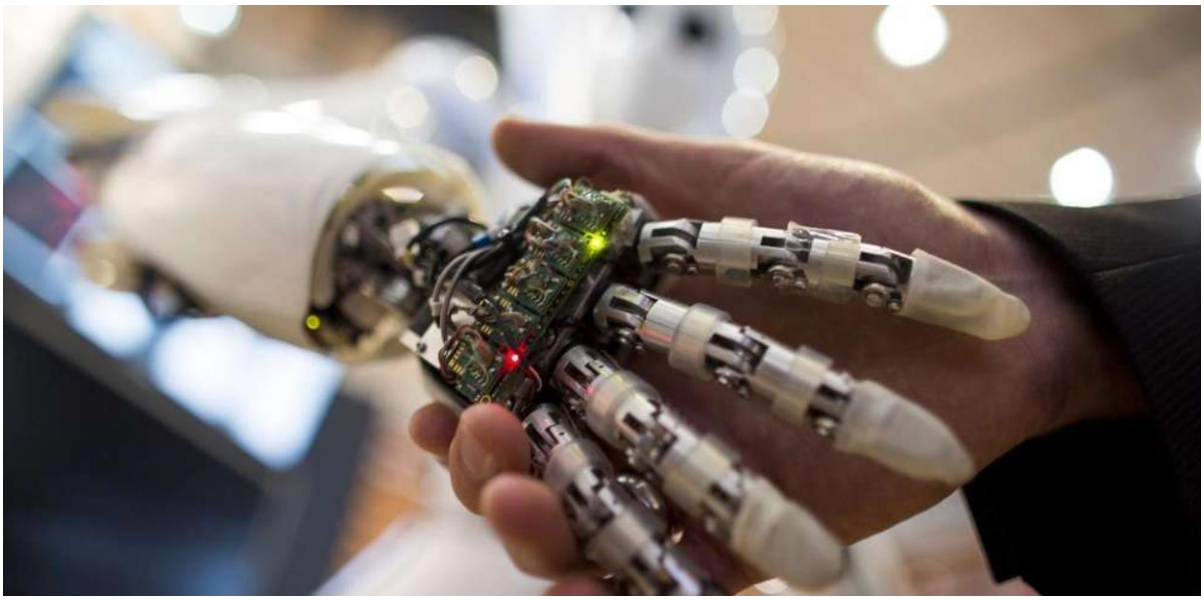
L’intelligence artificielle peut-elle vraiment déraper et devenir dangereuse ?

Thierry Berthier, membre de la Chaire de cyberdéfense & cybersécurité Saint-Cyr, Sogeti, Thales

Décembre 2016 - Article IV.14



L'intelligence artificielle est devenue un sujet inévitable. Mais en France, les pouvoirs publics le maîtrisent encore mal, selon le spécialiste en cyberdéfense Thierry Berthier. AFP PHOTO / CARSTEN KOALL

VIDÉOS – Entre méfiance, fascination et fantasme, l’intelligence artificielle continue à se développer rapidement. Ses applications sont nombreuses. Pour le pire ou le meilleur ?

Entre les scénarios catastrophes véhiculés par le cinéma hollywoodien et la crainte de voir des emplois supprimés au profit d’une automatisation des tâches, l’intelligence artificielle (IA) suscite autant de méfiance que de fascination. D’autant qu’elle évolue de plus en plus rapidement.

Dernier exemple en date, des scientifiques de l’université d’Oxford et de DeepMind, une filiale de Google, ont mis au point une intelligence artificielle, LipNet, **capable de lire sur les lèvres bien mieux**

qu'un être humain. Selon l'article publié dans le revue New Scientist, là où un professionnel de la lecture n'est parvenu à déchiffrer que 12,4% des mots à partir de vidéos, l'IA a atteint un score de 46,8%.

Vidéo : How easy do you think lipreading is ?

Plus tôt cette année, en mars, c'est cette fois le **joueur de Go considéré comme le plus fort d'Europe qui s'est incliné face à AlphaGo, un programme informatique**développé par Google DeepMind (encore lui). Un peu plus anecdotique mais largement médiatisé, des chercheurs du programme Google (encore eux) Brain ont publié le 24 octobre dernier un article dans lequel ils expliquent que deux intelligences artificielles (Bob et Alice) sont parvenues à créer un cryptage leur permettant de communiquer entre elles sans qu'une troisième (Eve) ne puisse déchiffrer leurs messages.

