



IA monstres : le grand frisson !

Au sujet de la lettre ouverte signée par Elon Musk et réclamant un moratoire de 6 mois du développement des IA génératives (GPT etc.) ...

2 avril 2023

Catégorie : **Intelligence Artificielle**

Tags : **éducation, éthique, futur, hubris, loi, machine, société, technique**

Personnages : **Raja Chatila, Clément Delangue, Tristan Harris, Elon Musk, Stuart Russell, Jaan Tallinn**

I think people should be happy that we are a little bit scared of this.

Sam Altman, CEO de OpenAI, société à l'origine de GPT¹

Matinale

Vendredi 31 mars dernier, c'était au tour de la Matinale de France Inter de contribuer à l'effervescence mondiale créée par la lettre ouverte et inquiète réclamant un moratoire de six mois dans le développement de systèmes d'IA générative tels que ChatGPT, Midjourney ou DALL.E (« *Pause Giant AI Experiments: An Open Letter* »²).

L'IA atteint en effet pour la première fois l'objectif rêvé dans les années 1960 par ses créateurs et auguré par la littérature de science-fiction : confondre l'humain et faire société avec lui. Si un grand média national comme Radio France donne voix à cette inquiétude, c'est que ces machines intelligentes parviennent enfin à ce que nous identifions comme le stade « 3 » de leur déploiement, celui nécessitant des moyens techniques et financiers considérables ([PageRank](#), [Parcoursup](#) et autres « [machines morales](#) »). Elles sont désormais *accessibles à tout le monde* et donc, pour autant que nous leur prêtions une existence propre, font *déjà* société avec nous : elles peuvent en principe travailler, passer des examens, produire de la « vérité », de l' « art », etc.

¹ Victor Ordonez , Taylor Dunn, Eric Noll / ABC News – 16 mars 2023 – [OpenAI CEO Sam Altman says AI will reshape society, acknowledges risks: 'A little bit scared of this'](#)

² Future of Life Institute – [Pause Giant AI Experiments: An Open Letter](#)

Lors de cette émission, Raja Chatila, roboticien, professeur émérite d'intelligence artificielle et d'éthique des technologies à Sorbonne Université, justifie ainsi son paraphe³ :

Là, il fallait lever un carton rouge. Quelque chose est en train de se passer, et il fallait en prendre conscience collectivement. Il ne s'agit pas de faire peur, il s'agit de dire une certaine réalité. Un certain nombre d'institutions sont en train de déployer des systèmes qui sont des basés sur l'apprentissage automatique de très grande quantité de données, et qui sont arrivés à un point de développement de reproduire des textes qui semblent rédigés par des humains. Mais ces textes ne sont pas porteurs de vérité, il est difficile de distinguer le vrai du faux. Le deuxième problème, c'est l'absence totale de sens. Ces systèmes ne comprennent pas ce qu'ils écrivent ou ce qu'ils disent.

Rien de bien nouveau donc, mais il faut reconnaître que, présentée ainsi, cette situation précipite une angoisse familière : l'angoisse du changement. Cette angoisse est habituellement réduite par les injonctions *positives* à la disruption, à la destruction créatrice et au solutionnisme technologique. Mais dans les cas des IA génératives, nous semblons réagir différemment. Il faut donc changer de méthode.

Sans douter une seconde de la soucieuse sincérité de M. Chatila ni de celle de la plupart des 2500 signataires de cette lettre ouverte (décompte au 1^{er} avril 2023), ceux-ci jouent une partition très classique de l'étape « 4 » : le ripolinage des artefacts technologiques hyperdéployés à

³ Radio France / Matinale de France Inter – 31 mars 2023 – [Le patron de Twitter Elon Musk et des centaines d'experts réclament une pause dans l'intelligence artificielle. Ils réclament un moratoire jusqu'à la mise en place de systèmes de sécurité](#)

coups de milliards et de gigawattheures en *machines morales*, étape ultime de leur crantage dans nos « système de croyances ». Cette lettre contribue ainsi à l'avènement *effectif* des IA génératives dans la société des humains.

