



Opinions Libres

le blog d'Olivier Ezratty

Ces startups qui veulent bidouiller le cerveau : Neuralink

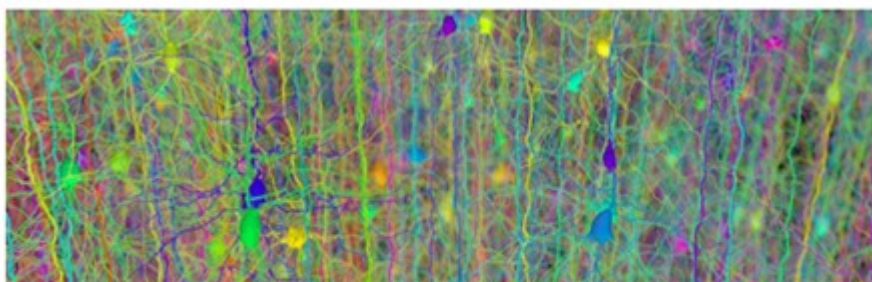
Ecrire dans le cerveau, le relier à une intelligence artificielle ou l'émuler dans une machine n'étaient jusqu'à présent que des lubies de transhumanistes singularistes à la Ray Kurzweil. Ce dernier envisageait ce genre de capacité dès "**The Age of Spiritual Machines**" publié en 1999 et disponible en accès libre sur **Google Drive**. Il récidivait en 2014 avec **How to create a mind**.

Seulement voilà, la donne commence à changer. Des startups se sont lancées pour mettre cela en pratique. Récemment créé par Elon Musk, **Neuralink** ambitionne de relier l'IA au cerveau humain via des nano-électrodes. **OpnWatr** veut de son côté "lire" dans le cerveau avec un capteur d'imagerie new wave et prétend pouvoir également "écrire" dans le cerveau, qui plus est, de manière non invasive.

Dans le même temps, **The Age of Em**, un **livre sorti début 2016**, se lance dans une prospective sociétale globale d'un monde inéluctable dans lequel une bonne partie de la vie se manifesterait sous forme d'avatars de cerveaux humains exécutés dans des machines. Bref, comme si le monde de Matrix était un lendemain certain. L'émulation du cerveau dans les ordinateurs occupe pas mal de monde, dont des chercheurs européens rassemblés dans le controversé **Human Brain Project**, financé à hauteur de 1 Md€ issus par le programme Europe 2020 et piloté par le Suisse **Henry Markram**, de l'EPFL Lausanne.

Ces scénarios font froid dans le dos par leurs implications diverses. On peut se rassurer en se disant qu'il ne s'agit que de délires technologiques dont l'aboutissement est très incertain voire non plausible. Il est cependant bon se regarder de près ce qu'il en est, et la manière dont l'état de l'art évolue. Un peu de *fact-checking* scientifique et technologique ne fait pas de mal ! Pour se rassurer... ou pas sur ces divers travaux d'apprentis sorciers, souvent issus des USA !

Il n'est pas toujours bien évident de se faire une opinion sur ces questions. Tout d'abord, d'un point de vue scientifique et technique avec le degré de faisabilité de bidouillage du cerveau, notamment dans le cortex qui gère la mémoire et le raisonnement. Les avis sont partagés. Ainsi, en octobre 2015, le biologiste américain **Kenneth Miller** expliquait **dans le New York Times** que le téléchargement de votre cerveau n'était pas pour demain. Il s'appuyait dans son raisonnement sur la complexité extrême du fonctionnement du cerveau et des neurones. Il dépend d'une soupe complexe, ne serait qu'avec des milliers de protéines différentes qui les activent. **Michael Cerullo**, **John Smart** et **Keith Willey** lui **répondaient** le même mois en mettant en avant divers travaux, plus ou moins aboutis, permettant d'être plus optimistes.



