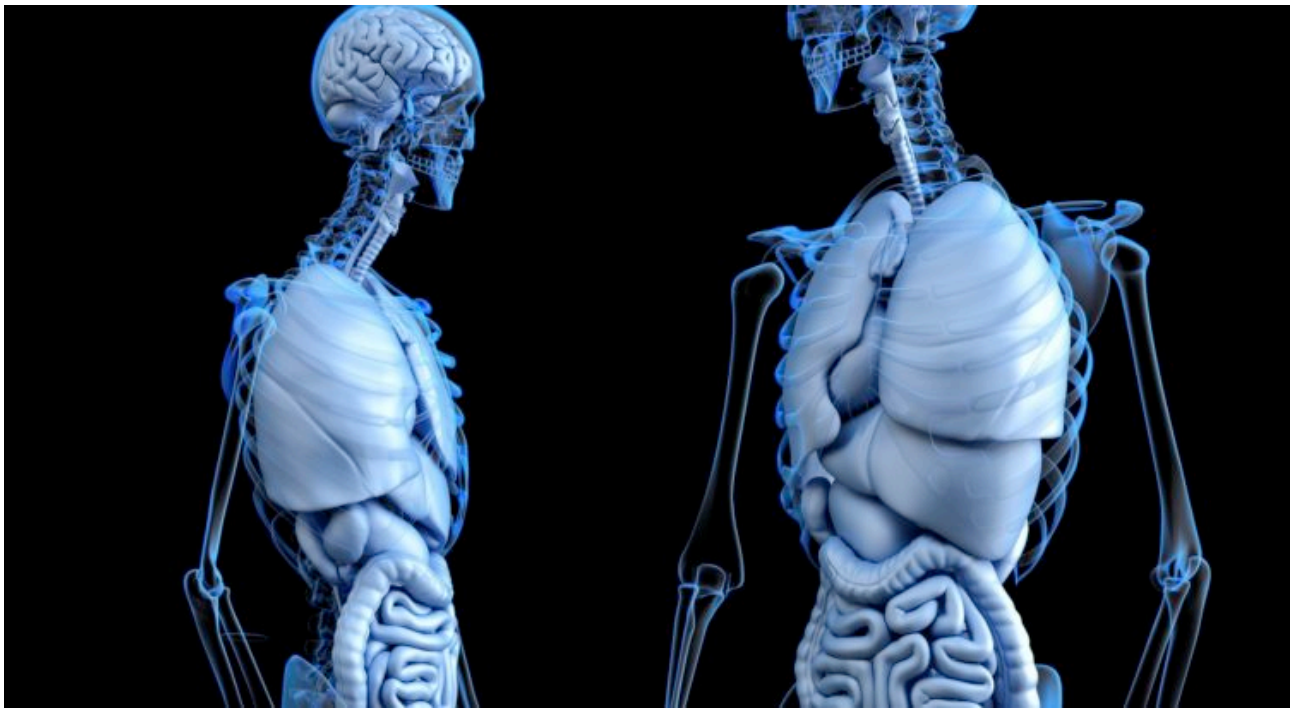
LT-24nov2020-Cybersécurité | Le blog de Solange Ghernaouti–Les superpouvoirs de l'intelligence artificielle

24 novembre 2020 Solange Ghernaouti



Fruits d'une logique d'optimisation et de rationalisation économique, inscrits dans une approche néolibérale, les dispositifs d'intelligence artificielle sont déterminés par ceux qui savent capter et exploiter les données, c'est à dire les acteurs hégémoniques de l'informatique et de l'Internet. L'intelligence artificielle est révélatrice d'un modèle de société imposé par les acteurs les plus forts. Elle est en passe de devenir le principal déterminant de l'organisation et de la gouvernance algorithmique du monde. La possibilité de bénéficier de certaines de ses applications ne devrait pas occulter les menaces et défis liés aux nouvelles formes de pouvoirs que l'intelligence artificielle procure à ceux qui la maîtrisent et qui imposent son usage, comme l'a très bien analysé le philosophe Eric Sadin dans son ouvrage "*L' Intelligence artificielle ou l'enjeu du siècle".**

Les capacités de l'intelligence se déclinent en différents types de pouvoirs dont le trait commun est la surveillance informatisée des personnes.

Le pouvoir prédictif qui suggère et propose, ce qui permet d'influencer et de manipuler des individus.

Le pouvoir d'énoncer la vérité et de l'imposer, ce qui exclut toute réalité autre que celle captée et déterminée par l'intelligence artificielle ainsi que toute prise en compte de ce qui n'est pas programmé.

Le pouvoir injonctif qui ordonne de faire, qui oriente les actions et comportements des individus et qui les contrôle en temps réel. Dépossédées de leur libre arbitre, les personnes sont tenues en laisse électronique et pilotées à distance. Elles deviennent les marionnettes de chair et de sang des robots informatiques.

Le pouvoir coercitif contraint et pénalise les comportements déviants par rapport à une normalité imposée par les concepteurs des dispositifs d'intelligence artificielle.

Quelle utilité sociale ?