Vidéo : Match 1 - Google DeepMind Challenge Match: Lee Sedol vs AlphaGo

Des avancées récentes de plus en plus nombreuses qui entraînent immanquablement cette crainte, en particulier chez les Français : et si les intelligences artificielles prenaient leur autonomie, que feraient-elles ensuite ? Pourraient-elles finir par nous nuire ?

"Point Godwin technologique"

Une question qui interpelle Thierry Berthier, maître de conférence en mathématiques à l'université de Limoges et spécialiste en cyberdéfense, qui voit dans cette idée d'une "dérive malfaisante de l'IA", "un point Godwin technologique".

Tout n'est pas rose pour autant. Et en parlant de point Godwin, il aura fallu à peine plus de 24 heures à Tay, l'intelligence artificielle de Microsoft lancée le 23 mars sur Twitter pour l'atteindre. Ce "chatbot" était supposé s'adresser aux Américains de 18–24 ans en reproduisant leur langage. Et plus on parlait avec Tay, plus elle était censée apprendre et étoffer son langage. Problème, elle s'est immédiatement heurté à une vaste opération de détournement menée par les "trolls" du forum 4Chan. Objectif : **transformer Tay en "adolescente" raciste, sexiste, négationniste et antisémite**. Mission accomplie. Dès le lendemain, Microsoft a dû la désactiver.

Il aura fallu 24 heures pour transformer Tay, l'IA de Microsoft, en néo-nazie virtuelle.

Si cette expérience a permis de pointer les dégâts que l'on pouvait faire lors de la phase d'apprentissage d'une intelligence artificielle, elle a aussi démontré que le pouvoir de nuisance de Tay était relativement limité. Certes, elle a tenu des propos inacceptables, mais on est encore loin de la (quasi)fin du monde dépeinte dans "Terminator". Car comme le relève Thierry Berthier, Tay, comme les autres IA actuelles, est **très spécifique, sectorisée**. "Elles sont douées pour apprendre dans leur propre domaine (lire sur les lèvres, jouer au Go), mais elle ne savent rien faire d'autre."

IA "fortes" vs IA "faibles"

C'est ce que le spécialiste en cyberdéfense appelle des "IA faibles", autrement dit "des machines simulant des comportements humains sans conscience d'elles-mêmes". Par opposition aux "IA fortes", c'est à dire des "machines produisant un comportement intelligent, capables d'avoir conscience d'elles-mêmes en éprouvant des sentiments et une compréhension de leur propre raisonnement".

Ces dernières restent pour l'instant de la science-fiction, comme de nombreux spécialistes du domaine le soulignent, à l'image notamment d'Omar Mubin, de l'université de Sydney et Eduardo B. Sandoval, de celle de Canterbury. Eux aussi le rappellent :

"La différence entre les "Terminator", "Blade Runner" et autres "Transformers" et les robots actuels tient au degré de conscience, d'autonomie, d'apparence physique. Beaucoup de robots fictionnels […] se montrent capables d'éprouver des émotions et d'autres facultés humaines"

Vidéo : The Terminator Movie Trailer

Nous en sommes bien loin dans le monde réel. Pas de scénario apocalyptique lié aux intelligences artificielles dans un futur plus ou moins proche alors ? Pas de l'avis de Thierry Berthier. Il n'empêche, face à l'ampleur du débat, avec deux autres chercheurs experts en IA et en stratégie, Jean-Gabriel Ganascia et Olivier Kempf, il s'est penché sur l'hypothèse d'une "dérive malveillante d'une IA" dans des conditions réelles, actuelles. Si les trois chercheurs réservent les détails de leur étude à la Revue Défense Nationale, ils ont tout de même **mis le doigt sur une potentielle situation de crise**.

Reste à en déterminer la faisabilité, mais "dans ce scénario en 4 ou 5 étapes, plusieurs intelligences artificielles considérées comme faibles et ne pouvant rien faire séparément pourraient, mises bout à bout, entraîner l'établissement d'un contexte de crise", explique Thierry Berthier. Elles ne le feraient pas consciemment évidemment, mais "la complexité croissante des systèmes experts, des plates-formes d'aide à la décision s'appuyant sur des processus d'apprentissage rend aujourd'hui possible des mises en résonance conduisant potentiellement à des situations critiques. Il s'agit alors bien d'une forme faible d'alerte lancée par Hawking et Musk".