Voyons ceci de plus près, non sans un petit détour par Elon Musk pour commencer.

Elon Musk

Elon Musk, personnage vigoureux scanné dans [Elon Musk, vassal spécial](#), semble être l'un des instigateurs de cette lettre ouverte. Pas grand monde n'est dupe de certaines de ses motivations. Il se trouve ainsi que ChatGPT est la créature de OpenAI, une structure qu'il a coprésidée à partir de 2015 puis lâchée en 2018 (faute de résultats !). Petit souvenir de 2016, bras croisés, regard perçant, parfait⁴ :

Inside OpenAI, Elon Musk's Wild Plan to Set Artificial Intelligence Free

OpenAI wants to give away the 21st century's most transformative technology. In the process, it could remake the way people make tech.



MICHAL CZERWONKA/REDUX

⁴ Wired / Cade Metz – 27 avril 2016 – [Inside OpenAI, Elon Musk's Wild Plan to Set Artificial Intelligence Free](#)

La réaction pour le moins crispée d'Elon Musk face au succès de ces IA génératives est parfaitement expliquée dans l'excellent article de Matt Novak qui conclut⁵ :

Musk était parfaitement satisfait de développer des outils d'intelligence artificielle à une vitesse fulgurante lorsqu'il finançait OpenAI. Mais maintenant qu'il a quitté OpenAI [racheté par Microsoft] et qu'il l'a vu devenir le chef de file d'une course aux technologies les plus pointues pour changer le monde, il veut que tout s'arrête pendant six mois. Si j'étais un parieur, je dirais que Musk pense pouvoir pousser ses ingénieurs à mettre au point leur propre IA avancée dans un délai de six mois. Ce n'est pas plus compliqué que cela.

Nous sommes prêts à parier avec lui et même à doubler la mise : les motivations d'Elon Musk ne se résument pas à un simple tournoi entre pairs dont il aurait perdu une manche. C'est bien plus que cela. Musk est viscéralement engagé dans la transformation *de l'humanité* (c'est-à-dire de lui-même...). Selon ce transhumaniste radical, la planète terre et l'humain « naturel » sont obsolètes. Par conséquent, l'IA en tant que simple *technologie* au service de l'humanité, au même titre que la voiture, l'électricité ou le numérique, est un trompe-l'œil et surtout un défi pour son projet « spéciste » à l'égard des machines. Pour résister à la machine qui advient et qui nous terrorise, l'humain doit en quelque sorte « ingérer » la machine. C'est le motif-même de son projet Neuralink, qui pose des problèmes éthiques bien plus redoutables que les IA génératives, à condition que ces dernières soient simplement envisagées *comme des*

⁵ Matt Novak / Forbes – 29 mars 2023 – [Elon Musk's AI History May Be Behind His Call To Pause Development](#)

outils et non pas comme des acteurs sociaux. Si Neuralink décroche un jour la timbale de la « phase 3 », nous promettons de signer une lettre ouverte !

Invité de la Matinale, Clément Delangue, co-fondateur de la start-up française « Hugging Face », a pris la position qui nous semble la plus juste, dans la ligne de cette éthique du dévoilement qui nous est chère (nous soulignons) :

Je n'ai pas signé la pétition. Ce que l'on voit ici, c'est que cette pétition ressemble à une opération marketing un peu dirigée par Elon Musk. Il a été dépassé par l'intelligence artificielle. L'un des problèmes avec cette pétition, c'est que les solutions proposées sont très peu pratiques ou applicables. Ce qui est important, et ce que montre ce débat, c'est qu'aujourd'hui il faut plus de transparence et d'éducation au sujet de ces systèmes.

Mais cette posture beaucoup moins anxiogène fait aussi beaucoup moins d'effet...

