

L'Intelligence artificielle et le Covid-19 : qu'offre et n'offre pas l'IA

mercredi 6 mai 2020, par [KRIVINE Hubert](#) (Date de rédaction antérieure : 6 mai 2020).

La crise du Covid-19 comme la crise économique subséquente sera probablement l'événement majeur du 21^e siècle. Et l'Intelligence artificielle (IA) dont on attendait monts et merveilles - au moins en terme de prévisions - est restée muette. Il n'y a là rien d'étonnant pour nous : l'IA ne peut pas prévoir ce qui ne s'est jamais produit [1].

La machine, quand elle procède par induction (c'est-à-dire par généralisation), ne peut prévoir que sur la base - certes, gigantesque - de ce qui s'est déjà passé. En caricaturant un peu, pour la machine, ce qui arrivera est déjà arrivé ou proche de l'avoir été, et sans *big data*, on ne l'aurait jamais décelé.

Il n'existe aucune définition consensuelle de l'intelligence ni de l'homme, ni de la machine. Les apprentissages de la machine et ceux de l'humain ont des similitudes et des différences. Il y a des limites à la prévision par induction : quelle que soit la taille des *big data*, les données du monde réel sont sans commune mesure avec la taille de leur stockage. Ce que l'impuissance de l'IA illustre est que la carte (numérique) n'est pas le terrain. Par contre l'être humain, au-delà de l'induction, peut faire preuve d'intuition et d'imagination. Ce qui a produit des Einstein et des Jules Verne.

Dans les sciences dites « dures », comme la physique, le chercheur peut raisonnablement isoler son objet d'étude du reste de l'univers et de son histoire. Prenons deux exemples caricaturaux. *i)* Les atomes d'oxygène existent tous identiques à eux-mêmes et ce, depuis des milliards d'années. *ii)* Les jeux de go ou les échecs ont un nombre de parties possibles gigantesques, infiniment supérieur au nombre d'atomes de l'univers, mais ils peuvent être isolés de tout contexte. En ce sens, et en ce sens seulement, ils posent un problème « simple », que l'IA sait aborder avec succès. La carte, ici l'échiquier ou le *goban*, est bien le terrain.

Rien de tel pour les sciences du vivant : en biologie, il est impossible de faire abstraction de l'environnement et de son évolution ; en médecine par exemple il n'y a pas deux patients identiques. Enfin un virus ne peut être compris indépendamment de son histoire et de son interaction avec les cellules qu'il infecte. Voilà pourquoi la démarche par induction de l'IA trouve ses limites.

L'IA ne pouvait donc pas prévoir la pandémie comme elle peut le faire pour la consommation de margarine, l'achat de voitures ou le développement de tumeurs cutanées, mais le déclenchement de l'épidémie ayant eu lieu, on pouvait s'attendre à une bonne prévision de son développement mondial. Or, les instituts utilisant l'IA l'avaient mal anticipée : seules quelques [2] (rares) entreprises avaient, avant l'OMS, mis en garde sur l'extension d'une telle catastrophe.

Pourtant l'expansion d'autres épidémies (dengue, chikungunya, choléra, paludisme) avaient été souvent [3] mieux prévue par l'IA. Voilà qui souligne la forte dépendance des résultats à la qualité des *big data*. D'une part il semble maintenant établi que le gouvernement chinois - même s'il avait remarquablement vite transmis le séquençage du virus -, avait truqué les données

épidémiologiques ; d'autre part, la connaissance des échanges internet chinois bridés par la censure est moins aisée donc moins riche et certainement biaisée. Les données brutes sont un mythe : elles doivent toujours être contextualisées.

