



HAL
open science

Pour ou contre être gouvernés par une intelligence artificielle ?

Marie Coris, Serge Ahmed

► **To cite this version:**

Marie Coris, Serge Ahmed. Pour ou contre être gouvernés par une intelligence artificielle?. The Conversation France, 2020. hal-04073190

HAL Id: hal-04073190

<https://hal.science/hal-04073190>

Submitted on 18 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License



“Procès” hors-norme à la macropole de Bordeaux en 2050. Sébastien Morin, Author provided

Pour ou contre être gouvernés par une intelligence artificielle ?

Publié: 9 janvier 2020, 21:44 CET

Marie Coris

Enseignant-chercheur économie de l'innovation, laboratoire GREThA, Université de Bordeaux

Serge Ahmed

Research Director in Neurosciences, Université de Bordeaux

Nous ne serions plus vraiment satisfaits par nos dirigeants. Est-ce à dire que nous accepterions de laisser une intelligence artificielle nous gouverner ? C'est la question qu'ont eue à trancher quelque 250 citoyens convoqués dans la salle plénière de l'Hôtel de Région de Nouvelle-Aquitaine le 20 novembre dernier, à l'occasion d'un « faux » procès co-organisé par la Compagnie Primesautier Théâtre et l'université de Bordeaux (festival FACTS).

Bordeaux Macropole, nous sommes en 2050

Pour aiguïser la curiosité et susciter les débats, un duo de chercheur-artiste (Marie Coris pour l'université de Bordeaux, Antoine Wellens et Virgile Simon pour la compagnie Primesautier Théâtre) a inventé une histoire abracadabranqueste, situant l'action à Bordeaux Macropole, en 2050. Chaque lundi depuis le 7 octobre, le quotidien *Sud-Ouest* s'est fait le relais de ce « polar ».

Tout allait merveilleusement bien depuis l'élection, en 2047, de Martin Nasmushe à la tête de la Macropole. Plus de pollution, de bouchons, du travail pour toutes et tous et une douceur de vivre unanimement partagée par des macropolitains heureux et épanouis. Et patatras.

Après le « bug » de Martin Nasmushe à l'occasion de ses vœux de rentrée, tout ce bonheur s'écroule progressivement quand on découvre que ce qui gouvernait la Macropole n'était pas un homme... mais une intelligence artificielle (à laquelle Martin avait connecté son jumeau, Edi). Martin suicidé, c'est sur la culpabilité de son assistante de l'ombre (Marguerite Faustus) que les Bordelais ont dû se prononcer. Mais surtout, ils ont dû choisir entre le maintien et la déconnexion de l'IA.

Un faux procès pour de vraies questions

Au-delà de la mise en scène, la formule du « faux » procès a permis d'organiser une consultation citoyenne sur une question d'actualité. Car c'est bien aujourd'hui que se pose la question de la société que nous souhaitons pour demain. Pourquoi voulons-nous des IA et pour quoi faire ? Prenons l'exemple du véhicule autonome. On se demande souvent comment s'assurer qu'il fera les bons choix face à l'urgence (écrasera-t-il l'enfant qui traverse la route ou plutôt le groupe d'octogénaires sur le bas-côté ?), on se pose la question de la responsabilité mais se demande-t-on si (et pour quoi) on voudrait des véhicules autonomes ?

Toute l'originalité du projet est que tous les témoins sont des chercheurs et universitaires (tous bordelais), venus « faire état de leur savoir et rien que de leur savoir ». De la philosophie aux sciences computationnelles, en passant par la robotique, les lettres, les neurosciences, le droit, l'économie et, bien entendu, toutes les sciences du numérique, tous les savoirs étaient à la barre.



Un témoin à la barre. Sébastien Morin, Author provided

C'est quoi l'intelligence ? Et l'intelligence artificielle ? Où en est-on et que peut-on en attendre ? Comment faire la part entre fantasmes, mythes et réalité ? Quels sont les risques de bug ? Qui serait responsable en cas de bug ? Comment sont récupérées et traitées nos données, par qui pour quoi ? Comment une machine apprend-elle ? Peut-elle apprendre, quelle est la part de l'humain et l'œuvre de la machine ? Et au fait, c'est quoi une prise de décision, quelle est la part de libre arbitre ? Et d'ailleurs, le libre arbitre existe-t-il ?

Impossible de résumer ici les trois heures de débat à découvrir ci-dessous, dont certains éléments se retrouvent dans les contributions de deux des témoins, Frédéric Alexandre et Nicolas Rougier. Retenons ici les deux principaux arguments retenus, respectivement « à charge » et « à décharge ».

La fable du bonheur

Si les questions des biais et du manque de transparence des IA se sont évidemment retrouvées au cœur des débats, le principal argument plaidant pour la déconnexion de l'IA est à chercher ailleurs. Un monde dans lequel une IA parviendrait à créer et maintenir, dans la durée, une société heureuse est-il possible ? Peut-on imaginer qu'une IA réussirait ce qu'apparemment nous aurions échoué à faire pendant des milliers d'années, c'est-à-dire, en gros, depuis que nous aurions fini par nous domestiquer nous-mêmes il y a environ 10 000 ans ?