Elon Musk n'en n'est pas à son coup d'essai. Utilisant déjà en 2015 son Future of Life Institute comme véhicule de propagande, il nous faisait déjà le coup de la lettre ouverte et inquiète⁶ signée, entre autres, par le regretté physicien Stephen Hawking (mais qu'est donc allé faire cet éminent physicien dans ce véhicule ?). Ce dernier déclarait déjà ceci en 2014 à la BBC (inspirant peut-être les inclinations de Musk)⁷ :

Les humains, limités par une évolution biologique lente, ne pourraient pas rivaliser et seraient supplantés par l'IA.

⁶ Future of Life Institution – [Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter](#)

⁷ Michael Sainato / Observer – 19 août 2015 – [Stephen Hawking, Elon Musk, and Bill Gates Warn About Artificial Intelligence](#)

Raisonnement très physicien à l'appui de « Ah, mon Dieu, c'est horrible ! ». Mais on en a vu d'autres...

Le Bien et le Mal

Effective altruism is about doing good better

Parmi les acteurs remarquables du Future of Life Institute et signataires de ces lettres ouvertes, nous reconnaissons trois personnages déjà croisés ici : Jaan Tallinn, milliardaire estonien et « *altruiste efficace* »⁸ (mentionné ici comme prototype pour une recension ultérieure du courant moral de l'« *Effective Altruism* » – voir aussi la mention de Jaan Tallinn dans [Un futur sans nous](#)), Tristan Harris ([Tristan Harris et le marais de l'éthique numérique](#) où il est question entre autres de l'éthique de dévoilement) et enfin Stuart Russell ([Being Stuart Russell – Le retour de la philosophie morale](#)). Tous cautionnent des formes diverses et variées de ce *conséquentialisme* qui inspire les signataires les plus sincères.

Il est tout à fait remarquable qu'en France tout au moins, pour ce que nous en savons, aucune pensée propre et renouvelée ne vient plus faire rempart ni simplement discussion contre cet « utilitarisme tech » auxquels même les chercheurs se sont habitués. On s'émeut, on signe et puis on continue dans les laboratoires. Les traces du « retour » de cette philosophie morale, c'est-à-dire d'une axiologie du « bien » et du « mal » (catégories désormais numérisables) sont perceptibles dans le texte. Nous lisons ainsi :

⁸ Wikipedia – [Jaan Tallinn](#)

Faut-il laisser les machines inonder nos canaux d'information de propagande et de contre-vérité ? Devrions-nous automatiser tous les emplois, y compris ceux qui sont gratifiants ? Devons-nous développer des esprits non humains, qui pourraient un jour être plus nombreux et plus intelligents, nous périmer et nous remplacer ? Devrions-nous risquer de perdre le contrôle de notre civilisation ?

Bien sûr que non ! Tout ceci, c'est *mal*, évidemment. Mais tout ceci existe bien déjà, parfois depuis longtemps, sans l'IA. En effet, qui « *inonde nos canaux d'information de propagande et de contre-vérité* » (c'est mal), si ce n'est *quelqu'un*, certes équipé aujourd'hui d'outils numériques et smart ? Qui cherche inlassablement à « *automatiser tous les emplois, y compris les plus gratifiants* » (c'est mal), si ce n'est, depuis le XIXème siècle, le capitalisme industriel servi par des mégamachines ? Qui cherche inlassablement à « *développer des esprits non humains* » (c'est mal), si ce n'est, au moins, le complexe militaire, creuset de l'IA moderne rappelons-le, pour d'évidentes raisons ? Qui peut penser sérieusement que nous puissions « *perdre le contrôle de la civilisation* » (c'est mal) à cause de l'IA, alors que d'autres forces bien humaines s'en chargent bien plus efficacement ?

Bref, qui à part nous-mêmes ? Peut-être devrions-nous instaurer une pause planétaire de 6 mois le temps de régler tout ça ?

Par pitié, régulez-nous !

