

Intelligence artificielle et Covid-19, ou les limites du « solutionnisme technologique »

Par Nicolas BRAULT

Enseignant-chercheur en histoire et philosophie des sciences,
Interact UP 2018.C102, Institut polytechnique UniLaSalle

Dès le début de la pandémie de Covid-19, le recours à l'intelligence artificielle et aux données massives apparaît aux yeux de nombreux scientifiques et responsables politiques dans le monde comme la solution ultime pour gérer la pandémie et ses conséquences. Or, après plus de deux ans et quelque quinze millions de morts, l'intelligence artificielle n'a pas tenu ses promesses. En réalité, elle ne pouvait pas les tenir, et ce pour deux raisons : d'abord, en raison de problèmes liés à la validité des données et de leur traitement algorithmique, et au risque de biais inhérent aux données massives ; ensuite, du fait que le paradigme épidémiologique de la pandémie n'est pas le bon, et qu'il convient de lui préférer celui de la « syndémie », la pandémie reflétant les inégalités économiques et sociales tout en les aggravant. Ainsi, ni une solution purement biomédicale ni une solution purement technologique ne peuvent fonctionner : il faut adopter une approche biosociale qui permet aux sociétés de vivre avec le virus, tout en protégeant les populations les plus vulnérables.

Introduction

Alors que la première vague épidémique de Covid-19 frappait le monde, de nombreux articles scientifiques (Arora *et al.*, 2020 ; Bragazzi *et al.*, 2020 ; Ho, 2020 ; Peng *et al.*, 2020 ; Vaishya *et al.*, 2020) décrivaient de façon enthousiaste comment l'intelligence artificielle allait permettre aux pays touchés de gérer la pandémie de façon optimale. Qu'il s'agisse des diagnostics, des pronostics, de la prévention, de la prédiction des vagues épidémiques, de thérapeutique, du traçage des contaminations, du développement de médicaments et de vaccins ou encore de la réduction de la charge de travail du personnel soignant, le recours à l'intelligence artificielle et aux données massives (ou *Big Data*) allait permettre de régler si ce n'est la totalité, du moins la majorité des problèmes engendrés par la pandémie de Covid-19, ressuscitant ainsi le mythe antique de Panacée¹.

Pourtant, plus de deux ans et cinq vagues épidémiques après le déclenchement de cette pandémie, qui a entraîné une surmortalité estimée entre 15 et

18 millions² de personnes dans le monde (Wang *et al.*, 2022), le bilan de l'utilité de l'intelligence artificielle dans cette pandémie apparaît largement contrasté. En effet, comme le souligne l'épidémiologiste et historien de la médecine, Joël Coste, « les données massives et l'IA n'ont apporté ni connaissance pertinente ni dispositif efficace pour la gestion de la crise » (Coste, 2020). Selon lui, il y a deux principales raisons à cet état de fait : « l'absence de données disponibles (sur l'origine géographique des sujets, l'activité des médecins généralistes, l'activité du secteur médicosocial...) et l'impossibilité d'un traitement rapide de certaines d'entre elles (les causes médicales de décès) » (Coste, 2020). En d'autres termes, soit les données n'étaient pas produites ou disponibles, soit elles étaient produites et disponibles mais pas en mesure d'être traitées rapidement.

Dans cet article, nous voulons montrer que même si les données avaient pu être produites ou traitées rapidement, cela n'aurait pas fondamentalement changé le cours de la pandémie, en France ou ailleurs, et cela pour au moins deux raisons distinctes mais liées. La première raison est intrinsèque à l'intelligence artificielle

¹ Panacée (mot issu du grec ancien, traduit par « remède universel ») est la fille du dieu grec de la médecine Asclépios et d'Hygie, la déesse de la santé, de la propreté et de l'hygiène, laquelle est aussi la sœur de Panacée.

