Institut Maçonnique Européen

Plateforme IA Alliance – Consultation janvier 2019

Commission européenne - Dialogue Article 17

Détail de notre contribution, sur les lignes Directrices Ethiques pour une IA de confiance

Contexte.

Résumé du document

Mi-juin 2018, la Commission européenne a nommé 52 experts au sein du nouveau groupe de haut niveau sur l'intelligence artificielle (le High Level Expert Group on Artificial Intelligence). Le groupe, composé de représentants du monde universitaire, des entreprises et de la société civile, avait pour vocation de soutenir la mise en œuvre de la communication de l'UE sur l'intelligence artificielle publiée en avril 2018.
Le groupe d'experts de haut niveau formulera des recommandations sur la manière de relever les défis et les opportunités à moyen et à long terme liés à l'intelligence artificielle. Les recommandations alimenteront le processus d'élaboration des politiques, le processus d'évaluation législative et l'élaboration d'une stratégie numérique de prochaine génération. Le groupe préparera également un projet de lignes directrices sur l'éthique qui s'appuiera sur les travaux du groupe européen d'éthique des sciences et des nouvelles technologies et de l'Agence des droits fondamentaux de l'Union européenne dans ce domaine. Les lignes directrices couvriront des questions telles que l'équité, la sécurité, la transparence, l'avenir du travail et plus largement l'impact sur le respect des droits fondamentaux, y compris la protection de la vie privée et des données personnelles, la dignité, la protection des consommateurs et la non-discrimination. Le projet de lignes directrices sera finalisé d'ici la fin de l'année et présenté à la Commission au début de 2019.
C'est donc dans le cadre de ses fonctions, mais également des délais imposés par Bruxelles, que le AI HLEG (High Level Group on Artificial Intelligence) a publié le premier draft sur les lignes directrices sur l'éthique pour une IA de confiance (voir en fin de document les définitions proposées par le HLEG).

Nous avons convenu avec les SS qu'il fallait mettre en exergue de chaque réponse aux trois questions posées, *le ou les points clés qui sont développés*. Pour la conclusion nous avons repris celle du rapport que la TRGM a exposé le 18/06/18.

Texte de la contribution.

I. Le sujet de la définition : Pour ne pas confondre Programme informatique et Conscience

Point clé : éduquer à mieux connaitre et appréhender l'Intelligence Artificielle en tant qu'outil en évitant les qualifications ambiguës

Dans le draft, les termes « raisonner » et « raisonnement » sont utilisés pour désigner le mode de fonctionnement de l'IA. **Or,** l'on doit se poser la question de différencier les processus de l'humain et les évolutions technologiques, **il faudrait pour cela éviter d'utiliser le mot « intelligence » pour des algorithmes.**

- **L'usage du terme « raisonnement » sous-tendrait l'avènement d'une « IA forte » capable de copier le fonctionnement du cerveau humain. Or à ce jour, le cerveau humain et les processus cognitifs ne sont pas modélisable et il n'existe aucune théorie proche d'offrir un modèle même approché – soit du cerveau, soit de la « rationalité ».**
- L'Intelligence artificielle dépasse déjà nos aptitudes dans notre capacité à calculer, mémoriser ou discerner des détails. En outre, l'IA peut déjà être considérée d'égale à égale sur une traduction, un raisonnement spécialisé, voire une détection d'émotions.
- Dans la pensée il y a une conscience, dans un logiciel informatique il y a une action et pas de conscience. La déduction n'implique pas une conscience. La notion de finitude est très importante dans le développement de la conscience humaine tandis que la notion d'infinitude est induite dans le programme informatique.
- **Notre groupe de travail reste divisé en ce qui concerne la nature d'une IA en devenir : nouvelle forme de conscience, fruit d'une complexification de la matière résultante de la densification des connections dans les ordinateurs quantiques, ou évolution technologique exponentielle** associée à une cognition augmentée.
- **Nous aimerions donc souligner l'importance de la terminologie employée pour désigner n ensemble de technologies** et recommander en particulier de cesser d'employer le terme d'Intelligence artificielle.
- La question sur laquelle nous sommes toutes d'accord : c'est comment la contrôler et rester vigilantes ?

II. Réaliser une IA digne de confiance

Points clés :
L'éthique de l'Intelligence artificielle doit faire partie du dispositif de gestion des risques de la responsabilité du conseil d'administration (risques sociétaux)
Il est de la responsabilité sociale de l'entreprise de donner au citoyen les outils de compréhension et de préservation de la liberté de conscience face aux IA qu'elle déploie

Le document rédigé par le HLEG est très théorique et fixe le cadre de référence sans vraiment expliciter comment vérifier et mesurer le respect de ce cadre.
La mesure de l'éthique est effectuée en regard des principes normatifs des droits de l'homme. Mais Le rapport fait peut-être l'impasse sur des sujets de bien collectif, de bien individuel, qui peuvent comporter une certaine relativité. Il existe un risque que la façon de définir la norme entraine une convergence et nuise à la diversité, qui peut également être considérée comme un bien. Il est complexe de mesurer les effets sociétaux induits par ces mécanismes à grande échelle.

