Intelligence artificielle forte, quatre raisons de douter

Les perspectives de l'Intelligence Artificielle inquiètent - les ordinateurs pourraient-ils atteindre à une intelligence similaire à celle de l'être humain ? Les robots pourraient-ils devenir indépendants de leurs créateurs, voire hostiles, et aller même jusqu'à les remplacer ?

Aussi important que puisse être dans l'avenir l'impact de l'IA sur la productivité et sur l'emploi, il y a en réalité au moins quatre raisons de douter qu'une Intelligence Artificielle "forte", c'est-à-dire similaire à l'humaine - "quelqu'un dans la machine" - soit pratiquement ou même théoriquement possible.

Les rêves ou les cauchemars des théoriciens de la "Singularité" et du remplacement par les machines ne sont même pas pour après-demain, et probablement pour jamais.

De l'astronome Martin Rees s'inquiétant que les robots ne prennent bientôt le pouvoir, l'astrophysicien Stephen Hawking craignant que limités par une évolution trop lente les êtres humains s'avèrent incapables de rivaliser avec l'intelligence artificielle ou l'entrepreneur Elon Musk décrivant le risque que l'humanité n' "invoque un démon" et ne mette sa propre existence en danger en construisant une intelligence artificielle supérieure... jusqu'au mouvement trans-humaniste espérant des progrès de l'intelligence artificielle non seulement la multiplication des possibilités humaines mais rien de moins que l'immortalité - ainsi l'un des interlocuteurs de Mark O'Connell dans "To Be a Machine" prophétisant la possibilité de transférer son esprit dans une machine, laquelle ne serait pas soumise à la mort.

Ces craintes et ces espoirs découlent tous d'une source unique : la perspective de doter prochainement un ordinateur d'un esprit similaire à celui de l'être humain, qui lui serait bientôt supérieur - ce qu'il est convenu d'appeler une "intelligence artificielle forte".

Cette perspective est-elle véritablement réaliste ?

Avant de développer l'argumentation, commençons par sa conclusion - le résumé pour décideurs :

> *Construire à base d'ordinateurs une intelligence artificielle forte – c'est-à-dire égalant ou dépassant l'esprit humain, « quelqu'un dans la machine » – est fort probablement impossible pour raison de principe, car au moins trois questions fondamentales pourraient chacune à elle seule en exclure la possibilité, et la réponse définitive à chacune de ces questions est à ce jour inconnue.*
>
> *Si toutefois aucune des trois réponses ne s'avéraient faire obstacle et si ce projet était donc théoriquement possible, sa difficulté inhérente – hors de toute proportion avec la difficulté à fabriquer par exemple de simples robots autonomes utiles dans la vie courante ou l'industrie – bref la question pratique, assurerait qu'elle ne pourra de toute façon être qu'un projet à très long terme, comparable par exemple avec ce qu'est pour l'astronautique le vol interstellaire.*
>
> *L'IA forte dans dix ou vingt ans des transhumanistes théoriciens de la "Singularité" n'est que balivernes.*

Alexis Toulet – 14 septembre 2017

En cette époque où l'humanité fait face aux prodromes d'une crise gigantesque, vague scélérate additionnant fragilités du système financier et entrée dans l'âge des limites notamment en énergie fossile sur fonds de catastrophe écologique en cours incluant un dérèglement climatique aux conséquences de long terme très menaçantes, il est très agréable de découvrir - pour une fois ! - que tel nouveau monstre menaçant sortant du brouillard... n'est finalement qu'un banal épouvantail et un jouet pour faire peur aux enfants.

En l'espèce, la menace d'une conscience artificielle née de l'informatique parvenant à supplanter les humains.

Un peu de contexte…

Pour commencer, le mythe de la création prochaine d'esprits artificiels est vivace depuis les années 1950. Ce qui, comme on dit, ne nous rajeunit pas. Il est en général annoncé pour le prochain coin de rue, dans quelques petites années. Puis, lorsque les prédictions ne se sont - à l'évidence - pas réalisées, d'autres reprennent le conte en toute bonne foi, et roulez jeunesse ! Jusqu'à la prochaine fois.

Cependant, les réalisations de la discipline "IA" ne sont pas du tout à la hauteur de ces craintes et espoirs tonitruants. Ce n'est pas qu'elles soient inexistantes, ni négligeables, loin de là ! Simplement, la reproduction informatique – on pourrait dire le mime – d'activités humaines généralement considérées comme intelligentes ne mène pas à l'apparition d'une conscience artificielle. La carte serait-elle par nature différente du territoire ? La simulation, différente de la réalité ?