WHY BRAIN EMULATION IS COMING SOONER THAN MANY THINK – A BPF RESPONSE TO DR. MILLER’S NYT EDITORIAL ON BRAIN EMULATION

By Michael Cerullo, John Smart and Keith Wiley Posted October 27, 2015 in Brain Preservation

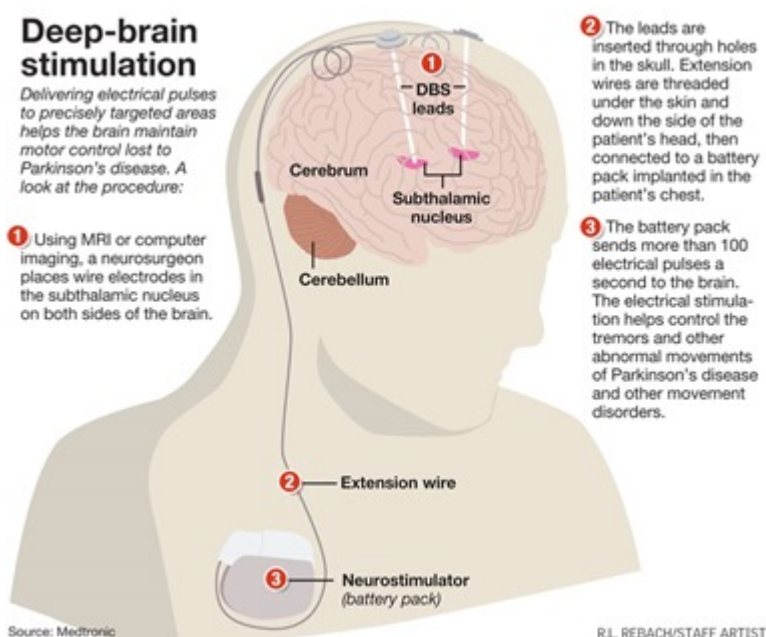


Mon propos sera d’examiner – en plusieurs parties – ces différents projets consistant à lire dans un cerveau vivant ou mort, puis à écrire dans un cerveau vivant ou à simuler son fonctionnement. Je vais essayer d’évaluer leur faisabilité, leur terme, leurs implications. Et surtout, identifier les applications non dystopiques de ces technologies.

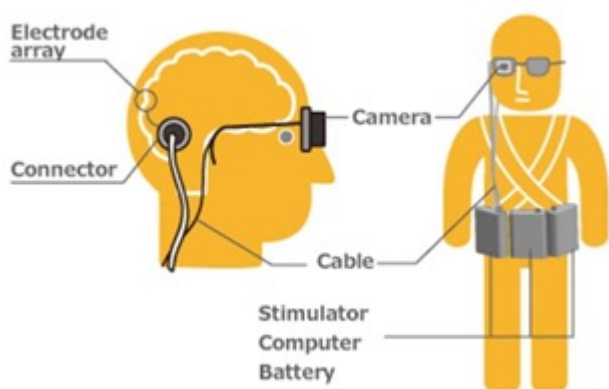
Les niveaux de bidouillage du cerveau

Commençons d’abord par segmenter ces différents projets prométhéens :

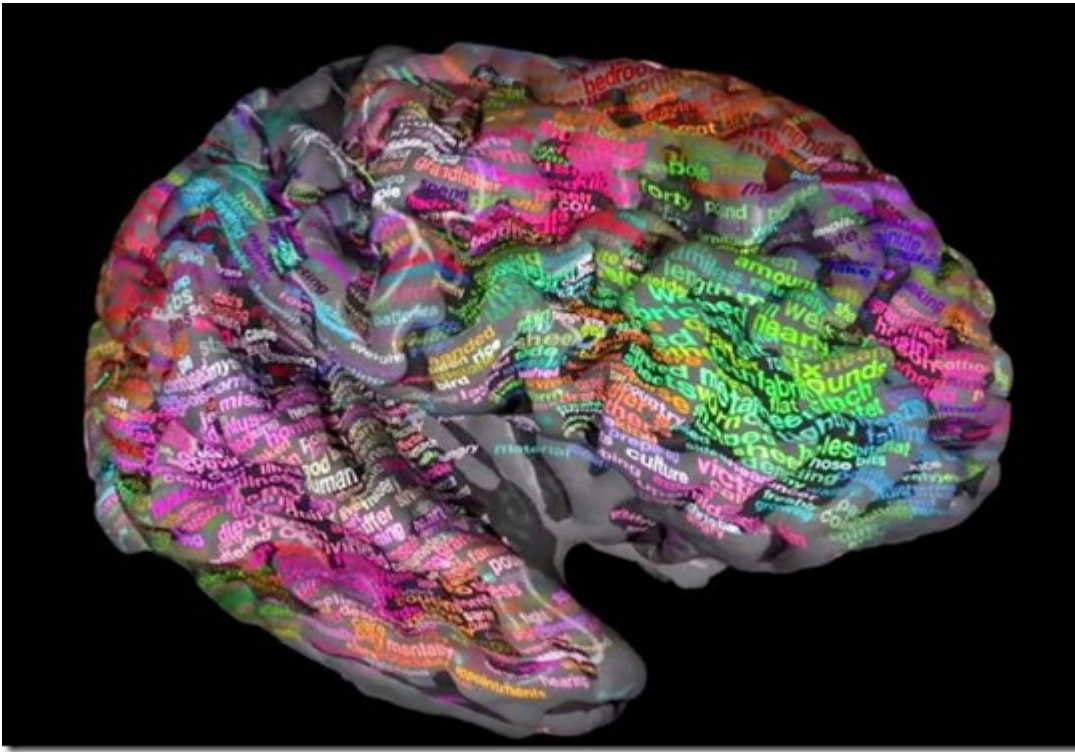
- **Activer des zones cérébrales** dysfonctionnelles ou désactivées, comme pour le contrôle de la marche chez des patients atteints de la maladie de Parkinson. C’est l’état de l’art actuel, sous l’appellation de stimulation cérébrale profonde, pratiquée sur plus de 100 000 patients dans le monde, via un système fourni notamment par l’américain Medtronic. La méthode qui a notamment été mise au point par le neurologue français **Alim Louis Benabid** exploite des électrodes placées au cœur du cerveau, dans les noyaux sous-thalamiques. Elles désactivent des neurones de ce noyau en agissant sur leurs axones pour permettre un fonctionnement normal du cortex moteur. En gros, cela ne fait que débrancher sur commande – avec un boîtier externe – une prise du cerveau, pour permettre au patient de bien contrôler ses mouvements comme la marche.