Nous pensons que pour adresser de façon concrète les problématiques éthiques, il faut examiner concrètement le rôle des acteurs qui sont les sociétés, les individus et les états ou organes de régulation.
En ce qui concerne le point II, réalisation d'une IA digne de confiance, les responsabilités reposent essentiellement sur les sociétés qui mettent en œuvre l'IA. Donc, elles doivent mettre

en place la gouvernance appropriée, en interne et avec leurs clients. Elles doivent rester à l'écoute des règles et des évolutions proposées par la société civile et par les tiers de confiance.

A- **Les sociétés qui utilisent l'IA** sont confrontées à 2 types d'enjeux : l'utilisation de l'IA dans les produits et l'utilisation de l'IA dans l'organisation du travail

L'IA va bouleverser les processus internes et l'organisation des entreprises : cela devrait se faire dans le respect des salariés et du sens donné au travail de chacun. En allant plus loin dans le découpage des tâches, l'IA créerait un néo-taylorisme déresponsabilisant ou au contraire permettrait aux personnes de s'épanouir par des organisations plus agiles.

L'utilisation de l'IA pour la commercialisation ou dans les produits peut également avoir des impacts sociétaux.

Aussi, les sociétés doivent prévoir des algorithmes construits, dès les phases de conception et de pré-lancement, pour respecter les principes éthiques (« Ethic by Design ») et qui sont constamment suivis et vérifiés.

Le respect des principes s'accompagne de la mise en œuvre des mesures suivantes :

1) Garantir la Précision des algorithmes : c'est l'analyse technique qui permet d'évaluer leur fiabilité, notamment le risque d'erreurs dans le système et le risque de préjudice pour les utilisateurs. Il est alors nécessaire de prévoir le processus de détection et correction des erreurs ;
2) Créer des outils permettant une compréhension suffisante des utilisateurs : **explicabilité** ;
3) Mettre en place des Tiers de confiance pour vérifier les algorithmes sur la base d'un échantillonnage ou par des jeux de tests spécifiques : **auditabilité** ;
4) Créer des normes relatives à l'**impartialité** : pour juger de l'absence de biais vis-à-vis de groupes ou de catégories de population, il faudrait inclure un algorithme d'exploration de données qui tient compte de l'équité. Mais la vision éthique d'une société comporte des éléments relatifs à la culture et à la période ;
5) Introduire les risques liés à l'utilisation de l'IA, qu'ils concernent les changements d'organisation ou l'impact social des produits et des services dans la cartographie des risques ESG (Environnementaux, Sociétaux et de Gouvernance) pour la partie des indicateurs sociétaux. Définir une chaine de responsabilité, comme pour la RGPD (Règlement Général sur la Protection des Données), pour donner des réponses rapides en cas de problème. Cette chaine de responsabilité doit remonter jusqu'au conseil d'administration via l'inclusion dans le rapport obligatoire sur la RSE (rapport de responsabilité sociale d'entreprise).
   Toutes ces mesures pourraient être contrôlées en interne par un responsable de l'éthique des algorithmes.

   Par rapport à leurs utilisateurs, les créateurs d'IA doivent mettre en place les conditions d'une utilisation éclairée de l'IA, par les moyens suivants :

B- L**e citoyen / utilisateur de l'IA :** il doit disposer des outils de compréhension qui permettent de préserver sa liberté de conscience et d'identifier des biais. Ces outils sont :

   1) **d'une part des interfaces qui rendent transparents**, compréhensibles et auditables les facteurs qui ont conduits à la proposition, des interfaces qui permettent d'éviter la manipulation mentale en donnant un recul et une diversité d'offres ;

   2) **D'autre part la formation des citoyens** : il n'en est pas fait état dans le document alors que c'est une des clés pour une utilisation éthique de l'IA. Elle pourrait être partiellement prise en charge par les sociétés qui déploient l'IA. Cette formation pourrait être complétée par une plateforme européenne adaptée à chaque état comprenant la formation par des MOOC, des cahiers de biais avérés…Il s'agit d'un enjeu majeur de politique publique

   3) Information obligatoire pour les services et produits sur la nature des algorithmes utilisés et les risques identifiés induits, notamment relatifs aux données personnelles utilisées et aux facteurs pouvant biaiser ou influencer le comportement, avec des exemples illustratifs à la portée de tous. Cette information est particulièrement cruciale dans le cas d'IA embarquées dans des Robots Humanoïdes.

III. Evaluer l'IA digne de confiance

Points clés :
L'évaluation des risques liés à l'IA repose sur la vigilance du citoyen et des associations professionnelles.
Il faut créer un régulateur indépendant garant du respect des principes éthiques, ayant accès à toutes les données et doté des moyens de contrôle et qui peut être saisi par les citoyens.