Naturellement, la puissance des ordinateurs, jusqu'ici très inférieure à celle d'un cerveau humain, constitue une explication possible de l'échec à ce jour à produire une "IA forte". Peut-être tout simplement les ordinateurs n'étaient-ils pas encore assez performants ? Voilà qui pourrait amener à penser que l'augmentation exponentielle des capacités de traitement informatique mettra en revanche bientôt à portée le Saint Graal d'une conscience artificielle.

Réalisant une simulation précise du fonctionnement physique des neurones, l'un des plus grands calculateurs début 2014 a pu simuler le fonctionnement de 1% d'un cerveau humain pendant une seconde... mais le calcul lui a pris 40 minutes. De ce point de vue, les plus puissants ordinateurs actuels sont très loin du compte. En revanche, en se limitant à une simulation logique en réseau de neurones, la puissance nécessaire à un "cerveau humain" en temps réel est déjà largement disponible. Il suffit d'examiner les ordres de grandeur : 10^11 neurones et 1,5. 10^14 synapses, effectuant des "calculs" à un rythme inférieur à 10^2 par seconde, soit au maximum 1,5.10^16 opérations par seconde nécessaires - et probablement beaucoup moins - alors que le plus grand superordinateur était en juin 2017 le Sunway TaihuLight chinois, lequel peut effectuer 9,3. 10^16 opérations par seconde. Soit au moins 6 fois plus que nécessaire pour simuler complètement et en temps réel le fonctionnement logique du réseau neuronal d'un cerveau humain.

Cependant, même compte tenu du potentiel de cette puissance de calcul, quatre questions doivent être posées, **quatre obstacles majeurs barrent le chemin de la construction d'une IA forte ou conscience artificielle**. Les trois premières sont fondamentales, il s'agit de la possibilité théorique elle-même que des êtres humains puissent réaliser un tel « objet pensant » à base d'ordinateurs. La quatrième est tout simplement la question pratique, à supposer que les trois questions de possibilité théorique soient décidées dans un sens favorable – et ce n'est pas la moindre.

Les trois questions fondamentales, pour commencer

C'est que pour que le projet de construire une conscience artificielle par voie informatique ait ne serait-ce qu'un sens, trois conditions sont nécessaires :

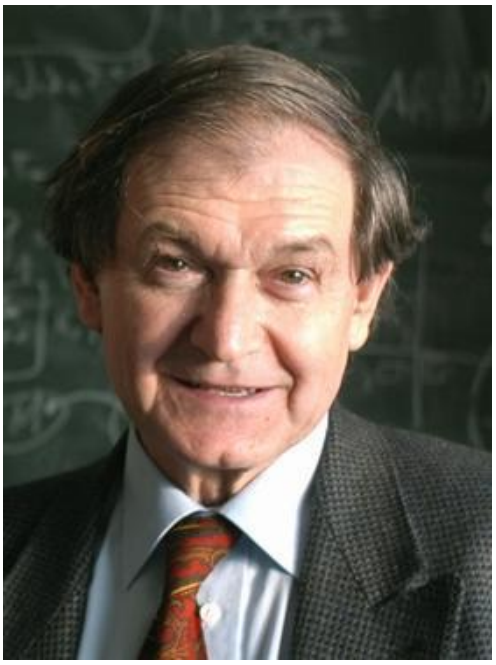
1 - Il faut que la conscience, telle qu'elle se manifeste par exemple dans la tête de tout un chacun, soit entièrement compréhensible en termes matériels. C'est là une position philosophique matérialiste.

