REVUE MÉDECINE ET PHILOSOPHIE

Philosophie de l'intelligence artificielle

#4 (2) / 2020 - 2021
REVUE SEMESTRIELLE

Sommaire

PHILOSOPHIE DE L'INTELLIGENCE ARTIFICIELLE

2020 - 2021, VOLUME 4 (2)

REVUE SCIENTIFIQUE
SEMESTRIELLE
INTERDISCIPLINAIRE
DE MÉDECINE ET PHILOSOPHIE

Cofondateurs et codirecteurs de publication :
Brice Poreau
Christophe Gauld

Adresse mail :
medecineetphilosophie@gmail.com

Site de la revue :
medecine-philosophie.com

ISSN :
2679-2069

e-ISSN :
2650-5614

Imprimeur :
Centre Littéraire d'Impression Provençal.
04 91 65 05 01
contact@imprimerieclip.com

Crédit Photo :
Enzo Zeylstra
Instagram @enzozey

Revue semestrielle gratuite,
en libre accès
Mentions légales consultables sur
le site internet

Fondation Après-Tout
Sous l'égide de Fondation de France
Accompagner les professionnels au service de l'humain

L'intelligence artificielle en santé et la reconnaissance du principe de Garantie Humaine

REVUE MÉDECINE ET PHILOSOPHIE

David Gruson*

*Directeur du Programme Santé Groupe Jouve, Fondateur Ethik-IA, Membre de la Chaire Santé ScPo Paris

RÉSUMÉ

La diffusion d'une vague d'innovations sur l'intelligence artificielle en santé modifie très profondément les pratiques médicales. Elle soulève également des enjeux majeurs de régulation des enjeux éthiques associés. La crise COVID19 a montré tout à la fois l'ampleur de ces opportunités et de ces risques. Une voie de passage pragmatique doit donc être aménagée pour permettre la mise en œuvre pratique de ces innovations au service des patients. L'objectif est donc celui d'une régulation positive pour préserver les valeurs essentielles de notre système de santé et plus particulièrement les fondements d'une médecine personnalisée. C'est le sens de la Garantie Humaine de l'IA, principe reconnu par l'article 11 du projet de loi de bioéthique qui doit entrer en vigueur dans les prochaines semaines.

MOTS-CLÉS : Intelligence artificielle, Machine Learning, Garantie Humaine, Comité consultatif national d'éthique, Bioéthique, Data Management, RGPD.

DOI : 10.51328/101

L'intelligence artificielle en santé : perspectives d'avancées et besoin de régulation positive

Le déploiement de l'intelligence artificielle est source d'améliorations potentielles littéralement extraordinaires pour notre santé. Ce saut qualitatif possible pour notre système de soins s'accompagne d'un effet de levier potentiel majeur pour la croissance de la France. Il induit également des transformations possiblement radicales des métiers du champ sanitaire et médico-social. Nous vivons actuellement une véritable révolution des cas d'usage de recours à l'intelligence artificielle. La technique la plus mature est, de loin, la reconnaissance d'image par apprentissage machine. Elle trouve d'ores et déjà à s'appliquer largement en radiologie, dermatologie, ophtalmologie ou encore en oncologie. Dans toutes ces disciplines, les performances de diagnostic des algorithmes dépassent désormais fréquemment celles de l'Humain.

Dans *La Machine, le Médecin et Moi*[1], je montrais à quel point notre système de santé était en train de connaître une transformation profonde sous l'effet de l'émergence progressive d'un véritable data management en santé.

La collecte massive de données – ce que l'on a pris l'habitude d'appeler un peu vite « big data » – constitue une condition *sine qua non* au déploiement de ce *data management* et de l'intelligence artificielle. Cette dernière se fonde, en effet, sur des algorithmes qui nécessitent la mobilisation de données fiables et en nombre suffisant pour dégager des calculs robustes de probabilités permettant d'appuyer les orientations de l'intelligence artificielle.

Cette donnée-requérance ne constitue pas un principe nouveau : les *datas* constituent la matière première, la base de l'alimentation de tout programme informatique. Le déploiement actuel du pilotage par les données et de l'IA sur un large spectre fait simplement changer l'enjeu

[1] Editions de l'Observatoire, novembre 2018.

d'échelle.

Une « course aux données de santé » s'est donc engagée au niveau mondial avec les signaux de plus en plus visibles d'une compétition exacerbée. Pour pouvoir approvisionner les algorithmes, ces données doivent être médicalement et techniquement fiables mais également en volume suffisant pour permettre à l'IA de s'appuyer sur des régularités statistiques robustes. Cette compétition internationale pour les données de santé est naturellement marquée par un facteur temps. Le premier fabricant de solutions de *data management* et d'IA qui sera parvenu à élaborer une solution en santé opérationnelle alimentée par une masse de données fiables et en nombre suffisant aura acquis un avantage sans doute décisif sur ses concurrents. Avec naturellement des perspectives financières colossales à la clé.

Pour autant, cette diffusion rapide de l'intelligence artificielle en santé est aussi génératrice d'un besoin de régulation éthique. La crise sanitaire que nous traversons l'a, à nouveau, fortement montré. Dans la réponse au COVID19, certains pays – en particulier en Asie – ont eu plus largement recours que d'autres à l'IA, au pilotage par les données et aux technologies numériques. Des dispositifs de reconnaissance faciale ainsi que l'utilisation de thermomètres connectés ont permis la surveillance de la température et l'identification de personnes à risque d'être positive au COVID19. Les données de géolocalisation ont été largement utilisées pour connaitre les flux des personnes et bloquer certains déplacements. Des robots ont été introduits dans certains hôpitaux, pour accompagner voire renforcer les équipes médicales, assurer une présence auprès des patients et répondre à leurs besoins, décontaminer certains services… Le *Data Tracking* a été mis en œuvre sans restriction et un choix d'efficacité a été fait au détriment de la protection des données de santé.

Mais au-delà des recours à ces dispositifs fortement visibles et dont les usages ont été pour certains contestables au regard des principes de nos sociétés démocratiques, le recours à l'intelligence artificielle en temps de crise épidémique porte aussi la promesse d'accélération de certains procédés relevant aussi bien du diagnostic que de la recherche. Des méthodes diagnostiques reposant sur la reconnaissance d'images par apprentissage machine permettraient un diagnostic beaucoup plus rapide et efficace sur la base de clichés de tomodensitométrie. L'IA induit aussi un potentiel d'apport majeur concernant l'identification d'éventuels traitements efficaces.

Les options radicales, choisies par certains pays, pour répondre au risque collectif au détriment de la protection des libertés individuelles, semblent très éloignés des principes fondateurs du règlement général sur la protection des données (RGPD) en Europe et, plus largement, des valeurs essentielles de notre médecine personnalisée. Cette gestion de crise illustre l'importance et l'intérêt de la Garantie Humaine de l'IA, consistant dans la mise en place d'une supervision humaine lors du recours à un algorithme d'intelligence artificielle.

La Garantie Humaine de l'intelligence artificielle en santé : clé essentielle d'une régulation éthique positive

Les principes au fondement de notre médecine personnalisée en France et en Europe peuvent entrer en confrontation avec un certain nombre de principes éthiques. Dans les deux tomes de la fiction d'anticipation *S.A.R.R.A.*[2], je décris la dystopie d'une Europe dominée par les géants numériques américains et chinois – PanGoLink et FU-TECH – et qui n'a d'autre recours, face à une pandémie d'Ebola, que de jeter aux oubliettes tous les principes qui l'avaient amenée à établir le règlement général sur la protection des données.

L'avis 129 du Comité consultatif national d'éthique[3], émis dans le cadre de la préparation de la révision bioéthique, a identifié ces risques éthiques et notamment ceux d'une délégation pratique de la décision du médecin et du consentement du patient à l'IA. Le principal danger est donc celui de la perte d'un recul critique des soignants et des soignés avec, en arrière-plan, une mécanique algorithmique fondée sur la loi du plus grand nombre, cette dernière pouvant aller à l'encontre d'intérêts d'individus ou de groupes d'individus.

Pour éviter une perspective aussi sinistre, le principe de Garantie Humaine de l'IA est issu d'un mouvement de propositions académiques, citoyennes mais aussi de professionnels de santé. La reconnaissance de ce principe dans l'article 11 du projet de loi bioéthique et sa concrétisation était le sens même du projet porté depuis le départ par Ethik-IA. Ce texte doit désormais entrer en vigueur au début de l'année 2021. Il comprend deux normes nouvelles : l'information du patient sur le recours à l'IA dans sa prise en charge, d'une part, et le principe de Garantie Humaine de l'IA lui-même d'autre part. Le concept de « Garantie Humaine » peut paraître abstrait mais il est, en réalité, très concret. Dans le cas de l'IA, l'idée est d'appliquer les principes de régulation de l'intelligence artificielle en amont et en aval de l'algorithme lui-même en établissant des points de supervision humaine. Non pas à chaque étape, sinon l'innovation serait bloquée. Mais sur des points critiques identifiés dans un dialogue partagé entre les professionnels, les patients et les concepteurs d'innovation.

La supervision peut s'exercer avec le déploiement de « collèges de garantie humaine » associant médecins, professionnels paramédicaux et représentants des usagers. Leur vocation serait d'assurer *a posteriori* une révision de dossiers médicaux pour porter un regard humain sur les options thérapeutiques conseillées ou prises par l'algorithme. L'objectif consiste à s'assurer « au fil de l'eau » que l'algorithme reste sur un développement de *Machine Learning* à la fois efficace médicalement et responsable éthiquement. Les dossiers à auditer pourraient être définis à partir d'événements indésirables constatés, de critères prédéterminés ou d'une sélection aléatoire. Nous avons tenu, avec Ethik-IA, la première session pilote de collège de garantie humaine sous l'égide de l'Union française pour la santé bucco-dentaire (UFSBD) dans le cas d'une solution d'IA appliquée dans le domaine des soins bucco-dentaires (protocole innovant de l'article 51 de la loi de financement de la Sécurité sociale). Il est aussi à relever que le principe de garantie humaine a reçu, en 2020, des concrétisations dans des cadres très significatifs :

- D'une part, la Garantie humaine a été incorporée dans la grille d'auto-évaluation des dispositifs médi-

[2] S.A.R.R.A., une intelligence artificielle (juin 2018) ; S.A.R.R.A., une conscience artificielle (mars 2020), Editions Beta Publisher

[3] https://www.ccne-ethique.fr/fr/actualites/lavis-129-contribution-du-ccne-la-revision-de-la-loi-de-bioethique-est-en-ligne

caux intégrant de l'IA publiée par la Haute Autorité de Santé au mois d'octobre dernier ;

- D'autre part, le principe de garantie humaine fait actuellement l'objet de discussions et prolongements dans le cadre de la task-force dédiée par l'OMS à la régulation de l'IA en santé en vue de l'émission d'une recommandation dans le courant de ce premier semestre 2020.;
- Enfin, le principe a été repris dans le Livre Blanc sur l'IA publié par la Commission européenne le 19 février 2020.

Conclusion

Ces reconnaissances successives et de plus en plus larges du principe de Garantie Humaine de l'intelligence artificielle représentent autant d'avancées considérables dans le sens d'une logique de régulation positive du numérique et de l'intelligence artificielle en santé. A l'issue d'une année 2020 éprouvante à bien des égards pour notre système de santé et celles et ceux qui le font vivre au quotidien, il s'agit là sans doute d'une lueur d'espoir pour l'innovation technologique au service des patients

Nietzsche critique du transhumanisme. Fécondité d'un anachronisme philosophique

REVUE MÉDECINE ET PHILOSOPHIE

Arnaud Sorosina*
*HiPhiMo, Paris I

RÉSUMÉ Certains penseurs transhumanistes se sont réclamés de Nietzsche pour défendre l'idée que le transhumanisme est généalogiquement lié à sa pensée, en particulier à la notion de surhumain. Or, l'examen des engagements idéologiques propres au projet transhumaniste montre que, loin de leur être appariée, la philosophie de Nietzsche constitue par anticipation la réfutation radicale des ambitions transhumanistes, sur le fondement d'une évaluation de la forme de vie nihiliste qui préside à leur expression.

MOTS-CLÉS : transhumanisme, Nietzsche, valeurs, modernité, longévité, nihilisme.
DOI : 10.51328/102

Dans les lignes qui suivent, nous aimerions montrer, en quelque sorte de manière borgésienne, dans un exercice de philosophie-fiction, comment une pensée apparemment datée historiquement trouve précisément sa fécondité dans sa capacité transhistorique à ne pas être de son temps.

C'est de bonne guerre, d'abord dans la mesure où nous voulons aborder quelques aspects du transhumanisme à la lumière d'un penseur qui n'a pas connu ni le mot ni la chose, mais qui se présentait comme un penseur inactuel, comme est appelé du reste à le devenir tout penseur dont les idées ne font pas partie de la *cognitio ex datis* (et par suite de l'histoire factuelle), mais de la *cognitio ex principiis* (et par suite de l'histoire, intemporelle, des pensées irréductibles à leur contexte d'émergence).

C'est de bonne guerre à plus forte raison dans la mesure où un certain nombre de transhumanistes de ces dernières décennies se sont réclamés de Nietzsche pour en faire l'étendard de leurs propres conceptions[1], tant il est vrai que moins on possède d'enluminures nobiliaires pour donner de la substance à une pensée faible, plus on a besoin de s'inventer un pedigree idéologiquement fantasmé pour doubler du secours de l'autorité généalogique les insuffisances discursives de la seule raison argumentative.

Le lecteur aura deviné que le ton du présent article n'est pas celui communément admis par la « science normale », précisément parce que, prenant au sérieux la critique nietzschéenne de la volonté de vérité formulée dans le §344 du *Gai Savoir*, l'auteur de ces lignes voudrait mettre à l'épreuve cette critique en assumant explicitement un *ethos* polémique insoumis non pas aux bienséances de l'argumentation, mais aux postures d'objectivité de chercheurs qui, après avoir adultéré le sens d'un texte qu'ils exproprient de sa portée, s'estiment fondés à l'inféoder à des causes qui constituent une véritable aliénation de sa nature. C'est du moins ce que nous souhaiterions démontrer : quelle que soit notre adhésion ou non aux critiques que Nietzsche aurait à formuler à l'encontre des postulats axiologiques et doctrinaux du transhumanisme, nous voudrions montrer que, constitutivement, la philosophie de Nietzsche est l'ennemie la plus résolue de ses tenants et aboutissants.

Le transhumanisme : une « idée moderne »

Le premier credo transhumaniste qui paraît explicitement incompatible avec la pensée de Nietzsche réside dans la conviction optimiste selon laquelle le degré de développement de la raison instrumentale serait un critère de mesure de la valeur d'une théorie, sinon d'une civilisation. Cette idée, qui fait partie de ce que Nietzsche appelle les « idées modernes », va de pair avec la conviction selon laquelle les progrès technologiques seraient les garants

[1] Voir par exemple K.-G. Giesen (2004). Giesen voit dans l'idée de transhumain une réactivation de la figure nietzschéenne du surhumain.

du bonheur humain, dans la mesure où ils actualiseraient le projet cartésien d'une optimisation médicale de la santé humaine, cette dernière étant envisagée sommairement comme conservation de soi et augmentation du confort. Le §4 de *l'Essai d'autocritique* ajouté en 1886 pour la réédition de *La Naissance de la tragédie* (1872) prend précisément le contrepied d'une telle évaluation de la raison suffisante, qui apparaît aux yeux de Nietzsche comme une suffisance de la raison. L'étiologie de l'optimisme des Lumières, dont participe le projet transhumaniste, conduit ainsi à soutenir que son fondement psychoaffectif réside dans une aspiration quiétiste à l'ataraxie, que Nietzsche trouve à l'œuvre dans l'optimisme des écoles socratiques – aspiration qui préside à toutes les eschatologies séculières de la modernité, avec leurs lendemains qui chantent et leurs projets de perfectionnement de l'humanité[2]. C'est ce dont témoigne encore la section du *Crépuscule des idoles* où Nietzsche prend à partie « ceux qui veulent améliorer l'humanité ». L'idée d'*enhancement*, à ce titre, n'est que la version technicienne de cette volonté morale d'amélioration.

Il est vrai que Nietzsche promeut l'intensification de la puissance, si l'on entend par là le sentiment cénesthésique et affectif d'un approfondissement de nos catégories de perception de la réalité, notamment quant à notre grille d'évaluation esthétique. Mais c'est en ce cas le corps propre, ce que la phénoménologie nomme la « chair » (*Leib*), et certainement pas le corps-objet (*Körper*), qui se trouve invoqué comme le nouveau centre de gravité qui doit faire l'objet d'une revitalisation, cette dernière se jouant dans un épanouissement intensif du Soi quant à ses modes de subjectivation (disponibilité esthésique, créativité artistique), et non dans une augmentation extensive de ses modes d'objectivation (efficacité, longévité).

De ce point de vue, on prend la mesure de l'intensification de la vie, du point de vue nietzschéen, en fonction des fonctions propres à la vie telle qu'elle se définit dans ses manifestations phénoménologiques et non pas biologiques[3]. Le propre de la vie, c'est d'être un centre d'activité qui s'approprie des forces en les digérant pour leur insuffler un sens et une valeur en vertu d'une affectivité singulière. La vie, à travers un individu par exemple, est ainsi une puissance *herméneutique* d'individuation, et par suite une capacité à produire des comportements et des œuvres dépositaires d'une *Stimmung*, d'une tonalité affective à nulle autre semblable, comme ce peut être le cas, disons, des *Nocturnes* de Chopin.

Ce que Nietzsche appelle la grande santé (par exemple dans le §382 du *Gai Savoir*) apparaît dès lors comme la capacité d'un vivant ou d'une forme de vie partagée à résorber une lésion qu'elle a souvent provoquée elle-même pour conquérir sa puissance à partir d'un entraînement qui l'exerce à intensifier sa capacité de cicatrisation – cette métaphore servant à décrire les processus psychoaffectifs d'intégration d'une blessure, de façon telle que le Soi intensifie sa force d'unification à mesure qu'il se rend poreux à une altérité qui constitue son aliment.

Or, c'est très exactement l'inverse que promeut le transhumaniste, si l'on nous permet d'utiliser l'article défini pour en dessiner le portrait conceptuel comme s'il s'agissait d'un type. Les déterminations essentielles que nous associons à ce terme apparaîtront au fil de notre démonstration.

Commençons par celle qui apparaîtrait comme la plus fondamentale du point de vue nietzschéen : la vie apparaît au transhumaniste comme un ensemble de fonctions anatomo-organiques non pas vécues, mais objectivées dans l'espace de leur manifestation corporelle. Dès lors, il mesure la vigueur de la vie à l'aune de critères quantitatifs qui en évaluent les performances dans la terminologie libéral-darwinienne de l'efficacité, de l'adaptabilité et de la longévité.

Par conséquent, la vie apparaît au transhumaniste comme un ensemble de tâches techniques qui peuvent et doivent être prolongées, complétées et améliorées par des techniques objectives étrangères à la vie même, à moins que, dans ce modèle qui technicise la vie, la technique n'apparaisse alors comme une vie plus réussie, plus efficace et plus durable – auquel cas le posthumain[4] hybridé serait une forme plus authentique de la vie même, à savoir la réalisation la plus vigoureuse de l'homme-machine.

Il ne nous appartient pas ici de critiquer le monisme techniciste (même s'il est passablement suranné et nous reconduit au matérialisme sommaire des iatromécaniciens galiléens[5]) qui prétend réduire la vie à ses composantes mécaniques. Remarquons simplement que Nietzsche défend une forme de monisme diamétralement opposée, puisque c'est un monisme vital qui invite à considérer la physique des objets inertes et la mécanique des objets techniques comme une proto-forme de la vie, comme de la vie à l'état le plus infinitésimal, autant dire le moins vital qui soit[6].

Anatomo-pathologie du transhumanisme

Dans cette perspective, que révèle l'idéal transhumaniste à un examen généalogique qui obéirait aux critères nietzschéens de l'évaluation, sinon l'expression d'une forme de vie si diminuée qu'elle n'envisage même plus sa réparation ou sa résilience à partir de sa propre puissance – *mais à partir de puissances étrangères à la vie* ?

L'individu des démocraties libérales modernes imaginait encore son bonheur à partir d'excitants physiologiques congrus à sa forme de vie. Même s'il faisait contresens en interprétant comme curatives des méthodes qui le confortaient dans ses propres travers pathologiques, ses pharmacopées artificielles étaient destinées à pallier la vacuité de sa forme de vie par des moyens propres *à lui rappeler qu'il était vivant*. Il veut « vivre des expériences », « vivre intensément », mais est incapable de trouver en lui-même les ressorts physiologiques de cette intensification. Raison pour laquelle il fait appel à des

[2] Sur ce point, Nietzsche est incontestablement à l'opposé des aspirations posthumanistes, y compris à l'époque où il revendique un renouveau de la culture allemande : « Lutte contre l'idée que le but de l'humanité se situe dans l'avenir [...]. L'humanité n'est pas là pour elle-même, son but est dans ses sommets, les grands artistes et les grands saints, donc ni devant ni derrière nous. » (*Fragments posthumes* (désormais FP), 1870-1871, 7 [100]).

[3] Il suffirait pour s'en convaincre de prendre la lecture de la section d'*Ainsi parlait Zarathoustra* intitulée « Du dépassement de soi », où Nietzsche oppose le Soi-volonté de puissance au petit « moi »-citadelle de la tradition métaphysique, dont l'individu-monade transhumain apparaît comme la dégénérescence néolibérale.

[4] Nous suivons à peu de choses près la distinction établie par J.-Y. Goffi entre transhumanisme et posthumanisme pour distinguer le projet d'amélioration transhumain de l'utopie eschatologique posthumaine qui en constitue la mouture la plus radicalement messianique. Voir J.-Y. Goffi, « Transhumanisme » in M. Kristanek (dir.), *Encyclopédie philosophique* [En ligne : http://encyclo-philo.fr/transhumanisme-a/].

[5] Voir sur ce point l'article canonique de G. Canguilhem, « Organisme et machine » in *La Connaissance de la vie*, Paris, Vrin, 1952.

[6] Voir le §36 de *Par-delà bien et mal*.

sources d'excitation et à des stimuli extérieurs pour se sentir vivant sur le mode passif d'une vitalité pour ainsi dire pathologiquement extorquée à lui-même : drogues, alcool, excitants divers sont autant de palliatifs qui maintiennent artificiellement stimulée la réactivité sensible d'une forme de vie en manque d'intensité[7].

Ce ne sont pas là encore, contrairement à ce que certains transhumanistes suggèrent hâtivement – en amateurs de raccourcis qui font passer des approximations pour des identités[8] –, des techniques d'augmentation transhumaines, dans la mesure où l'entrée dans la transhumanité ne se fait qu'au moment où une forme de vie renonce à elle-même jusqu'à vouloir son auto-abolition *sous une autre forme que l'auto-anéantissement.*

De ce point de vue, le suicide n'est pas *encore* l'expression la plus aboutie du nihilisme, comme Schopenhauer l'avait bien compris, car c'est toujours à l'aune d'un certain idéal de vitalité qu'une forme de vie se supprime elle-même. Dans le transhumanisme en revanche, il atteint son point de non-retour. En effet, l'idée d'une renaissance de la vie sous une forme qui n'est plus elle substitue à la vie, *dans le monde*, autre chose qu'elle-même. La vie cherche à guérir d'elle-même en renaissant sous une forme d'abord partiellement vivante (transhumanisme) et, à terme, non-vivante (posthumanisme). L'hybridation de la vie et de la technologie signifie fantasmatiquement que la vie a renoncé à revêtir les attributs bio-esthétiques qui la déterminaient comme vie, entretenant la croyance que la vie non-biologique est encore vivante, et même plus authentiquement que sous ses anciens avatars, désormais considérés comme obsolètes[9]. Voilà une chrysalide, inattendue et qui pourtant porte à son paroxysme le processus d'épuisement séculier que recouvre l'histoire – narrée dans la *Généalogie de la morale* – des métamorphoses de l'idéal ascétique.

Dans une perspective nietzschéenne, le transhumanisme appartient donc à la structure même de l'histoire du nihilisme, dont il vérifie la cohérence quasi-téléologique jusqu'à lui donner sa forme la plus aboutie. Reste à vérifier que, du point de vue typologique, le transhumanisme est bien l'*accomplissement nihiliste* du dernier homme dont Nietzsche a déterminé fameusement le type.

La sécularisation technicienne de Dieu

Le transhumanisme serait ainsi un avatar du christianisme qui s'ignore ? La thèse paraît audacieuse, mais à partir de l'étiologie nietzschéenne de la culture, la généalogie médicale de la science positiviste comme christianité séculière conduit à un diagnostic qui anticipe celui de Jacques Ellul, lorsque ce dernier explique que « [l'homme] reporte son sens du sacré sur cela même qui a détruit tout ce qui en a été l'objet : la technique[10] ». La technoscience transhumaniste, à la lumière de Nietzsche, n'est rien d'autre que la pérennisation d'une collusion ancienne entre science et technique, dont elle n'est que la conséquence la plus aboutie, s'il est vrai qu'après l'arraisonnement de la nature hors de l'homme, il ne restait plus à ce dernier qu'à opérer l'arraisonnement de la nature en l'homme, sinon de la nature de l'homme. La destruction technicienne des idoles religieuses aboutit à un christianisme désenchanté qui, ou bien ne parvient pas à faire son deuil des anciennes transcendances (pessimisme, nihilisme), ou bien renaît sous d'autres formes, que l'on dirait « crypto-chrétiennes ». En solde de tout compte, la destruction des anciennes idoles a pour condition, dans la modernité, l'érection d'idoles nouvelles. Simplement, des idoles matérielles ont remplacé les idoles immatérielles, si bien que les hommes modernes semblent rejouer l'épisode de l'adoration du veau d'or, dans une conjoncture que Nietzsche appelle symboliquement la fête de l'âne – antithèse du pessimisme, qui consiste à dire oui (*ya* – « hi-ha ») à tout, ce qui est exactement la définition de la vulgarité :

> *« La position de la religion à l'égard de la nature a été, autrefois, inverse : la religion correspondait à la conception vulgaire de la nature. / Aujourd'hui, la conception vulgaire est la théorie matérialiste. Par conséquent, ce qui subsiste actuellement de religion se doit de parler en matérialiste au peuple*[11]*. »*

Cette science matérialiste qui se prend pour la valeur-étalon d'une civilisation est l'une des ombres de Dieu les plus délétères et les plus insidieuses qui soit, aux yeux du philosophe. L'éloge nietzschéen de la science ne vaut que pour la science comme méthode, lorsque la science envisage sa propre activité comme le moyen d'autre chose qu'elle-même, et met sa volonté de vérité au service de la culture. En revanche, Nietzsche réprouve la théodicée dissimulée des scientifiques qui, faisant de la recherche scientifique (et, plus tard, du progrès technoscientifique) une fin en soi, assurent à l'idole monothéiste des modes de survie séculiers d'autant plus fourbes qu'ils ne le sont pas toujours, comme c'est le cas du positivisme de Comte, qui finit par s'afficher ouvertement comme « Religion de l'humanité[12] ».

Or, le transhumanisme est bel et bien une religion séculière, par son culte de la science (et conformément à l'histoire qu'il se donne[13]) : l'humanisme évolutionnaire de Julian Huxley entendait précisément s'élever au statut de religion sans révélation, en tant qu'idéologie où les sciences de la nature et les techniques eugéniques seraient censées remplacer la dogmatique religieuse. Or, c'est très exactement ce genre de sécularisation scientifique du religieux par le programme évolutionniste de naturalisation de la morale que Nietzsche prend à partie, dès la première *Considération inactuelle* (1873).

[7] Sur la généalogie de cette intensité vitale envisagée comme une valeur typiquement moderne, voir Garcia (2016).

[8] Un certain nombre de transhumanistes soutiennent que nous sommes déjà des humains augmentés, par le port des lunettes, des lentilles et ainsi de suite. Si tel était le cas, alors il faudrait considérer que nous sommes déjà augmentés du fait d'être ontologiquement diminués par rapport aux animaux qui portent leur monde technique avec eux : dans la mesure où toute activité technique humaine est originairement exogène, *homo habilis* était déjà en ce cas un transhumain. Il apparaît donc que l'argument, destiné à précipiter l'avènement d'un futur cyborg en invoquant non pas tant son imminence que sa réalisation anticipée, est une pétition de principe si absurde qu'elle porterait à considérer que nous avons toujours été transhumains. La plasticité sémantique de cette « transhumanité » invite en l'occurrence à penser que nous avons moins affaire ici à un concept qu'à un idéologème manipulable *ad libitum*.

[9] Nous exprimons ici encore notre désaccord résolu à l'encontre de Sorgner (2017), qui fait de Nietzsche un admirateur de la science – ce qui formulé en ces termes n'est ni vrai ni faux – et par conséquent de la technologie – ce qui est un contresens total. Tuncel (2017, p. 3) a raison d'observer que la science n'a absolument aucune valeur en soi, mais seulement en fonction du type de vie qu'elle promeut. Or, le transhumanisme promeut la forme de vie pour laquelle Nietzsche exprime son mépris le plus profond.