- **Connecter les entrées du cerveau** avec des capteurs numériques, notamment au niveau de la vue (**exemple**). Cela fait partie des techniques les plus avancées du jour. Cela peut passer par des électrodes placées sur le nerf optique ou sur le cortex visuel (ci-dessous, **source**), situé à l'arrière du cerveau, pour envoyer une image à très basse résolution au patient et lui redonner le sens de la vue. L'image provient notamment d'une caméra filmant ce qui est devant le patient. La plasticité du cerveau lui permet de s'adapter à cette nouvelle source de vision. On peut imaginer intervenir de la même manière au niveau des autres sens comme l'ouïe. Ces différents moyens d'agir sur le cerveau concernent des informations brutes qui sont ensuite interprétées par le cerveau, par exemple, pour reconnaître les éléments d'une image ou les phonèmes dans le son.



- **Lire dans le cerveau.** L'état de l'art actuel qui s'appuie sur des EEG (électro-encéphalogrammes), de l'IRM fonctionnelle et avec des électrodes placées sur la dure-mère du cerveau (qui permet comme dans le projet de Clinatec à Grenoble de détecter les zones actives du cortex et notamment celles du cortex moteur pour détecter la manière dont le cerveau souhaite piloter les principaux gestes comme la marche ou la préhension d'objets). Ces techniques sont regroupées sous l'appellation de BMI pour **Brain Machine Interfaces**. Les primitives détectées sont généralement assez rudimentaires. En effet, les mouvements sont ensuite contrôlés par le cervelet qui joue le rôle du coprocesseur de mouvements. Contrairement à la partie cognitive, notre motricité dépend d'un nombre relativement limité de primitive, de mouvements. L'IRM fonctionnelle permet maintenant de cartographier à haut niveau un certain nombre de fonctions cognitives. Elle permet par exemple de jouer le rôle de détecteur de mensonges plus précis que les capteurs habituels reposant sur la conductivité de la peau et l'évolution du rythme cardiaque. On peut imaginer détecter la pensée sur des primitives simples comme les lettres de l'alphabet ou des chiffres. Cf notamment le "Brain Dictionary" qui cartographie l'activation sémantique des zones corticales en fonction des mots entendus (**vidéo**). Chaque mot active simultanément différentes zones du cortex. La startup **OpnWatr** ambitionne de réaliser une IRM fonctionnelle très précise du cortex via un système de capteurs tenant dans un bandage ou un simple bonnet. Nous verrons cela dans le second article de cette série.