A. Le citoyen, l'individu et les associations professionnelles sont au cœur de la détection des biais et des dérives. Ils doivent disposer d'un système de remontée des anomalies à un régulateur national. Ils doivent pouvoir agir en tant que collectif, formant par exemple des **comités citoyens** qui pourraient tester les IA avec une diversité de profil, ou en tant qu'associations professionnelles composées de personnes expertes, à même de comprendre les éventuelles dérives rapportées dans leurs domaines.

B. Un organe de régulation des IA à l'image des régulateurs de la Banque et de l'Assurance pourrait traiter les alertes citoyennes : L'état / l'Europe a le pouvoir d'imposer des lois et de mettre en place **des organes de régulation nationaux et supra nationaux.**
Nous avons la conviction qu'il faudrait mettre en place un système de régulation qui puisse agir en tant que tiers de confiance, ayant accès à toutes les données pour en analyser la qualité et les biais, qui puisse vérifier l'application des principes éthiques, comme dans le cas des banques pour les évaluations de risques systémiques.
Le régulateur pourrait être saisi par tout citoyen qui identifie un biais et serait ainsi le centralisateur et le médiateur des plaintes relatives aux IA.
La question des biais se pose différemment : nous savons que les données humaines sont naturellement biaisées. Il faut alors rectifier les algorithmes en fonction de critères choisis pour corriger les injustices induites par les biais humain. Qui fixe ces critères et comment est un sujet éminemment politique qui souligne bien l'importance du point 2 de formation des citoyens et la nécessité de les impliquer dans le processus.
Par exemple :

- Dans le futur, hommes et femmes devraient être sur un même pied d'égalité. Peut-on envisager une politique pro-active pour rectifier ce biais du langage de façon opérative ? L'IA offre le moyen de rectifier les biais humains par l'éducation des algorithmes. Il faut alors prendre de la distance par rapport à une photographie de l'existant, par exemple e**n introduisant d'autres biais qui réintègreraient le féminin mais aussi des valeurs de notre société future. Ne pas sexuer les chatbots par ex. ;**
- Prévoir une brigade « anti-manipulation mentale » en charge d'identifier et prévenir l'usage de l'IA pour manipuler de groupes d'individus dans des buts criminels, terroristes, ou de privation de leur libre arbitre (ex. manipulation d'élections, recrutement par des réseaux terroristes) ;
- Imposer une procédure de « débranchement » des algorithmes et imaginer comment débrancher les algorithmes de surveillance, par exemple chinois qui ne disposeraient pas de ces normes
- Aborder le sujet en termes de biais souhaités et apporter des éléments de quantification

Commentaires

L'Institut Maçonnique Européen de la Grande Loge Féminine de France espère que ces bouleversements et progrès scientifiques seront réalisés au bénéfice de l'Homme et que seront préservés dans les temps à venir la Liberté, l'Egalité et la Fraternité qui fondent notre société civile.

Notes complémentaires :

Nos SS ont répertorié les définitions proposées par les experts du AI HLEG ; ces définitions sous-tendent leur travail, *notamment :*

Définitions

Intelligence artificielle ou IA :

L'intelligence artificielle (IA) fait référence à des systèmes conçus par des humains qui, face à un objectif complexe, agissent dans le monde physique ou numérique en percevant leur environnement, en interprétant les données collectées, données structurées ou non structurées, raisonnant sur les connaissances tirées de ces données et décidant de la (des) meilleure (s) mesure (s) à prendre (selon des paramètres prédéfinis) pour atteindre l'objectif. Les systèmes d'intelligence artificielle peuvent également être conçus pour apprendre à adapter leur comportement en analysant la manière dont l'environnement est affecté par leurs actions précédentes.

En tant que discipline scientifique, l'intelligence artificielle comprend plusieurs approches et techniques, telles que l'apprentissage automatique (dont l'apprentissage en profondeur et l'apprentissage par renforcement sont des exemples spécifiques), le raisonnement automatique (planification, ordonnancement, représentation et raisonnement des connaissances, recherche et optimisation), et la robotique (qui comprend le contrôle, la perception, les capteurs et les actionneurs, ainsi que l'intégration de toutes les autres techniques dans des systèmes cyber physiques).

Partialité (biais) :

Une partialité est un préjudice pour ou contre quelque chose ou quelqu'un qui peut entraîner des décisions injustes. On sait que les humains sont biaisés dans leur prise de décision. Comme les systèmes d'IA sont conçus par des humains, il est possible que les humains leur injectent leurs préjugés, même de manière non intentionnelle.

But éthique :
l'intelligence artificielle doit garantir le respect des droits fondamentaux et de la réglementation applicable, ainsi que le respect des principes et valeurs fondamentaux.

IA centrée sur l'homme :
le développement et l'utilisation de l'IA ne doivent pas être considérés comme un moyen en soi, mais dans le but du bien-être humain.

IA digne de confiance :
Une IA digne de confiance comporte deux éléments : (1) But éthique », et (2) être robustes et fiables sur le plan technique.

Pour le Groupe de travail sur les impacts sociétaux des nouvelles technologies (Congrès de Paris) et l'IME,
Dominique BONETTI, chargée de mission.