[10] *La Technique ou l'enjeu du siècle* [1954], Economica, 1990 p. 132.

[11] *FP novembre 1882-février 1883, 4 [221].*

[12] *Idem*

[13] Comme l'a rappelé G. Hottois (2017, p. 37).

Voilà pourquoi l'idée d'auto-transcendance que promeuvent les transhumanistes n'a rien de commun avec l'autodépassement nietzschéen, qui signifie une *autodestruction* du christianisme non seulement comme dogme, mais surtout comme morale : transvaluation de toutes les valeurs[14]. À cet égard, une partie non négligeable des transhumanistes ne font jamais que poursuivre l'œuvre de John Desmond Bernal dans *The World, the Flesh and the Devil*, qui considère l'auto-transcendance de l'homme, par l'intermédiaire de la technoscience – si du moins le terme a un sens à propos des techniques de transformation et de sélection du début du XX^{e} siècle, si on les compare à ce que la convergence NBIC promet d'accomplir –, comme la forme sécularisée de l'aspiration judéo-chrétienne à la perfection et à l'éternité[15]. Naturellement, les transhumanistes contemporains prennent leur distance à l'égard du mysticisme de Julian Huxley ou de l'enracinement chrétien de Bernal, mais il importe de reconnaître que leur *profession d'athéisme* dissimule une *profession de foi* involontairement reconduite par la force de séduction des valeurs chrétiennes qu'ils véhiculent encore sous des dehors modifiés : la *philosophie de l'histoire* qu'ils sont bon an mal an amenés à adopter reconduit ce procès de sécularisation dans lequel Nietzsche voyait l'un des plus grands dangers de la modernité[16].

Ce n'est du reste pas un hasard si Bernal et Haldane se firent membre du parti communiste, dont le messianisme était propre à donner une consistance politique à leur vision eschatologique de la science, tandis que Huxley défendait pour sa part un socialisme progressiste. On comprend dès lors pourquoi Nietzsche est si cavalier dans le traitement qu'il réserve à ces idéologies politiques, puisque les différences doctrinales entre l'anarchisme, le socialisme et le communisme ne l'intéressent guère, du moment qu'ils expriment tous à leur manière l'idéal ascétique et surtout, l'aspiration au dernier homme qui, on le comprend désormais, est cet homme qui cligne de l'œil, d'un bonheur entendu, lorsqu'il s'entend prononcer le mot « transhumain »[17].

L'évolutionnisme transhumain et l'évolutionnisme surhumain

Il ne faut donc pas s'en laisser conter par les prétentions objectivistes de l'évolutionnisme transhumaniste, lorsque celui-ci s'aventure à subsumer l'histoire humaine sous ses explications dont la grandiloquence ne doit dissimuler ni le fumet social-darwinien, épistémologiquement suspect, ni les tours de passe-passe qui voudraient établir l'idéologie néo-libérale sur des fondements moraux reposant tout entier sur une immense pétition de principe.

En effet, l'évolutionnisme qui constitue l'assise du projet transhumaniste est, comme dirait Bergson, un faux évolutionnisme. Non pas tant parce qu'il reconstitue l'évolution avec des fragments de l'évolué – comme Bergson en fait reproche à Spencer[18] –, que dans la mesure où il prétend dessiner l'image de l'homme futur à partir des idéaux de l'homme actuel. En somme, c'est un évolutionnisme à deux vitesses – épistémologique et morale –, en ce sens qu'il ne tient pas compte du fait que les valeurs sont elles-mêmes soumises à l'évolution historique qui transforme les corps. L'axiologie qui définit un fait, une action ou une vie comme « humains » ou digne d'être vécus n'est pas intemporelle, comme Nietzsche n'a cessé de le reprocher aux utilitaristes et aux généalogistes anglais de la morale, qui croient encore, chrétiennement, à l'anhistoricité de leurs critères d'évaluation (par exemple l'idée qu'une vie réussie est une vie qui serait la plus longue possible). Or, si l'on tient compte du fait que la manière d'évaluer est sécrétée par le type d'organisation pulsionnelle des corps, il n'y a plus aucun sens à vouloir statuer sur ce que *devrait être* la physiologie de l'homme de l'avenir, ou du transhumain, ou même bien sûr du surhumain. Raison pour laquelle Nietzsche détermine l'orientation tendancielle du surhumain sans jamais spécifier son contenu, puisqu'il ne s'agira pas d'une forme fixe et que la tâche de créer de nouvelles valeurs ne peut pas lui incomber. L'évolutionnisme transhumaniste laisse ainsi intacte l'anhistoricité des valeurs, lors même qu'elle prétend en établir le fondement évolutif.

Nietzsche anticipe très clairement les errements du transhumanisme sur ce point lorsqu'il prend à partie la *Natural History of Morals* anglo-saxonne pour montrer qu'elle rend compte de l'histoire de la morale sans se rendre aucunement attentive à l'*historicité des valeurs* dont elle prétendait pourtant rendre compte[19]. C'est encore un point sur lequel la pensée évolutionniste darwinienne et post-darwinienne – dans son anthropologie philosophique, c'est-à-dire depuis *La Filiation de l'homme* (1871) –, n'est pas allée jusqu'au bout de son entreprise épistémologique. En effet, elle a renversé la logique explicative de la religion, tout en sécularisant les valeurs religieuses en se contentant de les naturaliser pour en proposer une fondation (pseudo-)scientifique. Voilà pourquoi, dès l'époque où il rédige *Humain, trop humain* (1878), et où son allégeance aux méthodes scientifiques anglo-écossaises est pourtant la plus résolue, Nietzsche maintient tout de même certaines réserves, lorsqu'il rappelle que « beaucoup, sans s'en rendre compte, prennent même pour la forme stable dont il faut partir la toute dernière figure de l'homme, telle que l'a modelée l'influence de certaines religions, voire de certains événements politiques. Ils ne veulent pas comprendre que l'homme est le résultat d'un devenir[20] ». Ici, il n'est pas seulement question de l'homme sur le plan physiologique – sur ce point, Nietzsche partage le diagnostic de l'orthodoxie évolutionniste, quoi qu'il soit en complet désaccord sur le principe de sélection[21]. Il est surtout question de l'homme *moral* et du fait que les types d'évaluation ont eux aussi évolué, quoique cette évolution ait été précipitée par les transformations historiques et non par l'évolution naturelle en tant que telle.

Parvenus à ce point, une conséquence de la critique nietzschéenne de l'évolutionnisme de son temps devrait inviter à reconsidérer les lectures cavalières des textes où il en appelle à un dépassement de l'homme. De fait, dans

[14] Voir la Généalogie de la morale, III, §27.
[15] Voir notamment surtout Tirosh-Samuelson (2012).
[16] Voir par exemple les §343 et 344 du *Gai Savoir*.
[17] Bien sûr, les transhumanistes défendent l'idée que « l'évolution autonome du surhomme » est un « développement ouvert », depuis R. Ettinger (1972, p. 14). Mais cela n'enlève rien au fait que les idéaux et les valeurs transhumanistes, elles, sont bien intouchables : utilitaristes et proactives. Si donc il n'existe aucun groupe en principe auquel il faudrait remettre les clés de l'avenir pour qu'il décide du type (post)humain à promouvoir, le fait est que les transhumanistes, quant à eux, occupent bel et bien cette place.
[18] Voir Bergson ([1907], 2001, 363-369).
[19] Voir notamment *Généalogie de la morale*, I, §1-4 ; II, §11-12 ; III, §26.
[20] *Humain, trop humain*, I, § 2.
[21] Comme en témoigne le §224 de *Humain, trop humain*.

la mesure où c'est en vertu d'une image figée du bonheur humain que les transhumanistes forgent l'image de la vie transhumaine (sans souffrance ni maladie, sinon sans finitude), c'est en somme et par anticipation *contre un pseudo-dépassement* évolutionniste de l'homme que Nietzsche élève l'idée du surhumain, suggérant que la « sur-espèce » humaine n'est pas une authentique pensée de l'altérité, mais une pseudomorphose. L'hybridation technicienne de l'homme modifie peut-être la nature physiologique de l'homme, mais elle ne change rien à l'essentiel, en maintenant la structure fondamentale de la morale réactive qui interprète toute souffrance comme un mal dont il faudrait se débarrasser – comme une existence coupable dont les transhumanistes rejettent la faute non plus sur le péché originel, mais sur la nature elle-même, comme en témoigne la « Lettre à dame nature » de Max More.

Finalement, la *Selbstüberwindung* (l'auto-dépassement nietzschéen) est l'exact opposé du *self-enhancement* ou du *self-improvement* transhumaniste, s'il est vrai du moins que ces expressions ne sont que les derniers surgeons de la lignée généalogique d'un évolutionnisme phtisique. De Spinoza à Darwin et Spencer, la démarche scientifique et philosophique demeure surdéterminée par une axiologie qui ne dit pas son nom, et qui donne le primat à la *persévération dans l'être* ou encore à *la conservation de soi*, qui indique symptomatiquement que *c'est encore l'identité du même qui s'affirme dans la recherche moderne du dépassement de soi* – comme en témoigne le narcissisme éhonté promu par les théories du développement personnel. Par où il faut comprendre que le « dépassement de soi », comme formule phare de la dogmatique aussi naïve que manipulatrice du coaching de vie contemporain, n'est en aucune façon le dépassement du moi et de ses intérêts, mais leur extension au moyen du combat contre ce qui est interprété comme ce qui le limite et l'empêche d'étendre ses intérêts et sa volonté de durer *en restant le même* quant à ses aspirations.

Dans la mesure où le surhomme, en tant qu'image, renvoie à une dynamique et à un devenir, l'idée nietzschéenne du *dépassement de soi* n'a absolument rien à voir avec l'idée d'*enhancement*, puisqu'elle *vise l'exacte antithèse d'une forme accomplie et définitive*, ainsi que nous aurons à y revenir : « Je considère nuisibles tous les hommes qui ne savent plus être les adversaires de ce qu'ils aiment : ils corrompent ainsi les meilleures choses et les meilleures personnes[22] ». Ces hommes, ce sont les derniers hommes en quête de simple survie, prêts qu'ils sont à sacrifier *zoè*, la vie qualifiée, distanciée du procès vital dont elle participe[23], sur l'autel de *bios*, la vie nue qui est celle de son expression statistique : « espérance de vie ».

A la lumière de ces remarques, il ne devrait plus paraître plus douteux que l'homme augmenté, ou le transhumain – quel que soit le sens qu'on lui donne du moment que l'on s'accorde sur les buts du projet qu'il nomme –, serait pour Nietzsche encore humain, bien trop humain – et un humain finalement bien régressif. Bien loin de surmonter l'homme en l'homme, le transhumanisme est l'exaspération de sa petitesse. Ce qui importe, au contraire, c'est de « Créer un être supérieur à ce que nous sommes, voilà *notre* être. *Créer par-delà nous-mêmes*[24] ! » En définitive, ce que veut Nietzsche, c'est dépasser l'humanité *morale*, et il a besoin à cette fin de dépasser la *forme de vie* (éthique et éthologique) de l'homme, c'est-à-dire sa *psycho-physiologie* (PBM, §23), au moyen d'une restructuration pulsionnelle[25].

Dès lors, lorsque nous entendons Zarathoustra proclamer que l'homme est un pont et non un but, il ne saurait être question de la même chose que lorsqu'un posthumaniste comme Esfandiary soutient que le transhumain est encore une figure de transition. Cet humain transitionnel est celui qui recourt à la technologie pour accéder au posthumain[26], ce qui n'est précisément pas le cas chez Nietzsche, dont on sait combien il méprise les solutions de facilité que se donne l'homme pour optimiser son utilité et son bien-être. C'est sans doute une des (nombreuses) raisons pour lesquelles Nietzsche mépriserait profondément le proactivisme technoscientifique de Max More et le libertarianisme transhumaniste en général[27].

Douleur et intelligence transhumanistes ; souffrance et pensée surhumaines

Précisément parce que le transhumanisme est une négation de la *chair*, il passe à côté du problème de l'esthétique physiologique qui est au cœur de la réflexion nietzschéenne (mais, par-delà Nietzsche, de notre humanité) : non pas conquérir un corps supérieur *techniquement*, mais un corps d'une *intensité de vie superlative*. Relativement à ce qu'ils *valorisent*, les idéaux respectifs de Nietzsche et des transhumanistes – on ne retiendra ici que le traitement qu'ils réservent à la *souffrance*, d'une part, et à l'*intelligence*, de l'autre, si nous avons raison de considérer qu'il s'agit des deux aspects principaux du messianisme technoscientifique – sont diamétralement et irréconciliablement opposés.

1/ La question de la souffrance, d'abord – rabattue sur celle de la douleur par le matérialisme réducteur des transhumanistes.

Nietzsche envisage la souffrance comme un élément incontournable de la création (voir par exemple le §318 du *Gai Savoir*). Dans cette perspective, le surhumain nietzschéen ne *supplante* pas l'humain, comme s'il s'agissait d'annihiler l'homme pour le remplacer par autre chose. C'est *au sein* de son corps *vécu* qu'est appelée à se faire la transition vers le surhomme, dans la mesure où les révolutions véritables se font *au-dedans* de nous, dans la chair souffrante et non dans le corps endolori. C'est la raison pour laquelle Nietzsche fait l'apologie de la souffrance, certainement pas par dolorisme ascétique, mais dans la mesure où la souffrance participe de l'économie ontologique qui garantit la cohérence éthique de l'*amor fati*. C'est une manière de renvoyer dos-à-dos le pessimiste et l'optimiste, car l'un comme l'autre se morfondent face à la souffrance, même si le premier en fait une fatalité, tandis que le second souhaite et croit à sa possible résorption[28]. Plutôt donc que d'abroger la

[22] *FP* novembre-février 1882-1883, 5 [1] 3.

[23] Voir sur ce point Byung-Chul Han (2015, p. 46 sq.), qui invite à penser la chrématistique et le capitalisme comme l'excès propre à la vie nue qui, fétichisant sa durée, fait de la médecine le thaumaturge, et de la mort la perte absolue, invitant ainsi à penser l'idolâtrie des *thérapies* comme la sécularisation sotériologique de la *théologie*.

[24] *FP* novembre-février 1882-1883, 5 [1] 204, c'est Nietzsche qui souligne.

[25] Voir par exemple les *Fragments posthumes*, hiver 1883-1884, 24 [16].

[26] Voir Goffi (2015).

[27] Voir encore, parmi d'autres textes, la critique nietzschéenne du libéralisme formulée dans le §38 des « Incursions d'un inactuel » dans le *Crépuscule des idoles*, où se trouve défendue une conception agonistique de la liberté diamétralement opposée à sa conception transhumaniste.

[28] Voir par ex. *FP* automne 1881, 13 [4].

souffrance, comme y rêvent les transhumanistes, il faut au contraire la rechercher pour la dominer, la rechercher comme un moyen d'intensifier sa propre puissance de contraindre l'adversité du réel au service de sa volonté de puissance :

« *La* volonté de souffrir *: il vous faut de temps à autre vivre dans le monde, vous les créateurs. Il faut que vous touchiez* presque *à l'anéantissement – pour ensuite* bénir *votre labyrinthe et votre égarement. Faute de quoi vous ne pourrez* créer, *mais seulement* dépérir[29]. »

« *Créer, c'est se délivrer de la souffrance. Mais la souffrance est nécessaire à ceux qui créent. Souffrir, c'est se transformer ; la mort est présente dans toute mise au monde*[30]. »

2/ Les prétentions de la « super-intellingence », ensuite.

Outre qu'il cherche à séduire en nous les pulsions bourgeoises de repli sur soi en prétendant abolir la négativité de l'existence pour faire de nous des hommes positifs, le transhumanisme vante les mérites d'une intelligence assistée qui viendrait compléter l'augmentation somatique de notre bonheur d'un perfectionnement intellectuel de nos compétences cognitives ainsi que d'un affûtage de nos performances sensorielles.

Redisons-le pour les besoins de la démonstration : là où le transhumain s'éparpille dans la *res extensa* envisagée comme *res technica*, comme le révèle symptomatiquement son appel incessant à la *ratio* calculatrice, informatique, numérique[31] et statistique[32], le surhumain est la forme de vie la plus intense, qui considère la raison suffisante comme une raison outrecuidante qui a oublié qu'elle n'était qu'un petit îlot d'un ensemble plus vaste, que Nietzsche appelle le Soi, c'est-à-dire la grande raison envisagée comme raison éprouvante et approbatrice, raison qualitative qui sécrète les évaluations à partir de préférences psychophysiologiques. Aussi Nietzsche n'est-il jamais plus opposé au transhumanisme que sur la question de l'intelligence, faculté pragmatique tout à fait secondaire et sans aucune valeur en elle-même, dans la mesure où elle est un moyen et non une fin. C'est un point sur lequel il a retenu la leçon de Schopenhauer : la véritable intelligence n'est pas computationnelle ou combinatoire, puissance de calcul et de traitement des informations ; c'est une activité créatrice, c'est-à-dire artistique et, finalement, *esthétique* :

« *Connaître est un désir et une soif : connaître est une création. L'amour du corps et du monde est la conséquence de la connaissance en tant qu'elle est une volonté. En tant qu'elle est créatrice, toute connaissance est une non-connaissance. Tout* percer à jour, *ce serait la mort, le dégoût, le mal. Il n'y a même aucune forme de connaissance qui ne soit d'abord création*[33] »

L'extension transhumaniste de nos sens n'est donc absolument pas une *amélioration* de notre connaissance, de ce point de vue, mais la négation de sa spécificité. Les lentilles de contact ou les microscopes augmentent la résolution de l'œil, mais elles ne le rendent pas meilleur, à moins de considérer que la valeur peut être rabattue sur la *performance* – ce qui est précisément ce que Nietzsche conteste : « notre œil a une vision fausse ; il raccourcit et resserre : est-ce là une raison pour rejeter la vue et dire qu'elle n'a aucune valeur[34] ? ».

La connaissance, en effet, est irréductiblement sélective : c'est une puissance de *hiérarchisation* qui doit laisser certains éléments au second plan, parce qu'il est tout à fait contreproductif de tout connaître. L'homme moderne, qui veut une société de transparence sans aucune part d'ombre pour sa volonté de savoir, est *saturé* d'informations, qui sont tout le contraire de la connaissance[35]. Cette dernière, en effet, fonde toujours un savoir qui se découpe sur le fond d'un non-savoir, ainsi que Nietzsche l'avait déjà montré dans la deuxième des *Considérations inactuelles* : le sens historique est une pulsion de connaissance hypertrophiée qui croit pouvoir faire de la connaissance une fin en soi, syndrome d'un *mal d'archives*, dont l'idéal encyclopédique d'une gravure générale de tout le savoir humain sur un *hardware* numérique n'est que la résultante contemporaine. L'intelligence artificielle a ceci d'obscène qu'elle ne sait pas dissimuler, ni oublier activement. Sa volonté de savoir est sans fin, là où la santé exige surtout de vouloir ne pas savoir[36].

L'humanité de l'avenir : la vie dans la durée ou la durée de la vie ?

Quelle image de l'avenir se dégage à partir de là, qui autorise à opposer frontalement le « surhumain » nietzschéen à la fois au transhumain et au posthumain ?

Le réagencement pulsionnel que Nietzsche exige quant à notre manière d'envisager la souffrance et la pensée a certes pour but de rehausser l'humanité (§287). Mais si l'on y regarde de près, c'est plus exactement certains de ses échantillons les plus admirables de créativité qui remplissent cet office, et s'ils le remplissent, c'est *à partir des forces propres à ce qui est encore humain en eux*, et certainement pas en substituant à l'humanité existante un Homme Nouveau. C'est en ce sens que, dans le *Gai Savoir*, Nietzsche appelle de ses vœux une « humanité de l'avenir » (*Le Gai Savoir*, §337) qui adviendrait à partir d'une connaissance historique de ses formes actuelles et passées.

Qu'est-ce alors que le surhumain, s'il ne participe pas d'une stratégie de planification volontariste qui entend intervenir dans la réalité humaine pour arraisonner l'homme lui-même au moyen de sa propre puissance technique ? C'est parce que l'humanité décline en vertu du mouvement de *Selbstüberwindung* propre à l'histoire du nihilisme européen qu'il est nécessaire d'assurer sa relève à ce moment précis de l'histoire, de manière à proposer, *parmi les voies possibles* de son futur, une alternative au dangereux faux-dilemme proposé par les hommes modernes : le triomphe du dernier homme ou celui du nihilisme.

Laissons de côté le nihilisme pour revenir sur cette figure du dernier homme, qui n'est plus une fiction anticipatrice aujourd'hui, mais un processus en cours d'actualisation. Ce que Nietzsche devine déjà à l'œuvre dans l'humanité européenne du XIXe siècle, c'est l'aspiration du dernier homme à devenir le *nec plus ultra*

[29] *FP* novembre 1882-février 1883, 5 [1] 225.
[30] Id., 226.
[31] Voir Stark (2016).
[32] Voir Rey (2016).
[33] FP novembre-février 1882-1883, 5 [1] 213..
[34] *FP* novembre 1882-février 1883, 4 [194].
[35] Voir par ex. Han (2015, p. 80-83 en particulier ; 2017, p. 69-76).
[36] Voir le *Crépuscule des idoles*, « Maximes », §5.

de l'évolution, jusqu'à déjouer l'obsolescence programmée à laquelle une marâtre nature l'a injustement voué. À l'idéal quantitatif qui, en vertu d'une idéologie statistique, nous fait expéditivement assimiler l'idée d'espérance de vie à celle de progrès, Nietzsche oppose une exigence qualitative d'un *approfondissement* de l'expérience vécue, car c'est cette expérience intensive, irréductible à toute calculabilité, qui fait l'épaisseur authentiquement temporelle de notre vie, dans la mesure où l'intensité d'une expérience *dilate le sentiment de la temporalité* jusqu'à élever certains instants exceptionnels à l'intemporalité, tant ils se survivent à eux-mêmes, comme immémorialement, au-dedans de nous mais au-delà de nous-mêmes, ainsi qu'en témoigne l'intuition fulgurante de l'éternel retour.

Ainsi, l'idéal transhumaniste d'une vie à la durée potentiellement indéfinie figure très clairement parmi les idées qui répugnent à Nietzsche, tout à fait disposé à troquer des centaines d'années de vie grégaire pour un instant créateur de la vie de Goethe[37]. Il paraît clair que l'allongement artificiel de la vie participe, dans une perspective nietzschéenne, de l'idéal de l'homme le plus méprisable, puisque Zarathoustra oppose au surhomme « le dernier homme [...] qui vit le plus longtemps[38] », et qui voudrait que tout le monde eût accès à cette éternité inerte[39]. Ainsi, loin d'ouvrir la temporalité humaine à la contingence de l'avenir, le dernier homme transhumain fermerait l'historicité humaine sur elle-même. Le surhumain n'est donc pas plus une figure eschatologique que téléologique, dans la mesure où Nietzsche défend une conception discontinuiste de l'histoire (*GM*, II, §12) qui empêche d'y voir l'ascension d'un quelconque processus monolithique, comme en témoigne un texte posthume capital pour notre propos :

> « *L'humanité n'a pas plus de but que n'en avaient les sauriens, mais elle a une évolution : c'est-à-dire que son terme n'a pas plus d'importance qu'un point quelconque de son parcours ! Par conséquent on ne saurait définir le bien en en faisant le moyen d'atteindre le « but de l'humanité ». Serait-ce ce qui prolongerait l'évolution le plus longtemps possible ? Ou ce qui la porterait à son point le plus haut ? Mais cela présupposerait derechef un critère pour mesurer ce point le plus haut ! Et pourquoi le plus longtemps possible ? Ou le minimum de déplaisir dans l'évolution ? C'est à cela qu'aujourd'hui tout aspire – mais cela signifie aussi l'évolution la moins puissante possible, un auto-affaiblissement général, un terne adieu à l'humanité antérieure*[40] »

Le portrait comparatif pourrait se poursuivre *ad libitum*, mais il devrait apparaître désormais, sur des fondements philologiques et philosophiques plausibles, que les transhumanistes sont bien, du point de vue nietzschéen, non seulement les héritiers, mais les plus draconiens metteurs en scène de la haine de la vie propre aux « contempteurs du corps » décrits par Zarathoustra. Des contempteurs d'autant plus dangereux qu'ils dissimulent leur haine de la vie sous le nom de l'Amour de la Vie éternelle *ici-bas* – selon une rhétorique qui les apparente à des Chrétiens manqués. Au moins, ces derniers avaient la décence de reconnaître que la « vraie » vie était ailleurs, dans quelque au-delà du monde. Les transhumanistes, eux, croient cette vraie vie accessible *au sein même du monde*, dans quelque *au-delà* du corps de chair, dans un corps désincarné, technicisé ou virtualisé dans sa chimérique numérisation. C'est cet au-delà séculier qui fait encore participer, qui fait même *à plus forte raison* participer l'idéal transhumaniste à l'ascétisme des arrière-mondes. Dernières en date de l'incessant défilé des ombres de Dieu dans les Cavernes (*GS*, §108), les ombres transhumaines et posthumaines[41] sont le pur produit de cette pulsion supraterrestre qui anime les corps décadents aspirant à se venger de la vie à l'âge de la dépression généralisée – *en croyant trouver le remède à leur dépression dans l'exaspération de sa cause majeure, le narcissisme de la conservation de soi.*

[37] Voir *Considérations inactuelles* II, 8 [§6], qui dit plus exactement : « j'échangerais bien des charretées entières de vies jeunes et ultramodernes contre quelques années de ce Goethe "épuisé", pour pouvoir encore participer à des entretiens comme ceux qu'il eut avec Eckermann, et me préserver ainsi des enseignements d'actualité dispensés par les légionnaires de l'instant présent. »

[38] *Ainsi parlait Zarathoustra*, prologue, §5, v. 97.

[39] C'est un point sur lequel Sorgner (2017, section 7) concède qu'il existe une divergence entre le transhumanisme et Nietzsche, mais c'est justement un point absolument crucial, qui indique que le fondement de l'anthropologie nietzschéenne est à l'antipode de l'anthropologie (utilitariste ou hédoniste) transhumaniste. En toute rigueur, du reste, le transhumanisme devrait *renoncer à toute anthropologie*, en tant qu'elle s'attache à postuler une nature humaine.

[40] FP automne 1880, 6 [*59*].

[41] Il appartiendrait à un examen de détail d'établir la connivence idéologique de ces deux versants du transhumanisme, que nous n'avons pas pu suffisamment distinguer ici.

RÉFÉRENCES

Nietzsche, Œuvres complètes, Paris, Gallimard, 1966- Kritische Studienausgabe [KSA], Berlin, de Gruyter, 1966-1977. The Complete Works of Friedrich Nietzsche, Stanford, Stanford University Press

H. Bergson (1907), L'Évolution créatrice, Paris, PUF, 2001

G. Canguilhem (1952), « Organisme et machine » in La Connaissance de la vie, Paris, Vrin

Jacquels Ellul (1954), La Technique ou l'enjeu du siècle, Paris, Economica, 1990

R Ettinger, (1972) Man into Superman

T. Garcia (2016, La Vie intense. Une obsession moderne, Paris, Autrement

K.-G. Giesen (2004), « Transhumanisme et génétique humaine », L'Observatoire de la génétique, 16 : https://iatranshumanisme.files.wordpress.com/2015/08/no-16.pdf)

Jean-Yves Goffi, (2015),« Aux origines contemporaines du transhumanisme » in Ethique, politique, religions, Paris, Classiques Garnier

(2017), « Transhumanisme » in M. Kristanek (dir.), l'Encyclopédie philosophique, URL: http://encyclo-philo.fr/transhumanisme-a/

Buyng-Chul Han (2012), Agonie des Eros, Berlin, Matthes Seitz

(2013), Im Schwarm, Berlin, Matthes Seitz, 2013, trad. fr. Dans la nuée. Réflexions sur le numérique, Paris, Actes Sud, 2015

(2013), Transparenzgesellschaft, Berlin, Matthes Seitz, 2013, trad. fr. La Société de transparence, Paris, PUF, 2017

Gilbert Hottois (2017), Philosophie et idéologies trans/posthumanistes, Paris, Vrin

Olivier Rey (2006), Quand le monde s'est fait nombre, Paris, Stock

Stefan Lorenz Sorgner (2017) « Beyond Humanism : Reflections on Trans- and Posthumanism » in Y. Tuncel, op. cit.

