

Convergence entre intelligence artificielle et neurotechnologies : une réalité et un futur souhaitable ?

Par Alexis GÉNIN
Institut du cerveau (ICM)

De grandes entreprises et acteurs étatiques ont fait des neurotechnologies un champ d'action aussi stratégique que la conquête spatiale. Effectivement, l'accélération de la miniaturisation des technologies implantables et la puissance prédictive des outils d'intelligence artificielle convergent pour rendre prochainement possibles des développements proches de ceux prédits par les romans d'anticipation. Dans ce contexte, les développeurs de technologies médicales sont eux-mêmes dépassés par de nouveaux entrants pour qui le malade n'est plus qu'un « client » comme un autre aux côtés des soldats et des hommes pressés à la recherche d'augmentation de leurs performances. Faut-il suivre ce chemin techno-centré et tenter, avec des moyens financiers plus faibles, de copier l'approche des États-Unis et de la Chine ? Ou faut-il innover vraiment avec notre recherche en biologie et le dynamisme de nos jeunes entreprises innovantes, pour oser la voie différente, plus durable, du développement de neurotechnologies responsables ?

Parler de neurotechnologies est devenu un *must*. *Neuro-coaching*, *neuro-marketing*, *neuro-design*, l'abondance des néologismes en vogue est révélatrice de l'importance que prend tout moyen d'intervention nouveau et externe dans le domaine mystérieux de nos pensées.

La question de la convergence de cette (de ces) neurotechnologie(s) avec les outils dits d'intelligence artificielle pose en premier la difficulté d'adresser simultanément deux des plus grands poncifs du début des années 2020 en évitant les sables mouvants des lieux communs. Cette première épreuve n'est malheureusement que le prélude à une seconde, la question de savoir si ces neurotechnologies peuvent encore être abordées sous le double angle du savoir biologique et de l'entrepreneuriat responsable, ou si les actions bruyantes des géants de la *Big Tech* ont déjà fait de ces considérations des vieilleries démodées.

Car, à la différence des autres secteurs de la santé (au sens large), dans lesquels les acteurs industriels agissent en s'appuyant sur un socle de culture scientifique et médicale orienté vers les malades en tant qu'objets biologiques complexes, dans celui des neurotechnologies, de puissants entrants issus du commerce en ligne, des réseaux sociaux et du monde des véhicules électriques brisent tous les cadres avec la promesse hâtive de « libérer la puissance du cerveau ». Vite, bien sûr, car il s'agit dans

chaque cas du rêve d'hommes pressés. Les neurotechnologies seraient, pensent-ils, la clé pour sortir de cette grotte platonicienne dans laquelle l'humanité est enfermée du fait de ses limites biologiques ; secondairement, et de façon moins affichée, elles doivent aussi devenir de puissants leviers pour accroître les ventes de leurs produits et monétiser de nouvelles sources de données collectées sur les consommateurs.

Si la faisabilité de ces promesses reste incertaine, leurs moyens de mise en œuvre ne sont plus totalement une science-fiction. Le considérable essor de la puissance informatique et la miniaturisation croissante des technologies ont fait entrer la cybernétique dans la réalité médicale : des membres artificiels peuvent déjà – avec beaucoup d'entraînement et des résultats variables – être pilotés par des électrodes implantées dans le cortex moteur ; des robots semi-autonomes à visage humain exploitent les connaissances en neurosciences pour mimer l'existence d'une conscience comparable à la nôtre, reproduire des émotions, dialoguer avec esprit. Depuis plus de quarante ans, la stimulation intracérébrale profonde rend aux malades de Parkinson la capacité de marcher, et celle du nerf vague la capacité de maîtriser notre rythme cardiaque. Des publications scientifiques récentes laissent même espérer que le volume de l'hippocampe, une des premières régions cérébrales affectées par la maladie d'Alzheimer, pourrait être préservé par la stimulation intracérébrale de la région du



Une personne paraplégique équipée d'un exosquelette retrouve la possibilité de se mouvoir.

« Une des observations les plus frappantes, par exemple, des premiers exosquelettes pilotés par l'activité cérébrale de patients tétraplégiques a été de constater que le fait pour le malade de s'observer en train de marcher "comme si" sa moelle épinière n'avait pas été sectionnée induit une réactivation des processus de repousse neuronale. »

fornyx⁽¹⁾. Ceci ferait pour la première fois d'une neurotechnologie un « disease modifier » n'agissant pas seulement sur l'expression symptomatique de la maladie mais aussi sur son substrat organique.

