« Il ne faut pas oublier la complexité de nos cerveaux »

Entre idolâtrer l'intelligence artificielle et la maudire, il existe un juste milieu que le chercheur en neurosciences Lionel Naccache tente d'expliquer.

Propos recueillis par Olivier Tesquet

« Il y a certaines tâches *que l'on ne devrait pas confier aux ordinateurs, indépendamment du fait que l'on puisse ou non leur confier ces tâches.* » L'instruction a 50 ans. Elle est signée Joseph Weizenbaum, juif de gauche rescapé de l'Allemagne nazie, qui inventa le premier chatbot en 1966 puis passa le reste de sa vie à s'inquiéter des dangers de l'intelligence artificielle. Un paradoxe qui nous poursuit encore. Dans une ultramodernité où l'IA est à la fois le bac à sable de brillants esprits mathématiques et le siège des discussions les plus irrationnelles, comment s'y retrouver ? Maintenant qu'elle a cessé d'être une expérience de laboratoire pour devenir la béquille de centaines de millions de personnes, comment la regarder ? Est-ce un alter ego ? Si Lionel Naccache, neurologue à l'hôpital de la Pitié-Salpêtrière, chercheur en neurosciences à l'Institut du cerveau et professeur à Sorbonne Université[1], considère que ChatGPT est *« une technologie tragique »*, il tient une intéressante position médiane, mi-fascinée, mi-effrayée, qui se garde bien d'être définitive. Et si les industriels de l'IA rêvent à voix haute d'une « intelligence artificielle générale » capable de concurrencer nos propres capacités cognitives, lui considère que nous en sommes loin.

Qu'est-ce que la fascination pour l'IA dit du rapport que nous entretenons avec notre propre intelligence ?

Dans mes échanges, en tant que chercheur et neurologue, j'observe deux tendances ambivalentes : d'un côté, ce que j'appelle la « cerveaulâtrie », en adoration devant les neurosciences comme si c'était l'alpha et l'oméga de notre vie mentale ; et de l'autre, ceux qui crieraient plutôt « vade retro cerebras » et envisagent les neurosciences comme une agression de leur subjectivité et de leur libre arbitre. Spinoza écrivait que *« nul ne sait ce que le corps peut »*, et il ne faut pas oublier la complexité de nos corps et de nos cerveaux. Le cadre matérialiste que j'ai adopté repose aussi sur l'observation clinique de malades : au quotidien, je vois des patients qui, à la suite de maladies dégénératives ou de lésions cérébrales, présentent des perturbations parfois profondes de pans entiers de leur existence subjective : leur sens de l'humour, leur mémoire, leur jugement, etc.

Vous écrivez que ChatGPT est *« une technologie tragique »*. Pourquoi ?

Le ressort d'une tragédie, c'est que le destin de tous les personnages est joué à l'avance. Or, les modèles génératifs de langage actuels tels que ChatGPT sont capables d'atteindre une forme de perfection dans du déjà fait, du déjà dit, du déjà pensé. Ils arrivent à reproduire de manière époustouflante des choses déjà formulées : ce sont des dizaines de milliards de paramètres qui sont ajustés par apprentissage sur des corpus massifs afin de prédire le mot d'après. En revanche, générer de l'inédit, c'est beaucoup plus difficile…



Photo générée par l'IA issue de la série *La Quatrième Mémoire* (2025), du plasticien Grégory Chatonsky.

Un monde parallèle pour les artistes

Qu'elles soient nées de requêtes adressées, via les intelligences artificielles, aux profondeurs abyssales de la mémoire du Web, ou de celles des archives photos de la ville du Havre, les installations faites de textes, d'images et de sons du plasticien Grégory Chatonsky suggèrent des passés alternatifs comme une « quatrième » mémoire, celle que nous aurions pu avoir : *« Une œuvre posthume de mon vivant : non pas comme un monument dédié à ce que j'ai déjà fait, mais comme une machine qui poursuit ce que j'aurais pu faire. »* L'artiste participera à l'exposition « Le monde de l'IA », au Jeu de Paume, à Paris, qui va explorer la rencontre entre ces nouvelles technologies et les univers des artistes invités. Et comme l'écrit ChatGPT sollicitée par Quentin Bajac, historien de la photo et directeur du musée, pour participer à la rédaction de la préface du catalogue : *« Ensemble, ouvrons la porte à un dialogue fascinant sur l'avenir de l'art à l'ère numérique. »* ▷ *Laurent Abadjian*

« La Ville qui n'existait pas, épisode III, 2025-2049 », dans le cadre d'Un été au Havre.

« Le monde selon l'IA », du 11 avril au 21 septembre, Jeu de Paume, Paris 8e. Catalogue *Le Monde selon l'IA*, éd. JBE, 304 p., 39 €.