² D'après les chiffres de l'Organisation mondiale de la santé : <https://www.who.int/fr/news/item/05-05-2022-14.9-million-excess-deaths-were-associated-with-the-covid-19-pandemic-in-2020-and-2021>

et aux données massives : ces données et leur traitement algorithmique sont en effet sujets à de nombreux problèmes (validité des données, et donc validité des modèles) et à des biais (sous- ou sur-représentation de certaines catégories d'âge, de sexe, relatives à l'origine ethnique, etc.) qui tendent à fausser les résultats. Cela est particulièrement vrai pour les applications de téléphonie mobile (et la mHealth, en général), sur lesquelles nous entendons nous focaliser ici. La seconde raison est externe et tient à la nature de cette pandémie : en effet, selon R. Horton, cette pandémie doit d'abord être considérée comme une « syndémie », qui révèle « les interactions biologiques et sociales importantes pour le pronostic, le traitement et les politiques de santé publique » (Horton, 2020) et qui reflète les inégalités économiques et sociales en même temps qu'elle les aggrave, le recours à l'intelligence artificielle ne faisant d'ailleurs bien souvent que renforcer ce phénomène.

En d'autres termes, le recours à des algorithmes ou à des applications apparaît comme une allégorie du « solutionnisme technologique » (Morozov, 2014) à l'œuvre dans nos sociétés contemporaines : tout problème (politique, moral ou sanitaire) devient un problème essentiellement technique. Or, les sociétés humaines, notamment lorsqu'elles sont confrontées à une pandémie, ne se laissent pas si facilement réduire ni analyser en termes algorithmiques. Pour autant, doit-on considérer que toute tentative de contrôle *via* l'intelligence artificielle est vouée à l'échec ? Ou bien doit-on conclure que le problème ne vient pas de l'instrument, mais plutôt de l'approche adoptée par les êtres humains par rapport au virus : un virus qu'il ne s'agirait pas de contrôler, mais plutôt avec lequel il faudrait composer, voire cohabiter ?

Usages et mésusages de l'intelligence artificielle en temps de pandémie

Si un secteur a particulièrement bénéficié de la pandémie de Covid-19, c'est bien évidemment celui du numérique : alors que la capitalisation boursière des GAFAM a explosé durant la pandémie³, un grand nombre d'applications permettant le traçage des contacts, l'autodiagnostic, le *reporting* médical ou bien encore le contrôle du respect de la quarantaine ont été mises en place dans au moins 55 pays dans le monde d'après le Conseil de l'Europe⁴. En France, l'application TousAntiCovid a ainsi été mise à disposition *via* des applications mobiles le 2 juin 2020 (sous le nom StopCovid). L'idée d'une application de traçage paraît en effet à la fois simple et séduisante : plutôt que de mener des enquêtes épidé-

miologiques de terrain longues et fastidieuses, il suffirait ainsi non pas d'interroger les personnes, mais d'interroger en quelque sorte leurs téléphones portables. En activant la fonctionnalité *bluetooth* desdits téléphones, il serait ainsi possible de repérer toutes les personnes ayant été en contact proche avec un individu contaminé (la portée du Bluetooth variant entre quelques mètres et vingt mètres), et donc de remonter les chaînes de contamination ou de détecter d'éventuels *clusters*. De plus, il serait possible (sans même passer par une application tierce) d'utiliser la fonction de géolocalisation des téléphones portables et, par ce biais, aussi bien de retracer les trajets de leurs utilisateurs que d'identifier les autres personnes que ceux-ci auraient pu croiser, ou encore de vérifier que les personnes respectent leur quarantaine ou le confinement. Ces données pourraient tout à fait être croisées avec d'autres données personnelles comme l'identité de la personne, ses caractéristiques socio-démographiques ou encore son dossier médical. Bien évidemment, des obstacles éthiques et juridiques de taille surgissent immédiatement, qu'il s'agisse des libertés individuelles, du droit à la vie privée ou encore du secret médical. Pourtant, au-delà des problèmes juridiques et éthiques posés par l'utilisation de ce genre de technologies, le recours à ces dernières et à des algorithmes pour analyser les données produites soulève un problème épistémologique plus profond, car consubstantiel à ces technologies, qui peut se résumer au risque de biais.

