



Google DeepMind

GOOGLE DEEPMIND

L'INTELLIGENCE HUMAINE EST-
ELLE DÉJÀ DÉPASSÉE ?

SOMMAIRE

SOMMAIRE	2
INTRODUCTION	3
GOOGLE DEEPMIND	4
▪ GOOGLE DEEPMIND : LA MACHINE SURPASSE UNE NOUVELLE FOIS L'INTELLIGENCE HUMAINE	4
▪ LES MACHINES À APPRENDRE : RÉSEAUX NEURONAUX ET APPRENTISSAGES EN PROFONDEUR	6
▪ DES PROGRAMMES INFORMATIQUES QUI APPRENNENT À JOUER	7
▪ LES PERFORMANCES DERRIÈRE LA VICTOIRE D'ALPHAGO	8
▪ DES MACHINES QUI SE SOUVIENNENT, LE DÉFI DES INTELLIGENCES ARTIFICIELLES À VENIR	9
À VOIR PLUS LOIN	10
▪ BILL GATES SE MÉFIE DE L'INTELLIGENCE ARTIFICIELLE	10
▪ L'INTELLIGENCE ARTIFICIELLE INAMICALE : RÉALITÉ OU FANTASME ?	12
▪ COMMENT DONNER UNE CONSCIENCE À UN ORDINATEUR ?	16

INTRODUCTION

Près de 20 ans après avoir battu Gary Kasparov, champion du monde d'échecs en titre, un nouveau coup a été porté par la machine à la suprématie de l'intelligence humaine. L'équipe de développement de DeepMind, le programme d'intelligence artificielle de Google, a annoncé la victoire de son programme AlphaGo contre le joueur professionnel et champion d'Europe de go en titre, Fan Hui. Sommes-nous en train d'assister à la naissance d'une nouvelle ère ? Que vaut vraiment l'intelligence artificielle face au cerveau humain ?

GOOGLE DEEPMIND

GOOGLE DEEPMIND : LA MACHINE SURPASSE UNE NOUVELLE FOIS L'INTELLIGENCE HUMAINE



En 1997, DeepBlue, le super-ordinateur d'IBM, spécialement programmé pour l'occasion, battait Gary Kasparov, champion du monde d'échecs en titre. 19 ans après ce premier acte, un nouveau coup a été porté par la machine à la suprématie de l'intelligence humaine dans leur art de prédilection, les jeux de plateau.

Le jeu de go, jeu millénaire dont les origines légendaires se confondent avec les fondements de la civilisation chinoise, gardait la réputation d'être irréductible au calcul informa-

tique, même si des programmes récents approchaient d'un niveau respectable.

Au long d'une divulgation médiatique soigneusement préparée par la publication d'un article dans la revue Nature[1], l'équipe de développement de DeepMind, le programme d'intelligence artificielle de Google, a annoncé la victoire de son programme AlphaGo contre le joueur professionnel et champion d'Europe de go en titre, Fan Hui, dans un match en cinq manches disputé au mois d'octobre, où le champion a chaque fois été dominé.

Ce résultat, inespéré autant que spectaculaire, prend appui sur la technologie des réseaux neuronaux, inspirée du fonctionnement cognitif du cerveau humain. Elle vise spécifiquement à fonder le fonctionnement de la machine sur un processus d'apprentissage, inscrit dans un « entraînement », pour répondre à un objectif donné.

Cette nouvelle prouesse issue des laboratoires du géant californien laisse présager d'applications numériques nombreuses.

En arrière-plan se joue une bataille qui promet d'être homérique avec Facebook, attelé au développement d'une technologie similaire et pour l'heure distancé.

Fort de son avance, Google impose son calendrier : prochaine étape, son programme AlphaGo affrontera Lee Sedol, joueur de go coréen au sommet de la hiérarchie mondiale, dans une nouvelle série de rencontres prévues au mois de mars.

[1] Silver D. *et al.*, 2016, « Mastering the game of Go with deep neural networks and tree search », *Nature*, vol. 529, 28 janvier 2016, p. 484-489.

Par Etienne Monin

03/02/2016

LES MACHINES À APPRENDRE : RÉSEAUX NEURONAUX ET APPRENTISSAGES EN PROFONDEUR

L'intelligence artificielle adopte l'architecture d'un « réseau neuronal », organisé en couches de neurones interconnectées. Le programme DeepMind de Google appréhende un environnement donné et l'interprète au travers de son système neuronal pour y répondre.

Son algorithme décompose l'analyse d'un ensemble de données complexes couche après couche pour extraire de cette information brute un schéma structuré, « créer des règles à partir de rien ».

Le programme initial mis en place au Google X Lab en Californie en 2011 consistait en un réseau d'un million de neurones simulés, dotés d'un milliard de connexions, répartis entre 1000 ordinateurs. Il a montré la performance des réseaux neuronaux dans l'identification d'images et la reconnaissance vocale, suivant la méthode de « l'apprentissage en profondeur ».

