

INTÉGRER DU HASARD DANS LES ALGORITHMES DE L'IA ?

Publié le 31 janvier 2020



par Laetitia Theunis

L'intervention du Professeur Alexei Grinbaum n'est pas passée inaperçue au colloque de l'ESPHIN ([Espace Philosophique de l'Université de Namur](#)). Physicien et philosophe, chercheur au sein du [laboratoire LARSIM/CEA](#) de l'Université Paris-Saclay, il travaille sur les fondements de la physique et l'éthique dans le cadre des nouvelles technologies. Selon lui, les décisions morales prises par les machines doivent résulter du hasard. Cela pose la question du paramétrage des voitures autonomes.

Qui sacrifier ? Les piétons ou les passagers ?

Déjà affaiblies par les [problèmes d'hallucination de l'intelligence artificielle \(IA\) embarquée](#), les voilà depuis quelques années sous les feux de la rampe concernant les questions éthiques soulevées face à un accident inévitable. Depuis 1967, elles sont connues sous l'expression « dilemme du chauffeur de trolley ».

Dans la voiture autonome, un père et son fils. Soudain, quatre enfants surgissent et traversent la route. Impossible d'éviter l'accident. La voiture autonome, doit-elle éviter les piétons, quitte à se fracasser contre un mur et tuer ses passagers, ou doit-elle percuter les enfants pour sauver le père et son fils à son bord ? Qui l'algorithme doit-il sacrifier ?

Moralement, il est exclu d'individualiser les gens

Il est important de noter qu'il est nécessaire que les variables amenant à ce choix soient calculables

par l'algorithme.

« Certains préfèrent sacrifier les personnes les plus âgées, d'autres les moins âgées, voire les moins intelligentes. Pour d'autres, il convient de sauver la vie des amis du conducteur sur facebook. Mais un comité d'éthique de philosophes allemands, avec leur approche très kantienne, a dit qu'il était inadmissible, pour des questions morales, d'individualiser les gens selon n'importe quel critère. Que ce soit l'âge, l'intelligence ou autre », explique Alexei Grinbaum.

Mettre à mort par référendum est moralement condamnable

Le Massachusetts Institute of Technology (MIT) a développé son projet [Moral Machine](#). Réalisé sous la forme d'un jeu vidéo, il met en scène des piétons et des passagers de tous types (enfant, adulte, chat, chien, délinquant, homme, femme, femme enceinte, médecin, sans-abri, personne obèse, cadre d'entreprise, piéton traversant hors des clous) qui sont mis à mort par un référendum en ligne.

Basés sur près de 40 millions de décisions émanant de pays du monde entier, les [résultats](#), désormais à disposition des gouvernements et des instances régulatrices, laissent pantois sur la diversité des choix moraux des internautes. Ils varient grandement selon les continents et se regroupent en trois familles.

« Dans le monde occidental, on préfère ne rien faire. Mais si l'on doit vraiment choisir, il faut sauver la vie du plus grand nombre de personnes. En Amérique du Sud, les citoyens ont plébiscité le fait de sauver la vie des personnes ayant le statut social le plus élevé. En Asie, on veut sauver la vie des personnes qui respectent la loi », résume le philosophe.

Face à ce panel bigarré d'avis, « les algorithmes doivent-ils déterminer l'origine ethnique des passagers de la voiture autonome et ajuster leur comportement en fonction de cela ? Pour les ingénieurs, c'est impensable et inadmissible, car c'est impossible à réaliser en pratique », explique Alexei Grinbaum. « De plus, décider par référendum qui mettre à mort, c'est moralement condamnable. »

Le tirage au sort permet de se soustraire au champ du jugement ...

Pour le philosophe, la solution réside dans le hasard. Il prend comme exemple l'écriture automatique de Google Smart. Ce logiciel complète les phrases que l'on a commencé à écrire dans Google.

« Il y a un an, Google prend conscience d'un problème éthique en anglais avec les pronoms personnels 'he' et 'she' : le logiciel se trompait parfois, et décrivait une dame en employant 'he' et inversement. Suite au mouvement « metoo », de nombreuses personnes ont écrit à Google pour dire que c'était un problème. Il a alors éteint cette fonction. Depuis novembre 2018, le logiciel ne dit plus « she » ou « he », mais remplace ces deux pronoms personnels par « it ». Il a donc ajouté de l'inhumain dans les échanges écrits. »

Selon Alexei Grinbaum, la solution aurait été d'avoir recours au hasard. « Google aurait dû passer un certain seuil de conflit avéré, tirer au hasard le pronom personnel et informer l'utilisateur que ce choix était effectué au hasard. »

Et d'extrapoler, « le hasard est au centre de la solution des questions d'éthique autour des nouvelles technologies, cela permettrait de les inscrire dans la continuité de l'éthique de notre culture. »

... mais les programmes ne savent pas créer du hasard

Gros bémol : les programmes informatiques les plus développés ne sont pas capables de créer véritablement du hasard.

Pour le philosophe, ce point n'est pas crucial. « Il sera suffisant d'utiliser le pseudo-hasard que savent fabriquer les systèmes aujourd'hui, notamment via les générateurs standards de nombres

aléatoires. L'important est que l'utilisateur perçoive le hasard généré comme du vrai hasard et qu'il ait confiance en la méthode. Pour ce faire, il n'y a pas besoin d'aller jusqu'au générateur quantique de nombres aléatoires. »

Quid, par exemple, de l'utilisation de robots armés ? Tirer au hasard pourrait alors être utilisé comme écran permettant de cacher des non-prises de responsabilité de la part de militaires ou de politiques. « Psychologiquement et politiquement, ma solution est difficilement acceptée ; Et pourtant, j'ai l'impression que c'est la seule qui soit juste moralement », conclut Alexei Grinbaum. La réflexion est ouverte.