V. Stark (2016), Le Navigateur obsolète, Paris, Belles Lettres, 2016

H. Tirosh-Samuelson (2012), « Transhumanism as a Secularist Faith », Zygon, 47/4, p.710-734.

Y. Tuncel, (2017), « Introduction » in Y. Tuncel (dir.), Nietzsche and Transhumanism. Precursor or Enemy ?, Cambridge, Scholar Publishing

Recommendation Algorithms, a Neglected Opportunity for Public Health

REVUE MÉDECINE ET PHILOSOPHIE

Lê Nguyên Hoang[1], Louis Faucon[1] and El-Mahdi El-Mhamdi[2] *
[1] Ecole Polytechnique Fédérale de Lausanne, Switzerland, Emails: firstname.lastname@epfl.ch,
[2]École Polytechnique, France Email: elmahdi@elmhamdi.com

ABSTRACT

The public discussion on artificial intelligence for public health often revolves around future applications like drug discovery or personalized medicine. But already deployed artificial intelligence for content recommendation, especially on social networks, arguably plays a far greater role. After all, such algorithms are used on a daily basis by billions of users worldwide. In this paper, we argue that, left unchecked, this enormous influence of recommendation algorithms poses serious risks for public health, e.g., in terms of misinformation and mental health. But more importantly, we argue that this enormous influence also yields a fabulous opportunity to provide quality information and to encourage healthier habits at scale. We also discuss the philosophical, technical and socio-economical challenges to seize this immense opportunity, and sketch the outlines of potential solutions. In particular, we argue that it would be extremely helpful if public and private institutions could publicly take a stand, as this may then generate the necessary social, economical and political pressure to massively invest in the research, development and deployment of the potential solutions.

KEYWORDS : recommandation algorithms, public health, artificial intelligence.
DOI : 10.51328/103

Introduction

Artificial Intelligence (AI) promises major advances in medicine and public health, from advancing our knowledge of molecular biology (Senior et al., 2020; Gupta et al., 2020) to monitoring the progress of large-scale pandemics (Cavlo et al., 2020); from treatment development (Ong et al., 2020) to scalable diagnosis instruments (Rajpurkar et al., 2017). However, such developments also raise ethical concerns, especially in terms of software security, privacy and misuse (Fernández-Alemán et al., 2013). Moreover, one may argue that this line of work has been somewhat under-delivering, at least in contrast to the massive investments and hype that accompany the "AI and health" slogans (Shortliffe, 2019).

On the other hand, AI algorithms have been widely deployed on highly influential large scale platforms such as Facebook, YouTube or Twitter. The website Statista (Clement, 2020) reports that, in 2019, "the average daily social media usage of internet users worldwide amounted to 144 minutes per day". Moreover, what social media users are exposed to seems to be extremely dependent on the AI algorithms that the big tech companies use for content recommendation. YouTube Chief Product Officer[1] Neal Mohan reported that 70% of YouTube views result from algorithmic recommendation, as opposed to user's search, user's subscription feeds or external links (Solsman, 2018).

The algorithms designed to provide such recommendations are called *recommendation algorithms.* Recommendation algorithms typically survey the content published on their platforms and the activity of the platforms' users to organize users' news feeds, and to suggest new content, accounts and groups to consume, follow and join. Critically, such algorithms are *customized*. They provide tailored recommendations to different users, which

[1] As of January 2018, when the interview in the references was conducted.

makes them challenging to study, especially for external researchers (Aral, 2020).

In a widely debated experiment involving 689,003 Facebook users, Kramer et al. (2014) showed that a tiny modification of the Facebook newsfeed algorithm sufficed to slightly change users' behaviors within a single week. Namely, by simply removing 10% of negative posts on the Facebook newsfeed, their analysis revealed that users started posting more positive contents. Yet, Hohnhold et al. (2015) also showed that it usually takes weeks, if not months, to observe important user behavior changes on Google search after a modification of advertisement placements. These two studies, among others, suggest that large-scale human behaviors can be significantly modified by what social media algorithms choose to expose billions of their users to. Given the scale of the problem, Milano et al. (2020) argue that society at large should now be regarded as an important stakeholder of what social media algorithms recommend.

This evidently raises important ethical concerns, especially in terms of political manipulation and misinformation, as will be discussed in the next section of this paper. However, in most of this paper, we will mostly stress the fact that social media should also be regarded as an urgent opportunity to be seized by global health actors. Indeed, for many diseases, including obesity, COVID-19 and mental health, the information that patients are exposed to and the habits that they adopt are arguably some of the best available treatments. In fact, many health agencies have already massively invested, or are asked to massively invest, in public service announcements to mitigate these diseases (Nestle and Jacobson, 2000; COCONEL Group, 2020). Yet, such announcements have arguably failed to fully take advantage of the opportunities offered by social media.

In this paper, we argue that it is urgent that a lot more attention be paid to such opportunities, both by computer scientists and big tech companies, but also and equally importantly by philosophers, doctors and public health agencies. We believe that, to seize such opportunities, it is critical for all of these entities to recognize the importance of recommendation algorithms for global health, so that added social and legal pressures are put on social media companies. In fact, we argue that it would be extremely helpful if, for instance, such entities could publicly declare that making recommendation algorithms beneficial for public health has become a top healthcare priority.

We also present partial solutions to improve global health through recommendation algorithms, and call for further academic efforts to research, test, audit, analyze, question, correct, develop, secure, legislate, debate and deploy such solutions. Clearly, this is no easy task, but this is why massive efforts should be invested in researching solutions as soon as possible; and why advocating for the importance of recommendation algorithms seems extremely helpful.

The paper is organized as follows: In Section 2, we discuss the scale of the infodemic and why it is still arguably a neglected aspect of public health despite recent efforts. In Section 3, we develop an argument on how quality information can and should be used as a medical intervention. In Section 4, we review the impact large scale information systems could have on mental health. In Section 5, we present a series of easily implementable solutions. We conclude in Section 6.

Infodemic

A blend of *information* and epidemic, the term infodemic gained popularity during the ongoing COVID-19 pandemic[2] (Galloti et al., 2020) as misinformation campaigns around the disease gained momentum. The WHO, the UN, the UNICEF, the UNDP, the UNESCO, the UNAIDS, the ITU, the UN Global Pulse and the IFRC[3] (Joint Statement, 2020) started dedicating special efforts to battle the infodemic. In particular, WHO hosts a page dedicated to COVID-19 misinformation[4].

But the infodemic did not begin during the COVID-19 crisis. For instance, vaccine hesitancy gained enough ground that the WHO listed it (WHO, 2019) among its top ten public health crises as of January 2019 (one year before the COVID-19 outbreak). Johnson et al. (2020), after analysing discussion groups involving 100 million Facebook users, warns that their "theoretical framework reproduces the recent explosive growth in anti-vaccination views, and predicts that these views will dominate [the public opinion landscape] in a decade".

Misinformation also affects other areas of medicine, such as cancer (Loeb et al., 2019), alternative medicine (Collier, 2018) or nutrition (Myrick and Erlichman, 2020). Disturbingly, some influencers with misleading and dangerous health information are widely recommended by recommendation algorithms. For instance, despite having been reported hundreds of times back in 2016 for dangerous misinformation (Schepman, 2016; Olivier, 2020), and despite YouTube's claimed will to fight dangerous health misinformation, some YouTubers with misleading health-related content have gained over 500,000 subscribers, and accumulated millions of views in 2020 alone.

The COVID-19 pandemic has arguably made health misinformation even more problematic, especially as public health became all the more intertwined with political agendas (Biancovilli and Jurberg, 2020). Indeed, Bradshaw and Howard (2018) report that there are already important politically-motivated investments to bias public opinion. Concerningly, Vosoughi et al. (2018) provide evidence that some misinformation spreads much faster than some reliable information on social media. Unfortunately, such a phenomenon does not seem restricted to social networks with recommendation algorithms. More recently, the French COVID-19 conspiracy documentary *Hold Up* went viral on Vimeo, and through sharings of extracts from the documentary on all sorts of social medias. Its success allowed it to raise over 180,000 euros on the crowdfunding platform Ulule, and over 150,000 euros on the participative financing platform Tipeee. Machado et

[2] An interesting but not so surprising fact is that the term itself did not have its own wikipedia page before the Covid19 pandemic https://en.wikipedia.org/w/index.php?title=Infodemicaction=history, a page which is mostly a redirection to the one on the misinformation arroud the COVID-19 pandemic https://en.wikipedia.org/wiki/Misinformation-related-to-the-COVID-19-pandemic.

[3] "WHO" stands for "World Health Organization", "UN" for "United Nations", "UNICEF" for United Nations Children's Fund, "UNDP" for "United Nations Development Programme", "UNESCO" for "United Nations Educational, Scientific and Cultural Organization", "UNAIDS" for "United Nations Programme on HIV and AIDS", "ITU" for "International Telecommunication Union" and "IFRC" for "International Federation of Red Cross and Red Crescent Societies".

[4] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters.

al. (2019) report a large amount of junk news on public WhatsApp groups.

We stress the fact that misinformation need not be false information, or information from "fake news" sources (Grinberg et al., 2019). Factual evidence can deeply mislead, e.g., by telling the story of an individual who survived from a disease after they adopted some alternative medicine treatment, or of a child who sadly died a few months after they received some vaccine (Nisbett and Borgida, 1975). In fact, even statistical factual data can "lie", e.g., because of cherry-picking (Morse, 2010), misinterpretation (Kerr, 1998) or confounding variables (Simpson, 1951; Wagner, 1982). And while double-blind randomized controlled trials provide more robust and reliable signals, they too are malleable and can be hacked to provide misleading conclusions, as evidenced by the reproducibility crisis (Baker, 2016) and as argued in the case of drug testing by Stegenga (2018). This has led to harsh criticisms of today's dominant *null hypothesis statistical test* method (Amrhein et al., 2019), and a call for the research, development and use of more reliable statistical approaches and ways of phrasing research conclusions (Wasserstein et al., 2019).

Given the political motivations and financial incentives to spread some information rather than others (Kahan et al., 2013), and the cost of thorough fact checking, of sound reasoning, of exhaustive literature surveying and of querying multiple experts, in the absence of quality human or algorithmic content moderation, it seems that low quality information should be expected to dominate (Aral, 2020). This raises serious concerns for global health. It seems urgent to promote a lot more quality health information. Interestingly, recommendation algorithms could be a formidable asset to do so.

But reliable information about disease prevention and treatment may not be what is most urgent to recommend. Interestingly, the *Healthy People 2030* project by the *US Office of Disease Prevention and Health Promotion* added "attaining health literacy" as one of its foundational principles and overarching goals (ODPHP, 2017). They defined health literacy as "people's capacities to find, understand, and use health information and services for informed decisions and actions". But attaining such health literacy requires repeated exposures to quality pedagogical explanations of what a reliable information search entails. Unfortunately, such explanations are currently mostly drowned within a flood of junk news. The help of recommendation algorithms to dig out and promote such pedagogical contents seems essential.

Overall, instead of merely a threat, the predominance of recommendation algorithms could also be regarded as a great opportunity to drastically improve global health information; which could then drastically improve global health. Unfortunately, thus far, the enormous potential of recommendation algorithms to do good seems very neglected (Hoang, 2020a). Typically, the Netflix documentary *The Social Dilemma* depicts a very negative view on recommendation algorithms, and barely suggests that they could be a powerful asset to do a vast amount of good. It seems urgent to also underline the great public health opportunity offered by recommendation algorithms, if designed at least partly to improve global health.

Quality Information Saves Lives

Perhaps no story highlights the importance of quality science information better than Ignaz Semmelweis' failure to convince his colleagues of the importance of hygiene. In the 1840s, Semmelweis imposed a hand-washing policy in his clinic, before delivering babies. He then observed a drastic reduction of the childbed fever death rate of the new mothers. Semmelweis had discovered the staggering effectiveness of hygiene. Unfortunately, Semmelweis failed to communicate his findings effectively. Instead, he presented flawed explanations[5] (Tulodziecki, 2013). After years of rejections, Semmelweis became increasingly angry and even accused some of his colleagues of murder (Dykes, 2016). This led to a failure to standardize hygiene.

But producing quality information is merely the first necessary step. To exploit this information, it then needs to be effectively communicated. The COVID-19 pandemic arguably illustrates some failures to communicate quality information effectively. Indeed, before sufficiently compelling data allowed to conclude which COVID-19 vaccine should be widely approved and recommended (Zimmer, 2020), the best treatment available was arguably prevention through adequate behaviors. This includes hygiene, physical distancing and wearing masks, as well as the acceptance of more drastic measures such as tracing, isolating and lockdown. Arguably, there is a lot of room for improvement on this front, especially in Western countries.

The importance of quality communication has been long recognized for other health concerns, such as addictions, nutrition or lack of physical exercise, among others. Regulations forced tobacco and alcohol industries to include a warning against risks in their advertisements, while massive investments have been made to promote healthier diets[6]. In France, the slogans "eat five fruits and vegetables per day" or "antibiotics should not be [automatically self-prescribed]" have been memorized by millions of individuals.

Unfortunately, the effectiveness of such communications is unclear. In fact, Werle and Cuny (2012) designed a randomized controlled trial, whose results revealed a negative effect of the spots "eat five fruits and vegetables per day" on teenagers' food consumption. The authors suggest that broadcasting such spots after an advertisement for unhealthy hedonic food seems to have increased the acceptability of the hedonic food. More generally, we should not exclude the possibility that clear and factual messages backfire, especially if they aim to affect the audience's beliefs or behavior. Vogelsanger (2018) shares similar concerns in the context of "climate preaching", which may appear accusatory to climate denialists and reinforce their denial.

More generally, determining what makes a message effective is arguably a very challenging research endeavor. Garcia-Retamero and Cokely (2017) found notable differences between the effectiveness of infographics, and

[5] In particular, Tulodziecki (2013) argues that Semmelweise put too much emphasis on cadaveric material being the only cause of childbed fever. Yet this claim was inconsistent with childbed fever in hospitals and with the seasonality of childbed fever, which is why Semmelweis' views got rejected by his colleagues.

[6] While anecdotal, the video below shows Amazon's Alexa algorithm recommending fast food twice to a hungry user. Billions of users of Alexa, OK Google or Siri may be nudged towards unhealthy or healthy food, because of the way such algorithms are designed. https://twitter.com/so_sroy/status/1325392314739662850

showed notable differences. Crucially, when such messages are spread at scale, even a 1% difference in, say, user engagement, ends up having a huge impact.

What makes this line of research all the more challenging is that the reception of a message may strongly depend on the recipient's world view. Concerningly, Kahan et al. (2013) showed that politically motivated reasoning could make individuals diverge in their analysis of univocal but tricky numerical data. This means that a purely factual piece of information can be misleading for a subpopulation. Curiously, even the individuals who are more scientifically educated, often failed to correctly analyze the data if the data contradicted their intuitions; in fact, in such a case, educated individuals then performed just as poorly as uneducated individuals. Intriguingly, however, Kahan et al. (2017) later found out that, as opposed to intelligence and data, scientific curiosity seems to successfully make individuals with diverging political identities converge on factual considerations. More empirical data on the effectiveness of different messaging to different audiences seems critical.

Unfortunately, collecting data on the effectiveness of healthcare messages, especially in their actual context of diffusion, is extremely hard. But interestingly, social media platforms seem to be in a particularly fitting position to do so. Indeed, such platforms constantly collect massive amounts of data about what contents users are exposed to, how much time they spend watching such contents, and what the users do after being exposed to the contents. To research the effectiveness of different health messages, it seems critical that health agencies work with such platform providers. If done correctly, major progress may be achieved in domains like obesity, pandemic prevention or vaccination, especially if techniques like multi-armed bandit optimization are used to optimize the search for the most effective communication contents (Berry et al., 2010).

Results from social psychology seem important to integrate too. Typically, self-affirmation theory (Badea and Sherman, 2019) has consistently shown that individuals were more likely to accept contradictory views if they first affirm their values or successes that are not questioned by the contradictory views. Interestingly, for instance, Shermann et al. (2000) showed that subjects were significantly more receptive to articles on AIDS risks, and more willing to buy condoms, if they first underwent such a self-affirmation exercise. The self-affirmation exercise consisted of writing an essay describing why the subject's most important value is so important to them, and a time when it played a particularly important role. Meanwhile, subjects in the no-affirmation group had to do so for a value of little importance to them. Remarkably, the fraction of subjects who purchased condoms went from around 25% for the no-affirmation group to around 50% for the self-affirmation group.

Interestingly, experiments run by Werle et al. (2011) also give evidence of the importance of customizing the information to be communicated. In particular, the experiments showed that, when targeting teenagers, highlighting the social risks of obesity seems more effective. To make the experiment realistic, the authors tested the effect of a repeated exposure to prevention messages in a brochure containing diverse unrelated topics. They then asked subjects to fill forms, and to choose a thank-you snack. They found out that 65% of the subjects exposed to the social argument chose the healthy snack, as opposed to 55% for subjects exposed to the health argument. This strongly suggests that an effective health campaign must deliver several arguments and must customize the argument to deliver to the target audience. As it turns out, recommendation algorithms are precisely designed and optimized to perform such a customization.

Perhaps most importantly, recommendation algorithms can promote important health messages at scale, by reaching billions of individuals. Moreover, previous randomized controlled experiments on voting turnouts (Bond et al., 2012), positive messaging (Kramer et al., 2014) and ad blindness (Hohnhold et al., 2015) have already highlighted the effectiveness of algorithms at affecting users' feelings, beliefs and behaviors. Right now, this immense power is not used for good, and arguably has very undesirable consequences. If recommendation algorithms were designed for good, their enormous impact could improve the health of millions of individuals. Realizing this may change our discourse on recommendation algorithms, and what we demand from them, formally or informally. This seems critical to seize the opportunities presented by recommendation algorithms.

Mental Health

In this section, we propose to focus on the particular challenge of mental health, partly because it has been recognized as a growing concern (American Heart Association, 2019), and partly because, as we will see, the impact of social networks on mental health has gained a lot of attention lately.

In fact, on May 14, 2020, the World Health Organization argued that "substantial investment [is] needed to avert mental health crisis". Depression and anxiety were increasing, while many mental health services were interrupted because of the COVID-19 pandemic. What is more, the sudden isolation imposed by the lockdowns and physical distancing measures was increasing pre-existing concerns about the negative impacts of technology abuse on mental health — a trend that was underlined by the increasing popularity of the term *doomscrolling* (Watercutter, 2020).

In fact, according to Heron (2017), suicide is a leading cause of death among young people of age 15 to 34 in the United States, second only to unintended injuries. Worldwide, suicides add up to nearly one million deaths per year. It is noteworthy that the exposure to suicidal stories seem to increase suicide risks (Yıldız et al., 2018; Chan et al., 2018; Swedo et al., 2020), albeit stories about suicidal ideation without suicidal behavior may actually decrease suicide risks (Niederkrotenthaler, 2010). The former case is called the Werther effect, while the latter is known as the Papageno effect (Scherr and Steinleitner, 2015).

In any case, Carlyle et al. (2017) point out that content with the hashtags suicide and suicidal trigger more engagements than others. This suggests that (1) the Instagram recommendation algorithm might favor such contents and (2) users may be incentivized, consciously or not, to post such contents. This led the authors to conclude that "public health and mental health professionals should consider increased involvement on this platform". In particular, more research seems needed to better distin-

guish the contents that will likely increase suicide risks from those that will not, and to design algorithms that will promote the latter over the former.

But such dramatic mental health issues are not the only concerns to be had. Ironically, social media seem to also create loneliness and depression (Hunt et al., 2018), which increase risks for other diseases such as cancer or Alzheimer (Cacioppo and Cacioppo, 2014). Anxiety can also accompany the recurrent exposure to bad news, while anger may result from repeated exposure to aggressive opinions. In fact, according to Hubert (2014), in France, 18% of middle school students declare themselves victims of some cyber aggression. Jackson (2019) even argues that the negative bias of classical and social media may be causing *learned helplessness*, that is, users may feel that many challenges are beyond hope, which may hinder their willingness to do good. Finally, our digital experience may be causing a drastic reduction of our attention span, as suggested by measures made by González and Mark (2004) and Mark et al. (2016) on workers' average duration of online screen focus.

Interestingly, social media could be an important part of the solution to reduce the negative effects of technology abuse. As a starter, Eichstaedt et al. (2018) showed promising results for depression diagnosis, which is well known to be hard without social media data if patients do not themselves decide to consult doctors. In particular, while the use of social media seems overall beneficial to users' mental health, Holmgren and Coyne (2018) link abusive scrolling to relational aggression and depression.

But social media could be doing a lot more. They could arguably help users' mental health by promoting therapeutic contents, or by suggesting users to consult a psychiatrist. They could help users nurture their curiosity and their happiness, by favoring enthusiastic contents every now and then.

Evidently, any attempt to do so will be filled with potential pitfalls. It thus seems critical that such attempts result from a close collaboration between social media companies, medical experts and health organizations (Ginsberg and Burke, 2017), as was done for instance in Litt et al. (2020) and Ernala et al. (2020). In fact, this is only one of many ethical, technological and socio-economical challenges (Hoang and El Mhamdi, 2019) that need to be faced to seize the fabulous sanitary opportunities provided by social media. But such opportunities seem large enough to justify massive investments to research, develop and deploy potential solutions to meet these challenges.

Challenges and Potential Solutions

Given the limits of today's algorithms, to combat misinformation and promote quality information, it seems critical to rely on the judgment of experts. Several systems have been proposed, notably for fact-checking (Shamlo, 2018) or ethical decision making (Lee et al., 2019). These proposed solutions reveal several challenges.

First, there needs to be a mechanism for assessing the quality of experts to know how much their judgments can be trusted. Second, interfaces should be designed to collect quality expert judgments effortlessly. Third, potentially conflicting judgments from multiple experts should be aggregated into a unique recommendation decision. And finally, such solutions need to be actually implemented by the large social media companies. We discuss below these challenges in further details.

Identifying experts

In practice, expertise is most often certified by the degrees the experts obtained. Frustratingly, few universities enable third party websites to automatically certify the fact that a given expert obtained a given degree from them (Federal Trade Commission, 2005). Another common proxy to assess an expert's expertise is to check their publication list. It is noteworthy, however, that some platforms like Google Scholar do not allow for easy scraping of their data[7] (Else, 2018), which arguably hinders the ease to automatically verify experts' expertise. We acknowledge, however, that such proxies can be misleading (Waltman and Van Eck, 2012).

Besides, no single expert should be considered perfectly reliable, especially when they are discussing topics outside their domain of expertise. As an example, physics Nobel laureate Ivar Giaever opposes the scientific consensus on anthropogenic climate change. More generally, it seems desirable to aggregate the views of a large number of experts, rather than to take the view of a single expert for granted. One simple solution to do so would be to accept any individual with an email address from a trusted institution to register as an expert. However, to which extent should a given individual be regarded as an expert, especially on transdisciplinary questions with a moral dimension, is a question on which an agreement seems challenging to reach.

Designing an adequate interface

The role of the interface through which expert judgments are queried is arguably a very neglected research direction. After all, experts are typically busy people; it is often difficult to obtain enough of their attention to collect inputs from them. An appealing interface, which is effortless to use and which asks informative and yet easy-to-answer questions, seems critical to get the most out of experts.

One interesting proposal by Noothigattu et al. (2017) and by Lee et al. (2019) is to rely on comparison-based judgments. In this framework, the expert is repeatedly asked to choose one of two options. Social comparison theory (Festinger, 1954) argues that this better fits our natural judgment process. This can also avoid boundary effects, e.g. when users tend to give a maximal rating to too many items. Interestingly, the Bradley and Terry (1952) model allows to infer scores from such comparison-based judgments. In this model similar to the ELO system used to rank chess players, if an option is systematically preferred to another option, then the reconstructed scores of the former will be significantly larger than the score of the other option.

Aggregating potentially conflicting judgments

Unfortunately, we should expect experts to disagree on many topics, as evidenced by the lack of consensus in, say, moral philosophy. One solution to nevertheless reach a collective decision is to aggregate individual judgments

[7] It is unclear to which extent scraping is allowed by Google's Term of Service: "Google reserves the right to suspend or terminate your access to the services or delete your Google Account if [...] we reasonably believe that your conduct causes harm or liability to a user, third party, or Google — for example, by [...] scraping content that doesn't belong to you".

from multiple experts through some voting mechanism. Such voting mechanisms are the object of study of computational social choice theory (Brandt et al., 2016).

In particular, this field has highlighted the importance of properties such as strategy-proofness, which demands that honesty be an optimal strategy. This seems necessary to incentivize experts to provide high quality judgments. Interestingly, such aggregation of different experts' opinions also allows the system to be robust, because trust wrongly placed in a particular expert would be compensated by judgments from other experts. Some solutions such as *majority judgment* (Balinski and Laraki, 2011) have, to some extent, such properties.

Socio-economical challenges

Recently, the Tournesol framework has been proposed by Hoang (2020b) and aims to combine all the partial solutions discussed above. But even if a platform like Tournesol successfully identifies quality contents to recommend, this identification will have little impact if it is not used by actual large-scale recommendation algorithms.

It is noteworthy, as well, that any intervention must anticipate the risks of an exodus of social platform users to other less moderated platforms, such as *4chan*, *Parler* (Culliford and Paul, 2019) or *Bitchute* (Trujillo et al., 2020). To achieve this, the reliability of the content should not be the only feature that should matter. It seems critical as well that the content be engaging. More generally, it seems important that social media platforms strike a happy balance between entertaining users and delivering reliable information.

While implementing such ideas may conflict with their short-term priorities, interestingly, the social media leaders have publicly claimed their increased desire to make their platforms more beneficial to mankind (Wojcicki, 2020; Zuckerberg, 2018), which led to measurable improvements on their platforms, e.g., in terms of conspiracy theory recommendations (Faddoul et al., 2020) or added snippet links to Wikipedia or WHO. This is probably helped by increased social pressure and regulation threats.

However, to achieve more, additional support from health organizations seems greatly desirable. If they publicly declare that making recommendation algorithms robustly beneficial is a top global health priority, then we may expect an important increase of public and private investments in this research direction. Assuming that a compelling technical solution is then proposed, that social and legal pressures are large enough, there might then be a reasonable hope that such a solution will indeed be implemented by these social media companies.

Conclusion

In this paper, we argued for the importance of recommendation algorithms in improving public health. We discussed how both misinformation and social network addiction are public health challenges raised by these large scale systems, which may be aggravated by recommendation algorithms. But it is noteworthy that the absence of a recommendation algorithm, such as on direct messaging applications, still seems to expose us to such risks. Instead, we argued that recommendation algorithms should be regarded as an opportunity to have a large positive impact on public health. In particular, by better identifying which contents should be promoted at scale, provided that we could also convince or force large social media companies to promote such contents, we can vastly increase the reach of quality information. We identified several aspects of potential solutions, as well as numerous philosophical, technical and social challenges.

Our hope is that more computer scientists, technologists, companies, but also philosophers, medical doctors and health organizations, will research and promote such solutions as well, and that progress will be made to solve such challenges. In particular, we hope to have convinced our readers that it would be very valuable for the future of public health, if public and private institutions could publicly declare that recommendation algorithms have become a major risk and a massive opportunity for global health and treat these algorithms as such.

Acknowledgement

Authors would like to thank Konrad Seifert, Siméon Campos, Christophe Michel, Felix Grimberg, Mohamed Oubenal, Najib El Mokhtari, Stéphane Debove, Naomi Nederlof, Adam Shamir and Juliette Ferry-Danini for their useful feedbacks, which greatly improved the paper.

RÉFÉRENCES

Aral, S. (2020). The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health–and How We Must Adapt. Currency.

American Heart Association. (2019). Mental Health: A Workforce Crisis.

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance.

Badea, C., & Sherman, D. K. (2019). Self-affirmation and prejudice reduction: When and why?. Current Directions in Psychological Science, 28(1), 40-46.

Baker, M. (2016). Reproducibility crisis. Nature, 533(26), 353-66.

Balinski, M., & Laraki, R. (2011). Majority judgment: measuring, ranking, and electing. MIT press.

Berry, S. M., Carlin, B. P., Lee, J. J., & Muller, P. (2010). Bayesian adaptive methods for clinical trials. CRC press.

Biancovilli, P. & Jurberg, C. (2020). When governments spread lies, the fight is against two viruses: A study on the novel coronavirus pandemic in Brazil. Medrxiv.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415), 295-298.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. Biometrika, 39(3/4), 324-345. Bradsh

aw, S., & Howard, P. N. (2018). Challenging truth and trust: A global inventory of organized social media manipulation. The Computational Propaganda Project, 1.

Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (Eds.). (2016). Handbook of computational social choice. Cambridge University Press.

Cacioppo, J. T., & Cacioppo, S. (2014). Social relationships and health: The toxic effects of perceived social isolation. Social and personality psychology compass, 8(2), 58-72.

Calvo, R. A., Deterding, S., & Ryan, R. M. (2020). Health surveillance during covid-19 pandemic.

Carlyle, K. E., Guidry, J. P., Williams, K., Tabaac, A., & Perrin, P. B. (2018). Suicide conversations on Instagram™: contagion or caring?. Journal of Communication in Healthcare, 11(1), 12-18.

Chan, S., Denny, S., Fleming, T., Fortune, S., Peiris-John, R., & Dyson, B. (2018). Exposure to suicide behaviour and individual risk of self-harm: Findings from a nationally representative New Zealand high school survey. Australian & New Zealand Journal of Psychiatry, 52(4), 349-356.

Clement J. (2020). Daily time spent on social networking by internet users worldwide from 2012 to 2019. https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide

COCONEL Group. (2020). A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. The Lancet. Infectious diseases, 20(7), 769.