Bien que l'intelligence artificielle s'immisce dans toutes les activités humaines, l'utilité sociale de tels systèmes est peu questionnée ou démontrée. Par ailleurs, plusieurs problèmes, qui commencent à être bien documentés existent. Ils sont essentiellement relatifs aux données et aux algorithmes utilisés. Ainsi, la perte de contrôle par l'humain de la manière dont les données sont collectées et utilisées pour les apprentissages automatiques (*machine learning*) introduit des biais. De plus, la non transparence des algorithmes (l'opacité des mécanismes de *deep learning*), l'impossibilité de les vérifier, de certifier leur innocuité, sécurité et qualité, sont des facteurs de risques. L'opacité des algorithmes est renforcée par l'impossibilité pour l'humain d'interpréter et d'expliquer ce qu'ils font et de justifier la pertinence des résultats.

La boite noire au cœur de l'impossible confiance

Les algorithmes mise en œuvre par des dispositifs d'intelligence artificielle sont développés sous forme de boite noire (secret de fabrication de leurs propriétaires), excluant de facto la possibilité pour des acteurs externes indépendants, de pouvoir vérifier et valider leur qualité et leur sécurité. Dès lors, comment éviter que de tels systèmes soient inexplicables ou incompréhensibles pour les humains ?

Comment comprendre pourquoi et comment un algorithme d'intelligence artificielle se comporte comme il le fait, prend une décision et arrive à ses résultats ? Comment éviter des prises de décisions algorithmiques automatisées non transparentes ? Comment rendre transparents les raisonnements quand les propriétés sont cachées et résultent de choix opaques ?

En fait, sans être en mesure de comprendre les logiques internes des systèmes d'intelligence artificielle, sans pouvoir vérifier et expliquer leur mode de fonctionnement, comment faire confiance aux décisions prises par de tels systèmes ?

Supprimer l'obscurité de leurs comportements internes, assurer la transparence des données et des algorithmes devrait être une condition préalable à leur usage.

Une question de responsabilité et d'engagement contraignant

Les valeurs promues par l'UNESCO sont relatives au respect, à la protection et à la promotion de la dignité humaine, des droits de l'homme et des libertés fondamentales. Elles sont également liées aux besoins de protection de l'environnement et des écosystèmes, de diversité et d'inclusion, de vivre en harmonie et en paix.

S'il est aisé d'être en faveur de tels principes fondamentaux, alors les systèmes d'intelligence artificielle devraient être conçus afin de les respecter et de les rendre effectifs. Ainsi, les concepteurs, fabricants, distributeurs de dispositifs d'intelligence artificielle, doivent s'assurer, avant leur mise à disposition et leur utilisation, qu'ils ne portent pas atteinte aux valeurs fondamentales. Ils devraient être responsables de l'innocuité, de la sûreté, de la fiabilité et de la sécurité de leurs dispositifs.

Pour qu'un développement responsable de l'intelligence artificielle soit envisageable, il faudrait qu'il existe aux niveau national et international des mécanismes qui permettent de refuser l'usage de dispositifs d'intelligence artificielle qui portent atteintes aux valeurs énoncées par l'UNESCO, de dénoncer les dérives et les préjudices engendrés et de poursuivre en justice les entités responsables.

L’impérieuse nécessité de nouveaux droits fondamentaux

Des gardes fous sont nécessaires afin de garantir :

1. Le droit des individus à avoir des opinions et des comportements différents de ceux énoncés par une IA.
2. La tolérance des dispositifs d’IA envers ceux et celles dont les particularités (à la marge de ce qui est défini comme normal par les concepteurs d’IA).
3. Le droit de la liberté de l’humain de pouvoir se soustraire au pouvoir d’influence, d’orientation et de coercition de IA.
4. Le droit à la déconnexion.
5. Le droit de ne pas vivre sous surveillance informatisée.
6. Le droit de la personne à savoir si elle interagit avec une intelligence artificielle (qui simule l’humain).
7. Le droit à la transparence des prises de décisions effectuées par une intelligence artificielle.
8. Le droit de pouvoir recourir contre une décision prise par une intelligence artificielle.

Un récit commun préfabriqué par les promoteurs et vendeurs de systèmes d’intelligence artificielle et qui promeuvent leur adoption massive pour « résoudre tous les problèmes du monde » se développe. Il est élaboré dans la continuité de l’évolution du numérique et de la plateformatisation du monde. Taillé sur mesure pour l’optimisation des intérêts de certains acteurs qui tentent de persuader qu’il n’existe pas d’autres alternatives (TINA – *There is No Alternative*). Toutefois, la banalisation de l’intelligence artificielle n’est pas une fatalité, la manière dont elle est conçue et utilisée non plus.

Des limites sont à poser au développement sans limite des dispositifs d’intelligence artificielle. Des contre-discours sont à opposer aux récits évoquant une « intelligence artificielle bienfaisante » et des modèles technico-économiques sont à faire vivre.

Au delà de l’espoir, un horizon, une nécessité, celle de contribuer à l’origine du futur, de participer au pouvoir de création de la Technologie et d’être acteur de la métamorphose en cours.

L’aventure ne fait que commencer …

** Eric Sadin, “L’ Intelligence artificielle ou l’enjeu du siècle”, L’échappée 2018.*

droits de l'homme éthique Intelligence artificielle les pouvoir de l'IA Surveillance technopouvoir UNESCO utilité social

SOLANGE GHERNAOUTI

Docteur en informatique, la professeure Solange Ghernaouti dirige le Swiss Cybersecurity Advisory & Research Group (UNIL) est pionnière de l'interdisciplinarité de la sécurité numérique, experte internationale en cybersécurité et cyberdéfense. Auteure de nombreux livres et publications, elle est membre de l'Académie suisse des sciences techniques, de la Commission suisse de l'Unesco, Chevalier de la Légion d'honneur.