Cette approche, initiée dans les années 1980, vise à renforcer les interconnexions pertinentes entre couches neuronales simulées. Yann LeCun, aujourd'hui directeur de la recherche en intelligence artificielle à Facebook, produit à cette époque un programme d'identification des chèques bancaires.

La révolution numérique des années 2000 ouvre à de nouvelles expériences par la masse des données disponibles et la puissance de calcul accrue. Les projets menés autour de la reconnaissance vocale améliorent de 25 % la qualité dans le cas du logiciel Siri intégré à la plateforme IOS d'Apple.

L'algorithme d'identification d'images remporte en 2012 la compétition ImageNet : les programmes « apprennent » à partir d'un stock d'un million d'images étiquetées en caté-

gories. Un test les soumet à une série d'images inconnues dont ils doivent dire l'étiquette.

Le taux d'erreur passe avec ce nouvel algorithme de 25 % à 15 %. Pour cela, les « filets neuronaux profonds » - *deep neural nets* - exploitent une deuxième stratégie, l'apprentissage par renforcement, qui institue un système de décision, par des mécanismes de récompense et de validation, comme un score, à chaque expérience.

L'algorithme apprend par essai-erreur quelles actions maximisent sa récompense à n'importe quel moment en environnement complexe. Plus la quantité d'information traitée est importante, plus l'algorithme s'améliore.

Par Etienne Monin

03/02/2016

DES PROGRAMMES INFORMATIQUES QUI APPRENNENT À JOUER

Les jeux informatiques servent de terrain privilégié d'expérimentation de ces intelligences artificielles « apprenantes ». Un programme conçu à l'université d'Alberta (Canada) s'est révélé imbattable dans une variante du poker Texas Hold'em (partie en face à face) après 1500 parties jouées.

03/02/2016

L'algorithme, dit de « minimisation rétrospective du regret », construit un arbre de décision à partir des résultats de chaque partie, associés à un degré de « regret ». Il réévalue également les décisions prises dans les tours précédents.

Le bluff est généré par fréquences statistiques calculées à partir des observations. Ces technologies font envisager des applications pour l'aide à la décision dans les domaines médicaux, biologiques ou financiers.

L'innovation de l'algorithme DeepMind est de pouvoir évoluer dans des environnements différents et d'y former chaque fois par apprentissage un schéma d'action. Il apprend formellement à reconnaître les configurations favorables de pixels.

Les ingénieurs l'ont d'abord entraîné à jouer à 49 jeux d'arcades, tels Pong, Space invaders et Breakout. Dans chacun d'eux, l'algorithme a décidé d'un comportement jusqu'à jusqu'à totalement « maîtriser » le jeu. Demis Hassabis, co-fondateur de DeepMind, en fait la démonstration avec le jeu Breakout (une plaque mobile renvoie une balle verticalement pour casser des briques).

L'algorithme après 300 parties gagne à tous les coups. Au bout de 500 parties, il a identifié une stratégie optimale, renvoyant la balle de manière à creuser dans les briques par le dessus.

Par Etienne Monin

LES PERFORMANCES DERRIÈRE LA VICTOIRE D'ALPHAGO

Les 361 intersections du goban, le plateau de jeu, fournissent des milliards de configurations possibles des pierres blanches et noires dans une partie moyenne. L'algorithme de Monte Carlo en vigueur dans les programmes courants simule et évalue à chaque étape du jeu les positions gagnantes ou perdantes. Mais ces machines ne peuvent anticiper un grand nombre de coups, là où l'intuition humaine apprécie l'équilibre général et le potentiel local dans la disposition des pierres.

Des pierres de handicap (coups d'avance) leur sont encore attribuées dans les parties contre des professionnels. AlphaGo explore aussi les données de cette manière, en comparant la partie en cours à des échantillons de trente ans millions de positions de parties professionnelles inventoriées.

AlphaGo a en outre été entraîné en jouant virtuellement avec 50 autres ordinateurs, améliorant son système d'interprétation et de prédiction par l'apprentissage renforcé.

Fan Hui, son adversaire, a reconnu qu'il aurait pu croire affronter un humain au jeu parfois « étrange » mais « très fort ». Il a adopté un style plus agressif après la première partie, sans déstabiliser la machine au style plus défensif, « territorial ». Il abandonne dans la cinquième partie, ayant accumulé un retard de dix points.

Un commentateur professionnel coréen sur un portail de go juge qu'AlphaGo a surpassé son adversaire par des coups « solides et souples », sans commettre de graves erreurs.

L'enthousiasme de l'équipe de Google est d'autant plus grand que personne ne s'attendait à ce qu'un ordinateur parvienne aussi rapidement à ce niveau de jeu.

Pour son artisan Demis Hassabis, AlphaGo a acquis la capacité d'inventer par lui-même des choses nouvelles,

qui repoussent ce que les meilleurs humains avaient pu concevoir.