De la même façon, les approches non invasives dites de « neuro-feedback », qui permettent de visualiser et d'apprendre à contrôler une partie de l'expression de son activité cérébrale (plus précisément, du reflet partiel de celle-ci dans certaines zones des premières couches du cortex cérébral), peuvent être des aides précieuses pour mettre le malade – ou le sujet sain – en capacité d'agir lui-même sur son état de santé. La démonstration expérimentale réussie en a été faite, par exemple, pour le traitement des troubles de l'attention chez l'enfant⁽²⁾.

Ces développements technologiques, qui ont jusqu'à maintenant peu eu besoin de mobiliser la puissance prédictive des outils d'intelligence artificielle (IA), sont donc évidemment une extraordinaire opportunité d'amélioration de la médecine. De manière tout à fait intrigante, beaucoup des

résultats obtenus dévoilent des capacités de plasticité du tissu cérébral qui avaient été largement sous-estimées. Une des observations les plus frappantes, par exemple, des premiers exosquelettes pilotés par l'activité cérébrale de patients tétraplégiques a été de constater que le fait pour le malade de s'observer en train de marcher « comme si » sa moelle épinière n'avait pas été sectionnée induit une réactivation des processus de repousse neuronale⁽³⁾. La technologie a ici servi de déclencheur à un processus biologique quiescent, a révélé un potentiel. Un champ immense de la médecine reste donc à explorer sur la possibilité, notamment par des outils technologiques, de stimuler les capacités endogènes innées de l'humain à s'auto-réparer.

Les dérives éthiques des visions « techno-centrées »

La convergence des neurotechnologies et de la puissance prédictive des outils d'IA pourrait, théoriquement, être, elle aussi, un puissant outil au service de la médecine et de l'humain. La réalité observée aujourd'hui est malheureusement que la gigantesque opportunité économique sous-

(1) *Brain Res.* 2019, Jul 15; 1715:213-223, doi: 10.1016/j.brainres.2019.03.030. *Epub* 2019, Mar 26. *Brain Stimul.* May-Jun 2015;8(3):645-654.

(2) *Translational Psychiatry*, vol. 8, article number: 149 (2018).

(3) "Tetraplegia", *Nature Scientific Reports*, vol. 6, article number: 30383 (2016).

jacente à ces « neurotechnologies », ainsi que – parfois – une vision du monde dans laquelle l'être humain tel que nous le connaissons doit disparaître, sont une force puissante qui risque de nous diriger vers des applications bien moins souhaitables. Pour arriver à un post-humain à l'existence biologique réduite, le champ et les outils de la neurotechnologie ne sont plus vus comme des outils de réparation de l'humain blessé ni comme des moyens transitoires de « remise en capacité. » Ils deviennent une ambition de fusion de l'homme avec la technologie, par le recours à de larges faisceaux de micro-électrodes implantées, utilisables comme des outils de dopage cérébral pour acquérir un plus haut QI, une meilleure résistance à la fatigue et au stress, éprouver moins d'émotions, et pourquoi pas permettre une connexion directe du cerveau à Internet... Pour certaines entreprises de la première puissance mondiale, c'est une étape logique avant la conquête de Mars pour les acteurs gouvernementaux de la seconde, ces technologies sont vues comme un outil pour assurer sa domination mondiale, comme la conquête spatiale l'a été pendant la Guerre Froide. Si le site de la DARPA américaine n'évoque publiquement que la réparation des soldats blessés sur le terrain et le pilotage de drones *via* une interface EEG non implantée, du côté de l'armée chinoise, les programmes sur les « technologies d'augmentation de la performance humaine » et « d'intelligence hybride homme-machine » se multiplient de manière affichée⁽⁴⁾.

Chaque année, des centaines de millions de dollars sont investis par Neuralink, Facebook, Google ou les BATX chinois pour développer des solutions neurotechnologiques optimisées par des algorithmes IA. Au vu des investissements, et même si les premières démonstrations publiques sagement scénarisées ne dépassent pas de beaucoup l'état de l'art scientifique connu, le mythe d'une technologie cérébrale implantée, adaptable et autonome devient un futur réaliste. Il reste à se demander si celui-ci est également un futur souhaitable.