En effet, quel que soit le domaine d'application retenu pour analyser l'application de l'intelligence artificielle à la pandémie, un certain nombre de problèmes se posent, qui vont conduire à ce que les épidémiologistes appellent un biais, qu'ils définissent comme une « déviation systématique des résultats ou des inférences par rapport à la vérité, ou [comme] les processus conduisant à une telle déviation » (Last et International Epidemiological Association, 2001). Différentes sources de biais ont déjà été identifiées (Brault et Saxena, 2021) à propos de santé mobile (ou « mHealth »). Le problème principal concerne la qualité et la validité des données : en effet, les données peuvent, par exemple, être structurées ou encodées de manière différente en fonction du système d'exploitation (Android, Apple, etc.), ce qui rend problématique leur agrégation et leur analyse par un même algorithme (et, *a fortiori*, par des algorithmes différents). Ces données peuvent être aussi, en fonction des systèmes d'exploitation ou des applications utilisés, structurées ou seulement semi-structurées, voire même non structurées, ce qui rend là encore leur analyse difficile. De fait, comme dans tout système de données massives, certaines d'entre elles vont manquer : on sait, par exemple, que certaines personnes n'ont tout simplement pas de *smartphones*. Il existe ainsi des inégalités en termes d'équipements qui épousent à la fois la structure par âge de la population, mais aussi les inégalités sociales. Une étude de l'Insee⁵ a ainsi montré que si 94 % des 15-29 ans ont un *smartphone*, les 75 ans

³ L'acronyme GAFAM désigne les grands groupes technologiques américains : Google, Apple, Facebook, Amazon et Microsoft. À titre d'exemple, la capitalisation boursière d'Apple a progressé de 150 % entre mars 2020 et janvier 2022, tutoyant même la barre des 3 000 milliards de dollars en décembre 2021. À ce sujet, voir, par exemple, cet article du *Figaro* de janvier 2022, <https://www.lefigaro.fr/secteur/high-tech/la-folle-ascension-des-geants-americains-de-la-tech-20220105>

⁴ Voir : <https://www.coe.int/en/web/artificial-intelligence/ai-covid19>

⁵ LEGLEYE S., NOUGARET A. & VIARD-GUILLOT L. (2022), « 94 % des 15-29 ans ont un *smartphone* en 2021 », *Insee Focus*, n°259, 24 janvier, <https://www.insee.fr/fr/statistiques/6036909#titre-bloc-4>

ou plus (pourtant la catégorie de personnes la plus à risque de mourir du Covid-19) ne sont que 36 % à en posséder un. De même, dans la population générale, le taux d'équipement en *smartphones* passe de 53 à 92 % entre la population des moins diplômés et celle des plus diplômés ; tout comme il varie de 65 % pour les 10 % les plus pauvres à 95 % pour les 10 % les plus riches de la population française.

Pandémie, intelligence artificielle et syndémie

Or, d'après une autre étude de l'Insee (Barhoumi *et al.*, 2020), ce sont précisément les personnes en situation de précarité, mais aussi les plus âgées, qui ont été les plus touchées par la surmortalité due au Covid-19. En ce sens, il apparaît que les personnes les plus concernées par la pandémie sont en réalité aussi celles qui échappent le plus à une analyse en termes d'intelligence artificielle et de données massives, ce qui crée nécessairement un biais ou une distorsion systématique dans les résultats. Dès lors, si l'échantillon de données de départ implique une sous-représentation de certaines catégories de personnes, les plus pauvres et/ou les plus âgées, cette sous-représentation va se retrouver tant au niveau de la calibration que de l'entraînement du modèle, et donc au niveau de sa précision. Et l'erreur systématique va se répandre tout au long du processus algorithmique de l'analyse des données massives, et ce quelle que soit la taille de l'échantillon. En d'autres termes, espérer arriver à un contrôle de la pandémie *via* le recours à l'intelligence artificielle et aux données massives est largement illusoire à partir du moment où cette analyse est sans doute biaisée, l'intelligence artificielle ne faisant que refléter et, par là même, sans doute, qu'aggraver les inégalités économiques et sociales déjà présentes.