Le prochain affrontement avec le coréen Lee Sedol s'annonce ainsi ébouriffant. La communauté du go soutient son champion mais la balle est dans le camp de l'intelligence artificielle qui poursuit sa préparation depuis sa victoire d'octobre 2015.

Lee Sedol a annoncé « considérer ce match comme un moment important de l'histoire du go » mais demeurer « confiant dans le fait de pouvoir gagner. » Un programmeur de Deep Mind se demande quelles ruses il trouvera à opposer à l'algorithme, « qui ressemble à un mur ».

Par Etienne Monin

03/02/2016

DES MACHINES QUI SE SOUVIENNENT, LE DÉFI DES INTELLIGENCES ARTIFICIELLES À VENIR

Les techniciens ambitionnent désormais de construire un système d'intelligence artificielle « généralisé », qui tire de ses apprentissages établis dans des environnements préalables des moyens transmissibles à d'autres tâches.

La satisfaction de l'équipe Google est compréhensible, ses réalisations anticipent de dix ans les prédictions des spécialistes.

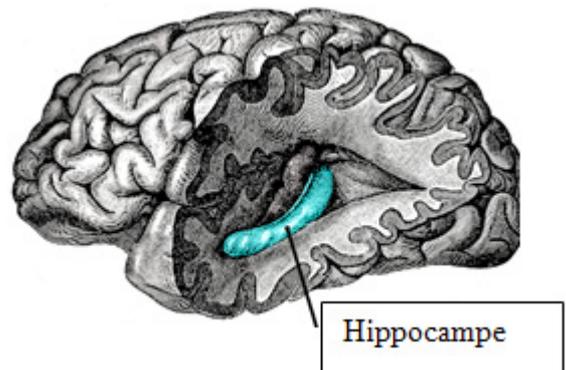
La centaine de chercheurs réunis à Londres a montré son efficacité depuis l'absorption de DeepMind en 2014, en prenant le meilleur sur le programme *Darkforest* de Facebook, lui aussi chargé de s'ouvrir la voie du succès au jeu de go.

Les spécialistes du secteur sont sur la brèche. Les résultats et codes révélés sont vite assimilés par les laboratoires en pointe qui les mettent à leur tour en œuvre.

Les techniciens ambitionnent désormais de construire un système d'intelligence artificielle « généralisé », qui tire de ses apprentissages établis dans des environnements préalables des moyens transmissibles à d'autres tâches. C'est aussi la capacité humaine à décomposer les problèmes en problèmes plus petits et à envisager les conséquences d'actions distantes.

Les chercheurs s'intéressent ainsi au fonctionnement de l'hippocampe, siège de la mémoire, pour reproduire les mécanismes de transfert d'expérience et ainsi créer des automates d'un nouveau genre.

Leur couplage avec d'autres technologies informatiques,



telle l'ordinateur quantique élaborée par la société D-Wave au Canada pour résoudre des problèmes mathématiques, elle aussi passée dans le giron de Google.

Par Etienne Monin

03/02/2016

À VOIR PLUS LOIN

BILL GATES SE MÉFIE DE L'INTELLIGENCE ARTIFICIELLE

Aux yeux de Bill Gates, l'intelligence artificielle est un danger à plus ou moins long terme. Mais faut-il vraiment en avoir peur ? L'être humain est-il condamné à l'obsolescence ?

En matière d'intelligence artificielle, Mr **Gates** n'est pas le seul à penser que la situation pourrait nos échapper. D'autres pointures telles que **Elon Musk**, l'emblématique patron de **Tesla Motors** et de **Space X**, ainsi que Stephen Hawking, professeur de mathématiques et physicien théoricien britannique dont la « vie extraordinaire » a été traduit récemment sur grand écran, ont le même point de vue... qui est que nous ne devons pas concevoir des machines au pouvoir de réflexion supérieur à celui de l'être humain.

La superintelligence artificielle est une menace pour l'humanité. Celui qui a cofondé Microsoft en 1975 a d'ailleurs exprimé son inquiétude lors d'une discussion sur le site web communautaire **Reddit**. Il précise « qu'au début les machines accompliront de nombreuses tâches pour nous, que cela devrait être positif si nous le gérons bien » mais qu'à un moment donné, probablement « dans quelques décennies, l'intelligence des machines sera assez forte pour devenir un sujet de préoccupation ».

Ce n'est donc pas une menace immédiate, rassurons-nous, l'intelligence artificielle est encore balbutiante. Non, il s'agit avant tout d'une mise en garde encourageant à la méfiance. Pour Bill Gates, le développement de l'intelligence artificielle doit être assuré avec une grande précaution.

Car si l'IA est encore limitée, il redoute le jour où les robots seront dotés d'une conscience. Il rejoint ainsi la position d'Elon Musk, qui dans ce souci d'encadrement de la recherche, s'est délesté de dix millions de dollars au profit de l'association Future of Life Institute.