- **Ecrire dans le cerveau** au-delà du contrôle de la perception des sens et au niveau cognitif. C'est ce que prétendent faire aussi bien Neuralink qu'OpnWatr dont nous parlerons plus loin. A ce stade de nos connaissances sur le fonctionnement du cerveau, cela reste une chimère, tout du moins dans l'espace-temps de référence d'une startup. A ce jour, il semble raisonnable d'écrire dans le cerveau via les sens uniquement. Cela permettrait déjà d'interagir avec un système cognitif artificiel. Par exemple, un casque OpnWatr pourrait détecter que l'on pense successivement à telle et telle lettre pour saisir une requête dans un moteur de recherche et des électrodes de Neuralink pour écrire le résultat de la recherche sur le cortex visuel en superposition de notre vue normale. Bref, l'écran de contrôle de Robocop (1987) !
- **Emuler le fonctionnement d'un cerveau dans un ordinateur.** C'est l'objectif de l'Européen Human Brain Project. Le projet fonctionne par étape, cherchant à simuler le système nerveux d'animaux en allant du plus simple (un ver) au plus compliqué (un cerveau d'homme). Le projet reconstruit des neurones dans des ordinateurs. En gros, ce sont des réseaux de neurones logiciels et/ou matériels qui reproduisent par approximation le fonctionnement d'un système nerveux. L'un des éléments clés est d'alimenter ces systèmes nerveux avec des entrées simulant le monde extérieur.
- **Copier le contenu d'un cerveau d'un défunt dans un ordinateur.** Des chercheurs planchent là-dessus. J'ai trouvé pas mal de littérature sur le sujet. L'idée est de congeler un cerveau et de le découper ensuite en fines lamelles qui sont examinées au niveau moléculaire pour détecter tous les nerfs, axones, synapses et mêmes vésicules dans les synapses. Et de copier ensuite tout cela dans un ordinateur pour faire tourner le logiciel de simulation de l'étape précédente. Il faut aussi pouvoir simuler le monde extérieur pour alimenter ce cerveau ce qui est loin d'être évident.

- **Copier le contenu d'un cerveau vivant dans un ordinateur.** C'est le projet le plus space de l'ensemble permettant de copier son cerveau pour lui faire vivre une ou plusieurs expériences synthétiques dans un ordinateur. Cette solution est exposée dans le livre "The age of Em" mais ne repose sur aucun fondement scientifique à ce jour, même avec l'outil de scan du cerveau sur lequel planche OpnWter.

Ecrire dans un cerveau avec Neuralink

Neuralink développe des électrodes miniatures pouvant être intégrées dans le cerveau au niveau de chaque neurone pour augmenter les capacités cognitives et piloter des terminaux. L'idée à long terme est d'associer le cerveau et l'intelligence artificielle, histoire d'éviter que l'intelligence artificielle générale dépasse l'homme. Il doit lui permettre de mieux la contrôler ! Rien que cela. Pour éviter d'être positionnée sur un terme trop long, la société prévoit que sa technologie permettra de traiter certaines pathologies neurodégénératives.

Elon Musk a créé l'équipe de Neuralink de toutes pièces fin 2016. Il a acheté le nom de la société à une startup dénommée **NeuraLink en janvier 2017** et pour \$10K, sans que les vendeurs connaissent l'identité de l'acheteur. Cette startup travaillait sur des implants cérébraux destinés aux patients atteints de traumatismes cérébraux mais son équipe n'a pas été reprise par Neuralink.

Elon Musk fait partie de ces rares entrepreneurs béatifiés de leur vivant. Douter de leurs aventures vous ringardise instantanément. Pourtant, il leur arrive de promettre monts et merveilles et de ne pas systématiquement délivrer, surtout lorsque la promesse relève de la science fiction ou de l'économie fiction. Il en va ainsi lorsqu'il annonce en 2016 qu'il transportera des touristes autour de la Lune en 2018. L'année prochaine. Bien entendu. Et encore, la difficultés n'est pas si technique que cela puisqu'on l'a déjà fait entre 1969 et 1972 avec le programme Appolo. Mais pour des touristes ?

Avec Neuralink, Musk met la barre bien plus haut qu'avec Hyperloop, Space-X ou Tesla. Ces dernières, bien que très ambitieuses et gourmandes en capital, s'appuient sur l'état de l'art scientifique du moment. Avec Neuralink, Elon Musk peut aussi bien perfectionner des techniques existantes de stimulation du cerveau pour traiter des maladies neurodégénératives (lésions cancéreuses, AVC, malformations) que s'aventurer dans un terrain au-delà des limites actuelles du possible avec l'alimentation directe du cortex cérébral en contenus, sans passer par les entrées/sorties classiques que sont nos sens (ouïe, vue, ...).