Collier, R. (2018). Containing health myths in the age of viral misinformation.

Culliford E. and Paul, K. (2019). Unhappy with Twitter, thousands of Saudis join pro-Trump social network Parler. Reuter.

Dykes, B. (2016). A History Lesson On The Dangers Of Letting Data Speak For Itself. Forbes.

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., ... & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. Proceedings of the National Academy of Sciences, 115(44), 11203-11208.

Else, H. (2018). How I scraped data from Google Scholar. Nature News QA.

Ernala, S. K., Burke, M., Leavitt, A., & Ellison, N. B. (2020, April). How well do people report time spent on Facebook? An evaluation of established survey questions with recommendations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-14).

Faddoul, M., Chaslot, G., & Farid, H. (2020). A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos. arXiv preprint arXiv:2003.03318.

Federal Trade Commission (2005). Avoid Fake-Degree Burns By Researching Academic Credentials.

Festinger, L. (1954). A theory of social comparison processes. Human relations, 7(2), 117-140.

Fernández-Alemán, J. L., Señor, I. C., Lozoya, P. Á. O., & Toval, A. (2013). Security and privacy in electronic health records: A systematic literature review. Journal of biomedical informatics, 46(3), 541-562.

Galloti, R., Valle, F. Castaldo, N., Sacco, P., & De Domenico, M. (2020). Assessing the risks of "infodemics" in response to COVID-19 epidemics. Nature Human Behaviour.

Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: a systematic review of health research and evidence-based design heuristics. Human factors, 59(4), 582-627.

Ginsberg, D. and Burke, M. (2017). Hard Questions: Is Spending Time on Social Media Bad for Us? Facebook Hard Questions.

González, V. M., & Mark, G. (2004, April). "Constant, constant, multi-tasking craziness" managing multiple working spheres. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 113-120).

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. Science, 363(6425), 374-378.

Gupta, H., McCann, M. T., Donati, L., & Unser, M. (2020). CryoGAN: A New Reconstruction Paradigm for Single-particle Cryo-EM Via Deep Adversarial Learning. BioRxiv.

Heron, M. P. (2017). Deaths: leading causes for 2015.

Hoang, L. N. (2020a). Science communication desperately needs more aligned recommendation algorithms. Frontiers Communication, Science and Environmental Communication.

Hoang, L. N. (2020b). Tournesol: Collaborative content recommendations. https://bit.ly/tournesol-app

Hoang, L. N. & El Mhamdi, E. M. (2019). Le fabuleux chantier : Rendre l'intelligence artificielle robustement bénéfique. EDP Sciences.

Hohnhold, H., O'Brien, D., & Tang, D. (2015, August). Focusing on the Long-term: It's Good for Users and Business. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1849-1858).

Holmgren, H. G., & Coyne, S. M. (2017). Can't stop scrolling!: pathological use of social networking sites in emerging adulthood. Addiction Research Theory, 25(5), 375-382.

Hubert, T. (2014). Un collégien sur cinq concerné par la cyber-violence.

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. Journal of Social and Clinical Psychology, 37(10), 751-768.

Jackson, J. (2019). You Are What You Read: Why changing your media diet can change the world. Unbound Publishing.

Johnson, N. F., Velásquez, N., Restrepo, N. J., Leahy, R., Gabriel, N., El Oud, S., ... & Lupu, Y. (2020). The online competition between pro-and anti-vaccination views. Nature, 1-4. (https://www.nature.com/articles/s41586-020-2281-1.pdf)

Joint statement by WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse, and IFRC, 2020, Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. url: https://www.who.int/news-room/detail/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation

Kahan, D. M., Peters, E., Dawson, E., & Slovic, P. (2013). Motivated numeracy and enlightened self-government. Behavioural Public Policy, 1, 54-86.

Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Hall Jamieson, K. (2017). Science curiosity and political information processing. Political Psychology, 38, 179-199.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. Personality and social psychology review, 2(3), 196-217.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788-8790.

Lee, M. K., Kusbit, D., Kahng, A., Kim, J. T., Yuan, X., Chan, A., ... & Procaccia, A. D. (2019). WeBuildAI: Participatory framework for algorithmic governance. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-35.

Litt, E., Zhao, S., Kraut, R., & Burke, M. (2020). What Are Meaningful Social Interactions in Today's Media Landscape? A Cross-Cultural Survey. Social Media+ Society, 6(3), 2056305120942888.

Loeb, S., Sengupta, S., Butaney, M., Macaluso Jr, J. N., Czarniecki, S. W., Robbins, R., ... & Langford, A. (2019). Dissemination of misinformative and biased information about prostate cancer on YouTube. European urology, 75(4), 564-567.

Machado, C., Kira, B., Narayanan, V., Kollanyi, B., & Howard, P. (2019, May). A Study of Misinformation in WhatsApp groups with a focus on the Brazilian Presidential Elections. In Companion proceedings of the 2019 World Wide Web conference (pp. 1013-1019).

Mark, G., Iqbal, S. T., Czerwinski, M., Johns, P., & Sano, A. (2016, May). Neurotics can't focus: An in situ study of online multitasking in the workplace. In Proceedings of the 2016 CHI conference on human factors in computing systems (pp. 1739-1744).

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. AI & SOCIETY, 1-11.

Morse, J. M. (2010). "Cherry picking": Writing from thin data, 20(1), 3.

Myrick, J. G., & Erlichman, S. (2020). How audience involvement and social norms foster vulnerability to celebrity-based dietary misinformation. Psychology of Popular Media, 9(3), 367.

Nestle, M., & Jacobson, M. F. (2000). Halting the obesity epidemic: a public health policy approach. Public health reports, 115(1), 12.

Niederkrotenthaler, T., Voracek, M., Herberth, A., Till, B., Strauss, M., Etzersdorfer, E., ... & Sonneck, G. (2010). Role of media reports in completed and prevented suicide: Werther v. Papageno effects. The British Journal of Psychiatry, 197(3), 234-243.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. Journal of Personality and Social Psychology, 32(5), 932.

Noothigattu, R., Gaikwad, S., Awad, E., Dsouza, S., Rahwan, I., Ravikumar, P., & Procaccia, A. (2018). A voting-based system for ethical decision making. In Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).

ODPHP (2017). Secretary's Advisory Committee on National Health Promotion and Disease Prevention Objectives for 2030: Recommendations for an Approach to Healthy.

Olivier, J. (2020). Le youtubeur Thierry Casasnovas (Regenere) dans le viseur de la justice pour "mise en danger de la vie d'autrui". Télé-loisir.

Ong, E., Wong, M. U., Huffman, A., & He, Y. (2020). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. BioRxiv.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Lungren, M. P. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.

Schepman, T. (2016). Thierry Casasnovas, le gourou du « tout cru », vous attend tranquille sur YouTube. L'Obs.

Scherr, S., & Steinleitner, A. (2015). Between Werther and Papageno effects. Der Nervenarzt, 86(5), 557-565.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... & Penedones, H. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706-710.

Shamlo, N. B., & Alavi S. (2018) AVOW: Expert-Sourced Fact Checking and Statement Evaluation.

Sherman, D. A., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. Personality and Social Psychology Bulletin, 26(9), 1046-1058.

Shortliffe, E. H. (2019). Artificial intelligence in medicine: weighing the accomplishments, hype, and promise. Yearbook of medical informatics, 28(1), 257.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society: Series B (Methodological), 13(2), 238-241.

Solsman, J.E. (2018). YouTube's AI is the puppet master over most of what you watch. CNET.

Stegenga, J. (2018). Medical nihilism. Oxford University Press.

Swedo, E. A., Beauregard, J. L., de Fijter, S., Werhan, L., Norris, K., Montgomery, M. P., ... & Sumner, S. A. (2020). Associations between social media and suicidal behaviors during a youth suicide cluster in Ohio. Journal of Adolescent Health.

Trujillo, M., Gruppi, M., Buntain, C., & Horne, B. D. (2020). What is BitChute? In Proceedings of the 31st ACM Conference on Hypertext and Social Media (pp. 139-140).

Tulodziecki, D. (2013). Shattering the myth of Semmelweis. Philosophy of Science, 80(5), 1065-1075.

Vogelsanger, P. (2018). Preaching, accusation, guilt and denial: Learning from Ignaz Semmelweis for climate communication. Klimaatelier.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151.

Wagner, C. H. (1982). Simpson's paradox in real life. The American Statistician, 36(1), 46-48.

Waltman, L., & Van Eck, N. J. (2012). The inconsistency of the h-index. Journal of the American Society for Information Science and Technology, 63(2), 406-415.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p< 0.05".

Watercutter, A. (2020). Doomscrolling Is Slowly Eroding Your Mental Health. Wired.

Werle, C., Boesen-Mariani, S., Gavard-Perret, M. L., & Berthaud, S. (2011). Social risk efficacy in preventing youth obesity. ACR North American Advances.

Werle, C. O., & Cuny, C. (2012). The boomerang effect of mandatory sanitary messages to prevent obesity. Marketing Letters, 23(3), 883-891.

Wojcicki, S. (2020). My mid-year update to the youtube community. YouTube Official.

World Health Organization (WHO). (2019). Ten threats to global health in 2019. https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019

World Health Organization (WHO). (2020). Substantial investment needed to avert mental health crisis. https://www.who.int/news/item/14-05-2020-substantial-investment-needed-to-avert-mental-health-crisis

Yıldız, M., Orak, U., Walker, M. H., & Solakoglu, O. (2018). Suicide contagion, gender, and suicide attempts among adolescents. Death studies.

Zimmer, C. (2020). First, a Vaccine Approval. Then 'Chaos and Confusion'. New York Times.

Zuckerberg, M. (2018). Facebook Post on interference and misinformation in elections. https://www.facebook.com/zuck/posts/10104797374385071

L'intelligence artificielle (IA), promesses et inquiétudes : une médecine anonyme est ce qu'il y a de pire

REVUE MÉDECINE ET PHILOSOPHIE

Serge Tisseron*

*Psychiatre, docteur en psychologie HDR, membre de l'Académie des technologies, membre du Conseil scientifique du CRPMS (Université de Paris), Président de l'Institut pour l'Etude des Relations Homme-Robots (IERHR) www.sergetisseron.com

RÉSUMÉ

L'intelligence artificielle appliquée à la médecine offre des opportunités considérables. De nombreux domaines sont concernés : la formation des intervenants, l'aide au diagnostic, la mise au point de nouveaux médicaments, la création de nouvelles formes de thérapies, notamment basées sur la réalité virtuelle, et la possibilité d'offrir à ceux et celles qui habitent dans des déserts médicaux une prise en charge rapide et efficace. Mais en même temps, plusieurs études montrent qu'à vouloir trop recourir à l'intelligence artificielle, la médecine peut être rapidement menacée de déshumanisation. Il serait catastrophique que sous prétexte d'efficacité, et de pallier les problèmes d'urgence posés par la pandémie du Covid 19, des pratiques nouvelles se mettent en place sans questionnement éthique suffisant, avec le risque qu'elles s'installent durablement après la crise.

Autrement dit, il est essentiel qu'une charte éthique claire précise les usages de ces technologies. Mais parallèlement, il est tout aussi indispensable de veiller à ce que la relation entre thérapeutes et patients ne s'appauvrisse pas sous l'effet de la sophistication technologique. Les patients sont des personnes, et les médecins aussi. Ils ont un nom. L'IA, elle, n'en n'a pas. Une médecine anonyme est ce qu'il y a de pire.

MOTS-CLÉS : médecine, technologie, éthique.

DOI : 10.51328/104

Introduction

La création de l'intelligence artificielle correspond dès ses origines à deux projets distincts, ou si on préfère à deux désirs. John McCarthy, en 1956, à la conférence de Dartmouth, la présente comme un résolveur de problèmes universel capable à terme de reproduire la polyvalence de l'intelligence humaine, d'où le choix du terme qu'il impose. Mais parallèlement, les travaux de Turing ont donné à l'IA une autre orientation : la création d'une machine capable de se faire passer pour un être humain. Dans les années 1960, ces deux approches sont complétées par une distinction entre IA « faible », également appelée « étroite », et IA « forte », également appelée « générale ». La première n'a aucune conscience d'elle-même, mais peut devenir plus performante que l'homme, comme le montre une simple calculette capable de réaliser en quelques secondes des opérations mathématiques hors des compétences humaines. En revanche, la seconde serait consciente d'elle-même, exactement comme l'intelligence humaine et aurait l'équivalent des « sentiments » humains. L'IA forte est aujourd'hui présentée comme très hypothétique, et probablement pas souhaitable…

En médecine, l'IA dont les compétences sont recherchées est évidemment une IA étroite appliquée à des problèmes spécifiques, et de nombreux domaines sont concernés : la formation des intervenants, l'aide au diagnostic, la mise au point de nouveaux médicaments,

la création de nouvelles formes de thérapies, notamment basées sur la réalité virtuelle, et la possibilité d'offrir à ceux et celles qui habitent dans des déserts médicaux une prise en charge rapide et efficace. Mais en même temps, plusieurs études montrent qu'à vouloir trop recourir à l'intelligence artificielle, la médecine peut être rapidement menacée de déshumanisation.

A. Les promesses de l'intelligence artificielle

L'aide au diagnostic

L'IA apporte déjà depuis quelques années de grands services dans les examens radiologiques, et la pandémie de COVID 19 a vu apparaître le bot[1] « Coronavirus Self-Checker » de Microsoft qui analyse les symptômes et recommande d'éventuels examens complémentaires.

Pour nous en tenir au domaine psychiatrique, un coach virtuel appelé Sim Sensei[2] est utilisé par l'armée américaine pour repérer les signes d'angoisse et les risques suicidaires chez les soldats, notamment ceux qui souffrent d'un PTSD. La machine aurait obtenu de meilleurs résultats diagnostics que les psychiatres et psychologues pris comme référence. Il semblerait que ce soit l'absence d'interférences humaines, c'est-à-dire de ressentis réciproques, qui fonderait la supériorité de cet avatar sur un psy humain pour recueillir les confidences de ces patients, souvent confrontés à des situations à forte composantes de honte et de culpabilité. Tout d'abord, derrière l'écran de son ordinateur ou de son téléphone, la peur d'être jugé s'estomperait. Il serait plus facile de parler à un algorithme. Par ailleurs, l'avatar n'a jamais de mimiques d'étonnement, et encore moins de réprobation. Il ne court donc pas le risque de renvoyer à son insu au patient la culpabilité ou la honte. C'est dans l'absence d'interférences humaines, c'est-à-dire de projections du thérapeute sur le patient et du patient sur le thérapeute, que résiderait la plus grande efficacité du robot. Mais un tel échange est-il bien vierge de toute projection ? Si celles qui concernent la gêne ou la honte à aborder certains sujets s'estompent, il semble bien que d'autres apparaissent, non moins problématiques…

La VR, quant à elle, est utilisée pour étudier les diverses formes de mémoire (Plancher et al., 2012) et l'état des fonctions exécutives, notamment chez les personnes atteintes de déficiences cognitives (Klinger, 2014) ou de schizophrénie (Josman et al., 2009). Depuis 2019, à Bordeaux, une psychiatre virtuelle[3] diagnostique les addictions et les éventuels troubles dépressifs chez des patients. Elle permet aux médecins de réduire la durée des consultations, de limiter les attentes et pourrait permettre à des patients en zone rurale d'avoir accès à un diagnostic.

L'aide au traitement

La réalité virtuelle est d'ores et déjà présente dans de nombreux domaines. Ces thérapies sont d'autant mieux acceptées qu'elles sont vécues comme une forme de jeu et qu'elles peuvent être adaptées à l'environnement quotidien du patient (Klinger et al., 2013). Il est possible de sélectionner une fonction à développer, d'adapter le protocole aux difficultés de chaque patient et de mesurer la progression des apprentissages tandis que les réussites obtenues renforcent la confiance en soi du patient. Les domaines concernés sont très nombreux, notamment les TOC, les inquiétudes chroniques et les diverses phobies (Malbos et al., 2017).

Sont également concernés les troubles alimentaires (Riva et al., 2004 ; Gutierrez-Maldonado Ferrer-Garcia, 2005), les troubles sexuels masculins (Optale et al., 2004) et les toxicomanies (Lee et al., 2004) ; Bordnick et al., 2005 ; Auriacombe et al., 2018), l'anxiété (Robillard et al., 2010 ; Freeman et al., 2017), les troubles douloureux (Matamala-Gomez et al., 2019), les délinquants violents (Seinfeld et al., 2018). Des expérimentations sont en cours dans le domaine de la psychose, notamment des hallucinations auditives avec la création d'un avatar que le patient crée, puis que le thérapeute manipule (Craig et al., 2017).

Dans le domaine des troubles mentaux, le jeu vidéo Sparx a été proposé pour aider à lutter contre la dépression chez les adolescents (Merry et al., 2012).

Facebook utilise même une intelligence artificielle capable de fonctionner comme psychothérapeute pour les adolescents déprimés. La chose s'appelle Woebot. Son modèle est celui des thérapies comportementales et cognitives (TCC). Comme les thérapeutes appartenant à cette école, il s'adresse aux patients déprimés en essayant de « redresser » les représentations erronées qu'ils sont censés se faire d'eux-mêmes et du monde. Par exemple, si un patient dit : « Personne ne m'apprécie », le thérapeute répond : « Je suis sûr que ce n'est pas vrai, il y a des gens qui vous apprécient, mais vous ne vous en rendez pas compte parce que vous êtes dans un « cycle de pensées négatives ». Réfléchissons ensemble. Il y a bien un domaine dans lequel vous réussissez, etc. » C'est ce qu'on appelle le remodelage cognitif.

A la différence des psychanalystes, ces thérapeutes ne cherchent donc pas à savoir si une raison particulière a pu distordre le jugement du patient. Ils ne prennent pas non plus en compte le transfert, c'est-à-dire la façon dont chaque patient appréhende son thérapeute différemment. Pour eux, tous les thérapeutes bien formés sont censés travailler exactement de la même façon et obtenir les mêmes résultats – bien que certains d'entre eux reconnaissent en privé que leur personnalité intervient dans les réponses du patient à la méthode. La théorie prescrit d'éliminer au maximum les interférences liées aux relations humaines. Le thérapeute n'est rien et le protocole est tout. L'objectif est d'inviter le patient à adopter une vision plus positive de son quotidien. Une étude menée auprès de 70 étudiants répartis en deux groupes montre que des échanges menés avec Woebot pendant 15 jours sont plus efficaces que la consultation d'un livre électronique (Fitzpatrick et al., 2017). La comparaison n'a pas été menée avec un thérapeute réel. Le but était seulement de voir ce qui peut pallier le mieux au manque de thérapeutes, pas de démontrer qu'il faudrait en augmenter le nombre. Cela est en effet exclu pour diverses raisons, dont l'argument budgétaire n'est pas le moindre…

[1] Le mot « bot » désigne de façon abrégée un robot. L'utilisation d'un préfixe permet de préciser son domaine d'action : par exemple, un chatbot est un robot qui « chatte », un sexbot un robot utilisé pour des activités sexuelles et un cobot un robot collaboratif utilisé dans une tâche le plus souvent professionnelle.

[2] Rizzo, A.-S. (2011). SimSensei. Consulté le 19 septembre 2020 sur ict.usc.edu/prototypes/simsensei/

[3] LCI (2020, janvier). Pourriez-vous vous confier à une psychiatre virtuelle ? Consulté le 19 septembre 2020 sur https://www.lci.fr/high-tech/video-pourriez-vous-vous-confier-a-une-psychiatre-virtuelle-on-l-a-testee-pour-vous-2143295.html

Les limites de l'IA

Les limites de la robotisation

En Chine, des machines sont déployées massivement pour faire face à la pandémie. Il existe des robots désinfectants, des robots livreurs, des robots patrouilleurs, équipés de haut-parleurs et de caméras, qui accostent les passants sans masque et scannent leur température à l'aide d'une caméra infrarouge, et en cas de température, déclenchent une alarme et envoient une alerte à la police, et mêmes des robots soignants (encore en expérimentation) pour effectuer des tâches variées telles que l'auscultation cardiaque et respiratoire, le prélèvement de salive et la distribution de médicaments. Et la Chine n'est pas le seul pays à en développer : l'Italie le fait aussi, et si la robotique française le pouvait, il est probable que nous le ferions aussi. Si mourir au milieu des robots est présenté aux malades en fin de vie comme la meilleure façon de protéger leurs proches et le personnel médical, il est peu probable qu'ils osent s'en plaindre. Pourtant, les médecins et les infirmiers qui ont pu se protéger en pilotant à distance ces machines programmées pour vérifier les paramètres vitaux ou déclencher des procédures indispensables au maintien en vie des malades auraient eu moins besoin de le faire si plusieurs d'entre eux n'avaient pas déjà été durement touchés par le virus faute de protections satisfaisantes, et si ces mêmes équipements ne continuaient pas à manquer à ceux qui restent. De façon générale, il serait catastrophique que l'introduction précipitée de tels robots sous l'effet de l'urgence fasse oublier l'indispensable réflexion éthique qui doit précéder leur mise en place.

Les limites de la télé présence

Durant toute la période du confinement, une grande partie de l'accompagnement des plus fragiles s'est faite à distance, qu'il s'agisse de télémédecine, de soutien psychologique ou de travail social… Grâce à ces technologies, de nombreuses personnes vulnérables ont pu continuer à être suivies et accompagnées. De même, des robots de télé présence permettent à des enfants hospitalisés plusieurs semaines en chambre stérile de supporter le sentiment d'isolement en leur permettant de rester en contact avec leurs amis et leur famille[4]. Pour autant, il faut se poser la question de la qualité de ce suivi. Donner des indications de traitement à distance peut s'avérer compliqué, et s'il s'agit d'annoncer une maladie grave, rien ne remplace la présence physique (Tisseron, 2020b). D'ailleurs, aux États-Unis, un médecin ayant utilisé un robot de télé présence pour annoncer à un malade qu'il lui restait cinq jours à vivre semble avoir précipité sa mort[5]. L'événement interroge l'opportunité d'utiliser de tels robots dans des situations où des manifestations d'empathie sont attendues (Tisseron, 2017).

Les limites des chatbots en psychiatrie

La pseudo présence par machine interposée peut se révéler terriblement anxiogène dans un moment où c'est une vraie présence qui est attendue. La voix d'Alexa sortant d'une enceinte connectée pour rappeler à des patients atteints de démence de prendre leurs médicaments provoquerait même chez beaucoup d'entre eux une profonde détresse (Tisseron, 2020a). C'est bien compréhensible. La voix de synthèse crée une situation paradoxale de pseudo présence qui peut se révéler terriblement anxiogène à chaque fois que c'est une vraie présence qui est attendue. Elle accroit le sentiment d'insécurité et de solitude en augmentant la douleur de la présence humaine espérée et qui fait défaut.

Le risque de la capture des données

Revenons à Woebot. Pourquoi Facebook s'est-il lancé dans cette aventure ? Est-ce parce que la firme a pris conscience de la misère psychologique de beaucoup d'adolescents ? Mais si c'était le cas, pourquoi n'aurait-elle pas décidé de financer l'ouverture de centres de consultations et de prise en charge des souffrances étudiantes sur les campus ? Il suffit ici de rappeler le modèle économique de Facebook pour le comprendre. Cette entreprise vit de la capture des données personnelles de ses usagers qu'elle utilise ou qu'elle revend. Cela assure déjà Woebot de bénéficier de beaucoup d'informations pour poser les bonnes questions à ceux qui décident de l'utiliser : il exploite tout ce que son utilisateur a mis de lui sur Facebook, ou que ses proches ont mis sur lui. Mais les confidences qui lui sont faites constituent en même temps autant de nouvelles données personnelles que Facebook va pouvoir exploiter. Autrement dit, Woebot n'est finalement qu'un râteau plus fin qui va permettre à Facebook de ratisser bien mieux nos données les plus intimes, au risque même d'y agréger des données médicales qui devraient rester confidentielles. Et que va faire Woebot si un étudiant lui confie participer à des viols collectifs ? Et qui sera responsable s'il apparaît que l'utilisation de Woebot aggrave la santé mentale d'un patient ?

La vulnérabilité humaine face au risque d'une confiance excessive dans la machine

Mais le risque le plus grand est certainement la fragilité humaine par rapport à des technologies capables de simuler les compétences humaines (Tisseron, 2015, 2020). L'informaticien Joseph Weizenbaum a attiré notre attention sur ce point dès les années 1960. Il mit au point un programme capable de simuler les propos d'un thérapeute rogérien. La machine, baptisée Eliza, reformulait systématiquement les propos de son utilisateur sous la forme de questions, et lorsqu'elle ne trouvait pas comment le faire, elle affichait le message : « Je vous comprends ». Or Weizenbaum découvrit que certains de ses utilisateurs passaient beaucoup de temps avec elle en disant avoir l'impression qu'elle les comprenait. Ils étaient bien convaincus que la machine était une machine, mais ils pensaient pourtant qu'elle leur accordait la même qualité d'attention qu'un humain. Weizenbaum parla alors de « dissonance cognitive ».

La tendance à attribuer aux ordinateurs des attributs sociaux similaires à ceux des humains, avec le risque de biais cognitifs, est désigné aujourd'hui comme paradigme CASA (*Computers As Social Actors*) (Nass et al., 1994 ; Gambino et al., 2020). Ce traitement anthropocentrique s'applique à la fois dans des environnements naturels et de laboratoire, même si les utilisateurs conviennent que leurs machines ne sont pas des humains et ne devraient pas être traités comme tels.

[4] https://www.bfmtv.com/tech/vie-numerique/des-robots-pour-redonner-le-sourire-aux-enfants-hospitalises_AN_201609290063.*html*(*Consultle*10/12/2020).

[5] https://www.zdnet.fr/actualites/un-robot-medecin-apprend-a-un-patient-en-phase-terminale-qu-il-va-mourir-39882041.htm (Consulté le 10/12/2020).

Ce fonctionnement psychique a été éclairci par Kahneman (2011). Deux modes de raisonnement sont engagés, et ils peuvent se contredire. Le système 1 est rapide et intuitif. Il nous amène notamment, par commodité, à adopter vis-à-vis de nos objets familiers les mêmes comportements que vis-à-vis de nos semblables. Par exemple, si mon ordinateur tombe en panne, je peux lui dire : « *Non, tu ne vas pas me faire ça quand même ! Pas aujourd'hui !* » Mais si je peux réprimander mon ordinateur, je n'attends pas de lui qu'il me réponde et je ne crains pas qu'il soit fâché. Parce que nous avons aussi un système 2 qui, contrairement au système 1, est lent et fait appel à la rationalité. Dans nos relations aux objets, il nous permet de ne pas confondre le monde animé et le monde inanimé. Seules les créatures vivantes sont capables de poursuivre leurs propres objectifs selon leurs propres moyens.

Jusqu'à maintenant, la distinction était facile : pas de risque de confondre un grille-pain ou une photo copieuse avec un être vivant. Le problème est que tout cela va changer avec les machines dotées de la voix. Nous serons beaucoup plus enclins à les intégrer à notre réseau relationnel exactement comme des humains, et donc à fonctionner avec elles sur un mode intuitif, en utilisant notre système 1. Or ce système est facilement victime de biais de raisonnement. Il nous entraîne à établir des causalités là où il n'y en n'a pas. Par exemple à nous dire : « La machine me dit que cette coiffure me va bien, ce doit être vrai. » Et le problème, c'est qu'en avoir conscience ne suffit pas forcément à nous en prémunir. Il est urgent de réfléchir à ces questions, et de poser un cadre éthique pour nous protéger de nos faiblesses face à aux machines, et notamment à celles qui sont dotées de la voix. Car l'homme qui parle à la machine parle en réalité toujours à l'homme qui est derrière la machine.

Une indispensable charte éthique

La relation que nous allons nouer avec ces machines est appelée à devenir centrale en psychologie, dans la mesure où nous interagirons avec elles comme avec des humains, tout en sachant que nous ne pourrons pas leur donner les mêmes droits moraux et les mêmes responsabilités qu'à des humains (Tisseron, 2015, 2018). C'est pourquoi, en novembre 2017, j'ai donné à l'Institut pour l'Etude de la Relation Homme Robots (IERHR), fondé en 2013, sa charte éthique[6]. Elle porte précisément sur cinq points.

Liberté respectée des usagers

Notamment en imposant que l'accord sur les conditions d'utilisation d'un robot – en termes de respect de l'intimité et de la vie privée – soit signé par l'utilisateur lui-même, et pas par un tiers, en particulier dans les institutions soignantes ; en faisant en sorte que chacun puisse déplacer et ranger son robot selon sa volonté ; qu'un dispositif visuel rappelle à quel moment chaque machine collecte et transmet les données personnelles de son utilisateur ; et que celui-ci ait facilement accès à l'interrupteur permettant de le déconnecter.

En effet, il est capital de laisser à l'usager d'un robot la liberté de le débrancher s'il le désire, ou de le mettre dans le placard. Les robots doivent être conçus dans ce sens. Il ne s'agit pas seulement de créer les conditions matérielles pour que cela soit possible, mais aussi d'en créer les conditions psychologiques. Il est essentiel que l'arrêt du robot ne provoque pas une mise en scène de sa mort subite, par exemple par chute brutale de sa tête sur sa poitrine, qui puisse dissuader les personnes fragiles psychiquement de le débrancher.