La mise en garde de Hawking et Musk

Il fait ici référence à la lettre ouverte publiée en juillet 2015 sur le site du Future of Humanity Institute (FHI) et signée par plusieurs milliers de chercheurs et personnalités, dont l'éminent astrophysicien britannique Stephen Hawking, le célèbre entrepreneur de hautes technologies Elon Musk -patron de Tesla et Space X-, le cofondateur d'Apple Steve Wozniak, le Prix Nobel de Physique Frank Wilczek et nombre de chercheurs, dont des Français. Tous entendaient alors mettre **en garde contre l'utilisation de l'IA dans le domaine militaire**, et en particulier contre les systèmes armés automatisés, qu'ils

considèrent comme "la troisième révolution dans la pratique de la guerre, après la poudre et les armes nucléaires."

Les "Big Dog", des robots militaires. CRÉDIT PHOTO : U.S. MARINE CORPS PHOTO BY LANCE CPL. M. L. MEIER

Mais en fin de compte, ce n'est ici pas tant l'intelligence artificielle en elle-même qui inquiète que l'usage qui peut en être fait. Par l'être humain qui plus est. "Ce ne sera qu'une question de temps avant qu'elles n'apparaissent sur le marché noir et entre les mains de terroristes, de dictateurs (..) et de seigneurs de la guerre voulant perpétrer un nettoyage ethnique, etc..", arguaient-ils notamment. Mais dans combien de temps ? Quelques années au pire ? Plusieurs dizaines d'années au mieux ? Plus ? Difficile à dire. "**Le développement des IA n'est pas linéaire**", souligne Thierry Berthier. "Il a débuté dans les années 1950, a notamment stagné dans les années 1980–90, avant de connaître une évolution rapide ces dernières années".

Selon lui, c'est dans le domaine social que l'IA peut en réalité faire le plus de dégâts dans un futur très proche, dans le présent même. En réalisant certaines tâches mieux que nous, mais sans fatigue, les machines vont immanquablement (continuer à) entraîner **la suppression d'emplois**. Daniel Bloch, directeur de recherche au CNRS, reconnaît aussi que cette "révolution technique a sûrement de quoi bouleverser les équilibres sociaux […]. De même qu'avec l'imprimerie, la qualification très recherchée des scribes, clercs et calligraphes a pu disparaître".

Vidéo : ABB Robotics - YuMi at ABB Elektro-Praga

L'IA, tout simplement inévitable

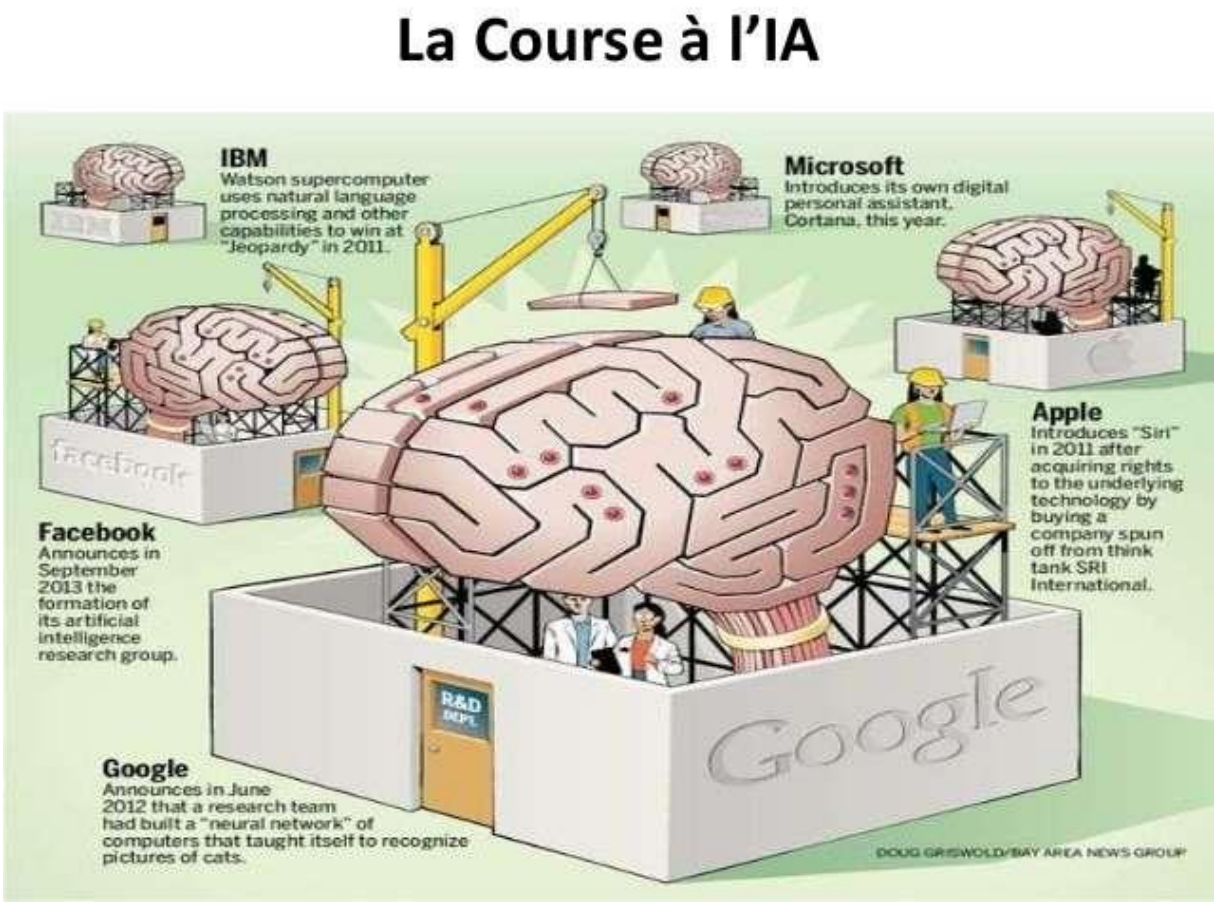
Si "pro" et "anti" s'écharpent autour de ces questions, tout n'est pas noir pour autant. Comme le souligne Daniel Bloch, de la même manière que l'invention de la calculatrice nous permet de ne plus avoir à "extraire au crayon et papier, la racine carrée d'un nombre", les IA "pourraient pour les optimistes, aider à construire des solutions inattendues". C'est aussi dans ce sens que va l'étude "Artificial Intelligence Life in 2030". Lancée à l'automne 2014 par l'université de Stanford, cette vaste étude propose de faire un état des lieux tous les cinq ans pendant 100 ans (en se concentrant sur l'Amérique du Nord, jugée plus représentative).