Il semble évident et nécessaire qu'une réflexion plus pous-

sée sur le sujet devrait être menée au niveau national et international. Scientifiques et entrepreneurs de pointe sont de plus en plus nombreux à penser qu'il faudrait instaurer un cadre éthique dans ce domaine, ajuster les lois, pour s'assurer simplement comme le dit le dirigeant de Tesla « que nous ne faisons rien de stupide ».

D'un point de vue extérieur cependant, il peut paraître assez étrange que ce soient ceux-là même qui conçoivent actuellement l'intelligence artificielle qui en ait le plus peur. Mais c'est aussi rassurant, d'une certaine manière, de voir que son développement n'est pas pris à la légère. Et que des arrangements sont mis en place afin d'éviter, comme le prédit funestement Stephen Hawking, que les humains ne soient dépassés par le développement de l'intelligence artificielle, « limités par leur lente évolution biologique », et ne soient en fin de compte devenus inutiles et remplacés comme nous remplaçons nos outils électroménagers devenus obsolètes.

Reste à savoir si cela représente une technique de communication, en amont, pour étouffer les craintes du public quant aux machines intelligentes à venir, étayées à la fois par ces sommités et par de nombreux films de science-fiction (Matrix, I. Robot, Blade Runner...). Car en affichant ainsi leurs préoccupations, ils démontrent aussi leur sérieux.

Alors faut-il vraiment avoir peur des machines autonomes comme le pensent Bill Gates, Elon Musk et Stephen Hawking ? À priori non, pas dans l'immédiat. La complexité de l'intelligence humaine est telle que nous ne sommes pas en mesure de faire émerger une conscience artificielle. L'architecture informatique est encore très basique et l'idée même qu'une intelligence artificielle se développe elle-même à un rythme exponentiel demeure pour le moment dans le domaine de la science-fiction.

En revanche, définir les effets de la robotisation sur l'emploi reste difficile. Sur ce point, les avis divergent. Les partisans estiment que sans une plus forte robotisation, l'industrialisation française coure à sa perte, menacée par les délocalisations des entreprises. La robotisation serait de fait un outil pour baisser les coûts de production qui nous permettrait d'atteindre une compétitivité égale à celle de la Chine. C'est en tout cas ce que pense le syndicat des machines et technologies de production (Symop.) Alors que les opposants craignent que la robotisation entraîne un plus haut taux de chômage, que les tâches à l'origine destinées aux humains soient effectuées par les machines ou les ordinateurs.

Seul l'avenir nous dira si notre confiance dans les machines n'était pas erronée et que cette crainte étayée par les sommités de l'innovation n'était pas exagérée.

Par Sébastien Tribot

23/03/2015

L'INTELLIGENCE ARTIFICIELLE INAMICALE : RÉALITÉ OU FANTASME ?

Les recherches en intelligence artificielle font de tels progrès que des scientifiques réputés, comme le physicien Stephen Hawking, voient dans l'I.A. un danger pour l'Homme. Pourtant, selon les spécialistes, nous sommes très loin de voir un jour surgir une I.A. inamicale.

“Terminator”, “2001, l'odyssée de l'espace”, “I, Robot” sont autant de films de science-fiction montrant comment l'intelligence artificielle deviendra un jour consciente d'elle-même, jusqu'à nous supplanter et menacer notre intégrité physique. Pourtant, dans la réalité, nous sommes loin, très loin de ce type de scénario.

Jean-Gabriel Ganascia, expert en intelligence artificielle et philosophe, chercheur au LIP6, le Laboratoire d'informatique de Paris 6, raconte l'histoire de l'I.A. : “ce terme est né en 1955, aux États-Unis. Il s'agissait d'étudier l'intelligence avec les machines, en simulant ses différents aspects”. Mais rapidement, déplore-t-il, “ce terme est devenu ambivalent, avec d'un côté le projet scientifique d'étude de l'intelligence avec des moyens artificiels (l'I.A. faible), et de l'autre, l'idée (entretenu par des chercheurs passionnés de science-fiction) de créer une intelligence avec les moyens de l'artificiel (l'I.A. forte)”.

Pour l'expert, l'idée qu'une machine “va tout à coup raisonner comme un homme et qu'un jour une intelligence artificielle générale aura conscience d'elle-même” n'est qu'un “pur fantasme”, qui n'existe pas sur le plan scientifique. Ainsi, il n'existerait, indique Jean-Gabriel Ganascia, qu'une seule intelligence artificielle : l'I.A. faible. Autrement dit, la simulation de l'intelligence et des capacités cogni-



tives de l'être humain, via des techniques informatiques (algorithmes, programmation).

“C'est cette I.A. que l'on retrouve dans les systèmes de recommandation, dans les voitures intelligentes, dans les drones militaires, dans les systèmes de reconnaissance d'empreintes, ou dans la traduction automatique”, explique le chercheur.