Transparence des algorithmes

De plus en plus de décisions qui ont des implications sur nos vies dépendent du résultat de systèmes algorithmiques : diagnostics médicaux, demandes de prêt ou d'assurance et peut-être, demain, décisions de justice. Ces algorithmes mettent en œuvre, le plus souvent de manière opaque, des critères de priorité, de préférence, de classement qui ne sont généralement pas connus des personnes concernées. Cette opacité peut faire craindre des discriminations, voire des manipulations. Pour répondre à ces craintes, on parle souvent de « transparence des algorithmes ». Il s'agirait de donner un accès libre à ces algorithmes et à leurs codes. Mais pour les utilisateurs, le fonctionnement d'un algorithme a peu d'intérêt. L'important est plus l'intelligibilité que la transparence. Autrement dit, il s'agit plutôt de donner aux usagers toutes les informations utiles pour qu'ils puissent en interpréter les résultats, et pour cela, il faudrait contraindre les concepteurs des algorithmes d'aide à la décision de produire, en plus des résultats attendus, des éléments d'explication. En outre, les citoyens doivent pouvoir débattre collectivement de ces questions qui relèvent en définitive de choix de société. Par exemple, comment trouver un juste équilibre entre respect de la vie privée et individualisation des traitements ? La loi informatique et libertés et le règlement européen sur la protection des données (RGPD) ont posé des premiers jalons dans ce domaine. Il est maintenant essentiel d'améliorer la formation des citoyens en matière de numérique.

Autonomie du patient

Les robots de prochaine génération capable d'identifier les émotions humaines et de simuler des émotions dans leurs intonations vocales, voire dans leurs mimiques, posent évidemment la question du développement d'un risque de dépendance. C'est pourquoi une question majeure de la robotique d'assistance, notamment pour les personnes âgées à domicile, réside dans le choix de machines qui encouragent ou non l'autonomie du patient. Les robots seront en effet de plus en plus capables de répondre aux attentes de communication les plus simples, telles que partager des conseils de cuisine, jouer à un jeu, ou coacher des exercices physiques. Mais un robot peut aussi informer son utilisateur sur les ressources humaines de proximité. Par exemple l'existence d'un club de quartier, ou de personnes elles aussi isolées avec lesquelles l'usager pourrait entrer en contact pour s'adonner à son activité préférée. Les fabricants de robots doivent se voir imposer de fabriquer des robots qui favorisent les relations entre les humains exactement de la même façon qu'ils sont obligés de fabriquer des robots qui ne mettent pas en danger la santé physique de leurs utilisateurs. Dans les deux cas, il y va de la santé, mentale d'un côté et physique de l'autre. Autrement dit, le modèle du robot de compagnie doit être le robot « humanisant » qui favorise les rencontres entre humains plutôt que le robot humanoïde capable de se substituer à un humain de compagnie (Tisseron, 2015).

[6] https://www.ierhr.org/charte-ethique/: :text=Notre-20charte-20ethiquetext=Pour-20s'assurer-20d'un,n-C3-A9cessaires-20pour-20les-20faire-20respecter.

Dignité : écarter le risque de confusion entre l'homme et la machine

Notamment en préconisant qu'une intelligence artificielle se présente toujours comme telle quand on interagit avec elle au téléphone ou sur Internet. L'IA qui se fait passer pour un humain, explicitement ou par défaut, devrait être interdite, tout comme les publicités toxiques qui prétendent nous vendre des robots ayant « des émotions », ou « du cœur ». Et pour atteindre cet objectif, il serait souhaitable qu'une partie de l'intérieur des robots soit toujours visible grâce à des protection transparentes plutôt qu'opaques, afin de rappeler leur caractère de machine. Et veiller aussi à ce que chaque protocole de soin donne lieu à une réflexion sur la pertinence du choix du robot pour le service attendu : l'utilisation des robots androïdes, qui ont une similitude d'apparence avec un humain et pourront de mieux en mieux en simuler un, devrait notamment être questionnée du point de vue du rapport bénéfices-risques.

Egalité de tous dans l'accès aux technologies innovantes

Cela passe notamment par le maintien d'une couverture de santé qui fasse bénéficier l'ensemble de la population de soins de qualité, en veillant à ce que les données recueillies sur chacun par les objets connectés n'aboutissent pas à l'élaboration de contrats « à la carte » en fonction des risques encourus, en particulier de la part des compagnies d'assurance. Les personnes dont les données révèlent des risques accrus de fragilité, de déficience ou de maladie doivent bénéficier de tous les services de façon identique au reste de la population.

Conclusion

Il faut nous garder de tout enthousiasme naïf, et plus encore du risque de croire que des machines pourraient bientôt remplacer des soignants. La technologie peut jouer un rôle positif dans l'amélioration du système de soins, mais les domaines dans lesquels elle pourrait être utilisée de façon efficace et apaisante ne sont pas encore cernés (Tisseron, 2020a). Veillons à faire en sorte que les robots ne remplacent jamais les humains, mais qu'ils permettent à ceux-ci de faire mieux, et dans de meilleures conditions, ce qu'ils faisaient jusque-là sans eux. Si certaines tâches peuvent être déléguées à des robots au sein des hôpitaux, cela ne peut pas être le cas de toutes. Alors que les effets de l'utilisation d'un robot dans l'annonce d'un diagnostic à un patient sont largement inconnus, et que certaines études font état d'un accroissement de l'inquiétude quand des conseils d'importance vitale sont donnés par une machine, il serait dramatique que la peur de la contamination et du manque de personnel soignant conduise à accepter certaines pratiques sans véritablement en questionner l'éthique.

Or, d'ores et déjà, certains services hospitaliers s'organisent d'une façon qui amène chacun à travailler comme une « IA » plus que comme un humain. C'est ce dont témoigne la lettre d'un patient publiée sur une liste internet de diffusion d'informations et d'échanges entre soignants[7]. Elle est intitulée : « *Une nuit aux urgences dans un grand hôpital parisien* ». Le patient dit l'avoir également envoyée au directeur et au médecin chef du service des urgences de l'hôpital concerné. Les mots qui viennent à la lecture de cette lettre posent avec acuité les problèmes dont souffre aujourd'hui la médecine, et dont l'IA, non seulement ne s'occupera pas, mais qu'elle risque même d'aggraver : technicisation, hyper-spécialisation, perte de sens, déshumanisation et finalement maltraitance. La technicisation se voit dans la façon dont les médecins finissent, avec l'IA, par gérer des dossiers plus que des patients : « *Tout le personnel était rassemblé dans une sorte d'aquarium vitré, entre les deux couloirs, et chacun semblait très occupé devant son écran. Il était très difficile de s'accorder le droit de les déranger. Quand une personne sortait de cette pièce, elle semblait courir vers une urgence, ou en revenir tout aussi pressée…* » Le personnel ne voit plus l'intérêt de se présenter : « *Je suis arrivé très vite en ambulance et quasiment dans le quart d'heure qui a suivi, j'ai été accueilli par une personne qui m'a fait un nouvel électrocardiogramme. Cette personne ne s'est pas présentée ni par son nom, ni par son statut. Une seconde personne est venue et ne s'est pas présentée non plus, et m'a demandé depuis quand j'avais mal. J'ai dit que c'était depuis une dizaine de jours mais que la douleur s'était aggravée depuis trois jours. Elle m'a dit que je n'étais pas un cas urgent. J'ai demandé pourquoi. Elle m'a répondu que c'est parce que la douleur durait depuis plusieurs jours. J'ai vu alors une troisième personne qui ne s'est pas présentée non plus, ni par son statut ni par son nom, et qui a regardé mon électrocardiogramme* ». Les médecins perdent tellement de vue les patients qu'ils se sentent agressés par leurs questions, comme si se développait chez eux une phobie du contact, produisant une attitude déshumanisante : « *Là, j'ai attendu quatre heures. Au bout de deux heures, je me suis levé car j'étais toujours allongé et je suis allé demander pourquoi une infirmière ne venait pas me faire la prise de sang qui permettrait de répondre à la question de mon état cardiaque, me permettrait de sortir, et libérerait une place. Une personne en blouse blanche, à qui je posais cette question a continué à marcher sans me regarder. Je lui ai dit de s'arrêter de marcher quand je lui adressais la parole et de me regarder quand elle me parlait. Cette personne était visiblement très décontenancée. Elle a dit "je ne sais pas, demandez à votre médecin". Je lui ai dit que je ne savais pas son nom, et que je ne l'avais vu que quelques minutes* ». L'hyper-spécialisation conduit à une perte de sens : « *Le fait d'être hospitalisé avec une suspicion de souffrance cardiaque a eu pour conséquence que tous les examens prescrits se sont polarisés sur ce seul objectif et que je n'ai eu aucun examen, ni aucun conseil, visant à déterminer l'origine de ma douleur thoracique. Je suis donc sorti avec la même douleur, sans aucune indication ni conseil, ni proposition de prendre rendez-vous dans d'autres spécialités pour mieux l'investiguer* ». Les malades une fois réduits à une suite d'informations au sujet de leur maladie, les situations de maltraitance pourtant évidentes ne sont plus identifiées et rien n'est proposé pour y remédier : « *J'étais entré aux urgences à 17H, et, si ma femme ne m'avait pas apporté quelques tranches de pain et du chocolat dans la soirée (sans que je n'en demande l'autorisation car je craignais qu'elle me soit refusée dans la mesure où j'étais dans une zone interdite à ceux qui n'ont pas de badge), j'aurais pu être à jeun depuis mon dernier repas de midi, soit 15H sans boire ni manger, sans aucune raison médicale. Personne ne s'en était soucié. L'infirmière qui m'a fait ma prise de sang, à laquelle j'ai posé la question, m'a répondu que rien n'était prévu* ».

Rappelons pour terminer cette réalité simple : les pa-

[7] Document consulté le 28 septembre 2020 sur une liste de partage d'informations médicales.)

tients sont des personnes, et les médecins aussi. Ils ont un nom. L'IA, elle, n'en n'a pas. Une médecine anonyme est ce qu'il y a de pire. Et c'est celle que risque de fabriquer l'IA, en réduisant les patients à leurs dossiers et les médecins à des machines à faire des diagnostics et à proposer des thérapeutiques.

Ouvrages de l'auteur sur le sujet :

L'emprise insidieuse des machines parlantes, Plus jamais seul (Les Liens qui libèrent) ; Petit traité de cyber psychologie (Le Pommier) ; *Empathie et manipulations, les pièges de la compassion* (Albin Michel Poche) ; *Le jour où mon robot m'aimera, vers l'empathie artificielle* (Albin Michel).

RÉFÉRENCES

Auriacombe, M., Moriceau, S., Serre, F., Denis, C., Micoulaud-Franchi, J.A., de Sevin, E., Bonhomme, E., Bioulac, S., Fatseas, M. Philip, P. (2018). Development and validation of a virtual agent to screen tobacco and alcohol use disorders, Drug Alcohol Depend.1(193), 1-6.

Bordnick, P.-S., Traylor, A.-C., Graap, K.-M., Copp, H.-L. et Brooks, J. (2005). Virtual reality cue reactivity assessment: a case study in a teen smoker. Appl Psychophysiol Biofeedback, 30(3), 187-193.

Craig, T.-K.-J., Rus-Calafell, M., Ward, T., Leff, J., Huckvale, M., Howarth, E., Emsley, R. Garety, P.-A. (2017). AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry. 5: 31-40

Fitzpatrick, K.-K., Darcy, A., Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health, 4(2):e19, DOI: 10.2196/mental.7785

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B. Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol. Med. 47, 2393–2400.

Gambino, A., Fox, J. Ratan, R. (2020). Building a stronger CASA: extending the Computers Are Social Actors Paradigm. 1. 71-80. 10.30658/hmc.1.5.

Gutierrez-Maldonado, J. Ferrer-Garcia, M. (2005). Assessment of virtual reality effectiveness to produce emotional reactivity in patients with eating disorder. In S. Richir B. Taravel (eds.), VRIC - Laval Virtual p. 131-138. Laval.

Josman, N., Elbaz Schenirderman, A., Klinger, E. Shevil, E. (2009). Using Virtual Reality to Evaluate Executive Functioning among Persons with Schizophrenia: A Validity Study. Schizophrenia Research, 115(2-3), 270-7.

Kahneman, D. (2011). Système 1 / Système 2 : Les deux vitesses de pensée. Flammarion, 2012.

Klinger, E., Kadri, A., Sorita, E., Le Guiet, J.-L., Coignard, P., Fuchs, P., Leroy, L., Du Lac, N., Servant, F. Joseph, P.-A. (2013). AGATHE: a tool for personalized rehabilitation of cognitive functions based on simulated activities of daily living. IRBM. 34:113-118.

Klinger, E. (2014). Les apports de la réalité virtuelle à la prise en charge des déficiences cognitives. In R. Picard (ed.), Réalités industrielles - Connaissances et systèmes technologiques pour la santé, p. 57-62. Les Annales des Mines.

Lee, J.-H., Hahn, W.-Y., Kim, H.-S., Ku, J.-H., Park, D.-W., Kim, S.-H., Yang, B.-H., Lim, Y.-S. Kim, S.-I. (2004). A functional magnetic resonance imaging (fMRI) study of nicotine craving and cue exposure therapy (CET) by using virtual stimuli. In CyberTherapy.

Malbos, E., Oppenheimer, R. Lacon, C. (2017). Se libérer des troubles anxieux par la réalité virtuelle : Psychothérapie pour traiter les phobies, l'inquiétude chronique, les TOC et la phobie sociale. Eyrolles.

Matamala-Gomez, M., Donegan, T., Bottiroli, S., Sandrini, G., Sanchez-Vives, M.V. Tassorelli, C. (2019). Immersive virtual reality and virtual embodiment for pain relief. Front. Hum. Neurosci. 13: 279.

Merry, S., Stasiak, K., Shepherd, M., Frampton, C., Fleming, T. Lucassen, M. (2012). The effectiveness of SPARX, acomputerised self-help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. British Medical Journal, 344, p.1-16.

Nass, C., Steuer, J. Tauber, E.-R. (1994). Computers are social actors, CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 72–78.

Optale, G., Pastore, M., Marin, S., Bordin, D., Nasta, A. Pianon, C. (2004). Male Sexual Dysfunctions: immersive Virtual Reality and multimedia therapy. Stud Health Technol Inform. 99:165-178.

Plancher, G., Barra, J., Orriols, E. et Piolino, P. (2012). The influence of action on episodic memory: a virtual reality study. Quarterly journal of experimental psychology (2006). 66.

Riva, G., Bacchetta, M., Cesa, G., Conti, S. Molinari, E. (2004). The use of VR in the treatment of Eating Disorders. Stud Health Technol Inform. 99:121-163.

Robillard, G., Bouchard, S., Dumoulin, S., Guitard, T. Klinger, E. (2010). Using virtual humans to alleviate social anxiety: preliminary report from a comparative outcome study. Stud Health Technol Inform, 154:57-60.

Seinfeld, S., Arroyo-Palacios, J., Iruretagoyena, G., Hortensius, R., Zapata, L.-E., Borland, D., De Gelder, B., Slater, M. Sanchez-Vives, M.-V. (2018). Offenders become the victim in virtual reality: impact of changing perspective in domestic violence. Sci. Rep. 8:2692.

Tisseron, S. (2015). Le jour où mon robot m'aimera, vers l'empathie artificielle, Albin Michel.

Tisseron, S. (2017). Empathie et manipulations, Espaces libres, 2020.

Tisseron, S. (2018). Petit traité de cyberpsychologie, Le Pommier.

Tisseron, S. (2020a). L'Emprise insidieuse des machines parlantes, plus jamais seul, Les Liens qui Libèrent.

Tisseron, S. (2020b). Facilités et pièges de la communication à distance : les leçons du confinement. In C. Byk (ed.), COVID-19 : vers un nouveau monde ? Une analyse de la pandémie à travers le regard des sciences sociales et humaines (p. 121-131). MA Editions.

Bio-objets : enjeux et perpectives de la civilisation *in vitro*

REVUE MÉDECINE ET PHILOSOPHIE

Céline Lafontaine*
*Professeure au département de sociologie de l'Université de Montréal, Canada

RÉSUMÉ

Cet article est un extrait du livre de l'auteure, "*Bio-Objets. Les nouvelles frontières du vivant*", à paraître au Seuil le 4 mars 2021. Céline Lafontaine est notamment l'auteure de *La société postmortelle*, *L'individu, la mort et le lien social à l'ère des technosciences*, Seuil 2008, et du livre *Le corps-marché. La marchandisation de la vie humaine à l'ère de la bioéconomie*, Seuil 2014.

MOTS-CLÉS : Bio-objets, *in vitro*.
DOI : 10.51328/105

Fruit de nombreuses années de recherche et d'expérimentation, la capacité d'isoler et de maintenir en vie des cellules en dehors d'un organisme figure parmi les plus grandes avancées biotechnologiques du siècle dernier. Associées à la rigueur et à l'objectivité, les images de culture *in vitro* sont désormais partout dans les médias de diffusion et de vulgarisation scientifique, au point où elles sont devenues le symbole même des sciences de la vie. Cette omniprésence de la culture in vitro dans l'imaginaire contemporain témoigne de sa naturalisation au sein de notre univers sociotechnique. Pourtant, avant de devenir l'icône de la biologie moderne, la culture *in vitro* a irrévocablement transformé notre conception de la vie, car elle a permis d'accroitre la plasticité biologique en rendant possible la modification et la reprogrammation de cellules humaines et non humaines (Landeker, 2007) . Alors qu'il s'agit d'un phénomène majeur de notre temps, ce n'est que tout récemment que les sciences sociales ont entrepris de se pencher sur les implications sociales, culturelles, économiques et éthiques de la prolifération d'objets biologiques devenus la matière première de l'industrie biomédicale.

Formulé dans le domaine des *science studies*, le concept de bio-objet vise précisément à décrire et à analyser les nouvelles formes de vitalité matérielle produites par les avancées biotechnologiques (Webster, 2012). À mi-chemin entre le biologique et l'artificiel, les bio-objets sont les descendants directs des technologies *in vitro*. Devenues banales par le fait de leur normalisation, les cellules *in vitro* représentent la forme la plus commune de bio-objets. Prélevées dans le cadre de protocoles de recherche ou de procédures thérapeutiques, les millions de cellules cultivées dans les cliniques et les laboratoires du monde globalisé constituent à la fois l'un des symboles les plus tangibles et les plus énigmatiques de la civilisation *in vitro* qui repose sur la production, le contrôle et la transformation de matière vivante dans tous les secteurs d'activités (agroalimentaire, biomédical, industriel). Forgé par le sociologue britannique Andrew Webster, le terme « bio-objet » renvoie aux phénomènes sociotechniques par lesquels des éléments vivants (cellules, gènes, gamètes, embryons, tissus, micro-organismes) sont isolés, modifiés et conservés artificiellement en vie afin d'en permettre des usages multiples (Webster, 2012). Ces dérivés d'organismes vivants partagent des caractéristiques communes qui autorisent à les réunir sous une même catégorie conceptuelle, malgré leur très grande diversité. Tout d'abord, les bio-objets ont en commun d'avoir été extraits de leur milieu organique et d'avoir été transformés par une série de procédures techniques afin d'être utilisés dans un contexte médical ou, plus largement, dans le cadre de recherches en biotechnologie (Dabrock, Braun, Ried et Sonnewald, 2013). Créées en tant qu'outils de recherche, à des fins de traitement ou dans une perspective industrielle, ces entités matérielles sont des créatures hybrides qui conservent certaines caractéristiques de leurs origines

organiques, permettant ainsi de les classer du côté du vivant. Par exemple, les cellules *in vitro* peuvent croître, se diviser, se développer et se reproduire : bref, elles sont biologiquement en vie. Elles diffèrent toutefois des organismes vivants, car elles possèdent une plus grande plasticité du fait qu'elles peuvent être congelées, stockées, conservées, décongelées, modifiées, hybridées et échangées (Landecker, 2005). Leur malléabilité technoscientifique confère aux bio-objets une valeur économique qui dépasse leur simple statut de marchandise destinée à la recherche, car plusieurs types de cellules ont désormais une valeur thérapeutique, comme les cellules souches qu'on utilse dans le cadre de la médecine régénératrice. À la fois objets tangibles et source d'anticipation, les bio-objets génèrent des espoirs immenses, tant du côté de la recherche que de l'économie (Lafontaine, 2015). Possédant un fort potentiel spéculatif, ces artefacts biologiques occupent une place centrale dans l'économie de la promesse qui caractérise désormais la recherche dans le domaine biomédical dont on annonce chaque jour les avancées miraculeuses (Lafontaine, 2014).

Au centre de débats et de revendications identitaires souvent contradictoires, voire conflictuels, notamment dans le cadre de l'industrie de la procréation assistée, les bio-objets déplacent et redéfinissent les contours du corps humain et de l'identité subjective en créant de nouvelles formes de fillations parentales et génétiques (Franklin, 2013). Parce qu'ils remettent en cause les frontières entre les corps, entre les espèces, entre vivant et non-vivant et entre nature et artifice, ils sont au centre de nombreuses controverses dont l'un des exemples les plus manifestes est celui entourant le statut des embryons humains (Metzler et Webster, 2011). En effet, les embryons *in vitro* sont considérés comme des personnes potentielles quand ils sont l'objet d'un projet parental, mais ils ont un statut de matériel biologique dans le contexte de la recherche. Qu'il s'agisse d'ovules, de sperme, de différents types de cellules souches ou de cellules hybridées et modifiées génétiquement, les bio-objets possèdent un statut scientifique, social et culturel fluctuant qui varie en fonction des contextes et des acteurs concernés. Cette ambiguïté fondamentale leur confère des qualités heuristiques particulières pour comprendre les nouvelles formes de socialité émergentes, mais aussi pour saisir la complexité de réalités sociales, matérielles et économiques inhérentes aux développements technoscientifiques. En plaçant au centre de la réflexion sociologique les dimensions matérielles et les enchevêtrements sociotechniques relatifs au déploiement des avancées biotechnologiques, le concept de bio-objet permet donc de mettre en lumière des réalités complexes rendues socialement invisibles du fait de leur banalisation dans le fonctionnement de la recherche. De plus, les perspectives ouvertes par cette approche conceptuelle instaurent une distance critique face à la thèse d'une génétisation des identités et d'une moléculariasation de la culture[1]. Il ne s'agit pas de nier l'influence de la génomique et du génie génétique dans la culture contemporaine, mais il m'apparaît nécessaire de prendre un recul théorique par rapport à l'essentialisme génétique, en montrant comment le projet même de maîtrise du vivant se révèle réducteur face à la matérialité fluide et insaisissable des bio-objets. Car l'une de leurs principales caractéristiques est justement de ne pas avoir d'identité déterminée et d'être malléable, ou modelable, à volonté. Malgré leur statut de produits technoscientifiques standardisés, ils ne peuvent donc pas être assimilés à de simples choses. En tant qu'objets biologiques, ils ont le potentiel de traverser les frontières organiques, en passant d'un corps à l'autre, de générer de nouvelles identités et de modifier notre conception de la vie elle-même (Tamminen et Vermeulen, 2012).

Les bio-objets qui s'accumulent dans les laboratoires et les biobanques du monde entier ne sont pas des entités fixes possédant un statut ontologique stable dans le temps. Au contraire, leurs identités biologiques sont construites et reconstruites en fonction des nouveaux usages technoscientifiques et médicaux (Palmer, 2009). En ce sens, ils constituent des objets-frontières, c'est-à-dire des objets qui traversent plusieurs mondes sociotechniques, articulés les uns aux autres (Trompette et Vinck, 2009). Les embryons *in vitro* passant du monde de la reproduction assistée à celui de la recherche, avec tout ce que cela suppose de contradictions au niveau des valeurs, des usages et des logiques identitaires qui leur sont associés, offrent un exemple clair de leur statut transfrontalier. Afin de rendre compte du caractère intrinsèquement fluide et changeant des bio-objets, Andrew Webster propose de centrer l'analyse sociologique sur ce qu'il nomme le processus de *bio-objectivation*, soit l'ensemble des procédés par lesquels ces entités biologiques sont d'abord fabriquées comme des objets par l'entremise d'un travail scientifique et technologique d'isolement et de standardisation, pour ensuite acquérir des identités particulières selon le contexte, les usages et les acteurs concernés (scientifiques, donneurs, patients, etc.) (Holmberg, Schwennesen et Webster, 2011). Par exemple, les pratiques matérielles, les représentations et les discours entourant un bio-objet ne seront pas les mêmes dans le cadre d'une recherche fondamentale que dans celui d'un protocole d'expérimentation clinique. Non seulement les règles qui encadrent la recherche et la clinique ne répondent pas aux mêmes exigences sanitaires et éthiques, mais la valeur qu'on attribue au matériel biologique varie considérablement d'un contexte à l'autre. Qu'il s'agisse de cellules souches embryonnaires ou de sang de cordons, de cellules souches pluripotentes induites ou de cellules souches autologues, chaque type de bio-objets possède une identité variable et fluctuante selon les usages, les promesses et les acteurs impliqués. Il va sans dire qu'un même bio-objet n'aura pas la même valeur selon que l'on est un donneur, un chercheur ou un patient qui espère guérir d'une maladie incurable.

Sur le plan symbolique, les bio-objets représentent des formes de *devenir* dans la mesure où ils concentrent en eux les potentialités du vivant, les multiples usages scientifiques réels ou anticipés ainsi que les promesses dont ils sont porteurs (Erikson, 2015). Figés dans le temps par des procédés de congélation et d'entreposage, ils incarnent l'espoir d'un monde affranchi de la maladie et le fantasme d'une vie biologique échappant aux diktats délétères du vieillissement nourrissant par le fait même une économie de la promesse. Plus concrètement, les bio-objets nous plongent dans une culture d'hybridation et d'intercorporalité au sein de laquelle les frontières corporelles sont sans cesse redéfinies et renégociées. Ce sont, en fait, des *créatures risquées*, car elles sont au centre d'une

[1] Je me réfère ici à la thèse développée par le sociologue Nikolas Rose selon laquelle le paradigme génétique est désormais la référence première de nos représentations de l'identité et de la socialité (Rose, 2007).

logique d'innovation qui traverse les corps et les espèces, engendrant par là même de nouvelles reconfigurations matérielles et culturelles dont on commence à peine à mesurer la portée civilisationnelle (Brown et Michael, 2004). On n'a qu'à penser aux possibilités ouvertes dans le domaine de l'édition génomique par la création l'outil CRISPR Cas-9, qui permet de modifier plus rapidement et plus directement des parties ciblées du génome, pour saisir l'ampleur des enjeux que soulève la production globalisée de bio-objets.

En regroupant sous une même catégorie analytique un ensemble d'entités hétérogènes, le concept de bio-objet rend possible l'élaboration d'une approche synthétique de réalités empiriques *a priori* très diffuses. Il permet d'appréhender la complexité d'un phénomène sociotechnique qui ne cesse de se déployer et de s'étendre à travers la démultiplication des entités matérielles issues de la culture *in vitro*. De manière plus profonde et plus globale, le concept de bio-objet nous force à réfléchir sur notre rapport au vivant à l'ère de l'Antropocène.

RÉFÉRENCES

Hannah Landecker, Culturing Life : How Cells Became Technologies, Cambridge, Cambridge University Press, 2007.

Andrew Webster, « Introduction. Bio-Objects: Exploring the Boundaries of Life » in Vermeulen, Niki, Tamminen, Sakari et Webster, Andrew (dir.), Bio-Objects. Life in the 21st century, Londres, Routledge, 2012

Peter Dabrock, Matthias Braun, Jens Ried et Uwe Sonnewald; « A primer to 'bio-objects': new challenges at the interface of science, technology and society », Systems and Synthetic Biology, vol. 7, 2013, p. 1-6.

Hannah Landecker, « Living Differently in Time: Plasticity, Temporality and Cellular biotechnologies », Culture Machine, vol. 7, 2005 ; disponible sur http://culturemachine.net/biopolitics/living-differently-in-time/.

Céline Lafontaine, « Régénérer le corps pour régénérer l'économie. La double promesse de la médecine régénératrice », in Audétat, Marc (dir.), Sciences et technologies émergentes : pourquoi tant de promesses ?, Paris, Hermann, 2015, p. 243-258.

Céline Lafontaine, Le Corps-Marché. La marchandisation de la vie humaine à l'ère de la bioéconomie, Paris, Seuil, 2014.

Sarah Franklin, Biological Relatives. IVF, Stem Cells, and the Future of Kinship, Durham, Duke University Press, 2013.

Ingrid Metzler et Andrew Webster, « Bio-objects and their Boundaries: Governing Matters at the Intersection of Society, Politics and Science », Croatian Medical Journal, vol. 52, no 5, 2011, p. 648-650.

Nikolas Rose, The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century, Princeton, Princeton University Press, 2007.