> *Incise – Non, cette condition n'a rien d'« évidente »*
>
> *La position opposée, c'est-à-dire l'existence de « quelque chose » de non réductible à la matière et qui serait intrinsèque à l'esprit ou à la personne humaine, apparaîtra suspecte à beaucoup pour une simple raison d'habitude.*
>
> *C'est que les explications de type surnaturel – esprits, fées, lutins ou dieux – ont évidemment reculé constamment dans les derniers siècles, la méthode scientifique permettant de comprendre toujours davantage de phénomènes toujours plus en profondeur, alors qu'ils avaient été autrefois considérés comme des mystères inaccessibles à l'esprit humain. Il est alors bien naturel de considérer qu'une tendance historiquement aussi bien établie continuera indéfiniment, et qu'elle permet d'apercevoir ce qui serait en définitive la vérité ultime : que l'ensemble de ce qui existe est matériel, donc soumis aux règles de la matière telles qu'elles sont progressivement dévoilées par l'effort scientifique humain.*
>
> *Il est bien évidemment loisible de choisir d'adopter une telle position. A qui la choisit, une position différente risque de n'apparaître motivée que par au choix : l'ignorance, le préjugé par exemple religieux, ou un sentimentalisme refusant d'admettre que moi aussi et non simplement le monde qui m'entoure, je pourrais n'être que matériel, et s'imaginant donc « par nature » différent.*
>
> *Mais il est également vrai que les tendances historiques les mieux établies peuvent rencontrer leurs limites, et surtout que la connaissance extraordinaire apportée par la méthode scientifique ne signifie pas nécessairement que celle-ci permet d'accéder à une vérité ultime. Car c'est bien ce que suppose la philosophie matérialiste, et ce passage de « la méthode scientifique a permis de constamment avancer dans la compréhension de la réalité » à « la méthode scientifique révèle la vérité ultime sur la réalité » n'est rien d'autre qu'un passage du fait à la croyance… qu'un « saut de la foi », s'il est permis d'être taquin.*
>
> *La position pleinement rationnelle à ce stade est en fait la position « agnostique », c'est-à-dire de ne pas conclure sur le matérialisme en tant que philosophie, parce que pour ce qu'on en sait aujourd'hui rien ne le démontre, et rien non plus ne l'interdit. Il faut donc rester ouvert à la fois à la possibilité qu'il soit pleinement justifié, et à celle qu'il ne le soit pas. Dans ce second cas l'IA forte pourrait être impossible pour raison de principe.*

2 - Si la première condition est remplie, il faut encore que le comportement de la matière impliquée dans l'émergence de la conscience soit entièrement compréhensible en termes calculables. C'est que les ordinateurs fonctionnent en termes calculables et déterministes – ce sont des machines logiques dites « machines de Turing », et le terme calculable veut d'ailleurs exactement dire « qui peut être calculé par une machine de Turing ». Il faut notamment, et pas seulement, que le hasard et l'indétermination décrits par la physique quantique n'aient qu'un rôle "spectateur" dans l'existence entre les oreilles d'un être humain vivant d'une « intelligence forte ». C'est une position déterministe et objectiviste

dans la compréhension de la conscience et de l'intelligence humaines. Qui là encore n'a rien n'évident - j'oserai dire, encore moins.

> *Incise – Et si cette condition avait **déjà** été démontrée fausse ?*
>
> *La question de savoir si le comportement de la matière à la base de la conscience est entièrement compréhensible en termes calculables, fait l'objet d'études et de discussions. Il faut toutefois signaler l'œuvre du physicien et mathématicien Roger Penrose, généralement reconnu comme l'un des plus grands esprits actuellement vivants, dans "*les Ombres de l'Esprit*".*
>
> *Aussi extraordinaire que cela puisse paraître, Penrose y a proposé une démonstration formelle du fait qu'une opération au moins de l'esprit humain – discerner la vérité mathématique d'une certaine proposition de logique avancée – est impossible si l'esprit humain est compréhensible en termes calculables et déterministes. Comme cette opération a bien lieu – elle est accessible à qui a un niveau licence en mathématiques, et Penrose guide le lecteur jusqu'à ce qu'il la réalise lui-même, ce qui rend sa démonstration particulièrement impressionnante – il s'ensuit que l'esprit n'est pas compréhensible en termes calculables. Donc ne peut être reproduit informatiquement. Il est naturel à partir de ce résultat de supposer que ce fonctionnement non calculable du cerveau humain ne se manifeste pas seulement dans l'opération mentale utilisée pour cette démonstration mathématique : si le cerveau humain possède cette caractéristique, elle est alors probablement générale à une partie importante de ses opérations, si ce n'est à la plupart.*
>
> *Dans la deuxième partie de son ouvrage, il propose une hypothèse – plus aventurée – sur une explication physique du caractère non calculable de l'esprit, basée sur des phénomènes quantiques intervenant notamment dans les microtubules des neurones et liés à la réduction du paquet d'ondes, phénomène postulé par la mécanique quantique, qu'elle ne cherche pas à expliquer, ce qui de l'avis de Penrose est insatisfaisant. Il s'agit là de physique spéculative, plus d'une idée de direction dans laquelle chercher la nouvelle physique – au-delà de la mécanique quantique donc – nécessaire à une compréhension scientifique de la conscience, si la démonstration de Penrose est correcte, que d'une quelconque précision sur ce que pourrait être cette physique.*
>
> *Dans ce livre, l'auteur répond à toutes les objections présentées à ses travaux précédents sur le même sujet. Le livre existe depuis une quinzaine d'années, et à ce jour personne n'a réussi à contrer son raisonnement ni à invalider la démonstration qu'il propose du caractère « non calculable » du fonctionnement du cerveau humain.*
>
> *S'il a raison, alors "*les Ombres de l'Esprit*" sera probablement considéré un jour comme un livre fondamental dans l'histoire scientifique. Et naturellement, le projet "Conscience artificielle" apparaîtra alors sans objet. Du moins dans sa version informatique : si l'hypothèse spéculative de Penrose s'avère judicieuse et est un jour développée, on ne peut exclure qu'il soit un jour possible de produire une conscience artificielle en se basant sur la physique nouvelle ainsi découverte. Mais alors, l'objet pensant produit ne serait pas un ordinateur... il en serait sans doute aussi différent par nature que l'ordinateur lui-même est par nature différent d'une machine à vapeur ou d'un marteau. Et naturellement les délais pour le réaliser seraient tout à fait indéterminés.*
>
> *Pour un résumé plus détaillé de ce livre, voir le commentaire de Jean Staune*



