique incluant tous est la garantie la plus viable de la survie de l'espèce humaine.**

Ce document propose 10 principes et points d'action spécifiques que les syndicats, les délégués syndicaux et les alliances mondiales doivent mettre en œuvre dans les conventions collectives, les accords-cadres mondiaux et les alliances multinationales. C'est en agissant ainsi que l'on pourra garantir les droits et l'influence des travailleurs à l'ère du numérique.

Table des matières

L'intelligence artificielle et le monde du travail.....	1
1 Réclamer que les systèmes d'IA soient transparents.....	4
2 Doter les systèmes d'IA d'une « boîte noire éthique ».....	4
3 Faire en sorte que l'IA serve les individus et la planète.....	4
4 Adopter une approche donnant les commandes à l'être humain.....	5
5 Veiller à une IA sans distinction de genre ni préjugés.....	5
6 Faire partager les avantages des systèmes d'IA.....	5
7 Assurer une transition juste et garantir le soutien des droits et libertés fondamentales.....	5
8 Établir un mécanisme de gouvernance mondial.....	6
9 Interdire l'attribution de responsabilités aux robots.....	6
10 Interdire la course aux armements d'IA.....	6

1 Réclamer que les systèmes d'IA soient transparents

Un système d'intelligence artificielle transparent est un système où il est possible de découvrir comment et pourquoi telle ou telle décision a été prise, ou dans le cas d'un robot, comment et pourquoi il a agi comme il l'a fait.

En particulier :

- a) Nous soulignons qu'un code source ouvert n'est ni nécessaire ni suffisant pour assurer la transparence ; la clarté ne peut pas être étouffée par la complexité.¹
- b) Pour les utilisateurs, la transparence est importante parce qu'elle assoit la confiance dans le système et la compréhension du système en leur fournissant un moyen simple de comprendre ce que fait le système et pourquoi.
- c) Pour la validation et la certification d'une IA/d'un SA, la transparence est importante parce qu'elle ouvre la possibilité de passer au crible les processus du système.
- d) Si des accidents surviennent, l'IA/le SA devra être transparent(e) et rendre des comptes à l'enquêteur afin que le processus interne qui a abouti à l'accident puisse être compris.
- e) Les travailleurs doivent avoir le droit de réclamer la transparence des décisions/résultats des systèmes d'IA et des algorithmes sous-jacents (voir principe 4 ci-après). Cela inclut le droit de faire appel des décisions prises par l'IA ou les algorithmes et les faire réviser par un être humain.
- f) Les travailleurs doivent être consultés sur la mise en œuvre, le développement et le déploiement des systèmes d'IA.
- g) Suite à un accident, les juges, les jurés, les avocats et les témoins experts impliqués dans le procès ont besoin de transparence et de redevabilité pour alimenter leur examen des preuves et leurs décisions.

Le principe de la transparence est un préalable qui permet de garantir que les autres principes soient observés.

Voir le principe 2 ci-dessous pour une solution opérationnelle.

2 Doter les systèmes d'IA d'une « boîte noire éthique »

La transparence intégrale d'un système d'IA devrait être facilitée par la présence d'un dispositif capable d'enregistrer l'information relative à ce système sous la forme d'une « boîte noire éthique » contenant non seulement les données pertinentes pour garantir la transparence et la redevabilité d'un système, mais aussi des données et des informations claires sur les considérations éthiques intégrées à ce système.

Appliquée aux robots, la boîte noire éthique enregistrerait toutes les décisions, les bases de la prise de décisions, les mouvements et les données sensorielles du robot hôte. Les données fournies par la boîte noire pourraient également aider les robots à expliquer leurs actions dans un langage compréhensible par les utilisateurs humains, ce qui encouragerait de meilleures relations et améliorerait l'expérience de l'utilisateur. La lecture de la boîte noire éthique devrait être simple et rapide.

3 Faire en sorte que l'IA serve les individus et la planète

Ce point inclut des codes d'éthique pour le développement, l'application et l'utilisation de l'IA afin que tout au long de leur processus opérationnel, les systèmes d'IA restent compatibles et

¹ (Pasquale 2015).

renforcent les principes de la dignité humaine, de l'intégrité, de la liberté, du respect de la sphère privée et de la diversité culturelle et de genre ainsi que les droits humains fondamentaux.

En outre, les systèmes d'IA doivent protéger et même améliorer les écosystèmes et la biodiversité de notre planète.

4 Adopter une approche donnant les commandes à l'être humain

Un préalable absolu est que le développement de l'IA soit responsable, sûr et utile, que les machines conservent le statut juridique d'outils et que des personnes morales conservent le contrôle sur ces machines et la responsabilité envers elles *en tout temps*.