Les industriels de l'IA sont hypnotisés par la création d'une intelligence artificielle générale, capable de « penser » de manière autonome. Qu'en pensez-vous ?

Il ne s'agit pas d'une diversion au sens d'un complot, mais avant tout de nos propres diversions. Nos ressources cognitives sont extrêmement limitées. Si on alloue toutes nos capacités à poursuivre une peur fantasmatique, on ne pourra plus prendre en compte les problèmes plus immédiats. Concentrons-nous sur les angoisses légitimes – qu'elles soient économiques, sociales, politiques – autour de systèmes déjà existants : les risques d'appauvrissement de nos subjectivités évoqués, mais également les risques d'inégalités, les risques politiques et sociaux inhérents à ces technologies par ailleurs utiles. Arrivera-t-on un jour à simuler une conscience artificielle ? Impossible de le dire, mais on en est loin, même si les progrès sont spectaculaires. Désormais, à l'intersection des neurosciences et de l'apprentissage profond [le *deep learning*, ndlr], on peut comparer ce qui se passe dans un vrai cerveau et dans un cerveau artificiel. Ainsi, on a déjà pu observer des ressemblances intéressantes, par exemple ce qui se passe dans telle ou telle région cérébrale, ce qui laisse à penser qu'on a réussi à simuler une partie des opérations accomplies par un cerveau neuronal.

À ce titre, quel regard portez-vous sur Neuralink, le projet d'implant cérébral d'Elon Musk, qui voudrait hybrider l'humain et l'intelligence artificielle ?

Il me semble essentiel d'adopter une posture méliorative, restaurer la mobilité d'un patient paralysé par exemple. C'est le principe de la médecine. Ce qui est plus problématique, et c'est très clair chez Musk, c'est qu'il s'agit pour lui de « déverrouiller » *(unlock)* l'être humain en bonne santé. Authentique Dr Knock 2.0, Musk tient un discours idéologique radical qui abolit toute frontière entre la condition saine et la condition pathologique. Tous les humains sont verrouillés, et nous allons les déverrouiller. Cela revient aussi à concevoir l'existence humaine comme une performance, qui change radicalement notre façon de vivre ensemble.

Quid de l'intelligence émotionnelle dans la création par l'IA d'un cerveau « parfait » totalement désincarné ?

Il existe un débat scientifique intense et ouvert à ce sujet. L'intelligence artificielle peut-elle embarquer une couche qui s'apparente à des émotions ? Le neurologue et neuroscientifique américain António Damásio opère une distinction utile entre les émotions, inconscientes, et les sentiments, qui correspondraient à l'appréhension consciente de ces représentations émotionnelles. Selon lui, la naissance d'un soi requiert une expérience émotionnelle ancrée dans le corps – on parle de cognition incarnée. C'est d'ailleurs l'une des visées de certains travaux de robotique : reproduire une situation perceptive et motrice, et y intégrer des paramètres internes qui équivaudraient à la pression artérielle, à un rythme respiratoire… À l'inverse, d'autres pensent qu'un cerveau artificiel pourrait théoriquement être le siège d'un « moi ». Mon hypothèse est à mi-chemin : je pense que l'émergence d'un soi conscient requiert un cerveau incarné, mais qu'une fois constitué, ce soi conscient peut s'affranchir d'une certaine manière de ces contraintes corporelles ●

1 *Sujet es-tu là ?*, de L. Naccache, éd. Odile Jacob, 304 p., 22,90 €.

Un accélérateur de découvertes archéologiques

« Comme l'a fait l'imagerie aérienne avant elle, l'IA est peut-être sur le point de révolutionner les découvertes archéologiques », estiment Masato Sakai, spécialiste de l'archéologie péruvienne, et son équipe. En ratissant, via une IA développée avec la firme IBM, les prises de vues au drone des plus de 600 kilomètres carrés du désert de Nazca, dans le sud du Pérou, les archéologues ont découvert, entre septembre 2022 et février 2023, soit en l'espace de six petits mois, trois cent trois nouveaux « géoglyphes » – des figures géantes tracées au sol, classées au patrimoine de l'Unesco, auxquelles viennent s'ajouter une myriade de motifs plus petits, de têtes coupées ou de chameaux domestiques. Une découverte qui montre tout l'intérêt de cette nouvelle technologie, dès lors que les experts l'emploient sur des corpus volumineux : en France, l'IA travaille actuellement à décrypter l'écriture en pattes de mouche des chanoines, sur les cent soixante-dix volumes d'actes de la cathédrale Notre-Dame.
▷ *Charlotte Fauve*



Un des trois cent trois nouveaux géoglyphes découverts en l'espace de six mois dans le désert péruvien de Nazca grâce à l'analyse de photographies aériennes par IA.