Roger Penrose - « Etant donné que la pensée inclut un élément non calculable, les ordinateurs ne pourront jamais faire ce que nous autres êtres humains faisons. » (source)

3 - Enfin, si les deux premières conditions sont vérifiées, il faut que cette conscience supposément compréhensible en termes matériels uniquement, et en termes calculables et déterministes exclusivement – il s'agit ici de la conscience qui se trouve dans le cerveau de l'inventeur – soit capable de concevoir le fonctionnement d'une autre conscience, celle que l'inventeur cherche à créer. Ce qui signifie que cette conscience présente dans son cerveau doit avoir la capacité de se comprendre elle-même ! En effet, si l'inventeur n'en était pas capable, comment pourrait-il déterminer les plans, méthodes et principes de la construction de l'IA ? Une troisième fois, cette position n'a rien d'évident – il est même permis de considérer qu'elle est la plus suspecte de toutes.

L'esprit peut-il se comprendre lui-même, donc contenir une description de lui-même ?

Les réponses à ces trois questions fondamentales sont à ce stade inconnues – sauf naturellement si la démonstration de Roger Penrose s'avère valide, auquel cas la seconde n'est pas vérifiée. Il existe des positions et des arguments philosophiques, naturellement, chacun avec leur validité. Mais il n'existe de réponse définitive au sens scientifique à aucune de ces questions. Peut-être cela changera-t-il un jour. En attendant, ces questions restent ouvertes.

Si la réponse à UNE SEULE des trois questions ci-dessus est négative, alors la création par voie informatique d'une conscience artificielle est irrémédiablement une chimère : on pourra reproduire sous forme informatique tel ou tel processus mental, ou en fournir un équivalent fonctionnel, on pourra créer des solutions logicielles pour traiter tel problème intellectuel particulier, parfois même mieux qu'un être humain – c'est d'ailleurs l'objet de la discipline IA, la vraie non la fantasmée, voir encore l'ouvrage sur le sujet du chercheur Jean-Gabriel Ganascia – on n'arrivera jamais à obtenir un objet avec qui on puisse sérieusement tailler une bavette, un objet qui serait quelqu'un. L'idée est alors à ranger dans le même rayon que les histoires de fées et du Père Noël.

Si et seulement si les TROIS réponses sont positives, alors la construction d'une conscience artificielle est théoriquement possible pour des êtres humains.

A ce sujet, on demandait la différence entre théorie et pratique. Un plaisantin répondit : "*En théorie, c'est la même chose. En pratique, non*"

Il est temps de parler des "menues" difficultés pratiques pour la création d'une conscience artificielle…

Ne nous étendons pas sur le fait que personne à ce jour n'a d'idée autre que très partielle et générale – moindrement testée en pratique donc – de comment au juste il faudrait s'y prendre. Si des pistes et réflexions diverses ont été proposées quant à l'aspect psychologique de la chose - voir par exemple pour le domaine francophone Paul Jorion dans Principes des systèmes intelligents (1989) ou Alain Cardon dans Un modèle constructible de Système Psychique (2011) *(texte complet en PDF)* parmi d'autres dans diverses langues - le fait même que ces pistes dont certaines sont anciennes n'aient pas permis d'aboutir à une réalisation concrète de type "quelqu'un dans la machine" montre que l'essentiel de la difficulté théorique reste devant nous. Ce que font les spécialistes en IA, ce qu'ils construisent dans la réalité, est bien différent, comme déjà dit. Et ce n'est pas faute d'essayer ni de réfléchir au moyen de construire une conscience.