La réponse semble être non. Qu'entend-on, d'ailleurs, par une société heureuse ? Le bonheur, c'est quoi ? L'absence de souffrance ? La présence de plaisirs ? L'absence ou la présence de désirs ? Qu'est-ce qui contribue au bonheur ? Les conditions matérielles qui concourent à une meilleure santé, un meilleur confort et une meilleure fortune ? Quels sont les critères du bonheur collectif ? Comment les définit-on, objectivement ? Et comment mesurer leur « efficacité » ?

« Le Procès de l'Intelligence artificielle » (Université Bordeaux Montaigne, 2019).

Quand bien même, comment l'IA aurait-elle eu vent de ces critères ? Aurait-elle réussi à les identifier elle-même ? Dans la masse quasi-infinie des données humaines, l'IA aurait-elle fini par percer à jour le secret du bonheur collectif et le sens de la vie ? Cela semble incroyable, presque impensable, sauf à admettre que le bonheur serait algorithmique.

Et quid de la comparaison sociale ? Imaginons qu'une autre IA, plus puissante, plus intelligente que celle qui gouverne Bordeaux Macropole parvienne à créer, dans une autre ville, une société encore plus heureuse. Par comparaison, les habitants de la Macropole se sentiraient alors moins heureux. Cette comparaison ferait naître chez eux de nouvelles attentes et donc de nouvelles frustrations. Pour tenter d'éviter ce problème, l'IA pourrait inventer et diffuser une nouvelle idéologie anti-comparaison sociale (« ne regarde pas si l'herbe est plus verte chez ton voisin »). Et si cela ne marchait pas, elle pourrait alors se tourner vers le développement de neurotechnologies afin de manipuler le cerveau des habitants de manière à bloquer leurs désirs de se comparer.

On peut être riche, vivre en bonne santé et confortablement et, pourtant, être insatisfait, frustré et malheureux. N'est-ce pas le propre de nos sociétés capitalistes et consuméristes que de créer et entretenir la frustration et le désir pour que la société continue à croître ? Sans cela, pas de frustration, pas de désir et, in fine, pas de croissance économique. Si une IA arrivait à créer une véritable société heureuse, on devrait donc s'attendre à ce qu'elle s'effondre économiquement ! Le bonheur serait-il antiéconomique ?

Humain, IA, même boîte noire

L'IA est une boîte noire, et alors ? Un être humain « a » aussi une boîte noire. Il s'agit tout simplement de son cerveau. En effet, quand on demande à une personne pourquoi il ou elle a pris ou prend telle décision dans tel contexte, elle est typiquement capable de fournir une réponse argumentée, généralement en fournissant les raisons de sa décision. La nature des raisons fournies dépend souvent étroitement de la manière dont la question posée a été comprise.

Par exemple, quand Alice demande à Bob pourquoi il a décidé ceci ou cela, elle ne s'attend pas à ce que Bob lui fournisse une réponse en termes neurobiologiques ou même computationnels. Elle s'attend à une liste de raisons reliées plus ou moins logiquement entre elles (la nature exacte des raisons pouvant varier selon les contextes, ou la nature de la relation entre Alice et Bob). Et pourtant, on pourrait discuter longuement sur le fait que les niveaux computationnels et neurobiologiques jouent un rôle crucial dans la décision prise par Bob. Le problème, c'est que Bob n'a pas accès à ces niveaux qui lui sont opaques. Il n'a pas un œil intérieur lui permettant de visualiser en temps réels les événements et processus neuronaux qui se déroulent dans son cerveau et qui vont déboucher à sa prise de décision. En d'autres termes, le cerveau de Bob est une boîte noire pour lui. Et alors ? Si, maintenant, Bob gouvernait la Macropole, on aurait un problème de transparence comparable à celui évoqué au sujet de l'IA.

Bob peut rendre compte et expliquer ses décisions mais ni lui, ni nous, ne savons comment ses décisions ont été réellement prises. Mais nous n'avons pas peur de cette opacité naturelle car nous n'en avons pas vraiment conscience. Rien ne nous empêcherait de créer une IA qui pourrait rendre compte de ses décisions en utilisant le type de raisons que les humains utilisent pour justifier leurs décisions. En somme, le problème selon lequel l'IA serait une boîte noire et que ce n'est pas bien pour la démocratie s'applique dans d'autres contextes où des décisions sont prises par des agents naturels. Notre peur de ne pas savoir comment l'IA prend ses décisions, même si nous les jugeons bonnes, serait donc peut-être due en grande partie à un préjugé. Celui de la peur. Peur de ce que nous ne connaissons pas, peur de ce qui nous échappe, peur surtout du fantasme de la boîte noire qu'on n'arriverait pas à comprendre.

Un verdict paradoxal

À l'unanimité, le jury populaire a reconnu l'assistante non coupable. En revanche, une nette majorité (les deux tiers) a décidé de débrancher l'IA. Paradoxe qui donne tout son sens à la plaidoirie de l'avocat de la défense : « l'opacité pour les humains OK, l'opacité pour l'IA non ! ».

Retour au gouvernement de l'humain par l'humain. En 2019. Qu'en sera-t-il en 2050 ?

Cet article est republié dans le cadre du Forum mondial Normandie pour la Paix organisé par la Région Normandie et dont The Conversation France est partenaire. Pour en savoir plus, visiter le site du Forum mondial Normandie pour la Paix.