Sakari Tamminen et Niki Vermeulen, « Bio-objects and generative relations », Croatian Medical Journal, vol. 53, no 2, 2012, p. 198-200.

Cecily Palmer, « Human and Object, Subject and Thing: The Troublesome Nature of Human Biological Material (HBM) », in Wahlberg, Ayo et Bauer, Susanne (dir.), Contested Categories: Life Sciences in Society, Londres, Routledge, 2009, p. 15-30.

Pascale Trompette et Dominic Vinck, « Retour sur la notion d'objet-frontière », Revue d'anthropologie des connaissances, vol. 3, no 1, 2009, p. 5-27.

Tora Holmberg, Nete Schwennesen et Andrew Webster, « Bio-objets and the bio-objectification process », Croatian Medical Journal, vol. 52, no 6, 2011, p. 740-742.

Lena Erikson, « Pluripotent Promises: Configurations of a Bio-object », in Vermeulen, Niki, Tamminen, Sakiri et Webster, Andrew (dir.), Bio-Objects: Life in the 21st Century, 2012, p. 27-42.

Nick Brown et Mike Michael, « Risky creatures: Institutional species boundary change in biotechnology regulation », Health, Risk Society, vol. 6, no 3, 2004, p. 207-222.

Intelligence Artificielle : impacts des représentations sociales de la notion « d'intelligence » sur le secteur de la santé

REVUE MÉDECINE ET PHILOSOPHIE

Adèle Ghiringhelli*
*Diplômée de l'Université de Montréal

RÉSUMÉ
L'objectif de cet article est de dévoiler sur quelle conception de l'être humain et de son intelligence s'est développé le projet idéologique de création d'une Intelligence Artificielle (IA) à l'image de l'Homme avant de devenir un simple produit technologique révolutionnaire pour tous les secteurs de l'économie en commençant par celui de la santé. Dissimulés aux confins des représentations sociales, les fondements de la représentation informationnelle de l'intelligence amènent à de réels choix scientifiques. Nous verrons donc à travers cet article que de concrètes applications d'IA introduites dans le domaine de la santé découlent 1) d'une représentation réductrice de l'Homme et de son intelligence ; 2) d'une pure ambition d'optimisation et 3) du reflet de notre monde à travers nos données. L'IA en médecine résulte de procédés sociohistoriques, anthropocentriques et philosophiques essentiels à analyser si l'on veut saisir toute la complexité de ce que l'IA met en lumière de la construction de notre réalité sociale et de notre rapport à nous-même.

MOTS-CLÉS : intelligence, Intelligence Artificielle, technosciences, constructivisme, réductionnisme, optimisation, données, médecine, santé.
DOI : 10.51328/107

Introduction

À la fois projet idéologique et objet technique, l'Intelligence Artificielle (IA) s'inscrit dans une promesse de linéarité infinie de l'évolution technoscientifique. Dans l'imaginaire collectif, le progrès technologique prend part au cours naturel de l'humanité à l'instar de l'évolution biologique de notre espèce. Bien que largement réfutable selon la perspective constructiviste que nous emprunterons au cours de cet article, ce cheminement de pensée permet de légitimer toute avancée technologique : elle serait bénéfique puisque inéluctable. Ce déterminisme technologique cultive, en outre, la conception selon laquelle l'individu n'a d'autre choix que de s'adapter pour enfin se conformer aux bouleversements qu'entrainent les technosciences. Cette idéologie technocratique est entretenue par un modèle occidental dominant qui permet difficilement d'entrevoir au-delà de cette configuration épistémique d'une évolution adaptative de la technique. C'est donc au nom du progrès, en tant qu'amélioration de la condition humaine par le biais de la technique, qu'est permis le développement de l'IA, comme bien des technosciences avec elle. Ces dernières font le plus souvent leur apparition au sein du domaine de la santé. Le caractère sacré de la médecine dans la société moderne permet de légitimer tout avancée. C'est donc par ce biais que la plupart des technologies se développent avant d'être prisées par les autres secteurs de l'économie. Si ces technologies servent un intérêt immédiat pour l'individu fragilisé, il est essentiel de se demander si elles servent réellement un intérêt pour l'humanité sur le long terme.

L'IA intervient déjà pour l'aide au diagnostic, le suivi des patients à distance et les traitements personnalisés. Elle émet des prédictions, s'incorpore dans les

prothèses dites « intelligentes » et il est même envisagé qu'elle puisse un jour nous opérer. L'IA fait ainsi progressivement son apparition dans divers domaines de la santé. Parmi bien d'autres en médecine prédictive, l'IDX-DR détecte les anomalies des photographies de rétines, tandis que l'application Thérapixel détecte le cancer du sein à partir de mammographies. OstoDetect cible l'emplacement précis de la fracture au niveau du poignet et rentre ainsi dans le domaine de la médecine de diagnostic. VizLVO, elle, a été créée pour l'aide à la décision médicale. Elle repère les occlusions d'artères dans les scanners cérébraux. (Collectif et al. 2020, 72) Sur treize applications d'IA de diagnostic médical autorisées par la FDA (Food and Drug Administration) et donc à ce jour commercialisées, une seule a été soumise à une étude rigoureuse (Bibault 2019). L'AEM (Agence Européenne des Médicaments) a des critères d'homologations plus stricts qu'aux États-Unis. Ceci-dit, la compétition internationale incite à aller toujours plus vite bien que moins surement. Outre les enjeux évidemment éthiques de cette mise en application souvent précoce, nous verrons que les enjeux sont aussi sociaux, culturels, anthropologiques et philosophiques.

En plus de rendre compte de l'altération causée par les nouvelles technosciences sur le rapport de l'individu à lui-même et au monde qui l'entoure, l'étude de l'IA constitue un point de départ pertinent à l'étude des représentations sociales de l'être humain. Il est donc primordial d'établir sur quelle conception de l'être humain et de son intelligence s'est développé le projet d'une reproduction artificielle de cette qualité reconnue intrinsèque à l'humain et d'élucider ce que l'IA nous apprend de notre représentation de l'intelligence humaine. Nous verrons dans un premier temps que ces applications découlent d'une représentation construite réductrice de l'Homme et de son intelligence. Si de nombreuses théories sur la particularité de l'esprit humain ont été pensées et envisagées au cours de l'histoire, c'est la première fois qu'un postulat est tenu pour acquis au point qu'il donne lieu à des applications concrètes qui révolutionnent le secteur de la santé, les moyens médicaux jusqu'à notre rapport individuel à notre bien-être. Dans un deuxième temps, nous nous étendrons sur le prisme de l'optimisation comme sortie voulue du système créé dans un but opératoire et non plus symbolique. Nous verrons pour finir que le phénomène du *Big Data* est aujourd'hui considéré comme le nouveau modèle de représentation du monde légitimant la concrétisation de l'IA tout en permettant, là encore, une certaine réduction de l'être humain et de son intelligence.

Une représentation sociale réductrice de l'homme et de son intelligence en IA

Dans la continuité de Galatée de Pygmalion, du Golem ou encore de l'ordinateur, l'IA s'inscrit dans la continuité de la genèse de créatures artificielles façonnées par l'Homme à l'image de l'Homme (Breton 1995, 7). La fascination de ce dernier pour son être ne date pas d'hier ; il y a longtemps déjà que la curiosité le pousse à déceler les confins de son espèce. Dans le processus de création, l'Homme fait appel à un imaginaire ancré dans un contexte spécifique. L'attention se porte sur ce qui est considéré comme caractérisant le mieux l'humanité. Si donc ces créatures divergent dans leurs formes et caractéristiques, c'est parce que les cadres de représentation dans lesquels elles ont été produites sont profondément différents. C'est pourquoi il est si pertinent d'analyser la créature artificielle de l'époque car elle révèle une conception toute particulière de l'être humain qui prend racine dans un cadre épistémologique bien spécifique.

S'il est évident qu'au regard de la créature artificielle de notre temps (l'IA) la composante la plus valorisée et intrinsèque de l'être humain est l'intelligence, il est important de comprendre que la représentation de ce concept est socialement construite. Les définitions de l'intelligence sont aussi nombreuses et diverses que les conceptions du monde dans lesquelles elles sont formulées. Aucune n'est univoque puisque toutes dépendent des représentations spécifiques à chaque société résultant des valeurs mises en avant au sein de celle-ci. Au sein de l'approche constructiviste, l'intelligence est comprise comme un phénomène social dont la construction historique et collective devrait être analysée en explorant les fondements psychosociaux des représentations qui lui sont associées. (Mugny et Carugati 2009, 163). Étudier le concept d'intelligence par le biais de la perspective constructiviste consiste donc à analyser par quels processus s'établit la construction symbolique d'une certaine conception de la réalité. Ainsi, au regard des représentations sociales, l'IA constitue un programme informatique doté de fonctions et mécanismes que les chercheurs en IA vont appeler « intelligence » sous le couvert d'une représentation communément acceptée dans un univers épistémologique bien spécifique.

Les fondements de l'IA en tant que discipline scientifique se sont construits sur une nouvelle vision du monde omnisciente appelé modèle informationnel dont la cybernétique est le tributaire. Au sortir de la seconde guerre mondiale, des rencontres rassemblant des scientifiques renommés dans leur champ disciplinaire respectif, appelées « conférences Macy », se tiendront dans le but de discuter d'une possible unification de différentes approches théoriques et méthodes sous un même modèle. Autrement dit, il est question de promouvoir un modèle théorique qui englobe tous les aspects du monde et pouvant expliquer tous les phénomènes. Ces conférences marquent un tournant scientifique qui se traduit par l'avènement de la cybernétique défini par Norbert Wiener (connu comme le père fondateur de cette discipline et membre phare de ces colloques) comme étant « la théorie entière du contrôle et de la communication, aussi bien chez l'animal que dans la machine » (Wiener et al. 1948, 70). Les sciences classiques s'appuyaient jusqu'ici sur une méthode analytique linéaire introduite par Descartes de décomposition des éléments de la nature afin de comprendre le fonctionnement du monde (Breton 1995, 108). Si Norbert Wiener conserve le même objectif de conquête de la nature, pour lui la méthode cartésienne qu'il appelle « fonctionnelle » doit faire place à la méthode « comportementale ». C'est la relation des objets entre eux et leurs interactions avec l'environnement sur lesquelles nous devrions focaliser notre analyse et non pas sur la propriété spécifique de chacun de ces objets. Autrement dit, c'est la transmission de l'information (l'interaction) qu'il pense essentielle à considérer et non pas le support matériel servant à la transmission (la propriété de l'objet). Ainsi, tout s'explique soudainement par le biais de la compréhension d'un processus global pour toute entité. Dès lors que les propriétés internes sont négligeables, mais que seul le mécanisme absolu est pris en considération,

on peut alors se permettre de reconnaître l'œuvre d'une même sorte d'abstraction en l'intelligence humaine et en l'intelligence artificielle.

Plus qu'un simple cadre théorique, le modèle informationnel constitue alors un véritable paradigme au sein duquel la façon de penser la vie et de se penser soi-même en tant qu'être humain est profondément bouleversée et subordonnée au principe de l'information (Lafontaine 2004, 43). « Cette opération [le désir de construire un être à l'image de l'Homme] est rendue possible par une double réduction, de l'humain à l'intelligence, et de l'intelligence au traitement de l'information. » (Breton, 1995: 138). La valeur de l'être humain résiderait donc en sa capacité à traiter l'information. Les mécanismes biologiques sont donc eux-mêmes réduits à des processus modélisés et donc formalisables dans un langage que la machine peut facilement manipuler. L'intégralité de l'intelligence humaine se résume en IA par la capacité d'analyse et d'observation, l'exécution d'opérations par la manipulation de chiffres et la compréhension de phénomènes complexes en faisant preuve de logique opératoire. Dans cette représentation logico-mathématique de l'être humain, ces facultés sont considérées comme substantielles à l'Homme. L'approche réductionniste empruntée par les chercheurs en IA leur permet d'affirmer qu'il serait possible de transmettre l'intelligence à la machine. Les chercheurs en IA ne sont pas en mesure de reproduire artificiellement l'intelligence mais sont capables de reproduire une version extrêmement simplifiée du mécanisme global de l'esprit, analogue au mécanisme de toute autre entité, naturelle comme artificielle.

Sans prise de conscience de la convergence des valeurs entretenues par le paradigme informationnel, l'analyse objective du phénomène que représente l'IA est inconcevable. L'IA s'est développée sur la base d'une représentation réductrice de l'humain et de son intelligence dont les fondements de cette assise ne sont que trop peu évoqués tant ils sont dissimulés aux confins des représentations sociales. De ce « paupérisme épistémologique » (Robillard 2019, 3) découle la mise sur le marché d'applications concrètes dans le secteur de la médecine ayant à terme des effets réels et concrets sur la santé de l'individu ; c'est pourquoi il est si essentiel de dévoiler le réductionnisme du modèle informationnel.

L'IA : d'un fantasme idéologique à une pure ambition d'optimisation

L'IA, voulant reproduire au plus près ce qui est considéré être l'intelligence humaine, s'est adaptée spontanément aux qualités et normes les plus valorisées (et donc constituant l'intelligence) d'une société en constante mouvance provoquant des divergences quant à l'approche de recherche à adopter. C'est pourquoi la méthode, le programme, les entrées apprivoisées ne sont pas les mêmes dépendamment de la sortie voulue du programme, soit de l'objectif de départ recherché. Le but de chacun de ces programmes étant de générer un comportement « intelligent », on cherche à travers l'IA à reproduire artificiellement ce qu'on pense caractérise le concept d'intelligence.

L'évolution de la représentation de l'intelligence évolue ainsi en trois temps correspondant aux trois phases principales de l'histoire de l'IA que nous survolerons rapidement ci-dessous.

Aux prémices de l'IA à l'après-guerre, l'étude de l'esprit est appréhendée sous la métaphore neuronale, branche aujourd'hui connu sous le nom de connexionnisme. Conformément à la perspective cybernétique, le mécanisme global du système neuronal devrait pouvoir se limiter en l'étude des connexions synaptiques entre deux neurones. L'essentiel à la compréhension du réseau de neurones résiderait en cette observation : lorsqu'une information est perçue, une réaction a lieu en fonction de cette dernière. C'est le principe de la rétroaction. La compréhension du processus élémentaire neuronal permettrait la modélisation du comportement humain lors de l'activité cognitive. L'intelligence, selon cette perspective, revient à la capacité «d'orienter et de réguler ses actions d'après les buts visés et les informations reçues» (Lafontaine 2004, 46), la définition même de la rétroaction.

À la fin des années 1950, l'ambition cybernétique d'une machine capable d'un ajustement adaptatif des entrées et des sorties semblait trop limitée (Benbouzid et Cardon 2018, 185). Le projet des pionniers de l'IA dite « symbolique » (McCarthy, Minsky, Shannon, etc.) est de doter la machine d'une capacité de raisonnement. L'intérêt est porté sur le processus cognitif et l'acquisition de connaissances plutôt que sur le comportement humain. Selon cette approche, la cognition (qui constitue l'ensemble des états mentaux) n'est pas réductible au simple niveau neuronal. Ce nouveau courant emprunte une approche computationnelle : il utilise la métaphore de l'ordinateur pour affirmer que l'esprit humain constituerait un système de traitement d'information. À la différence de la métaphore neuronale connexionniste, il ne suffit pas d'identifier le processus élémentaire du raisonnement, mais d'établir les règles de représentations internes qui abritent les connaissances en tant que telles et de les modéliser pour pouvoir les insérer dans le programme informatique afin qu'il ait accès au monde réel. L'information est donc associée à des symboles qui ont une réalité matérielle et une valeur sémantique de représentation qu'il convient de maitriser afin de donner à la machine des capacités abstraites et logiques. Il s'agit alors de mettre en avant la complexité des processus cognitifs de l'être humain qui ferait de lui une espèce à part. La simulation des processus cognitifs au sein de la machine doit donc être fondée sur ce qui fait la caractéristique de l'Homme et ce qui l'élève parmi les autres ; c'est-à-dire sa capacité de manipuler des représentations symboliques de hauts niveaux lui permettant de résoudre des problèmes d'une grande complexité. Le courant cognitiviste, et donc les chercheurs en IA symbolique, conçoit ainsi l'intelligence comme la capacité de résoudre des problèmes par la manipulation de symboles.

L'évolution de l'étude de l'esprit au travers des décennies met en avant un élément fondamental jusqu'alors délaissé par l'IA : l'apprentissage. Apparaît alors une nouvelle branche de l'IA appelé « apprentissage automatique » qui consiste à adapter son programme et ses résultats aux données qui lui sont transmises par un réseau d'entités élémentaires. Les chercheurs en IA entendent ainsi par apprentissage la confrontation entre les résultats d'une action (ou ce qui est attendu d'elle) et l'amélioration de cette action en vue d'atteindre l'objectif attendu. Grâce à la reprise des premiers travaux sur les réseaux de neurones artificiels combinés avec la puissance de traitement plus importante que jamais et aux méthodes d'apprentissage automatique, les chercheurs LeCun, Bengio sous la direction de Hinton sont capables de concevoir au cours des années 1990 des algorithmes

ayant des résultats inespérés. Le retour de l'approche connexionniste donne alors lieu à l'apprentissage profond (Deep Learning), une méthode développée sur la base de l'apprentissage automatique. Le but est d'optimiser la prédiction à partir d'un échantillon très important de données traitées de façon brute, c'est-à-dire sans passer par une modélisation explicite comme on retrouve en méthode computationnelle (IA symbolique). Cela dit, dépendant majoritairement de la quantité d'entrées à traiter (Bengio, Courville, et Vincent 2012, 1), les programmes d'apprentissage profond n'ont percé qu'à partir de 2010 grâce à l'extension du Web mettant à disposition des milliards de données. Aujourd'hui, c'est donc le potentiel d'optimisation sur objectif par l'intermédiaire du calcul statistique qui est recherché ; la visée étant de minimiser ou maximiser un comportement afin d'en soutirer le meilleur résultat possible selon l'objectif (Ghiringhelli 2020, 41). À l'image d'un retour d'investissement, les données de bases intégrées par le programme d'apprentissage profond seront optimisées à la sortie.

L'évolution de l'IA n'est pas seulement le fruit d'une évolution de représentation de l'homme et de son intelligence. Confrontée à de nombreux obstacles (divergences de théories, critiques philosophiques, manque de résultats concrets, puissance numérique trop faible, financements pas assez importants), l'IA comme discipline scientifique s'est transformée au cours du temps en produit technologique servant la machine économique. Contrairement à l'idée reçue, le regain d'intérêt pour l'IA dans les années 2000 n'est pas le résultat de découvertes théoriques récentes et révolutionnaires. Les chercheurs ont simplement réemprunté une méthode qui existait déjà dans un contexte qui lui permet de fonctionner sans aucun nouveau fondement théorique véritable. L'optimisation comme sortie voulue du logiciel d'IA provient davantage d'un délaissement de l'ambition idéologique de reproduction artificielle de l'Homme par l'Homme que d'une réelle représentation nouvelle de l'intelligence humaine (Ghiringhelli 2020, 49). Dans une optique d'efficacité et d'optimisation, ce qui est recherché en IA aujourd'hui n'est plus de comprendre le fonctionnement de l'esprit humain, mais de développer des programmes étant capables des mêmes résultats que l'humain quantitativement parlant. La quête de vérité s'est substituée à la quête d'efficacité. Comme l'énonce Michel Freitag, « Le monde n'est plus la totalité de ce qui est, mais l'ensemble de tout ce qu'on peut faire, prévoir, contrôler, transformer à volonté dans n'importe quel environnement. » (Freitag 2002, 389) L'IA classique tentait encore de mêler la connaissance pure, synthétique et proprement théorique, au caractère essentiellement instrumental de la méthode empirique. L'IA aujourd'hui ne focalise la recherche que sur la pratique, délaissant à jamais tout fondement conceptuel. Dorénavant, seul ce qui sert les intérêts corporatifs est valorisé et subventionné. Sous le prétexte d'un développement technologique pour le bien social (dans le domaine de la santé notamment), la conquête de l'IA par les entreprises et les nations poursuit le règne à la puissance économique.

« On est passé de la théorie à la simulation numérique via une modélisation fondée sur la théorie. Mais dans un stade ultérieur on glisse à des simulations qui s'éloignent de vraies justifications théoriques, mais qui maintiennent et augmentent leur efficacité prédictive en multipliant des ingrédients paramétriques, ajustés sur l'expérience. » (Balibar, 2012, 264)

Au sein de cette nouvelle configuration, il n'est plus question d'expliquer (comme la science le fait) mais de prédire (ce que permet la technique) (Balibar, 2012, 256). Nous remettons à la machine la responsabilité de choisir pour nous sur la base d'une généralisation de ce qui est probable et non certifié d'arriver. La généralisation permet en effet la prédiction : un des moyens ultimes de l'optimisation. Prédire, c'est prendre de l'avance ; soit gagner de l'argent en économisant du temps. Prédire c'est aussi éviter le pire, pouvoir anticiper et régler les problèmes avant même qu'ils arrivent ; ce à quoi le domaine de la santé promet de s'atteler en intégrant de plus en plus d'applications d'IA à leurs outils médicaux. Si l'objectif d'optimisation des applications d'IA peut paraitre dérisoire dans beaucoup de secteur de l'économie, elle n'est pas si anodine en médecine. Personne ne contestera l'intérêt de gagner du temps dans le secteur de la santé. Il est impossible de prédire ce qui va arriver, on peut toutefois tenter de prédire ce qui va probablement se passer dans un avenir plus ou moins proche et se fier à cette probabilité pour mettre en place des stratégies d'action visant à résoudre les problèmes de santé des individus. L'idée de laisser la santé de nos semblables entre les mains de probabilités peut toutefois sembler préoccupante.

Paradoxalement, il serait inexact d'affirmer que la sortie actuelle du système ne correspond désormais plus du tout à la représentation actuelle que l'on se fait de l'être humain et de son intelligence. Simple engrenage de la machine économique, celui-ci ne serait-il pas réduit dans notre société néolibérale à sa seule fonction, laquelle est de travailler pour consommer ? La crise sanitaire que nous traversons atteste cette présomption. En France, avant de prendre la décision de confiner à nouveau ses citoyens en octobre 2020 en raison de la haute propagation de la COVID-19, le gouvernement a fait le choix d'imposer un couvre-feu à partir de 21h jusqu'à 6h du matin, permettant seulement au citoyen de se rendre au travail. Par cette mesure, la réduction de l'individu à sa fonction purement opératoire n'est plus simplement établie de manière implicite, elle est exacerbée et énoncée haut et fort. Comme l'énonce le philosophe Lyotard, « soyez opératoires, c'est-à-dire commensurables, ou disparaissez » (Lyotard 1979, 8). Toutefois, sous le couvert de la crise sanitaire, cette instrumentalisation est toujours difficilement contestable.

Le Big Data ou nouveau modèle de représentation du monde réel

Notre utilisation d'Internet a bouleversé notre quotidien de manière si profonde et globale que la prolifération extrêmement massive et exponentielle de données (appelé *Big Data* à partir de 1997 selon l'*Association for Computing Machinery*) résultant de ce phénomène est appréhendée aujourd'hui comme la représentation la plus proche, précise et réaliste de notre monde. Cette appréhension de notre réalité, bien que largement questionnable, permet selon les chercheurs en apprentissage profond de franchir la limite insurmontable présentée par le philosophe Dreyfus dans sa critique envers l'IA symbolique selon laquelle: « le meilleur modèle du monde est le monde lui-même» (Dreyfus 2007, 247).

Chaque domaine de notre vie, ou presque, est aujourd'hui calibré par le numérique de telle façon que les données sont perçues comme la représentation de

l'environnement. Elles constitueraient en cela le modèle du monde réel. « Cette puissance nous échappe dans son objectivation et prend ainsi valeur de réalité première. » (Freitag 2002, 389) Ce serait par l'observation du monde, en grandissant dans celui-ci et en faisant l'expérience de ces modalités que l'on deviendrait intelligent. Si les données représentent le monde, il suffirait donc que le programme s'imprègne de ces dernières pour qu'il devienne intelligent à son tour. Les données censées représenter le monde réel sont partagées et donc sélectionnées par les utilisateurs du web. Ceux-ci procèdent à un tri de ce qui doit être partagé et ce qui ne doit pas l'être. Or, les réseaux sociaux nous dévoilent une image virtuelle et donc aseptisée, altérée, souvent magnifiée et contrôlée de la réalité. C'est une des raisons pourquoi il est mal fondé (puisque approximatif) d'appréhender les données comme une représentation fidèle du monde environnant.

Les fondements de cette réduction du monde réel au monde des données s'expliquent de la même manière que la réduction de l'intelligence humaine à un simple traitement de l'information. En ne considérant que le processus informationnel d'une entité sans prendre en compte ces qualités intrinsèques, il ne nous reste qu'une multitude d'unités numériques semblables dont la sémantique provient de leur multiplicité. C'est dans cette perspective qu'il est admis que l'ensemble de la quantité de données récoltées des activités numériques de chaque individu puisse donner lieu à une représentation fidèle du monde qui nous entoure.

« Si on réussissait à dépouiller entièrement un être ou une chose de ses qualités propres, le « résidu » qu'on obtiendrait présenterait assurément le maximum de simplicité, et, à la limite, cette extrême simplicité serait celle qui ne peut appartenir qu'à la quantité pure ; c'est-à-dire celle des « unités » toutes semblables entre elles, qui constituent la multiplicité numérique. » (Guénon 1945, 83)

Au-delà de la notion de données comme modèle le plus représentatif de notre monde, c'est la quantité de celles-ci qui importe. Il ne suffit pas d'avoir les données les plus adéquates à la situation, mais d'en avoir un nombre suffisamment important pour que le modèle soit le plus représentatif possible. La logique est la suivante : plus on a accès à un nombre de données important, plus on s'approche d'une expérience réaliste du monde dans lequel l'individu évolue. Or, l'Homme n'a pas besoin d'un accès illimité à des connaissances pour mener sa vie et agir à travers le monde. Il a en revanche besoin de connaissances qualitativement pertinentes à sa propre expérience. Là se trouve la limite de l'approche essentiellement quantitative des données : l'expérience que l'on a du monde ne se limite pas à un traitement d'information par captation de signaux. Une signification sémantique assignée à chacun de ces signaux est essentielle à une compréhension du monde environnant.

Cette appréhension du monde par le biais de nos données ne pourrait être plus équivoque dans le secteur médical. Notre santé est dorénavant l'objet d'un suivi numérique de plus en plus présent. À travers les applications mobiles, les objets connectés ou encore la télémédecine, nous nous trouvons dans l'ère de la santé connectée. Ces outils peuvent mesurer le rythme cardiaque, prendre la température du corps, constater le taux de glucide dans le sang, etc. pour ensuite alerter directement l'individu lorsque les mesures semblent anormales ; c'est-à-dire éloignées du pattern détecté par la récolte des données en temps normal. Elles posent ainsi un diagnostic directement accessible par le profane. Cette numérisation entend un suivi de soin opératoire par le biais de la récolte de données (quantitatives comme qualitatives) traitées et analysées quantitativement par l'IA. Nous avons donc affaire à une médecine quantifiée. Le suivi de santé numérique est à l'individu ce que le *Big Data* est au monde réel : une représentation quantifiable globalisante puisque simplifié d'un modèle. Les données sont alors considérées dans le milieu de la santé comme le reflet objectif d'une réalité (Thoër 2013, 13). L'IA établit des relations statistiques par le biais d'un traitement de données ; tandis que le médecin est capable de donner un sens à ces relations via un savoir accru et spécifique du sujet juxtaposé d'une connaissance du contexte psychosocial du patient établi par un dialogue frontal.

Au-delà du délaissement progressif du contact social au profit d'un suivi méthodique quantifiable, c'est l'introduction de l'IA comme vecteur actif (bien qu'encore intermédiaire) du processus de soins qui est important à soulever ici. L'admission récente de cet acteur interroge la sphère médiatique : « L'intelligence artificielle remplace-t-elle votre médecin ? (La Tribune 2016) », « L'intelligence artificielle, notre nouveau médecin ? (Sciences et Avenir 2019) », « L'intelligence artificielle peut-elle remplacer un vrai médecin ? (Femme Actuelle 2020) », « L'intelligence artificielle va-t-elle rendre les médecins obsolètes? (Pourquoi Docteur 2018) ». Si les réponses à ces questionnements sont pour la plupart négatifs, le fait même d'évoquer un dépassement démontre la conception alignée des deux entités (l'être humain vs la machine) dans une indifférenciation totale de leur nature. Ceci se justifie par un délaissement du dualisme nature/artifice à la base du modèle informationnel. (Esquivel Sada 2009) Une fois établis dans la même catégorie ontologique, la comparaison de leurs capacités est alors légitimée. Si l'on conçoit que la compétence du médecin ne s'étend pas au-delà de sa capacité à traiter des données (suivant la logique selon laquelle l'intelligence serait réduite à un simple traitement de l'information), alors, sous le critère ultime de l'optimisation, on peut dire que l'IA dépasse l'intelligence du médecin simplement parce qu'elle traite beaucoup plus de données en largement moins de temps que son rival.