Cet état de fait n'exclut cependant pas la possibilité que la discipline IA n'attende son Newton, son Galois, son Darwin ou son Einstein. Bref que la définition de la méthode générale ne soit à portée du prochain génie qui se penchera sérieusement sur la question. Le premier génie qui saura comprendre comment sa propre conscience fonctionne – rappelons que nous sommes dans l'hypothèse où la réponse à la question 3 ci-dessus serait positive, et où la chose ne serait pas une contradiction dans les termes.

Une difficulté plus grave se présente. C'est que une fois publiée la "*Théorie générale de l'esprit humain*" avec sa petite annexe "travaux pratiques – comment on fait" - par Jeannot Génie ou quel que soit son nom, il faudrait la construire pour de bon cette IA forte. Et là se situe un problème, **une difficulté... du format "mise en abîme"**.

Les passionnés d'affaires militaires comme les contrôleurs des programmes d'armement américains sont régulièrement entretenus des distrayantes nouvelles de l'avion de chasse F-35. Distrayantes pour l'observateur extérieur s'entend, non pour l'aviateur ni pour le contribuable américain. Non seulement le bouzin ne vole-t-il en effet que peu et

mal, mais surtout ses difficultés persistantes semblent bien résulter au fond d'un défaut de maîtrise de sa complexité, qu'un optimisme illuminé au moment de la conception initiale de l'engin a laissé croître au-delà de toute raison, comme d'ailleurs de toute nécessité. Si bien que sa complexité risque bien de s'avérer impossible à maîtriser par les équipes d'ingénierie, dont il est pourtant permis de penser qu'elles ne sont pas constituées d'amateurs, mais plutôt de certains des meilleurs des meilleurs, à la mesure des capacités financières de l'Oncle Sam à motiver ce genre de personnes pour travailler sur un projet essentiel à la perpétuation de la supériorité aérienne qui est un – sinon ce n'est le – pilier essentiel de sa suprématie militaire.

Ce phénomène est particulièrement criant s'agissant du logiciel embarqué et du système de maintenance informatisé du F-35 avec leurs 24 millions de ligne de code.

Pardon ?

Vous avez bien dit : 24 millions ? Une quantité aussi réduite d'instructions élémentaires, une complexité si ridiculement faible, tellement hors de proportion avec la complexité du fonctionnement d'un cerveau humain... et déjà les meilleures équipes au monde ne savent pas faire face !

Alors, quelles sont les chances que qui que ce soit arrive à appliquer les principes et méthodes découverts par Jeannot Génie, une fois qu'il aura eu l'obligeance de se présenter, à supposer qu'il le fasse jamais ?

Le système F-35 : les meilleures équipes, seulement 24 millions de lignes de code

Ça marchera bientôt, dites ?

Il est temps de conclure

Le projet "IA forte", c'est-à-dire une conscience artificielle « quelqu'un dans la boîte » basée sur un ordinateur :

- Soit est par principe impossible – c'est ce qui peut paraître le plus probable, les conditions pour qu'il en soit différemment étant plusieurs, chacune d'entre elle impérative, dont l'une déjà fortement mise en doute par ce qui ressemble fort à une démonstration formelle. Cependant la chose n'est pas définitivement prouvée

- Soit est théoriquement possible, auquel cas il sera peut-être réalisé une fois que ces deux conditions auront été remplies :

    1. Apparition du génie capable de comprendre le fonctionnement de sa propre conscience

    2. Evolution d'une civilisation capable de coordonner les talents, les intelligences, bref de faire travailler ensemble des êtres humains à un niveau fantastiquement au-delà de ce à quoi l'humanité est parvenue à ce jour. En somme, atteignant une capacité de « maîtrise de la complexité » bien au-delà de celle à laquelle nous sommes déjà parvenus. Au point de pouvoir réaliser des projets informatiques incluant à coup sûr des milliards, peut-être même des milliers de milliards d'instructions

Dans cette deuxième hypothèse, l'humanité réalisera peut-être en effet un jour une IA véritable, dont l'orientation vis-à-vis de l'humanité posera effectivement question. Ce jour n'arrivera pas du vivant d'aucun être humain d'aujourd'hui. Il est fort possible qu'il soit aussi loin de nous que ne l'était la construction d'un réacteur atomique au moment où Démocrite spéculait sur l'existence de l'atome au Vème siècle avant notre ère...