“L'I.A. est plus dangereuse que les armes nucléaires”

Pourtant, si l'on en croit des scientifiques réputés, en majorité anglo-saxons, l'I.A. forte, c'est pour demain. Elon Musk, président-fondateur de Tesla Motors et de SpaceX, estime ainsi qu'il faut rester “très prudent” à propos de l'intelligence artificielle. “Si je devais miser sur ce qui constitue notre plus grande menace existentielle, ce serait probablement ça”, a-t-il lancé en 2014 lors de l'[AeroAstro Centennial Symposium](#) du MIT. Pour lui, “L'I.A. est plus dangereuse que les armes nucléaires”.

De son côté, Stephen Hawking, physicien théoricien de renommée mondiale, [joue au prophète](#) : “le développement d’une pleine I.A. pourrait signifier la fin de la race humaine. Si elle finissait par devenir autonome, ce qui est probable, les humains, limités par leur lente évolution biologique, ne pourraient pas rivaliser”.

Avec Elon Musk et 700 autres scientifiques, Stephen Hawking a publié [une lettre ouverte](#) sur les dangers de l’I.A., qui se veut alarmiste : “en raison de son grand potentiel, il est important de se demander comment nous pourrions en profiter tout en évitant les pièges possibles”.

I.A. “superintelligente” et robots assistants

En 2014, le philosophe suédois [Nick Bostrom](#), connu pour être un des penseurs du [transhumanisme](#), a publié un ouvrage intitulé “Superintelligence”, dans lequel il analyse les bienfaits et les dangers de l’apparition d’une I.A. “[superintelligente](#)”.

Pour Nick Bostrom, si une telle entité venait à s’en prendre à l’homme, c’est surtout à cause de “l’orthogonalité” entre volonté et intelligence : pour poursuivre un but simple, une superintelligence choisira une solution simple. “Imaginons par exemple que nous lui demandions de fabriquer le plus grand nombre de trombones possible. Obnubilée par sa tâche, elle en viendrait à considérer que l’humanité est un obstacle à la production de trombones puisqu’elle peut décider à un moment ou à un autre que le compte y est. Logiquement, la super-intelligence se dirait que les humains ne sont après tout qu’un amas d’atomes qui, reconfigurés autrement, permettraient d’obtenir d’excellents trombones”, écrit le philosophe.

Les recherches en robotique faisant des pas de géant, le pire est-il à craindre ? Nous sommes encore loin d’avoir chacun un robot assistant à la maison, mais peu à peu, les robots commencent à investir notre quotidien.

Au Japon, dans l’usine de la [société Glory](#), en banlieue de Tokyo, [des humanoïdes-ouvriers font leur gym le matin](#) en même temps que leurs collègues humains. En Allemagne,

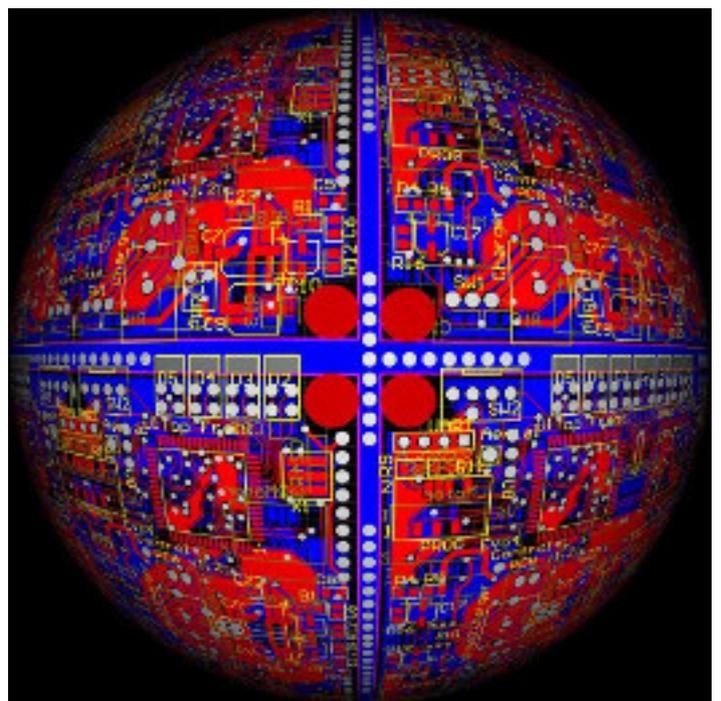
le robot de service “Max” aide des personnes âgées dans des maisons de retraite, comme le montre le documentaire édifiant d’Arte, “[Au coeur des robots](#)”.

Dans le domaine civil, nous n’en sommes encore qu’aux prémices d’une société robotisée. Mais en ce qui concerne le secteur militaire, c’est une autre histoire.