Implications de la convergence entre IA et *hard neurotechs*

Pour donner leur place à tous les arguments, il faut bien sûr rappeler que l'amélioration de soi par des outils externes fait partie de l'histoire du développement des sociétés : l'écriture a permis d'extérioriser et de transmettre pensées et souvenirs ; les Romains comptaient sur leurs esclaves pour mémoriser les faits utiles ; aujourd'hui, les *pacemakers*, les *smartphones*... sont aussi des augmentations des capacités du soi. Mais agir directement et physiquement sur le substrat de nos pensées – et donc de notre individualité – est une frontière qui n'a pas encore été franchie. C'est peut-être d'ailleurs la dernière, celle qui une fois franchie pourrait voir le concept de liberté – d'agentivité diraient les relativistes – disparaître. Outre ce danger de la déshumanisation, quatre autres vigies sont à dresser lorsque le sujet des neurotechnologies est

abordé, pour surveiller les quatre dimensions que sont la biologie, l'impact environnemental, la liberté individuelle et l'égalité d'accès.

La biologie, tout d'abord, nous enseigne que le cerveau est un organe dynamique, auto-régulé et en recherche constante d'optimisation énergétique, puisqu'il consomme à lui seul près de 20 % de notre énergie – pour 2 % du poids de notre corps. Si une technologie complète ou supplée un aspect quelconque de son fonctionnement, les zones cérébrales concernées sont alors appelées à se mettre au repos, voire à perdre progressivement leur capacité d'action autonome. Une première illustration frappante, suivie de bien d'autres, en a été donnée, il y a une vingtaine d'années, quand une étude de neuro-imagerie a permis de constater que le volume de l'hippocampe des chauffeurs de taxi londoniens avait diminué significativement avec l'arrivée des GPS⁽⁵⁾. L'hippocampe est une structure-clé des processus de mémorisation, en particulier de la mémoire des lieux. "Use it, or lose it", nous répète avec insistance la biologie. L'homme dit « augmenté » par les neurotechnologies risque d'être en réalité un homme profondément diminué qui aura perdu, si cette technologie faillit un jour, les capacités que celle-ci était censée stimuler.

Le second point pour ne pas faire de la convergence entre les neurotechnologies et l'IA – dans leur version présentée par les géants de la *Big Tech* – un futur souhaitable est celui de la consommation d'énergie. Le site OpenAI montre, par exemple, qu'entre 2012 et 2018, la consommation énergétique liée au travail de calcul des modèles d'IA a été multiplié par 300 000, soit un doublement tous les trois mois et demi⁽⁶⁾. Dans un article de 2019, intitulé "Energy and Policy Considerations for Deep Learning in NLP" et publié par le *MIT Technology Review*, Emma Strubell et ses collègues du département des sciences informatiques de la Massachusetts University ont par ailleurs estimé que le seul entraînement d'un modèle d'IA de type *natural language processing* a un impact énergétique correspondant à l'émission de 280 tonnes de CO₂ – l'équivalent des émissions de cinq voitures sur l'ensemble de leur durée de vie⁽⁷⁾. Or, d'après Nvidia, 80 à 90 % du coût énergétique d'un *neural network* informatique mobilisé en IA sont liées à son implémentation plutôt qu'à son entraînement. Appliqué aux neurosciences et dans un schéma à 1 000 électrodes comme celui promu par Neuralink, avec les fréquences d'acquisition habituellement utilisées pour de l'électrophysiologie fine, ce serait entre 1 000 et 3 000 Go de données qui seraient générées par jour et par sujet⁽⁸⁾. Pour assurer les *neuro-feedbacks* adaptés – c'est-à-dire rapides et continus –, cette masse de signal doit être traitée à très haute fréquence, en recourant au calcul inten-

(4) <https://www.eurasiareview.com/23022020-minds-at-war-chinas-pursuit-of-military-advantage-through-cognitive-science-and-biotechnology-analysis/>

(5) MAGUIRE Eleanor A. *et al.* (2000), "Navigation-related structural change in the hippocampi of taxi drivers", *PNAS* 97 (8), pp. 4398-4403, April 11, <https://doi.org/10.1073/pnas.070039597>

(6) <https://openai.com/blog/ai-and-compute/#modern>

(7) <https://arxiv.org/abs/1906.02243> ; <https://www.semanticscholar.org/paper/Evaluating-the-Energy-Efficiency-of-Deep-Neural-on-Li-Chen/6cd61b00eb6248ead7ea029d2daa9d79e68572cb>

(8) Données Institut du cerveau, Paris.

sif. Cela laisse imaginer, de manière peut-être simpliste, qu'une personne à qui l'on aura implanté une neurotechnologie de ce type deviendra à elle seule un problème écologique d'ampleur.