Conclusion

L'ambition fondatrice de l'IA en tant que domaine scientifique est de reproduire artificiellement l'intelligence humaine. Faut-il encore comprendre ce que l'on conçoit comme « intelligent » pour pouvoir façonnée cette créature tant fantasmée devenue aujourd'hui pur produit technologique. Nous avons vu que la définition du concept d'intelligence diffère en fonction du cadre épistémologique dans lequel il est compris. Or, c'est au sein du modèle informationnel que l'IA a pu être imaginée. Selon ce paradigme, l'intelligence, perçue comme la qualité intrinsèque de l'homme, est réduite à un simple traitement d'information plus au moins complexe en fonction des mouvances théoriques qui la traversèrent. Toutefois, nous avons vu que l'ambition de reproduction de l'esprit humain est aujourd'hui placée au second plan derrière une recherche d'optimisation. Bien que l'IA n'aurait pu être conçue en dehors du modèle informationnel, le

développement de la discipline et de ses programmes informatiques s'est fondé sur des approches et méthodes théoriques qui dépassent celles formulées par la cybernétique. Il est certain que le phénomène du *Big Data* s'est substitué à toute tentative de modélisation humaine fondée sur un courant de pensée théorique précis ; au point qu'elle constitue à ce jour le nouveau modèle de représentation du monde. Ainsi la valeur d'un système d'IA se trouve dans les résultats techniques et non plus désormais dans la réussite à modéliser le raisonnement humain au plus proche de sa configuration biologique.

Il apparait évident qu'une réflexion de fond est de rigueur avant l'acception complète par tous des applications de l'IA dans notre quotidien. D'autant plus dans le domaine de la santé dont les enjeux sont colossaux. Cependant, c'est dans ce même secteur que les arguments envers la progression rapide dénuée de considérations éthiques se font le plus entendre. Si l'objectif n'est que marketing et monétaire, on peut considérer plus facilement contester la mise sur le marché de ces technologies. Toutefois, dès que les objectifs touchent au domaine le plus sacré dans la société moderne (la médecine), la question d'une revendication ne se pose plus même si l'impact sollicite directement notre santé et pas seulement nos habitudes quotidiennes. Si des solutions existent permettant d'améliorer la santé de la population, pourquoi se questionner plutôt que d'agir dans l'immédiat ? Peut-être parce que comme l'a parfaitement formulé Hannah Arendt : « Il se pourrait [...] que nous soyons plus jamais capables de comprendre, c'est-à-dire de penser et d'exprimer, les choses que nous sommes capables de faire »(Arendt 1958, 36). Là est notre tâche en tant que chercheurs en sciences sociales : non seulement émettre une réflexion sur les œuvres de notre société mais aussi dévoiler ce déterminisme technologique à travers lequel nos actions dépassent notre compréhension de ces dernières. La question primordiale à se poser n'est plus de savoir si nous sommes capables de créer une IA. La question est de savoir si cela est souhaitable. L'IA en tant que produit social dévoile une reconfiguration de notre rapport au monde et à nous-même. En tant que productrices de nouvelles réalités sociales, elle engendre dans sa mise en application un bouleversement qui nous dépasse.

RÉFÉRENCES

Arendt, Hannah. 1958. Condition de l'homme moderne. Paris: Pocket.

Balibar, Françoise. 2012. « Prédire n'est pas expliquer ». Critique n° 778 (3): 255-65.

Benbouzid, Bilel, et Dominique Cardon. 2018. Machines prédictives.

Bengio, Yoshua, Aaron Courville, et Pascal Vincent. 2012. « Representation Learning: A Review and New Perspectives ». arXiv:1206.5538 [cs], juin. http://arxiv.org/abs/1206.5538.

Bibault, Jean-Emmanuel. 2019. « Comment réguler l'Intelligence Artificielle en Médecine ». Le Figaro, 2 mai 2019. https://sante.lefigaro.fr/article/comment-reguler-l-intelligence-artificielle-en-medecinefig-page.

Breton, Philippe. 1995. À l'image de l'Homme: du Golem aux créatures virtuelles. Paris: Seuil. Collectif, Nicholas Ayache, Alain Damasio, Yuval Noah Harari, et Cathy O'Neil. 2020. Nouvelle enquête sur l'intelligence artificielle: Médecine, santé, technologies: ce qui va changer dans nos vies. Flammarion.

Dreyfus, Hubert L. 2007. « Why Heideggerian AI Failed and How Fixing It Would Require Making It More Heideggerian. » Philosophical Psychology 20 (2): 247-48.

Esquivel Sada, Daphné. 2009. « Le «nanomonde» et le renversement de la distinction entre nature et technique: entre l'artificialisation de la nature et la naturalisation de la technique ». https://papyrus.bib.umontreal.ca/xmlui/handle/1866/2755.

Femme Actuelle. 2020. « L'intelligence artificielle peut-elle remplacer un vrai médecin? La réponse d'un spécialiste ». Femme Actuelle, 24 avril 2020. https://www.femmeactuelle.fr/sante/sante-pratique/lintelligence-artificielle-peut-elle-remplacer-un-vrai-medecin-la-reponse-dun-specialiste-2094202.

Freitag, Michel. 2002. L'oubli de la société: pour une théorie critique de la postmodernité. Collection Sociologie contemporaine. Laval: Presses de l'Université Laval.

Ghiringhelli, Adèle. 2020. « Analyse des représentations sociales du concept «d'intelligence» dans les discours sur l'Intelligence Artificielle. » Mémoire de recherche, Université de Montréal. https://papyrus.bib.umontreal.ca/xmlui/handle/1866/23698.

Guénon, René. 1945. Le règne de la quantité et les signes des temps. Paris: Gallimard. La

Tribune. 2016. « L'intelligence artificielle remplace-t-elle votre médecin? » La Tribune, 16 octobre 2016. https://www.latribune.fr/technos-medias/informatique/l-intelligence-artificielle-remplace-t-elle-votre-medecin-607123.html.

Lafontaine, Céline. 2004. L'empire cybernétique: des machines à penser à la pensée machine: essai. Paris: Seuil.

Lyotard, Jean-François. 1979. La condition postmoderne: rapport sur le savoir. Collection Critique. Paris: Éditions de Minuit.

Mugny, Gabriel, et Felice Carugati. 2009. Social Representations of Intelligence. Cambridge: Cambridge University Press.

Pourquoi Docteur. 2018. « L'intelligence artificielle va-t-elle rendre les médecins obsolètes? » www.pourquoidocteur.fr, 8 novembre 2018. https://www.pourquoidocteur.fr/Articles/Question-d-actu/27379-L-intelligence-artificielle-va-t-elle-rendre-medecins-obsoletes.

Robillard, Jean. 2019. « Qu'y a-t-il d'intelligent en intelligence artificielle? » manuscrit en circulation libre, paginé, 21.

Sciences et Avenir. 2019. « L'intelligence artificielle, notre prochain médecin? » Sciences et Avenir, 27 mars 2019. https://www.sciencesetavenir.fr/high-tech/sommet-start-up/l-intelligence-artificielle-va-t-elle-remplacer-le-medecin-132466.

Thoër, Christine. 2013. « Internet: un facteur de transformation de la relation médecin-patient? » Communiquer. Revue de communication sociale et publique, no 10 (décembre): 1-24. https://doi.org/10.4000/communiquer.506.

Wiener, Norbert, Ronan Le Roux, Robert Vallée, et Nicole Vallée. 1948. La cybernétique information et régulation dans le vivant et la machine. Paris: Éd. du Seuil.

Le stade prescriptif de la vérité : Hippocrate mis sous le joug du privé

REVUE MÉDECINE ET PHILOSOPHIE

Éric Sadin*
*Écrivain et philosophe

RÉSUMÉ
Cet article est un extrait du livre de l'auteur, "*L'Intelligence artificielle ou l'Enjeu du siècle*", publié chez L'échappée en 2018.

MOTS-CLÉS : intelligence artificielle, technocritique.
DOI : 10.51328/106

C'est un fait entendu. S'il existe un domaine qui à lui seul légitime l'existence de l'intelligence artificielle, dont on ne doute pas qu'il va bénéficier de toute sa puissance, dont nous allons tous finir par profiter, c'est bien celui de la médecine. Il relève d'un large consensus qu'elle devrait permettre à la recherche médicale de franchir des seuils d'une portée sans commune mesure historique. C'est un don qui nous est offert et ce serait faire preuve de mauvaise foi que de ne pas l'admettre. Cette perspective constitue l'argument irréfutable qui plaide au bout du compte pour ses développements. Et dans le cas où certains usages en vigueur, ou en passe de l'être, dans d'autres secteurs suscitent des inquiétudes ou des désapprobations, au moins les bienfaits annoncés sur ce terrain signalent que les choses sont complexes et qu'il convient de faire preuve de mesure dans l'appréciation des choses. Ce prétendu apport de l'intelligence artificielle à la médecine constitue le point de ralliement de ses thuriféraires réjouis, à l'occasion de tout débat, de pouvoir dégainer cette arme implacable.

C'est ce dont use de façon répétée et caricaturale, Yann LeCun, qui, alors qu'il travaille à sans cesse perfectionner la relation client chez Facebook, affirme à intervalles réguliers que «l'intelligence artificielle va sauver des vies», mais reconnaît qu'elle «représente aussi un danger[1] ». Alors, puisqu'elle est appelée in fine à «sauver des vies», à l'instar du pari de Pascal, l'étendue sans limites de ses promesses vaut la peine de nous y risquer. Tout étant, au sein de cette équation, ramené à l'exigence de la «vigilance» qui reporte aux calendes grecques la possibilité de nous déterminer d'ores et déjà et en conscience sur ces enjeux. Il faut savoir identifier les raisonnements destinés à tétaniser toute position divergente. Eh bien, puisqu'il est continuellement asséné que la médecine va tirer une infinité d'avantages des vertus augurées par l'intelligence artificielle, il convient alors d'aller voir de près comment les choses s'opèrent, au-delà des discours préfabriqués qui cherchent à paralyser toute entreprise critique.

L'introduction progressive, à partir des années 1990, d'instruments numériques destinés aux examens médicaux, dans la radiographie, la cardiologie ou l'ophtalmologie, parmi bien d'autres branches, cumulée à l'enregistrement des actes sur des serveurs a transformé la médecine en une pratique générant des volumes de données. Ces nouveaux usages ont notamment autorisé un suivi mémorisé des patients, ainsi qu'une connaissance approfondie de nombre d'états individuels et collectifs détenue par divers organismes, tels la Sécurité sociale, les ministères de la Santé ou l'Organisation mondiale de la santé (OMS). IBM, parmi d'autres sociétés, a vite voulu tirer parti de l'émergence d'une médecine devenue «informationnelle», mettant au point un programme dédié, Watson, qui répond à trois fonctionnalités majeures. La première consiste à récolter et à analyser toutes sortes de

[1] Morgane Tual, «Yann LeCun, de Facebook : "L'intelligence artificielle va sauver des vies", Le Monde, 23 septembre 2017. [lemonde.fr/festival/article/2017/09/23/yann-lecun-de-facebook-l-intelligence-artificielle-va-sauver-des-vies-mais-il-y-a-aussi-des-dangers$_5$190311$_4$415198.*html*]

données. Elles émanent des dossiers des personnes, se rapportent à l'évolution de leur situation, à l'efficacité des traitements sur chaque cas particulier et à leurs éventuels effets secondaires. En outre, le système s'enquiert des différents foyers d'infection dans le monde, il est également capable de «lire» et de synthétiser des articles scientifiques disponibles en ligne, affinant ainsi continuellement son niveau d'expertise. Il représente un outil d'accès à une multitude d'informations permettant aux médecins d'aller rechercher aisément celles correspondant à leurs besoins et de mieux déterminer certaines de leurs décisions.

La deuxième fonctionnalité qui, dans un second temps, s'est développée est bien plus troublante au regard de l'histoire de la discipline. Watson, tout comme d'autres programmes similaires, se voit doté de la faculté d'établir des diagnostics, pouvant déceler des tumeurs de la peau, par exemple, avec des degrés de précision supposés supérieurs aux humains. Une compétence émerge, qui correspond exactement à la fonction alèthéique de l'intelligence artificielle, à savoir révéler des états de fait généralement dissimulés à nos esprits. Un franchissement de seuil s'opère, qui voit des systèmes identifier d'éventuelles pathologies et appelés à exercer leur savoir-faire dans différentes spécialités[2]. Ce qui notamment les caractérise, c'est qu'ils ne sont pas issus, à l'origine, des recherches menées par le monde médical, mais de celles entreprises par des acteurs industriels. Ce sont eux qui initient et développent les protocoles, tâchant ensuite de les faire acquérir par les centres d'examens et de soin. Ici, une forme de disjonction est à l'œuvre entre le monde technico-économique et celui de la médecine, qui ne travaillent pas de concert, au sein de partenariats, mais qui voit le premier chercher à imposer ses «innovations» au second.

À vrai dire, il n'est guère nécessaire de procéder à des assauts continuels en vue de conquérir le corps médical qui, à l'instar des autres corporations, est majoritairement soumis à la doxa de l'amélioration de tous ses secteurs grâce à leur «transformation digitale», devant dorénavant s'opérer toutes affaires cessantes. Car, la médecine, au lieu de défendre le champ propre de ses prérogatives, de témoigner d'une nécessaire distance critique, de faire valoir l'exigence d'une patiente évaluation avant d'adopter des techniques qui l'engagent, fonce tête baissée en quelque sorte, supposant que se joue là une évolution jugée inévitable qui appelle de s'y raccorder avec entrain pour le bien supposé de la pratique et des patients. Il va de soi que le sempiternel cliché de la complémentarité, devant en l'occurrence être à l'œuvre dans le cadre du diagnostic, est appelé à en rester à de vaines formules dans la mesure où ce qui est voué à prévaloir, c'est la vérité indubitable énoncée par les systèmes. L'hypothèse, certes coûteuse, d'un contre-diagnostic automatisé représenterait a minima un contrepoids susceptible de la relativiser.

L'aura, promise à être toujours plus éclatante, octroyée à l'alètheia algorithmique, trouve sa conséquence directe dans une troisième fonctionnalité : celle d'établir des prescriptions, soit en fonction des diagnostics établis par les humains, soit par les systèmes eux-mêmes. Jusqu'à peu, la rédaction d'une ordonnance relevait de la stricte compétence des médecins, dont la connaissance des molécules faisait partie de la formation et qui pouvaient, en cas de doute ou dans certaines circonstances, se référer à des ouvrages, tel le dictionnaire Vidal, édité en France depuis 1914, et aujourd'hui accessible en ligne, qui répertorie les caractéristiques des médicaments produits par les laboratoires pharmaceutiques. Ces glossaires, offrant des sommes d'informations, permettent, le cas échéant, au médecin de mieux se déterminer, son pouvoir de décision lui revenant in fine. Ce qu'induit cette disposition automatisée, c'est, une fois encore, la part de libre appréciation qui s'estompe au profit d'une vérité littéralement prescriptive qui s'impose à la conscience humaine. Des dispositifs, élaborés par des compagnies privées, occupent le point nodal situé entre le corps médical et le monde pharmaceutique. De nouveaux acteurs s'immiscent au sein de cette relation historique, qui n'entendent pas revêtir le statut de seuls intermédiaires, mais celui d'interlocuteurs dont est supposée bientôt dépendre la plus exacte conformité entre un diagnostic et les traitements aptes à y répondre. Les sociétés pharmaceutiques ont depuis longtemps mis en œuvre des méthodes destinées à peser sur les choix des médecins en vue de les inciter à élire leurs produits. S'il existe des chartes déontologiques, on sait qu'ils sont souvent invités, à grands frais, dans des congrès professionnels ou bénéficient de largesses sous diverses formes. Néanmoins, l'intégrité demeure une vertu cardinale de la profession. À terme, elle deviendrait vaine, car elle se déplacerait sur les concepteurs de systèmes qui, en cas de suspicion à leur égard, pourraient toujours arguer de leur bonne foi, «protégés» par l'«objectivité mathématique» des équations. Une «lutte industrielle de la prescription» est annoncée, chaque entreprise prétendant s'arroger le rôle de plate-forme incontournable liant les différentes parties prenantes entre elles. Mais au-delà de cette volonté d'occuper une position tierce, une stratégie parallèle se met en place, cherchant à se situer tout au long de la chaîne de la santé, entraînant la dissolution de la place centrale depuis toujours tenue par le corps médical et ne s'embarrassant plus de toutes les exigences historiques patiemment définies au cours du temps par la discipline.

Car une des visées majeures de l'industrie du numérique consiste à faire main basse sur le domaine de la santé, envisagé, avec ceux de la voiture autonome, de la maison connectée et de l'éducation, comme les plus décisifs et pour lesquels elle entend se doter de tous les moyens nécessaires afin d'asseoir, à terme, une domination sans partage. Cette ambition appelle l'adoption de plusieurs axes stratégiques devant s'emboîter les uns aux autres. Le premier, situé à la base en quelque sorte, exige de collecter les plus grands volumes de données émises par les corps. C'est la raison pour laquelle les smartphones intègrent maintenant des mécanismes mesurant le nombre de pas effectués au quotidien par exemple, tout comme les bracelets, les lits, les miroirs, les balances, tous connectés, parmi bien d'autres dispositifs, dorénavant destinés, outre à répondre à leurs fonctions premières, à capter nos flux physiologiques. Jusqu'à récemment, la récolte et l'analyse de toutes ces informations autori-

[2] Par exemple, une équipe de Stanford (États-Unis) a conçu un logiciel destiné à identifier les tumeurs malignes cutanées les plus fréquentes – les carcinomes – et les plus redoutables, les mélanomes. Le système a été alimente par une base de 130 000 images représentant plus de 2 000 pathologies de la peau. Une équipe chinoise a mis au point un programme capable de diagnostiquer, avec une efficacité prétendument égale à celle d'un ophtalmologiste, une maladie rare, la cataracte congénitale. Cf. Lise Loumé, «Une intelligence artificielle capable de détecter les cancers de la peau», Sciences et avenir, 8 février 2017.

saient la formulation d'offres personnalisées ressortissant du marché du bien-être, certes susceptible de générer des profits colossaux, mais dont il est estimé qu'il relève d'une «articulation naturelle» d'y adjoindre un autre à la portée tout aussi colossale : celui de la prévention et du soin thérapeutiques.

Le deuxième axe stratégique requiert de s'attacher les compétences des médecins et des biologistes afin qu'ils soient partie prenante dans l'élaboration des expertises automatisées qui se trouveront de surcroît parées du prestige de leur savoir certifié. Le technolibéralisme, tout comme il l'a fait avec les ingénieurs et les programmeurs, entend maintenant s'inféoder les compétences médicales en vue de les intégrer dans les départements de recherche mis en place depuis le début des années 2010 consacrés à la santé, tels Google Health et Calico (Alphabet/Google), HealthKit, CareKit, ResearchKit (Apple), ou nombre de start-up œuvrant dans les «biotechs». Ils ont notamment développé des applications de diagnostic sur smartphone, via la prise de la température, l'analyse des fréquences vocales et de la toux, celle du visage et à terme de la sudation, fournissant également des kits d'analyse de sang. Car l'enjeu consiste à bientôt faire sauter la phase de la consultation, de la rendre obsolète, pour instaurer une pratique du suivi continu, une sorte de veille perpétuelle prodiguée à chacun affranchie du paiement traditionnel à l'acte et fondée sur le principe de l'abonnement, garantissant une attention assidue en toute circonstance.

Ainsi, le champ d'intervention s'opère sans rupture de faisceau : les acteurs industriels collectent les états des personnes via les appareils connectés et les applications dédiées, proposent des produits et des services de bien-être, peuvent préconiser des examens complémentaires appelés à être réalisés, soit auprès de leurs propres officines, soit auprès de celles ayant acheté les mots-clés en rapport, et entendent boucler la boucle en proposant eux-mêmes, en compagnie d'éventuels partenaires dépendant du domaine de la santé, des traitements thérapeutiques. Le médecin, l'hôpital, ainsi que d'autres qualifications, sont appelés à être détrônés, voire marginalisés, par l'émergence de nouveaux «entrants» prétendant s'ériger comme les interlocuteurs les plus aptes à s'assurer de notre parfait suivi, à nous alerter, en quasi-temps réel, relativement à l'imminence de risques ou de pathologies et à se charger de notre meilleure prise en charge, quels que soient les cas de figure. Par exemple, la start-up Forward ambitionne de mettre au point le «cabinet médical du futur». Il est conçu comme un espace destiné à récolter les informations les plus exhaustives à propos de chaque «client» qui lors d'une première inscription se voit soumis à une batterie d'examens : relevé du poids, de la taille, de la température, du rythme cardiaque, de la pression artérielle, prise de sang, séance de scanner, prélèvement d'un échantillon de salive en vue de procéder à un test ADN chargé d'estimer les risques de cancers d'origine génétique. Les conversations entre patient et médecin sont scrutées par une intelligence artificielle supposée capable de relever des points méritant d'être rajoutés à l'établissement des profils. L'entreprise fournit des instruments connectés, bracelets, balances, moniteurs de sommeil, capteurs électrocardiogrammes. Les données recueillies sont étudiées à distance par des robots dédiés pouvant, le cas échéant, en cas d'évolution anormale, programmer dans les meilleurs délais une visite en vue de procéder à des analyses et des tests préalablement identifiés. La société offre, contre un abonnement, un nombre en théorie illimité de consultations, d'éventuelles vaccinations, l'accès à des nutritionnistes, ainsi que la fourniture de médicaments génériques. Son fondateur espère installer ces cabinets dans plusieurs villes des États-Unis et à l'étranger, affirmant «vouloir tout rebâtir de zéro[3]».

Google a récemment inauguré une unité spécialisée dans les biotechnologies, Verily, qui en 2017 a recruté 10 000 volontaires se voyant équipés de multiples capteurs et devant être régulièrement soumis à divers prélèvements, afin de suivre sur une durée de quatre années l'évolution de leur état de santé et de déterminer les biomarqueurs aptes à signaler les signes avant-coureurs de pathologies. Cette initiative témoigne de façon emblématique de la volonté de s'emparer du champ de la médecine et d'abattre les structures qui jusque-là la régissaient, grâce au suivi permanent de chacun autorisant une gestion hyperindividualisée et tendanciellement prédictive de nos vies. La filiale a par exemple élaboré des lentilles de contact capables de mesurer le taux de sucre dans le sang de diabétiques, ayant conduit à l'établissement d'un partenariat avec le groupe pharmaceutique suisse Novartis, lui permettant ainsi d'être présente au long des différents processus, celui du diagnostic, de la prescription et enfin de la fourniture de solutions thérapeutiques.

Bien davantage que la seule intensification de l'immixtion du régime privé dans le champ de la médecine, c'est une vaste entreprise de confiscation qui est visée par l'industrie des données. Néanmoins, elle ne revêtira pas cette apparence, du moins dans un premier temps, vu que c'est une administration des soins placée sous les sceaux des plus hautes réactivité et efficacité qui est promise à être dispensée, à l'écart des salles d'attente, des hôpitaux saturés, et préservée des éventuelles inattentions ou faillibilité des médecins et des erreurs corollaires. Il faut saisir la régression qui s'opère dans le cadre de la relation liant le corps médical et le patient, dans la mesure où c'est un nouveau type, non dit, de verticalité qui s'instaure, imposant une vérité objective des expertises et des préconisations, ayant valeur d'énoncés prescriptifs supérieurement qualifiés. Ils évacuent de facto l'hypothèse de la pluralité des compétences aptes à formuler des jugements, sur la base de leur savoir et de leur expérience, ou celle de contre-évaluations qui s'avèrent nécessaires dans certains cas de figure, autant que la concertation impliquant toutes les parties impliquées lors de la définition des traitements. Voit-on encore que cette logique d'hyperindividualisation, soutenue par l'usage de l'intelligence artificielle, est appelée à affaiblir, tant dans les faits que dans les esprits, l'exigence humaniste de la solidarité, au profit d'engagements établis entre personnes et organismes noués de gré à gré? La société du contrat est vouée à se généraliser jusque dans le secteur médical, au détriment de l'établissement d'un régime commun, conduisant à ce que les couches de population les plus aisées puissent prioritairement profiter des services offerts. De surcroît, les sommes d'informations récoltées alimentent une connaissance sans cesse approfondie des personnes susceptible d'être exploitée à diverses fins, principalement marchandes.

Ces nouvelles pratiques se développent sans que ni le

[3] Jérôme Marin, «Bienvenue dans le cabinet médical du futur», Le Monde, 31 mars 2017. [lemonde.fr/economie/article/2017/03/31/bienvenue-dans-le-cabinet-medical-du-futur$_5$103599$_3$234.*html*]

monde médical ni la société réagissent et se mobilisent à la hauteur des enjeux. La doxa de l'amélioration des diagnostics et des traitements s'impose, occultant les procédés qui se mettent en œuvre. Ces soudaines mutations devraient nous enjoindre à les identifier précisément et à agir si nous tenons à ce que la pratique reste, autant que possible, fondée sur le principe historique, et éthique, du soin n'appelant pas indéfiniment la contrepartie d'un profit. Nous manquons d'une théorie critique du devenir de la médecine. Il conviendrait, avant toute chose, de nous emparer de notre droit *d'opérer des tris*, et de considérer que le diagnostic automatisé, s'il peut présenter des avantages en certaines circonstances, ne devrait être utilisé qu'avec parcimonie, faute de quoi on ouvre la boîte de Pandore à ce que l'industrie se positionne tout au long de la chaîne de la santé. Nous accorder ce droit, ou plus exactement, nous obliger à ce devoir, suppose d'affirmer haut et fort que la prescription automatisée, au vu de toutes les conséquences qu'elle entraîne, représente un franchissement de seuil qui doit être jugé inacceptable.

Puisque l'acuité de ces questions est indifférente au législateur, pire, attendu que nombre de ces évolutions se trouvent appuyées par des dispositifs juridiques, résultant généralement d'un intense travail de lobbying et ne recherchant in fine que la croissance économique, alors c'est à tous les acteurs concernés, ceux défendant le principe d'une médecine devant être mise à l'abri de logiques de marché, autant qu'à la société civile tout entière, de faire valoir de salutaires positions divergentes. Il est temps de discriminer avec conscience les phénomènes, et de statuer en commun, qu'à partir du moment où, une fois encore, des techniques et des procédés nous dessaisissent de notre pouvoir de décision, ils doivent alors être tenus comme irrecevables. Le corps médical, s'il tient à rester fidèle à ses valeurs, va devoir apprendre à mener des luttes, tout comme les ouvriers surent le faire au cours de la révolution industrielle et des années 1970 par exemple. Non pas en usant du droit de grève en l'occurrence, mais en étant assez néoluddite en quelque sorte, en usant, dans certaines circonstances, du droit imprescriptible de refuser certaines méthodes dans l'objectif résolu de défendre la pérennité de principes considérés comme intangibles. Si rien n'est entrepris, nous assisterons à l'avènement d'une médecine dont la prétendue qualité dépendra de la puissance alèthéique des systèmes et de la capacité du monde privé à se doter de tous les moyens logistiques et persuasifs nécessaires en vue de s'ériger comme l'interlocuteur majeur, éradiquant ainsi, le temps de moins d'une génération, le socle humaniste sur laquelle elle s'est constituée depuis l'Antiquité. Probablement voyons-nous ici l'image grossissante de ce qui est à l'œuvre dans d'autres domaines, et puisqu'il s'agit d'un sujet hautement sensible, supposé particulièrement attirer notre attention, au moins peut-on espérer que notre naïveté béate généralisée s'atténuera. Dans le cas contraire, elle relèvera alors moins d'une ignorance que d'une coupable irresponsabilité au regard des convictions qui nous fondent.

Une brève histoire des sciences computationnelles

REVUE MÉDECINE ET PHILOSOPHIE

Christophe Gauld*
*Psychiatre, CHU de Grenoble,
Doctorant en philosophie de la médecine, Paris 1 / Lyon 3

RÉSUMÉ

La médecine actuelle peut s'ancrer dans un paradigme nommé « computationnel ». La notion de computation vient du latin *computatio*, qui renvoie au calcul (ou opération). Selon Turing, les sciences computationnelles correspondent, au sens large, à toute science du calcul. De ce fait, l'intelligence artificielle, vaste domaine scientifique largement résumé actuellement à l'apprentissage automatique (dont fait partie l'apprentissage profond, ou *deep learning*), est une science computationnelle. Ainsi, une part non négligeable de la médecine actuelle s'intègre dans le paradigme des sciences computationnelles, dont les enjeux mêlent à la fois la philosophie, la pratique clinique, les technologies numériques, les méthodes d'analyses de données, la modélisation des maladies et des comportements et l'étude des phénomènes physiologiques. Nous cherchons dans cet article à décrire le terme de « computationnel » en le réinsérant dans le contexte historique des sciences cognitives. Nous veillerons à ne détailler aucun enjeu normatif ni éthiques, en retraçant uniquement une historiographie du concept.