Robots tueurs et machines “autonomes”

Certes, il n’existe pas encore de “[robots tueurs](#)” autonomes, capables de décider seuls s’ils doivent tirer ou pas. Mais les chercheurs des armées du monde entier, à commencer par la [DARPA](#), l’agence de recherche de l’armée américaine, travaillent à créer des “armes semi-autonomes” : ces avions ou frégates sans pilote, ces robots et ces machines nouvelle génération analysent une situation donnée, décident de la réponse à apporter, et en dernier ressort, un humain valide ou non la suite à donner.

En Corée du Sud, les [SGR-A1](#), des “robots sentinelles” armés de lance-grenades et de mitrailleuses, défendent les frontières. La décision finale [dépend d’un opérateur humain](#), mais rien ne dit que cette situation ne changera



pas un jour.

Alors que depuis deux ans, [Human Rights Watch](#) (HRW) mène une [campagne](#) pour stopper ces “robots létaux autonomes” (RLA), Elon Musk et Stephen Hawking ont publié cet été une autre [lettre ouverte](#), pour réclamer l’interdiction des armes autonomes et stopper une potentielle “course à l’armement” dans ce domaine.

Ces machines autonomes, qui existeront “probablement d’ici quelques années”, seraient là encore, pour Elon Musk et Stephen Hawking, “plus dangereuses que les armes nucléaires”. Pour [Hod Lipson](#), ingénieur en robotique et directeur du [Creative Machines Lab](#) de l’Université Cornell, à New-York, “combinée avec des robots physiques, l’I.A. pourrait effectivement être dangereuse”. D’où l’idée que nous devrions nous “préparer dès maintenant” à l’arrivée de l’I.A. forte, et penser aux moyens de la canaliser.

“Rendre une machine consciente, c’est de la pure science-fiction”

Jean-Gabriel Ganascia insiste : “il est facile de simuler l’intelligence et des émotions, mais rendre une machine consciente, c’est de la pure science-fiction”. Et de constater : “il y a derrière tout cela une croyance en la toute-puissance de la technologie, très proche d’une religion... peut être celle des [transhumanistes](#)”.

En revanche, l’expert estime que l’I.A. véritable, la “faible”, comporte des risques à ne pas négliger. “Comment réagira une voiture sans chauffeur confronté à un dilemme, comme choisir entre un carambolage et écraser un piéton ? Les machines demeurent imprévisibles, elles peuvent faire des erreurs”, explique-t-il.

Restent des risques “plus factuels”, comme l’utilisation de nos données par les fabricants des systèmes d’I.A. - en premier lieu, des robots de compagnie, qui en sauront beaucoup sur nous : “les données utilisées par les systèmes d’I.A. se propagent, se dupliquent, et sont forcément stockées quelque part pour être réutilisées. Il ne faut pas oublier le scénario d’une utilisation à mauvais escient de



ces nouveaux outils, par des personnes mal intentionnées”, remarque Jean-Gabriel Ganascia.

Au delà du scénario d’une intelligence artificielle inamicale façon Terminator, c’est bel et bien “l’impact de l’I.A. faible sur nos vies qui mérite d’être posé”, conclut-il. D’où son idée d’instaurer “un certain nombre de règles et principes éthiques” - pourquoi pas inspirés des lois d’Asimov.

Par Fabien Soyez, journaliste scientifique

25/12/2015

COMMENT DONNER UNE CONSCIENCE À UN ORDINATEUR ?

Le chercheur le plus connu du monde dans le domaine de la conscience dit que les machines pourraient un jour devenir conscientes d'elles-mêmes.

Pourquoi c'est important :

Les scientifiques ne tombent pas d'accord sur ce qu'est la conscience.

Un ver de terre est-il conscient ? Un bourdon ? Est-ce qu'un ordinateur qui peut jouer aux échecs ressent quelque chose ?

Pour Christof Koch, directeur scientifique du Allen Institute for Brain Science [Institut Allen pour la science du cerveau] à Seattle, la réponse à ces questions pourrait se trouver dans la trame constitutive de l'univers. La conscience, pense-t-il, est une propriété intrinsèque de la matière, tout comme la masse ou l'énergie. Organisez la matière de la bonne façon, comme dans le cerveau mammifère, et tada ! on peut ressentir.

M. Koch, aujourd'hui 57 ans, a passé presque un quart de siècle à essayer d'expliquer pourquoi, par exemple, le soleil nous fait chaud sur le visage. Mais après avoir écrit [trois livres sur la conscience](#), M. Koch dit que les chercheurs sont encore loin de savoir pourquoi la conscience survient, ou même de tomber d'accord sur ce que c'est. C'est un problème difficile (voir [Cracking the Brain's Codes](#)). C'est l'une des raisons pour lesquelles M. Koch a quitté son poste à Caltech [California Institute of Technology] en 2011 pour rejoindre un projet à 500 millions de dollars lancé par le milliardaire Paul Allen, co-fondateur de Microsoft.

Le but de l'Institut Allen est de construire un atlas détaillé de chaque neurone et chaque synapse du cerveau mammifère. Cela fournirait aux neurosciences un geyser* de données similaire à ce qu'a accompli le Projet Génome

Humain.