La société est cependant un peu plus avancée que la très controversée startup **Theranos** au moment de sa création. La créatrice de cette dernière, Elizabeth Holmes, avait un objectif technique mais pas la moindre idée de la manière de la réaliser. Créée en 2004, Theranos ambitionnait de réaliser des examens sanguins complets et low cost avec une seule goutte de sang. Elle a ensuite recruté des scientifiques qui ont travaillé sur divers procédés mais n'ont jamais réussi à les mettre au point. Même s'il ne s'agit pas pour autant de science fiction. Une société française comme **Archimej** est en fait assez proche de cet objectif ! Theranos avait trompé son monde en proposant ses tests sanguins par correspondance et en les réalisant avec des machines d'analyse biologiques traditionnelles de laboratoire, tout en diluant le sang des patients. Ce qui lui a valu de gros déboires avec la FDA et de générer l'une des chutes les plus drastiques de la Silicon Valley, la startup perdant une bonne partie de sa valorisation de \$9B.

Neuralink a pour sa part recruté une belle brochette de talents pour avancer. L'équipe pluridisciplinaire comprend **Paul Merolla** qui avait travaillé sur les chipsets neuromorphiques TrueNorth chez IBM, **Vanessa Tolosa**, une spécialiste des microélectrodes biocompatibles ainsi que **Max Hodak** et **DJ Seo**, des spécialistes des capteurs d'activité neuronale (**neural dust**) utilisés notamment pour capter les ordres de la zone motrice du cortex afin de permettre à des handicapés de piloter des exosquelettes.

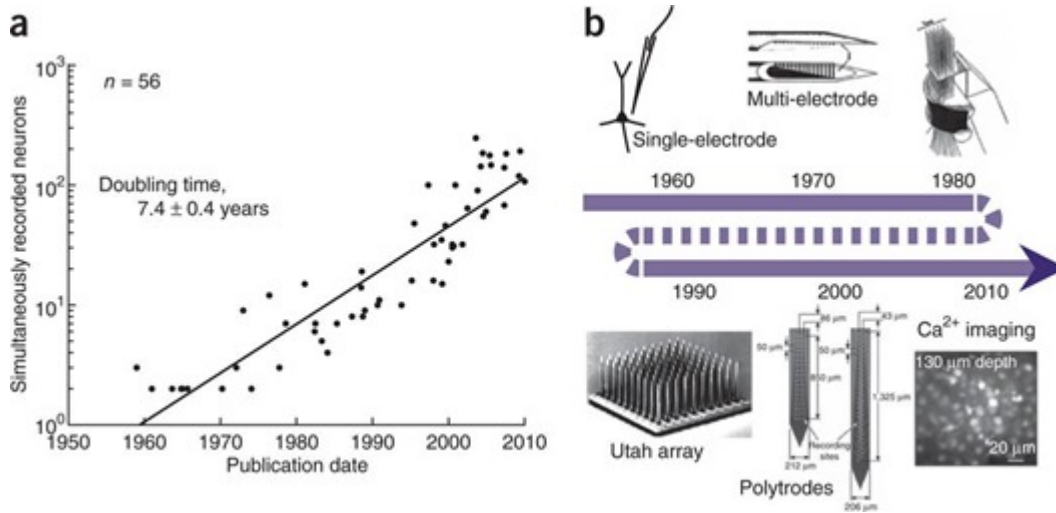
Le fameux blogueur **Tim Urban** de **Wait but why** connu pour ses longs articles de vulgarisation a publié en avril 2017 un **post de 135 pages** d'une seule traite qui bat largement les records de ce blog et détaille ce que prépare Neuralink. L'article est très bien fait et revient sur la connaissance établie du fonctionnement du cerveau. Il a même bénéficié de discussions avec Elon Musk en personne ainsi qu'avec l'équipe de Neuralink au complet [*au passage, Tim Urban est l'un des nombreux excellents intervenants de la **conférence USI** organisée depuis 2009 par Octo Technology et dont la prochaine édition a lieu à Paris les 19 et 20 juin 2017*].