MOTS-CLÉS : Intelligence artificielle ; computationnalisme ; sciences cognitives ; fonctionnalisme.
DOI : 10.51328/100

Introduction

Nous allons tenter d'éclaircir le terme de « computationnalisme » tel qu'il a été développé au sein des sciences cognitives. Nous ne chercherons pas à retracer finement l'histoire de ce terme et nous ne tenterons donc pas de cerner l'intégralité du champ des sciences cognitives. De plus, nous ne discuterons pas non plus de son ancrage dans la cybernétique, dans la logique symbolique, au sein du béhaviorisme ou de la psychologie de la première moitié du XXe siècle. Enfin, nous n'utiliserons volontairement pas le terme « d'intelligence artificielle » qui regroupe un champ large et mal défini de techniques, d'approches scientifiques, de technologies et dont le sens s'est transformé au cours du temps pour désigner aujourd'hui, dans la plupart des cas, l'apprentissage automatique (machine learning). Nous appuierons ces définitions sur la notion de syntaxe et de sémantique, que nous allons définir.

La computation (du latin *computare*, calculer) correspond à la manipulation de données selon des règles.

Le terme s'applique aujourd'hui de manière ubiquitaire, à différents domaines scientifiques et pour désigner de nombreuses approches différentes (Miłkowski, 2018). De fait, il est rarement précisé comment le terme de « computation » a évolué au fil du temps, entraînant de nombreuses confusions. Par la suite, nous allons voir qu'il existe au moins deux principales formes de computationnalismes et qu'il existe bien souvent une confusion entre les « sciences computationnelles » et « le computationnalisme » (tout comme il peut y avoir une confusion entre le large champ des sciences cognitives et du cognitivisme) (Dietrich, 1994).

Computationnalisme cognitiviste *versus* connexionnisme

Nous allons détailler une première acception du terme computationnalisme, qui possède un fort ancrage historique.

Rappelons avant tout que le champ des « sciences cognitives » décrit l'intégralité des sciences qui s'occupent du fonctionnement du cerveau et des comportements humains, tandis que le terme « cognitivisme » n'est qu'un des courant (longtemps le principal avec le béhaviorisme, au moins des années 1930 aux années 1950) de ce champ.

Dans l'histoire des sciences cognitives, le computationnalisme est une forme de cognitivisme : il désigne le fait de traiter de l'information (en concevant que le cerveau effectue des « calculs »). Nous n'entrons pas dans le détail, mais notons que le cognitivisme est lui-même une forme de fonctionnalisme (donc le computationnalisme est aussi une forme de fonctionnalisme). Il est important de comprendre que ces différents courants se distinguent du fait qu'ils intègrent 1) une syntaxe, c'est-à-dire un ensemble de règles et 2) une sémantique, c'est-à-dire des représentations du monde (Piccinini, 2004). Notons également que le « symbole » est l'unité de base de la syntaxe, et qu'il est habituel de considérer qu'il « contient » du sens, donc de la sémantique. Une science qui manie des symboles manie donc une syntaxe et une sémantique. Ainsi, le computationnalisme comme le cognitivisme sont symboliques.

Cette première acception du terme « computationnalisme » réfère donc uniquement aux processus de calcul du cerveau.

Le connexionnisme désigne un courant des sciences cognitives qui fait référence aux réseaux de neurones. Il s'agit de comprendre le cerveau comme un ensemble d'entités interconnectés et dont le calcul émerge de l'intégralité de ce réseau (sans avoir à le décomposer ou à le localiser au sein du cerveau). Le connexionnisme ne souhaite pas renoncer à la représentation, mais il la considère comme distribuée et non pas dépendante d'une seule entité physique. Le connexionnisme est né dans les années 1950 et ce courant s'oppose fortement au cognitivisme. Il n'est pas un fonctionnalisme (Churchland, 1981). Le connexionnisme datant d'avant les années 1990 – 2000, est uniquement syntaxique, car il manie des règles mais il ne manie pas de représentations (de sémantique). Cette première forme de connexionnisme est ainsi nommée subsymbolique, car il s'applique à des plus bas niveaux que le symbole : il concerne les réseaux et non de supposés « contenus mentaux ».

À noter cependant que les modèles computationnels ne sont pas connexionnistes, qui nécessitent au moins deux fonctions spécifiques (la rétropropagation de l'erreur et le *parallel processing*).

Computationnalisme connexionnisme versus cognitivisme

Après les années 2000, le terme de computationnalisme « change de bord ». Avec notamment les travaux sur les réseaux convolutionnels (et l'arrivée de l'apprentissage profond, ou *deep learning*), le connexionnisme manie des représentations (c'est-à-dire devient sémantique en plus d'être syntaxique). On ne parle pas de « symbole » cependant, car le traitement de l'information se fait à travers des réseaux et non par manipulation de symboles (de supposés contenus mentaux). Du fait de traitement de l'information à la fois syntaxique et sémantique, cette deuxième forme de connexionnisme (ou « néo-connexionnisme ») devient une forme de computationnalisme (et donc un fonctionnalisme).

Pour le dire autrement, le terme de « computationnalisme » désignait un traitement de l'information basé sur des symboles (le « cognitivisme ») : du fait des progrès du connexionnisme, le terme de « computationnalisme » s'applique désormais au traitement de l'information qui est basé sur des réseaux (le « néo-connexionnisme »).

Mais pourquoi ce terme de computationnalisme a-t-il pu « changer de bord » à ce point ?

Computationnalismes, cognitivisme et connexionnisme ne sont pas au même niveau

En fait, le terme de « computationnel » est insuffisant pour désigner une approche à lui seul. Il ne se réfère qu'au traitement de l'information permettant la réalisation de la fonction (d'où son accointance avec le terme de « fonctionnalisme ») (Marr, 1982).

Or, il peut exister différents « mécanismes » sous-jacents à cette réalisation d'une fonction (par exemple, le vol d'un oiseau). Le niveau de la « computation » est le niveau de la fonction tandis que le niveau du « mécanisme » est le niveau des règles qui rendent possible cette fonction (ces règles peuvent être nommées « algorithmes »). Depuis les années 2000, ces règles sont le plus souvent des réseaux, donc « connexionnistes », dans la plupart des modèles computationnels actuels. Mais ces règles peuvent correspondre à n'importe quel autre mécanisme (par exemple celles d'un automate cellulaire).

Ce constat permet de comprendre que le niveau computationnel, de la fonction, est indépendant des règles qui permettent sa réalisation (les algorithmes), et donc de l'implémentation physique de ces algorithmes. Il existe donc de nombreuses réalisations physiques d'une même fonction, problématique appelée « réalisabilité multiple » en sciences cognitives.

Computationnalisme et sciences computationnelles

Il existe une autre confusion, régulièrement retrouvée dans la littérature, entre « computationnalisme » et « computationnel ». Les sciences computationnelles développent des modèles computationnels, dont le but est d'étudier des systèmes complexes grâce à l'utilisation d'algorithmes (et donc de modèles mathématiques formels), de nombreuses variables et de ressources informatiques.

Par exemple, les neurosciences computationnelles sont nées très tard dans l'histoire des sciences cognitives (que l'on pourrait faire remonter au début du pragmatisme, du fonctionnalisme, du béhaviorisme et de la machine de Turing aux alentours des années 1930). Les neurosciences computationnelles sont nées avec les modèles de neurones biologique avec un article séminal en 1907 (LIF) (Brunel et Van Rossum, 2017), et introduisant donc les systèmes dynamiques non linéaires (à noter que le terme de computationnalisme, qui n'avait rien à voir avec le cerveau jusque dans les années 1980, est en réalité né avec McCulloch et Pitts en 1943, qui s'intéressaient à la

plausibilité biologique des réseaux, dont les thèses ont été confirmées par Hebb). Les neurosciences computationnelles sont devenues connexionnistes lors de l'essor de ce dernier en 1980 (notamment à la suite de la Conférence de 1985 organisée par Schwartz qui a introduit le terme de « neurosciences computationnelles »). Après les années 2000, elles se rattachent à l'apprentissage profond (et donc aux réseaux de neurones artificiels), aux systèmes complexes, et intègrent les modèles bayésiens du codage prédictif de Friston (dès 2005). Après les années 2010, les neurosciences computationnelles se rapprochent de l'apprentissage par renforcement. À noter également que la société des neurosciences a été créée relativement tard, dans les années 1970.

Du fait de l'évolution des connaissances, les sciences computationnelles sont actuellement principalement axées autour d'approches connexionnistes (et non cognitivistes). Le terme de « computationnel » est donc appliqué à une science en particulier, et semble plus restreint que le « computationnalisme ». Il correspond cependant au cadre théorique du computationnalisme, puisque les sciences computationnelles expliquent les fonctions (computationnalisme) par des algorithmes (connexionnisme).

Les neurosciences computationnelles se distinguent fortement des neurosciences « classiques » (cognitivistes) parce qu'elles cherchent à expliquer comment se déroule le traitement de l'information (niveau computationnel), autrefois réservé à la psychologie (tandis que les neurosciences classiques s'occupaient des mécanismes à l'origine de ce traitement de l'information).

Conclusion

Ainsi, nous avons que le terme de computationnel est plus restreint que le computationnalisme et que ce dernier a connu au moins deux acceptions au fil du temps (un computationnalisme « cognitiviste – symbolique » et un second « connexionniste – subsymbolique »). Ainsi, le cognitivisme renvoie à la computation selon des représentations et des règles maniant des symboles (les contenus mentaux), tandis que le connexionnisme renvoie à la computation selon des réseaux de neurones et des règles inscrites elles-mêmes dans ces réseaux. Nous avons donc isolé deux formes de computationnalismes et deux formes de connexionnismes. Avec certains auteurs comme Piccinini (2009), on pourrait aller plus loin en affirmant que toute forme de traitement de l'information est nécessairement connexionniste (selon une troisième forme de connexionnisme, au sens large), puisqu'il fait intervenir des réseaux d'entités.

Enfin, nous avons vu que l'évolution de ces acceptions a fortement été influencée par le niveau algorithmique, permettant la réalisation de la fonction par le biais de différents mécanismes. Une telle constatation permet d'affirmer que la question du traitement de l'information doit nécessairement se pencher sur les mécanismes sous-jacents – autrement dit, que la psychologie devrait pouvoir comprendre les neurosciences.

RÉFÉRENCES

Brunel, N., Van Rossum, M. C. (2007). Lapicque's 1907 paper: from frogs to integrate-and-fire. Biological cybernetics, 97(5-6), 337-339.

Dietrich, E. (1994). Computationalism. In Thinking Computers and Virtual Persons (pp. 109-136). Academic Press.

Churchland, P. M., Churchland, P. S. (1981). Functionalism, qualia, and intentionality. Philosophical Topics, 12(1), 121-145.

Friston, K. (2005). A theory of cortical responses. Philosophical transactions of the Royal Society B: Biological sciences, 360(1456), 815-836.

McCulloch, W. S., Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4), 115-133.

Miłkowski, M. (2018). From computer metaphor to computational modeling: the evolution of computationalism. Minds and Machines, 28(3), 515-541.

Piccinini, G. (2004). Functionalism, computationalism, and mental contents. Canadian Journal of Philosophy, 34(3), 375-410.

Piccinini, G. (2009). Computationalism in the Philosophy of Mind. Philosophy Compass, 4(3), 515-532.

Intelligence artificielle et médecine : l'apport de la philosophie et de l'éthique de la technique de Gilbert Simondon

REVUE MÉDECINE ET PHILOSOPHIE

Brice Poreau*

*MD/PhDPhD, Médecin, chef de service, Responsable des prélèvements du Rhône, Établissement Français du Sang Auvergne-Rhône-Alpes
Chercheur associé, Laboratoire S2HEP, Université Claude Bernard Lyon 1.

RÉSUMÉ

L'utilisation de l'intelligence artificielle dans le domaine de la santé pose la question des limites de celle-ci : jusqu'où devons-nous employer l'intelligence artificielle, la médecine pourrait-elle être déshumanisée et perdre son sens *princeps* ? Le philosophe Gilbert Simondon (1924-1989) propose durant la seconde moitié du vingtième siècle une philosophie de la technique. Il s'agit de repenser la technique, non pas comme aliénation, mais comme possibilité d'évolution, avec le développement de la technique. Cet article présente ainsi l'intelligence artificielle comme une technique que l'humain doit s'approprier afin de limiter les risques de déshumanisation et d'optimisation du potentiel de l'intelligence artificielle. La pensée de Gilbert Simondon est alors un outil philosophique et éthique permettant d'intégrer l'intelligence artificielle dans le domaine de la médecine, en conservant tout son sens éthique.

Summary

The use of artificial intelligence in the field of health raises the question of its limits: how far should we use artificial intelligence, could medicine be dehumanized and lose its original meaning? The philosopher Gilbert Simondon (1924-1989) proposed during the second half of the twentieth century a philosophy of technique. It is about rethinking technique, not as alienation, but as a possibility of evolution, with the development of technique. This article thus presents artificial intelligence as a technique that humans must appropriate in order to limit the risks of dehumanization and optimization of the potential of artificial intelligence. The thought of Gilbert Simondon is then a philosophical and ethical tool allowing the integration of artificial intelligence in the field of medicine, keeping all its ethical sense.

MOTS-CLÉS : éthique, intelligence artificielle, Gilbert Simondon, technique, médecine, Norbert Wiener.
DOI : 10.51328/109

Introduction

L'intelligence artificielle (IA) est un outil dont les débuts sont initiés dans les années 1940-1950, et qui prend un essor majeur dans les années 1970-1980 avec l'informatisation. L'un des objectifs est de permettre une simulation de l'intelligence humaine, comme l'apprentissage. Avec l'avènement de l'informatique, puis des nouvelles techniques de l'information et de communication, l'intelligence artificielle prend une place de plus en plus importante dans de nombreux domaines, dont

celui de la médecine. L'utilisation de données de masses nécessite par exemple une puissance de calcul exponentielle dont l'IA est en capacité de traiter (Hamet; Tremblay, 2017 ; Topol, 2019). Le développement de l'IA est également prégnant dans le cadre d'une médecine dite de précision (Schork, 2019).

Pour autant, l'un des risques qui est posé est l'utilisation d'une intelligence artificielle qui déposséderait le caractère humain de la médecine.

Durant la seconde moitié du vingtième siècle, Gilbert Simondon (1924-1989) expose une philosophie de la technique et une philosophie du vivant, utilisant les travaux de Norbert Wiener (1984-1964), qui nous apparaît tout à fait pertinente pour traiter la question éthique de l'humain dans l'IA comme outil.

En effet, Gilbert Simondon développe cette philosophie dans un contexte de l'émergence de l'informatique. Il y voit clairement un avenir majeur pour cette discipline. Mais, à l'instar d'aujourd'hui, se pose la question des limites de l'utilisation de cette technique. De plus, Gilbert Simondon a une approche très vaste. Celle-ci ne se limite pas à la technique, en effet, il crée une forme d'encyclopédisme (Barthélémy, 2016) qui revêt un caractère encore plus marquant actuellement avec l'utilisation de l'IA dans le domaine du vivant.

Partie 1 : Nobert Wiener et la cybernétique au centre de la philosophie de Simondon

Gilbert Simondon (1924-1989) est un philosophe français, ancien élève de l'École Normale Supérieure de 1944 à 1948, agrégé de philosophie et professeur de philosophie à Tours durant la première partie de sa carrière (Poreau, 2016). Il termine sa thèse d'État de philosophie en 1958. Celle-ci se focalise notamment sur le concept d'individuation, concept central de ses travaux (Chateau 2008 ; Barthélémy, 2016) et permet la création d'un lien entre la physique, la biologie et la psychologie (Simondon, [1964, 1989], 2005). Les thèses d'État comportaient alors également une thèse dite mineure. Dans le cas de Simondon, elle porte sur les objets techniques (Simondon, [1958] 2008).Gilbert Simondon exerce ensuite à Poitiers jusqu'en 1963 comme assistant professeur, puis professeur à la faculté des Sciences et Lettres. Il est nommé professeur à la Sorbonne et fonde le laboratoire de psychologie générale et de technologie dans les années 1980 (Chatelet 1994). Il donne des cours sur la communication, l'information et la perception (Simondon 2006, 2010).

Concernant le concept central de ses travaux, l'individuation (Fagot-Largeault, 1994 ; Perru, 2005 ; Poreau, 2013, 2016), voici ce qu'il définit : «L'individuation correspond à l'apparition de phases dans l'être qui sont des phases de l'être ; elle n'est pas une conséquence déposée au bord du devenir et isolée, mais cette opération même en train de s'accomplir » (Simondon, 2005). Ces phases sont liées à la fois dans les domaines physique, biologique et psychique, mais également les objets techniques. Le lien est notamment l'information. Sa philosophie, une forme d'encyclopédisme (Barthélémy, 2016), vise à dépasser les notions initiales de la cybernétique développées par Nobert Wiener (1894-1964) dans les années 1940. Comme le mentionne Jean-Hugues Barthélémy : «[Norbert Wiener] le mathématicien et père fondateur de la cybernétique est indéniablement l'interlocuteur central de l'ensemble de l'œuvre proprement philosophique de Simondon [...] » (Barthélémy, 2016). Les travaux publiés en 1948 par Wiener Cybernetics, or control and communication in the animal and the machine, mettent en exergue la théorie du contrôle ou de la commande (boucle de rétroaction), qui est lient ainsi le biologique, le physique et le psychique, comme le reprend alors Simondon, en en dépassant le concept originel. Il le mentionne dans son ouvrage Du mode d'existence des objets techniques : « La fonction dont nous tentons de tracer les grandes lignes serait celle d'un psychologue des machines, ou d'un sociologue des machines que l'on pourrait nommer le mécanologue. On trouve une esquisse de ce rôle dans l'intention de Nobert Wiener fondant la cybernétique, cette science de la commande et de la communication dans l'être vivant et la machine» (Simondon, 1958 [2008]). Il s'agit bien ici de lier le vivant et la technique, et de dépasser la vision initiale de Norbert Wiener : « Cependant, même si cette opposition du déterminisme divergent au déterminisme convergent ne rend pas compte de toute l réalité technique et de son rapport avec la vie, cette opposition contient en elle toute une méthode pour découvrir et pour définir un ensemble de valeurs impliquées dans les fonctionnements techniques et dans les concept au moyen desquels on peut les penser. Mais il est possible d'ajouter un prolongement à la réflexion de Nobert Wiener » (Simondon, 1958 [2008]).

Le développement de la physique, de la biologie et de la technique incluant la cybernétique puis l'informatique, fait de Simondon un philosophe dont la vision est très vaste : il étudie au-delà de chacune des disciplines dont certaines sont naissantes et montre dans son œuvre la nécessaire interdépendance entre elles. La transdisciplinarité est alors un atout indéniable pour lui permettre de revisiter la technique, bien au-delà d'une dichotomie de l'humain et de la technique. Ainsi, une philosophie, voire une éthique simondonienne émerge et peut contribuer à l'analyse de l'intelligence artificielle contemporaine, qui est issue en partie des travaux de Norbert Wiener. L'application à la médecine de l'intelligence artificielle nécessite une philosophie et une éthique, dont l'approche de Gilbert Simondon peut donc nous aiguiller.

Partie 2 : définition d'une éthique simondonienne de l'intelligence artificielle en médecine

L'intelligence artificielle est une technique. Développée voici plusieurs décennies, elle a évolué. Cependant, un risque identifié est que cette technique puisse dépasser l'homme, voir rendre l'humain non humain. Il est donc nécessaire de repenser la technique dans son ensemble. Simondon dans son approche philosophique, expose la nécessité de repenser la technique, à l'instar du vivant : « Il y a quelque chose de vivant dans un ensemble technique, et la fonction intégratrice de la vie ne peut être assurée que par des êtres humains ; 'être humain a la capacité de comprendre le fonctionnement de la machine, d'une part, et de vivre, d'autre part : on peut parler de vie technique, comme étant ce qui réalise en l'homme cette mise en relation des deux fonctions « (Simondon, 1958 [2008]). Il dépasse ainsi la dichotomie entre une technique, inventée par l'homme, et dont le contrôle lui échapperait, et l'humain. « L'homme est capable d'assumer la relation entre le vivant qu'il est et la machine qu'il fabrique ; l'opération technique exige une vie technique et

naturelle » (Simondon, 1958 [2008]).

Cette vision s'applique directement à l'intelligence artificielle (IA), et d'autant plus dans le domaine de la médecine. En effet, si nous reprenons les exemples contemporains d'utilisation de l'IA dans le traitement des données de masse, comme en génétique, il y a bien une vie technique et naturelle qui entoure le traitement de ces données, id est la nécessaire interprétation des résultats obtenus par l'application de l'IA, sans laquelle il n'y a plus aucun sens à l'IA.

Plus qu'une approche philosophique, c'est une approche éthique qui émerge de la pensée de Simondon. En effet, Simondon dépasse les dichotomies entre technique/humain, mais également entre culture/technique et humain/nature. L'interdépendance est totale. Ainsi, il présente la comparaison de l'étranger, quant à l'intégration de la technique dans la culture, la nature et l'humain. Il explique ainsi : « C'est la notion de machine qui est déjà faussée, comme la représentation de l'étranger dans les stéréotypies du groupe. Or, ce n'est pas l'étranger en tant qu'étranger qui peut devenir objet de pensée cultivée ; c'est seulement l'être humain. Le stéréotype de l'étranger ne peut être transformé en représentation juste et adéquate que si le rapport entre l'être qui juge et celui qui est l'étranger se diversifie, se multiplie pour acquérir une mobilité multiforme qui lui confère une certaine consistance, un pouvoir défini de réalité. Un stéréotype est une représentation à deux dimensions, comme une image, sans profondeur et sans plasticité » (Simondon, 1958 [2008]).

L'intelligence artificielle, une technique, un outil, notamment utilisé dans le domaine de la médecine, doit donc retrouver la profondeur et la plasticité indispensables pour concrétiser son lien avec l'humain. La spécificité de l'application de l'IA à la médecine est ce lien encore plus fort avec l'humain, centre de la médecine. Mais il s'agit d'un tout. La philosophie de Simondon montre que c'est une véritable relation qui doit être comprise comme telle entre la technique, dans notre cas, l'intelligence artificielle, dans un but qui est celui défini par l'homme et l'humain. Pour cela, la relation doit être objectiver : « Pour que la représentation des contenus techniques puisse s'incorporer à la culture, il faut qu'existe une objectivation de la relation technique pour l'homme » (Simondon, 1958 [2008]). Au-delà d'une philosophie de la technique, c'est une éthique que Simondon esquisse, normalisant ainsi le but, définissant un cadre, par la nécessité de cette objectivation. C'est une éthique du progrès qu'il définit, alors que les prémisses de l'IA sont créées (Simondon, 1959).

Gilbert Hottois confirme l'éthique chez Simondon qui raisonne encore plus aujourd'hui : « Le sens de celle-ci est. en fait, constitutif de l'entreprise simondonienne. Le sens moral étant le sens de l'individuation, c'est, en définitive, pour des raisons éthiques que l' élaboration d'une culture technique est, aujourd'hui, un devoir » (Hottois, 1993).

Comment, alors que l'utilisation de la technique qu'est l'IA augmente de façon très importante ces dernières années, peut-on mettre en place l'éthique simondonienne d'objectivation de la relation technique pour l'homme ? La réponse apportée par Simondon est l'éducation.

Conclusion

La nécessaire éducation à l'outil IA en général et en médecine en particulier Charbonnier relève en effet dans ses travaux de didactique la nécessité de l'éducation à l'objet technique. Simondon, très tôt dans sa pensée philosophique et éthique, défend ce point de vue : « La puissance synoptique de Simondon replace le projet d'un enseignement de la technique au sein d'une conception globale de la « formation humaine ». Il s'inscrit dans la lignée rare des défenseurs d'une vision politique et humaniste de l'enseignement de la technique au regard des enjeux sociétaux. Sa sensibilité et son intelligence de la technique l'amènent à vouloir lutter contre un déni culturel de la place du « manuel » et de la technique dans l'École » (Charbonnier, 2017).

Ainsi, dès la fin des années 1940 durant ses recherches et durant les années 1950, Simondon identifie le potentiel de la technique qui se développe et qui deviendra l'intelligence artificielle contemporaine. L'IA est employée dans de nombreux domaines. La relation de cette technique doit alors être objectivée pour comprendre cette relation à l'humain. Le domaine médical nécessite absolument, du fait de l'objet même de la médecine, cette objectivation, cette individualisation, comme le décrit Simondon pour les objets techniques, de l'IA.

C'est donc un cadre, une éthique qui s'accomplit autour de l'IA afin d'éviter tout risque d'utilisation au-delà de ce cadre. Plus que cela, c'est grâce à l'éducation, dès le plus jeune âge à l'IA, que cette relation sera possible, puisque c'est en comprenant les possibilités, le potentiel de l'outil IA que l'application, pensée par l'homme, pourra être appropriée, sans dérive. L'intelligence artificielle a un potentiel immense actuellement. Ce potentiel ne peut se concrétiser sans une approche philosophique d'une part, et éthique d'autre part. Simondon l'a bien perçu. Contemporain de Wiener et des premiers scientifiques ayant initié l'IA, il est un auteur indispensable pour permettre de consolider la philosophie et l'éthique de l'IA.

Comme pour de nombreuses questions éthiques, l'éducation est probablement la première action à mettre en place. Cette éducation est une transmission des connaissances, une transmission des clés de compréhension de l'objet technique, de tout outil, comme l'IA. Avec cette connaissance, cette compréhension, ce sont aussi les limites qui sont décrites et le cadre qui est posé. La technique évolue. Elle évolue actuellement rapidement dans le domaine de l'intelligence artificielle. En reprenant la philosophie de Simondon, il nous faut alors dépasser la dichotomie entre technique (IA) et humain. L'IA procède l'humain. Plus qu'un outil, c'est bien son utilisation, et essentiellement, la relation qui se crée qui doit être comprise dans sa globalité. Enfin, en reprenant l'éthique de Simondon, c'est bien l'ouverture d'esprit, l'esprit critique au sens d'une compréhension et d'une remise en cause de stéréotypies, qui fixent, avec l'éducation la relation fondamentale entre l'humain et l'IA, dont l'humain est l'essence même.

RÉFÉRENCES

Barthélémy, J.-H., *Simondon*, Les Belles Lettres, Paris, 2016.

Charbonnier S., « Présentation de l'article de Gilbert Simondon », *Recherches en didactiques*, 2017/1

(N° 23), p. 133-141. DOI : 10.3917/rdid.023.0133. URL : https://www.cairn.info/revue-recherches-en-didactiques-2017-1-page-133.htm

Chateau J.-Y., *Le vocabulaire de Simondon*, Ellipses, Paris, 2008.

Chatelet G., Gilbert Simondon. *Une pensée de l'individuation et de la technique*, Albin Michel, Paris, 1994.

Fagot-Largeault A., *L'individuation en biologie*, Gilbert Simondon. Une pensée de l'individuation et de la technique, Albin Michel, Paris, 1994.

Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism.* 2017 Apr;69S:S36-S40. doi: 10.1016/j.metabol.2017.01.011. Epub 2017 Jan 11. PMID: 28126242.

Hottois, G, *Simondon et la philosophie de la culture technique*, De Boeck université, Bruxelles, 1993.

Perru O., L'individuation chez Gilbert Simondon, *Bulletin d'Histoire et d'Epistémologie des Sciences de la Vie*, 2005, 12,159-172.

Poreau B., *De Pierre-Joseph Van Beneden à l'écologie et la médecine contemporaines : l'histoire du commensalisme.* Vrin, Paris, 2016.

Poreau B., *Simondon, philosophe du vivant ?*, Editions B. Poreau, collection Développons, Lyon, 2013.

Russell S, Norvig P. *Intelligence artificielle*. Pearson. 3Ème édition. Montreuil, France, 2010.

Schork NJ. Artificial Intelligence and Personalized Medicine. *Cancer Treat Res.* 2019;178:265-283. doi: 10.1007/978-3-030-16391-4$_1$1.*PMID* : 31209850; *PMCID* : *PMC*7580505.

Simondon G., *Du mode d'existence des objets techniques*, Aubier, Paris, 1958 [2008].

Simondon G., Les limites du progrès humain, *Revue de métaphysique et de morale*, 1959, 59 (3): 370-376.

Simondon G., *L'individu et sa genèse physico-biologique*, PUF, Paris, 1964.

Simondon G., *L'individuation psychique et collective*, Aubier, Paris, 1989.

Simondon G., *L'individuation à la lumière des notions de forme et d'information*, J. Millon, Grenoble, 2005.

Simondon G., *Cours sur la perception*, Editions de la transparence, Chatou, 2006.

Simondon G., *Communication et information, cours et conférences*, Editions de la transparence, Chatou, 2010.

Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019 Jan;25(1):44-56. doi: 10.1038/s41591-018-0300-7. Epub 2019 Jan 7. PMID: 30617339.

REVUE MÉDECINE ET PHILOSOPHIE

#4(2) / 2020 - 2021

Philosophie de l'intelligence artificielle

REVUE MÉDECINE ET PHILOSOPHIE

ISSN : 2650-5614