Mais M. Koch n'a pas abandonné sa recherche d'une grande théorie qui pourrait tout expliquer. En fait, il pense que la conscience pourrait être expliquée par ce qu'on appelle la théorie de l'information intégrée, laquelle affirme que la conscience est un produit des structures, comme le cerveau, qui ont à la fois la capacité d'emmagasiner une grande quantité d'informations, et une densité critique d'interconnexions entre leurs parties constitutives.

Pour M. Koch, cette théorie fournit un moyen d'évaluer des degrés de conscience chez les personnes ayant subi des dommages au cerveau, chez les espèces du règne animal, et même, dit-il, chez les machines. Nous avons interrogé M. Koch sur la conscience des ordinateurs pendant la conférence EmTech de MIT Technology Review.

La découverte du fondement biologique de la conscience sera-t-elle déshumanisante en aucune façon ? Et si ce n'était qu'une illusion ?

Je trouve ridicule cette vision de certaines personnes que la conscience est une illusion. Si c'est une illusion, alors c'est l'illusion la plus puissante que nous ayons. Et la plus célèbre déduction de la philosophie occidentale, c'était quoi déjà ? « Je pense. Donc je suis ». Le fait que nous possédions l'expérience consciente est la seule certitude indéniable que nous ayons.

Si les scientifiques découvrent le fondement de la conscience, quels genres de technologies pourraient en résulter ?

On pourrait avoir un test pour déterminer qui a une conscience et qui n'en a pas. Nous avons des débats passionnés dans ce pays sur l'avortement. J'aimerais avoir un moyen objectif de tester à quel moment un fœtus commence véritablement à avoir des sensations conscientes.

Ou bien si un patient [dans le coma] est conscient ou non. Souvent, on ne sait pas, en fait. Ce sont des questions que les gens se posent depuis la nuit des temps, mais une fois qu'on aura une théorie, et une théorie acceptée par le plus grand nombre, on pourra aussi y répondre. Et puis, si je voulais construire une machine qui soit consciente, ça me donnerait un plan directeur.

Donc vous pensez qu'un ordinateur peut être conscient ?

J'ai fait récemment un cours magistral au MIT sur la théorie de l'information intégrée, développée par Giulio Tononi à l'Université du Wisconsin. C'est une théorie qui fait une prédiction très claire : elle dit que la conscience est une propriété des systèmes complexes qui possèdent un répertoire « cause-effet » particulier. Ils ont une façon particulière d'interagir avec le monde, comme le fait le cerveau, ou en principe, comme pourrait le faire un ordinateur. Si vous arriviez à construire un ordinateur qui a les mêmes circuits et branchements que le cerveau, cet ordinateur serait aussi doué de conscience. Il pourrait ressentir et ça lui ferait quelque chose d'être cet ordinateur-là. Cependant, il n'en est pas de même des simulations digitales.

Si je construisais un modèle logiciel parfait du cerveau, il ne serait jamais conscient, mais une machine spécialement conçue qui imite le cerveau le serait ?

Exact. Cette théorie dit clairement qu'une simulation digitale ne serait pas consciente, ce qui est radicalement différent de la croyance fonctionnaliste de 99 % des gens du MIT ou des philosophes comme Daniel Dennett. Ils disent tous qu'une fois qu'on arrivera à tout simuler, il n'y aura besoin de rien d'autre, et ce sera conscient.

Je pense que la conscience, comme la masse, est une propriété fondamentale de l'univers. L'analogie, et elle est très bonne, est qu'on peut faire de très bonnes prédictions météo de nos jours. On peut prédire l'intérieur d'une tempête. Mais ce n'est jamais mouillé à l'intérieur de l'ordinateur. On peut simuler un trou noir par ordinateur, mais l'espace-temps n'y sera jamais déformé. La simulation n'est pas vraie.

C'est la même chose avec la conscience. Dans 100 ans, on pourra peut-être simuler la conscience sur un ordinateur. Mais l'ordinateur ne vivra, ne sentira rien. Nada. Ce sera tout noir dedans. Il ne pourra pas avoir de ressenti en tout état de cause, même s'il peut avoir notre intelligence et notre capacité de parler.

Je ne suis pas en train de dire que la conscience est une âme magique. Non. C'est quelque chose de physique. La conscience survient toujours sur du physique. Mais il faut un type de matériel informatique particulier pour l'instancier. [Ndlt : instancier = en programmation informatique, créer un élément à partir d'un modèle et d'un jeu de valeurs initiales de paramètres.] Un ordinateur fait de transistors, qui déplacent des charges à travers des portes électroniques, avec chacune de ces portes connectée à quelques autres portes, est une structure cause-effet très différente de celle que nous avons dans le cerveau, où on a un neurone connecté à 10.000 neurones afférents et projetant vers 10.000 autres neurones. Mais si vous construisiez l'ordinateur de façon appropriée, comme un ordinateur neuromorphique ([voir Thinking in Silicon](#)), il pourrait être conscient.