La lettre ouverte finit quand même par atterrir en implorant un contrôle social accru sur ces systèmes :

La recherche et le développement dans le domaine de l'IA devraient être recentrés sur l'amélioration de l'exactitude, de la

sécurité, de l'interprétabilité, de la transparence, de la robustesse, de l'alignement, de la fiabilité et de la loyauté des systèmes puissants et de pointe d'aujourd'hui.

Et ainsi de réclamer plus de réglementation et de contrôle, ainsi que des fonds publics pour la recherche en matière de sécurité technique de l'IA (« *technical AI safety research* »). On ne peut que souscrire, évidemment, puisqu'il s'agit de combattre le mal et en particulier de préserver notre « modèle démocratique » (bien) :

... [et] des institutions dotées de ressources suffisantes pour faire face aux bouleversements économiques et politiques spectaculaires (en particulier pour la démocratie) que l'IA provoquera.

Mais franchement, de quels épouvantables maux démocratiques *supplémentaires* l'IA est-elle capable quand *déjà*, ne supportant plus aucun risque, ne supportant plus les incertitudes de l'avenir, affolés, nous avons déposé notre liberté aux pieds de pouvoirs munis d'instruments de contrôle et de réglementations *préventives* ([Génération François Sureau](#)) ?

Le problème est plutôt que la recherche en IA (comme d'ailleurs la recherche en médecine, en biologie, en robotique...) échappe en grande partie aux capacités financières de la puissance publique civile et donc à un certain « bien commun by design », pour ainsi dire. Raja Chatila le reconnaît bien lui-même :

La recherche académique n'a pas les moyens d'être en compétition avec des entreprises de ce type-là [OpenAI, Google, etc.]. Ils ont tout, et dans les universités on est « petits joueurs ».

Pour la première fois peut-être, la plupart des États n'ont plus les moyens de mener ces recherches ni donc de donner la possibilité aux citoyens d'y participer *a priori*. On ne demande donc plus à cette puissance publique, bonne fille, et sous le mode de l'imploration, qu'une régulation *a posteriori* (Europe, championne du monde), et on lui octroie quelques miettes comme la « *recherche en matière de sécurité technique de l'IA* ».

Cela peut faire illusion et rasséréner les plus contrits. Mais les régulations demandées ne pourront évidemment être élaborées que par des cabinets d'experts, des structures de conseil et des spécialistes capables de comprendre les technologies en jeu, de sorte que la régulation reste un « business » avançant main dans la main avec la tech et partageant les mêmes réseaux d'influence. C'est ainsi que l'on peut passer sans nous de la phase 3 à la phase 4, avec le petit coup de pouce d'une lettre ouverte.

Art

Le vrai « péril » est que prenions vraiment peur et qu'ainsi tétanisés nous finissions par croire que l'IA et ses avatars sont bien plus que des *instruments* : des « choses » capables de provoquer une dangereuse rupture de civilisation. Ce discours alarmiste rend (préventivement) impossible toute discussion critique au sujet de la nécessité ou de l'inéluctabilité de ces technologies, posant qu'un polish éthique ou réglementaire, sur lequel les signataires nous invitent à nous concentrer, viendra à bout des *derniers* problèmes, quand bien même ces problèmes seraient « mortels » ! L'IA générative *doit* trouver sa place dans la société et cette lettre ouverte nous sommes en quelque sorte de faire notre part en préparant son lit réglementaire.

Nous invitons plutôt, à la suite de Clément Delangue, à *instruire notre regard*, aussi bien sur les techniques d'IA que sur les puissances qui les contrôlent. Depuis les millénaires de progrès technique, l'humain a toujours su vaincre l'effroi de son propre remplacement, c'est-à-dire de son absence de singularité, en se familiarisant avec les outils forgés par les artisans et en contrôlant leurs représentations par le langage (ce que nous nous efforçons de faire ici), puis par des formes ironiques et artistiques. Ce sont toutes ces représentations qu'il faut presser d'advenir.