En ce sens, la pandémie et l'intelligence artificielle fonctionnent sensiblement de la même manière, au sens où elles agissent comme un miroir grossissant des inégalités, en même temps qu'elles risquent de les accentuer. C'est pourquoi la notion de syndémie, théorisée par Merrill Singer (Singer *et al.*, 2017), est si pertinente : en effet, « peu importe l'efficacité d'un traitement ou la protection offerte par un vaccin, la poursuite [sur la voie] d'une solution purement biomédicale au Covid-19 est vouée à l'échec » (Horton, 2020). Tout comme est condamnée à l'avance la solution purement technique, ou comme le dit Morozov, le « solutionnisme technologique ». En réalité, ce qu'il convient de changer c'est l'approche, en abandonnant une stratégie à la fois scientifique (ou médicale) et politique de la santé afin d'adopter un modèle « biosocial » : la notion de syndémie permet ainsi de comprendre comment les problèmes de santé, qu'il s'agisse de maladies transmissibles ou chroniques, relèvent d'une « agrégation (*clustering*) de problèmes médicaux et sociaux au niveau d'une population » (Singer *et al.*, 2017), des problèmes qui sont eux-mêmes en « interaction syndémique » entre eux et avec d'autres facteurs environnementaux et psychologiques. Cette approche nécessite ainsi à la fois une politique de santé centrée sur les

populations et pas seulement sur les individus, mais aussi une politique globale de réduction des inégalités qui vise à « l'élimination des forces sociales de grande envergure » (Tsai *et al.*, 2017), comme les inégalités de genre ou les discriminations raciales. Cela nécessite d'adopter une approche multi-niveau de la santé, qui prenne en compte aussi bien les déterminants individuels que collectifs, et tout autant les facteurs environnementaux (physiques ou biologiques) que les facteurs économiques et sociaux. En d'autres termes, il s'agit d'une « éco-épidémiologie » que Susser et Susser (1996) appelaient de leur vœu il y a déjà plus de quinze ans. Étant donné la complexité de cette approche, l'intelligence artificielle pourrait constituer un outil pertinent pour analyser les données, à condition que les biais inhérents à celle-ci soient si ce n'est éliminés, du moins minimisés.

Conclusion

Ainsi, il apparaît que toute tentative de contrôle d'une épidémie par l'intelligence artificielle ne soit pas, par principe, vouée à l'échec. En réalité, ce n'est pas tant l'instrument qui est ici en cause que les conceptions épidémiologiques et scientifiques qui sous-tendent cette approche. Ainsi, de la même manière qu'une approche purement biomédicale de la pandémie est inopérante, une approche purement technologique, et ce quelle que soit la technologie retenue, l'est tout autant, et relève bien d'une forme de techno-solutionnisme. En effet, après plus de deux années de pandémie, il ne s'agit plus tant d'éliminer le virus que d'apprendre à vivre avec lui et donc de composer avec lui, notamment en évitant ses conséquences désastreuses en termes de morbidité et de mortalité. Cela nécessite d'adopter une approche beaucoup plus complexe de la santé et des maladies qui incluent de nombreux facteurs individuels et collectifs de natures différentes (facteurs physiques, biologiques, économiques, sociaux, psychologiques...). Dans ce cadre, l'utilisation de l'intelligence artificielle et des données massives peut s'avérer pertinente : comme la langue d'Esopo, l'intelligence artificielle peut être la pire des choses, en aggravant les inégalités ou en étant utilisée comme un instrument de surveillance des populations, ou, au contraire, la meilleure des choses, en permettant de prendre en compte les multiples dimensions de la santé et d'identifier leurs interactions complexes afin de fournir, dans le cadre d'une approche syndémique, le corpus « des preuves [apportées] à l'appui d'interventions appropriées afin d'améliorer la santé et le bien-être psychosocial des populations vulnérables dans le monde entier » (Tsai *et al.*, 2017).