Si je vous enfermais dans une pièce avec un ordinateur du futur, seriez-vous capable de déterminer s'il est conscient ?

Je ne pourrais pas de l'extérieur. Il faudrait que j'examine la mécanique.

Et avec le test de Turing ?

La question posée par le test de Turing est : « les machines peuvent-elles penser ? ». Mais finalement, c'est un test opérationnel pour l'intelligence, et non pour la conscience. Si vous avez une conversation intelligente avec un type dans une autre pièce, et qu'après une demi-heure vous ne pouvez pas décider s'il s'agit d'un ordinateur ou d'un humain, ben, là vous dites que l'ordi est aussi intelligent qu'un humain. Mais le test de Turing ne me dirait pas si l'ordinateur a ressenti quoi que ce soit. Je pourrais lui demander « Es-tu conscient ? » et la machine pourrait me répondre : « Mais oui, je suis pleinement conscient. Et pourquoi affirmez-vous que je ne le suis pas ? Je suis offensé ! » Mais je

ne pourrais pas vraiment savoir. Je serais obligé de dire : « Désolé, il va falloir que je te démonte pour savoir comment tu es fait et comment tu fais pour générer ces états physiques différents ».

N'y a-t-il pas une question-piège à poser, à laquelle seul un être conscient pourrait répondre ?

Très bonne question. Chez les humains, nous avons des tests pratiques pour évaluer la conscience. S'il vous arrive un accident grave et que vous atterrissez aux urgences, ils vous demanderont : « Pouvez-vous bouger les yeux ? Pouvez-vous bouger vos membres ? Pouvez-vous parler ? Si vous pouvez parler, savez-vous en quelle année nous sommes ? Savez-vous qui est le président ? »

Mais comment puis-je vraiment savoir que vous êtes conscient ? C'est le problème du solipsisme. Dans l'analyse précédente, je ne sais pas. Mais je sais que votre cerveau est très similaire au mien. J'ai passé plein de gens au scanner, et je sais qu'ils ont tous un cerveau, et que leur cerveau se comporte similairement au mien. Donc je peux faire la déduction parfaitement raisonnable que vous aussi, vous êtes conscient.

Mais plus ces systèmes différent de moi, plus il est difficile de faire cette démarche par déduction. Par exemple, prenez une abeille. Est-ce qu'elle ressent quelque chose du fait d'être une abeille et de voler dans les rayons dorés du soleil et de boire du nectar ? Je trouve qu'il est très difficile de savoir si une abeille est consciente ou non. Et un ordinateur est encore plus radicalement différent. Il n'existe aucun comportement duquel je puisse le rapprocher par analogie. Je serais obligé d'examiner la mécanique sous-jacente.

Pensez-vous qu'on construira un jour des machines conscientes ?

Je ne suis pas sûr de savoir pourquoi on voudrait faire ça. Mais il n'y a aucun doute dans mon esprit qu'on construira des machines intelligentes capables de passer le test de Turing bien avant de comprendre le fondement biologique de l'intelligence humaine. Et je pense qu'il y a des dangers

qui vont avec ça, que la plupart des gens, en optimistes insouciantes, ignorent complètement.

Quels dangers ?

Vous ne regardez pas les films de science-fiction ? « L'intelligence artificielle hors de contrôle », bien sûr. Pensez aux marchés financiers, à toutes ces machines d'échanges boursiers, tous ces flash-crashes. Les gens vont abuser de l'intelligence des ordinateurs, la maximiser aveuglément pour un quelconque but. Ça mènera à une concentration de plus en plus grande du pouvoir entre des mains de moins en moins nombreuses. On le voit déjà, ça va conduire à du chômage en masse. Et peut-être dans 30 ou 40 ans, je pense qu'il y a vraiment un risque existentialiste pour l'espèce, de l'ampleur des armes nucléaires ou d'une chute de météorite.

Et tout ça sans que les machines soient conscientes ? Dans les films, c'est précisément au moment où l'IA devient folle qu'elle gagne la conscience d'elle-même.

Ca, c'est parce que les gens veulent faire une histoire captivante. Si l'ennemi ne ressentait rien, s'il n'y avait rien en face, ça ne ferait pas un bon adversaire.

* ndlt : geyser de données, ou data firehose en anglais = lorsqu'un fournisseur de données sociales dynamiques (de type Twitter, par exemple) se met d'accord avec un utilisateur final pour lui envoyer la totalité des données dynamiques d'un secteur défini en temps réel, la puissance colossale du flux de données reçues par l'utilisateur final en temps réel peut être comparée par analogie à la puissance d'un geyser.

Source :

- Antonio Regalado, MIT Technology Review, 2 octobre 2014
- <http://www.technologyreview.com/news/531146/what-it-will-take-for-computers-to-be-conscious/>
- Traduit par Alexandra Bélieff

17/12/2014