On y apprend que l'objectif principal de Neuralink est de concevoir des électrodes aussi petites que possibles, à l'échelle du micron, qui seraient accolées au cortex sous la dure-mère et capables de détecter l'influx nerveux de la périphérie du cortex, à l'échelle des colonnes corticales qui contiennent une superposition d'une demi-douzaines de couches de cellules nerveuses reliées entre elles.

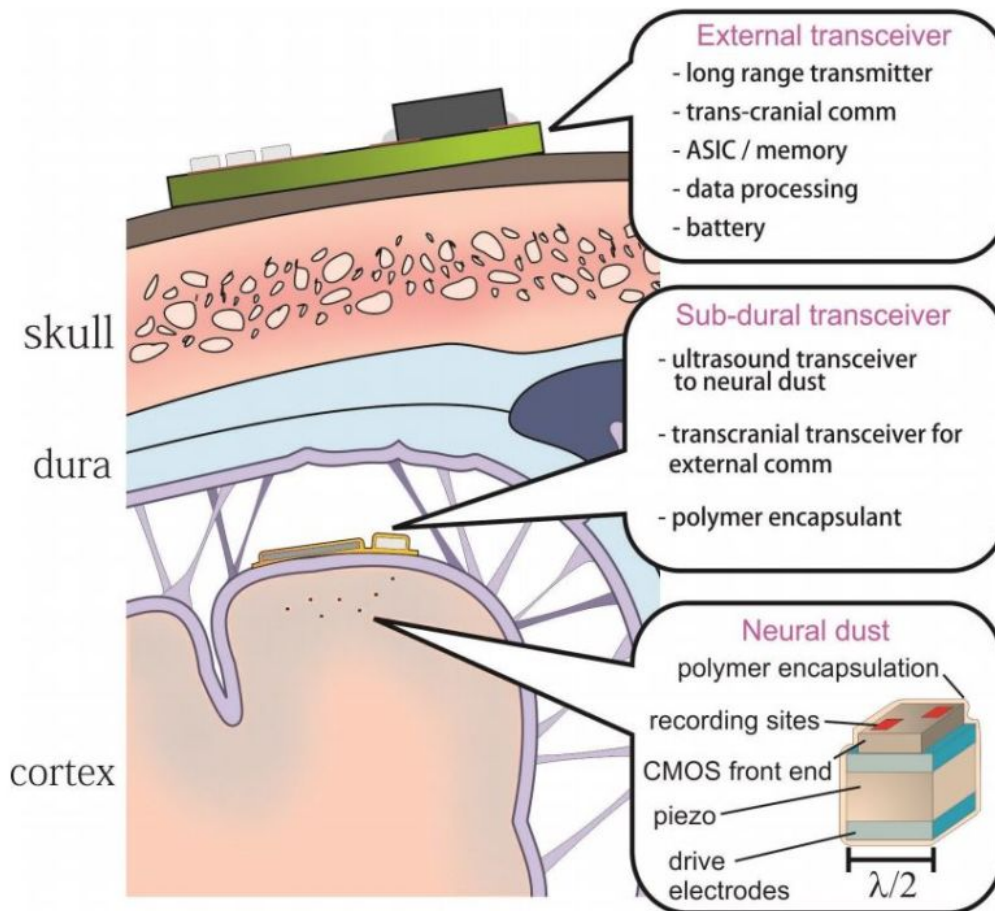
L'implantation d'électrodes dans le cerveau a déjà été **expérimentée**, non sans mal, par un neurologue américain, **Phil Kennedy**, qui s'est fait implanter quelques dizaines d'électrodes dans le cerveau en 2014. L'idée était de lire dans le cerveau pour capter les mots des pensées (cf **speech synthesizer**). L'expérience a duré quelques mois et a dû être abandonnée et les électrodes retirées.

L'état de l'art actuel est limité à quelques centaines d'électrodes mesurant l'activité de 500 neurones (sur les 85 à 100 milliards du cerveau...). L'équipe de Neuralink ambitionne de son côté d'implanter jusqu'à un million d'électrodes dans un cerveau. L'équipe compte sur les miracles de l'application de la loi de Moore à la miniaturisation des capteurs. Celle-ci est cependant plus lente que pour la miniaturisation des transistors.