Dans une maladie « standard », on passe de la connaissance de la maladie à celle du malade. Ici, c'est plutôt le chemin inverse qu'on est obligé de suivre. Grâce à une collaboration internationale, on dispose maintenant d'une masse colossale de données sur les patients ; pour transformer ces données en information sur la maladie, l'IA sera un outil indispensable. Par exemple le rôle protecteur de la nicotine, s'il est confirmé, renseignerait sur la nature des récepteurs cellulaires du virus. Autre exemple : il est établi qu'aux USA les minorités (latinos, noires) sont plus susceptibles d'être contaminées. Corrélation n'est pas causalité ; on pense tout de suite que leur état sanitaire moyen est plus détérioré que celui du reste de la population et qu'ils effectuent sans protection les tâches les plus risquées. Mais derrière cette évidence, ne pourrait-il pas se cacher un élément génétique d'influence plus marginale [4] ?

Une infinité d'autres questions restent ouvertes : rôle des comorbidités, de l'âge, du sexe, du groupe sanguin, des vaccinations antérieures, du tabac, etc. Tout ceci nécessite une analyse très fine portant nécessairement sur des cohortes gigantesques de données. Là encore une réflexion humaine doit établir ce que peuvent être les facteurs de confusion, c'est-à-dire des corrélations non causales. Il en va de même pour l'évaluation de l'efficacité des traitements puisqu'il s'agit d'une maladie qui régresse spontanément dans l'immense majorité des cas. Reste enfin la durée et l'efficacité de l'immunité procurée par la maladie. Question importante pour la possibilité de futurs vaccins. Pourquoi au niveau international les pays pauvres où le confinement semble impossible et où le système de santé est défaillant semblent, pour l'instant en tout cas, être les moins touchés ? La jeunesse de la population et l'absence de statistiques fiables sont-elles des explications suffisantes ? Par contre au niveau national, les secteurs les plus pauvres sont les plus affectés.

Pour découvrir un traitement [5], l'IA peut opérer de façon relativement empirique en se mettant en chasse de composés qui ont été employés avec succès sur l'homme pour d'autres maladies infectieuses. Elle peut de façon plus théorique tenter d'identifier la structure des cibles de protéines virales et trouver des sites thérapeutiques potentiels. Enfin on pourrait songer à des emplois plus classiques de l'IA pour l'interprétation de l'imagerie médicale [6] ou le traçage des malades. Bien entendu, ce dernier exemple est exemplaire du rôle ambigu que peut jouer l'IA : traçage ou flicage ?

Le traitement de la crise sanitaire a révélé les forces et les faiblesses de l'IA. Néanmoins, c'est certainement grâce à son exploitation intelligente qu'on pourra obtenir des temps de neutralisation du virus bien plus courts que pour la variole, la fièvre jaune, la rougeole ou la poliomyélite.

L'origine de la pandémie n'est pas « simple ». Il y a évidemment une cause biologique, l'existence d'un virus malfaisant. Mais pourquoi s'est-il développé, passant de l'animal à l'être humain ? La déforestation a joué un rôle. Son éradication non plus n'est pas non plus un problème « simple ». Il y a bien sûr un problème biologique (vaccins, traitement), mais aussi social et politique.

Comment les sociétés devraient-elles s'organiser pour contenir la maladie tenant compte que nous n'en sommes pas tous également victimes et coupables ? Ce n'est clairement plus un problème pour l'IA.

Hubert Krivine

P.-S.

- Article écrit pour la revue *Viento Sur*.

Notes

[1] Voir l'article « ¿Es inteligente la inteligencia artificial ? » *Viento Sur* du 5 juillet 2019. En français sur ESSF (article 39918), [Le Big Data, pour faire quoi ? - Agir sans comprendre ? Savoir sans science ?](#).

[2] Par exemple *Dataminr* à New York ou *Bluedot* à Toronto.

[3] À L'exception notable de *Google flu* censé prévoir l'épidémie de grippe et fermé en 2015.

[4] Ce qui expliquerait pourquoi des individus de condition sanitaire identique peuvent réagir si différemment.

[5] <https://www.theguardian.com/business/2020/mar/31/oxford-firm-to-screen-15000-drugs-in-search-for-coronavirus-treatment>.

[6] Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for COVID-19. IEEE Reviews in Biomedical Engineering, April 2020.