Références bibliographiques

- ARORA N., BANERJEE A. K. & NARASU M. L. (2020), "The role of artificial intelligence in tackling Covid-19", *Future Virology* 15(11), pp. 717-724. <https://doi.org/10.2217/fvl-2020-0130>
- BARHOUMI M., JONCHERY A., LOMBARDO P., LE MINEZ S., MAINAUD T., RAYNAUD E., PAILHÉ A., SOLAZ A. & POLLAK C. (2020), « Les inégalités sociales à l'épreuve de la

- crise sanitaire : un bilan du premier confinement », in *France, portrait social*, édition 2020, collection « Insee Références ».
- BRAGAZZI N. L., DAI H., DAMIANI G., BEHZADIFAR M., MARTINI M. & WU J. (2020), "How Big Data and Artificial Intelligence Can Help Better Manage the Covid-19 Pandemic", *International Journal of Environmental Research and Public Health* 17(9), p. 3176, <https://doi.org/10.3390/ijerph17093176>
- BRAULT N. & SAXENA M. (2021), "For a critical appraisal of artificial intelligence in healthcare: The problem of bias in mHealth", *Journal of Evaluation in Clinical Practice* 27(3), pp. 513-519, <https://doi.org/10.1111/jep.13528>
- COSTE J. (2020), « Données dites "massives" et santé publique : une mise en perspective historique », *Actualité et dossier en santé publique* 112, pp. 20-23.
- HO D. (2020), "Addressing Covid-19 Drug Development with Artificial Intelligence", *Advanced Intelligent Systems* 2(5), p. 2000070, <https://doi.org/10.1002/aisy.202000070>
- HORTON R. (2020), "Offline: Covid-19 is not a pandemic", *The Lancet* 396(10255), p. 874, [https://doi.org/10.1016/S0140-6736\(20\)32000-6](https://doi.org/10.1016/S0140-6736(20)32000-6)
- LAST J. M. & INTERNATIONAL EPIDEMIOLOGICAL ASSOCIATION (Éds.) (2001), *A dictionary of epidemiology*, 4th edition, Oxford, Oxford University Press.
- MOROZOV E. (2014), *Pour tout résoudre, cliquez ici : l'aberration du solutionnisme technologique*, Limoges, Fyp Éditions.
- PENG M., YANG J., SHI Q., YING L., ZHU H., ZHU G., DING X., HE Z., QIN J., WANG J., YAN H., BI X., SHEN B., WANG D., LUO L., ZHAO H., ZHANG C., LIN Z., HONG L., LI J. et al. (2020), "Artificial Intelligence Application in Covid-19 Diagnosis and Prediction", *SSRN Electronic Journal*, <https://doi.org/10.2139/ssrn.3541119>
- SINGER M., BULLED N., OSTRACH B. & MENDENHALL E. (2017), "Syndemics and the biosocial conception of health", *The Lancet* 389(10072), pp. 941-950, [https://doi.org/10.1016/S0140-6736\(17\)30003-X](https://doi.org/10.1016/S0140-6736(17)30003-X)
- SUSSER M. & SUSSER E. (1996), "Choosing a Future for Epidemiology: II. From Black Box to Chinese Boxes and Eco-Epidemiology", *American Journal of Public Health* 86(5), pp. 674-677.
- TSAI A. C., MENDENHALL E., TROSTLE J. A. & KAWACHI I. (2017), "Co-occurring epidemics, syndemics, and population health", *The Lancet* 389(10072), pp. 978-982, [https://doi.org/10.1016/S0140-6736\(17\)30403-8](https://doi.org/10.1016/S0140-6736(17)30403-8)
- VAISHYA R., JAVAID M., KHAN I. H. & HALEEM A. (2020), "Artificial Intelligence (AI) applications for Covid-19 pandemic", *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 14(4), pp. 337-339, <https://doi.org/10.1016/j.dsx.2020.04.012>
- WANG H., PAULSON K. R., PEASE S. A., WATSON S., COMFORT H., ZHENG P., ARAVKIN A. Y., BISIGNANO C., BARBER R. M., ALAM T., FULLER J. E., MAY E. A., JONES D. P., FRISCH M. E., ABBAFATI C., ADOLPH C., ALLORANT A., AMLAG J. O., BANG-JENSEN B., MURRAY C. J. L. et al. (2022), "Estimating excess mortality due to the Covid-19 pandemic: A systematic analysis of Covid-19-related mortality, 2020-21", *The Lancet* 399(10334), pp. 1513-1536, [https://doi.org/10.1016/S0140-6736\(21\)02796-3](https://doi.org/10.1016/S0140-6736(21)02796-3)