En doublant la densité des électrodes tous les 18 mois, il faudra attendre 2034 pour en assembler un million, un horizon bien lointain pour une startup ! Dans la pratique, le doublement n'a pour l'instant lieu que **tous les 7 ans** ! A supposer qu'il se poursuive, ce qui n'est évidemment pas forcément censé, il faudrait attendre 2225 pour capter l'activité de toutes les neurones du cerveau. Cela suppose évidemment qu'on puisse les implanter sur l'ensemble du cortex voire à l'intérieur même du cerveau, ce qui n'est pas une paire de manches. Neuralink envisage de robotiser l'implantation des neurones dans le cerveau voire de la rendre non-invasive, ce qui est un vaste programme. Cela passerait par l'utilisation de nano-chipsets qui seraient introduits dans le cerveau par la circulation sanguine.



Les électrodes de Neuralink devront être dotées de nombreuses capacités au-delà de leur miniaturisation extrême : être biocompatibles, isolées hermétiquement du cerveau, ne pas prendre de place grâce à des films très minces, être capables de capter des signaux et aussi de stimuler les neurones, de communiquer sans fil de manière bidirectionnelle, d'être rechargeables par induction, d'être implantée de préférence de manière non invasive. Ça fait un beau cahier des charges ! Une fois ces défis techniques relevés, probablement progressivement, Neuralink ambitionne de créer des solutions *mainstream*, allant progressivement du marché limité de l'homme diminué réparé au marché potentiellement plus vaste de l'homme sain augmenté.



Quelles sont les possibilités ouvertes par les projets de Neuralink ? L'objectif à long terme est de permettre d'upgrader le cerveau, notamment au niveau de sa bande passante. Sa bande passante

est aujourd'hui limitée par celle de ses sens. La captation de connaissances passe par les étapes de compression des idées en langage ou écrit et images et décompression de ces éléments en idées à la réception. En court-circuitant cette compression/décompression, Elon Musk pense que la communication des idées sera plus rapide. Une sorte de télépathie à haute fréquence qui s'affranchira des distances. Bref, un alignement de la vitesse des entrées sorties du cerveau sur celle de la pensée. Cet alignement ne permettrait cependant pas d'être très rapide en comparaison avec l'IA dont la vitesse n'est pas contrainte par la vitesse des flux neuronaux. Tout cela reste bien entendu très théorique car rien ne dit que ces pensées non structurées via le langage soient réellement compréhensibles sans être encodées par le biais du langage !

Pour accélérer cette communication d'idées, Neuralink définit quatre niveaux de liaisons :

- Le **niveau 1** : le cerveau sollicite directement le cloud en épelant une série de mots et le résultat s'affiche visuellement dans le cortex visuel au format texte.
- Le **niveau 2** : le cerveau sollicite le cloud de la même manière que dans le niveau 1 et la réponse devient directement une connaissance dans le cerveau sans passer par l'intermédiaire de la lecture.
- Le **niveau 3** : on a besoin d'une information et sans savoir d'où elle vient, elle devient disponible, qu'elle provienne de la mémoire ou du cloud. C'est une fusion totale entre le cerveau et le cloud. Dans une variante de ce scénario, on pourrait penser dans une langue et parler directement dans une autre langue que l'on ne connaît pas.
- Le **niveau 4** : c'est une variante du niveau 3 qui correspond à l'accès non pas à un fait ou une information mais à la compréhension globale de systèmes complexes.

L'équipe de Neuralink pense qu'il sera possible avec leur approche de couvrir les quatre niveaux, même si cela prendra du temps. En pratique, seul le niveau 1 semble réalisable avec notre connaissance actuelle du fonctionnement du cerveau. Et c'est peut-être mieux comme cela ! En effet, si les quatre niveaux étaient disponibles, la distinction entre homme et machine n'existerait plus.

D'autres scénarios sont envisageables au-delà de ce contrôle des entrées-sorties du cerveau. En stimulant les bonnes neurones, on peut contrôler la sécrétion d'hormones qui conditionnent l'appétit et la satiété (ghréline, leptine), l'humeur et la sensation de bien être, la dépression et l'anxiété (dopamine, sérotonine, ocytocine, ...), le niveau de stress (adrénaline, via la dopamine, CRH, ACTH) ou le sommeil (mélatonine). A ceci près que les zones du cerveau qui contrôlent la sécrétion de ces neurones sont profondes, dans ou proches du cerveau limbique et qu'il est bien plus délicat d'y placer des électrodes.