En pleine ère des « cyber » guerres, il est également important de constater que la puissance de calcul nécessaire à ces interfaces implique la mobilisation de relais informatiques distants qui vont devoir recevoir, traiter et restituer l'information sous la forme de commandes données au cerveau. Ces implants cérébraux seront donc structurellement actionnables, contrôlables depuis l'extérieur, en un mot *hackables*. Les neurotechnologies implantées, en prenant racine au cœur de nos processus de pensée et de décision, vont ainsi créer des portes d'entrée géantes pour analyser nos pensées, les monétiser sur le grand marché de la donnée et, potentiellement, les modifier à loisir dans leur substrat même.

Un dernier point appelle à faire preuve de vigilance vis-à-vis de ces *hard* neurotechnologies : à la différence de celles que l'on pourrait appeler « raisonnées », non invasives et qui – en visualisant la surface de l'iceberg de notre activité cérébrale – permettent d'apprendre à contrôler par soi-même certains processus mentaux, ces futures neurotechnologies implantées, qui ne pourront se déployer que par une convergence avec des outils d'IA, seront financièrement inabordables pour la majorité d'entre nous. À titre d'exemple, un dispositif actuel de stimulation intracérébrale profonde, alors qu'il ne comporte que 4 à 16 points de stimulation électrique sur une zone unique du cerveau et pâtit en outre d'une très faible adaptabilité des rythmes de stimulation, coûte près de 200 000 euros par patient sur une période de cinq ans⁽⁹⁾ (dont 50 à 80 000 euros pour la seule opération d'implantation du système.) Les systèmes à 1 000 électrodes annoncés avec la promesse d'apporter force et intelligence ne pourront donc être acquis d'abord que par les plus fortunés de nos concitoyens, puis principalement par le monde dit développé. Une profonde inégalité diviserait donc l'humanité entre, d'un côté, des « techno-dopés » capables de plus – la capacité à agir « mieux » étant encore une question – et les autres.

(9) *Cost DBS: Mov Disord* 2013, Jun. 28(6):763-71, doi: 10.1002/mds.25407. Epub 2013, Apr 10.

Retrouver la logique du vivant

Face à ces rêves d'augmentation quantitative de l'humain et aux visions centrées sur un « grand tout » technologique, la recherche en biologie et en santé, puissante et bien organisée en France grâce aux organismes de recherche mais aussi aux structures agiles et entrepreneuriales que sont les instituts hospitalo-universitaires (IHUs) et les instituts Carnot, offre la possibilité d'imaginer et de développer des utilisations raisonnées et – osons le terme – écologiques de la technologie. La compréhension fine des mécanismes de plasticité cérébrale, des façons d'agir sur les processus homéostatiques de notre cerveau, ouvre la perspective de développer des neurotechnologies outils, « au service de »... et de faire émerger puis grandir, grâce à cela, de très nombreuses entreprises innovantes. Dans le domaine du traitement des accidents vasculaires cérébraux par exemple, des essais cliniques sont déjà en cours à la Pitié-Salpêtrière (et ailleurs) pour soutenir le cerveau dans son processus endogène de récupération fonctionnelle, par la mise en œuvre de protocoles de stimulation électrique transcrânienne (tACS) : c'est la compréhension fine de l'organisation des réseaux corticaux, acquise par la recherche fondamentale, qui permet aujourd'hui d'orchestrer une musique électrique et d'animer intelligemment l'activité des zones saines du cerveau pour que celles-ci, par leur mobilisation, aident au processus de réparation en stimulant la plasticité synaptique, et potentiellement en « réveillant » des cellules quiescentes. De cette manière et de façon non invasive, une neurotechnologie pourrait faciliter la récupération du mouvement ou de la parole chez les malades, avec un mode d'action proche de ce que font, dans le domaine du cancer, les immunothérapies, en soutenant temporairement la biologie.

L'intégration dans ces processus de recherche du tissu très actif des *start-ups* qui se développent sur le territoire, permet d'espérer l'émergence en France d'une filière de neurotechnologies porteuse d'une vision respectueuse de l'humain biologique : celles-ci pourront alors converger de façon souhaitable avec les outils de l'intelligence artificielle pour se nourrir de la connaissance de la logique du vivant, et pour déployer des moyens d'intervention respectant celle-ci. Cela manifesterait la différence fondamentale entre la science de Socrate – un « savoir conscient de lui-même » –, et la froide et nue technologie.