Cela implique que les systèmes d'IA soient conçus et exploités de manière à se conformer au droit existant, y compris le respect de la sphère privée². Les individus devraient avoir le droit d'accéder aux données générées par les systèmes d'IA, de les gérer et de les contrôler, vu le pouvoir qu'ont ces systèmes d'analyser et d'utiliser ces données (voir principe 1 du document « Les 10 grands principes pour la protection de la confidentialité et des données des travailleurs »).

5 Veiller à une IA sans distinction de genre ni préjugés

Dans la conception et la maintenance de l'IA, il est vital que le système subisse des contrôles afin d'en éliminer les préjugés négatifs ou préjudiciables envers l'être humain, et de veiller à ce que tout préjugé (genre, origine, orientation sexuelle, âge) soit identifié et ne soit pas propagé par le système.

6 Faire partager les avantages des systèmes d'IA

Les technologies d'IA devraient profiter à un aussi grand nombre de personnes que possible et les autonomiser. La prospérité économique créée par l'IA devrait être répartie largement et de manière égale, au profit de l'ensemble de l'humanité.

Des politiques mondiales autant que nationales visant à surmonter le fossé numérique économique, technologique et social sont donc nécessaires.

7 Assurer une transition juste et garantir le soutien des droits et libertés fondamentales

Au fur et à mesure du développement des systèmes d'IA et de la création de réalités augmentées, les travailleurs et les tâches de travail disparaîtront. Il est vital que des politiques soient mises en place pour garantir une transition juste menant à la réalité numérique, y compris des mesures gouvernementales spécifiques pour aider les travailleurs qui ont perdu leur emploi à en trouver un nouvel.

Les systèmes d'IA joints au passage à l'économie numérique au sens large exigeront que les travailleurs à tous les niveaux et dans tous les métiers aient accès à la sécurité sociale et à l'apprentissage tout au long de la vie afin de rester employables. Il incombe aux États et aux entreprises de trouver des solutions permettant à tous les travailleurs, quelle que soit leur forme d'emploi, d'avoir le droit à l'un comme à l'autre et d'y accéder.

En outre, dans un monde où la précarisation ou l'individualisation du travail va croissant, tous les travailleurs, quelle que soit leur forme d'emploi, doivent avoir les mêmes droits sociaux et

² Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., . . . Winfield, A. (2011, avril). Principles of Robotics. The United Kingdom's Engineering and Physical Sciences Research Council (EPSRC). (publication sur le web).

fondamentaux forts. Tous les systèmes d'IA doivent inclure un système de contrôles pour déterminer si leur déploiement et leur généralisation s'harmonisent avec les droits des travailleurs tels que consignés dans les chartes des droits de l'homme, les conventions de l'OIT et les conventions collectives. Un algorithme « 8798 » reflétant les conventions fondamentales de l'OIT et inclus dans le système pourrait précisément y veiller. En cas d'échec, le système doit être arrêté.

8 Établir un mécanisme de gouvernance mondial

UNI recommande la création d'instances de gouvernance sur le travail décent et l'IA éthique, regroupant toutes les parties prenantes au niveau mondial comme au niveau régional. Ces instances devraient inclure les concepteurs d'IA, les fabricants, les propriétaires, les développeurs, les chercheurs, les employeurs, les juristes, les OSC et les syndicats. Des mécanismes de dénonciation des abus et des procédures de contrôle pour veiller au passage à l'IA éthique et à sa mise en œuvre doivent être établis. Ces instances devraient recevoir la compétence de recommander des processus et des procédures de conformité.

9 Interdire l'attribution de responsabilités aux robots

Les robots devraient être conçus et exploités dans la mesure du possible de manière à respecter la législation existante et les droits et libertés fondamentales, y compris le respect de la sphère privée. Ce point est lié à la question de la *responsabilité juridique*. Conformément à Bryson et al 2011, UNI Global affirme que la personne juridiquement responsable d'un robot devrait être identifiée. Les robots ne sont pas des parties responsables en droit.

10 Interdire la course aux armements d'IA

Les armes létales autonomes, y compris la cyberguerre, devraient être interdites.

Sources :

Le présent document s'est inspiré des principaux documents suivants, sur lesquels il s'appuie :

Joanna J. Bryson (2017) The meaning of the EPSRC principles of robotics, *Connection Science*, 29:2, 130-136, DOI: 10.1080/09540091.2017.1313817 <http://dx.doi.org/10.1080/09540091.2017.1313817>

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S.,...:Winfield, A. (2011, April). Principles of Robotics. The United Kingdom's Engineering and Physical Sciences Research Council (EPSRC). (publication sur le web).

<http://uk.businessinsider.com/stephen-hawking-elon-musk-backed-asimolar-ai-principles-for-artificial-intelligence-2017-2>

<https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>

<http://www.eesc.europa.eu/?i=portal.en.int-opinions.40538>https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

<https://futurism.com/experts-want-robots-to-have-an-ethical-black-box-that-explains-their-decision-making/>