Pour éviter les déboires d'une intelligence artificielle générale, certains mettent en avant l'impératif de la création d'un "kill switch", un bouton permettant d'éteindre une intelligence artificielle devenue dangereuse. Très bien. Mais si on connecte une AGI au cerveau pour en prendre le contrôle, elle pourrait elle aussi prendre le contrôle du cerveau et le hacker, soit d'elle-même, soit sous l'emprise d'une puissance maléfique d'origine humaine. C'est un peu l'analogie de la course à l'armement : on crée des armes comme des missiles, puis des armes anti-missiles, jusqu'à la dissuasion nucléaire qui est censée annihiler les rapports de force entre les pays qui disposent de cette arme. C'est une course infernale à l'armement.

Via la fondation **OpenAI** dont il est l'un des créateurs, Elon Musk veut créer une sorte de code de discipline de l'IA pour éviter qu'elle prenne le contrôle de l'espèce humaine. En reliant l'homme directement à l'IA, il crée pourtant les conditions d'une prise de contrôle de l'homme par celle-ci. C'en est presque contradictoire. L'arme d'Elon Musk pour éviter le dépassement de l'homme par l'IA est en fait un véritable cheval de Troie potentiellement bénéfique à une IA malsaine, surtout si mise entre de mauvaises mains. Un homme fortement "hackable" de l'extérieur deviendrait très vulnérable comme le sont nos systèmes informatiques aujourd'hui. Surtout s'il ne pouvait plus débrancher lui-même ses prothèses cognitives.

On peut alors se rassurer en se plaçant dans le camp des sceptiques comme Antonio Regalado qui publiait en avril 2017 **With Neuralink, Elon Musk Promises Human-to-Human Telepathy. Don't Believe It** dans la MIT Technology Review. Curieusement, son raisonnement est assez limité. Il s'appuie surtout sur la lente vitesse des progrès dans le domaine plutôt que sur des considérations scientifiques qui pourraient être mises en avant sur la complexité et la plasticité du cerveau. Il faut aussi compter avec **Kevin Kelly**, l'auteur de "The Inevitable", qui **estime** qu'une intelligence artificielle générale est un mythe parce que l'intelligence ne se mesure pas sur une simple échelle exponentielle unidimensionnelle.

Au sujet de Neuralink, voici un peu de lecture complémentaire avec **Neuralink wants to wire your brain to the internet - what could possibly go wrong?** de Chrostopher Markou dans The Conversation et **Neuralink, le projet d'Elon Musk pour augmenter nos cerveaux** de Guillaume Ledit dans Usbek et Rica, . Dans ce dernier article, Laurent Alexandre est cité en indiquant que « *l'idée est de faire passer les dispositifs par les vaisseaux du cou, de manière à ne pas ouvrir la boîte crânienne.* ». Je n'ai pas vu cela dans l'article très détaillé de Tim Urban dans **Wait but Why**, même si Elon Musk voudrait bien trouver une solution d'implantation non invasive de ses électrodes. A vrai dire, on ne sait pas encore comment le faire. Les faire passer par les vaisseaux sanguins est séduisant. On le fait couramment pour des stents et des cathéters, surtout pour les vaisseaux les plus gros. Dans le cas du cortex, cela supposerait que ces électrodes communicantes ne risquent pas de boucher les minuscules capillaires qui aboutissent dans le cortex ! Pas si évident !

Un organe du cerveau s'est d'ailleurs jamais évoqué dans tous ces projets : le cervelet. Son rôle est d'être notamment une sorte de coprocesseur du cortex pour gérer les mouvement appris et certains réflexes. Il gère une bonne part de la marche, de sports divers, de la conduite, du vélo et de la maîtrise d'instruments de musiques. Le cervelet est proche de la moelle épinière. Il est bien plus dense en neurones et axones que le cortex et il est plus difficile d'accès. C'est peut-être pour cela que ces savants ne s'aventurent pas trop dans son bidouillage. Donc, l'apprentissage automatique d'un geste ou d'un sport arrivera bien plus tard que l'alimentation de votre cortex en connaissances ! Vous êtes rassurés ou déçus ?

Dans le **prochain épisode**, nous ferons le tour d'une autre startup ambitieuse dans le domaine du bidouillage du cerveau, **OpnWatr** (OpenWater), dont la solution semble plus astucieuse que celle de Neuralink.

Cet article a été publié le 22 mai 2017 et édité en PDF le 19 février 2019.
(cc) Olivier Ezratty - "Opinions Libres" - <https://www.oezratty.net>