Dans un premier volet publié en septembre, les chercheurs se sont intéressés à huit grands domaines utilisant déjà, à des degrés divers, les intelligences artificielles : les transports, la domotique, la santé,

l'éducation, l'aide aux communautés à faibles ressources, la sécurité publique, l'emploi et le divertissement. Bilan : bien qu'elles soient accompagnées par la disparition de certains métiers et qu'elles posent des questions en matière de sécurité et de respect de la vie privée, les IA qui émergent depuis quelques années ont **un impact positif sur notre société et notre économie**. "Leur plus grand potentiel est, entre autres, de rendre la conduite plus sûre, d'aider les enfants à acquérir des connaissances et d'améliorer la vie des gens", souligne le rapport. "De fait, les bénéfices des IA dans les écoles, les maisons et les hôpitaux évoluent déjà à une vitesse fulgurante".

Vidéo : INNOVATION - Maison connectée : La domotique en plein essor ?

Les auteurs pointent notamment le fait que les gens ont déjà pris l'habitude de faire nombre de choses avec leur smartphone (téléphone intelligent) et même de lui parler. Dans les années, à venir, ces derniers devraient même leur permettre de surveiller leur état de santé, les alerter en cas de risque etc. Ils évoquent également le développement de véhicules autonomes. Bref, autant d'actions qui devraient de plus en plus impacter notre quotidien. Apple, Facebook Google, IBM, Microsoft, Tesla et même Hollywood… Les groupes les plus puissants du monde ne s'y trompent pas et misent déjà tout ou presque sur l'intelligence artificielle. Et même si son évolution a déjà prouvé qu'elle était chaotique et imprévisible (la preuve), **le sujet est devenu inévitable. Le débat doit donc devenir public**, estiment les chercheurs de Stanford.



Selon eux, plutôt que de s'inquiéter de scénarios catastrophes qui relèvent de la science fiction, ce qu'il faut désormais, c'est que les pouvoirs publics prennent des mesures permettant aux simples utilisateurs que nous sommes de comprendre comment fonctionnent ces intelligences artificielles,

comment les utiliser, voire les concevoir. Or, "**avez-vous entendu un seul de nos candidats à l'élection présidentielle se saisir de cette question** ?", interroge Thierry Berthier. Non.

Alexandra Tauziac, Sud Ouest