

# CES ROBOTS QUI POURRAIENT NOUS VOULOIR DU MAL...

**A**u regard des recherches qui sont menées dans certains laboratoires, Pepper est un mignon petit robot assez primitif. Les évolutions possibles de l'intelligence artificielle posent bien des questions, dont la plus cruciale n'est autre que celle de l'avenir de l'espèce humaine.

**Paris Match.** «Pepper» vous impressionne?

**Damien Ernst, chercheur en intelligence artificielle à l'Université de Liège.** Non, les modules d'intelligence artificielle (IA) implémentés dans ce type de machine sont largement en dessous du «state of the art». Le processus cognitif de cette machine n'est pas comparable à ce qui se passe dans un cerveau humain.

Pour dire les choses simplement: le robot a enregistré une relation «entrée-sortie». Les informations, sous forme de sons ou d'images perçues par ses différents capteurs, celles introduites via sa tablette, via le net ou une connexion Bluetooth, sont traitées par Pepper pour en extraire des «features», des caractéristiques de haut niveau, auxquelles il associe des réactions préprogrammées. Mais, depuis trois ou quatre ans, ce fonctionnement par extractions de «features» est en train de disparaître. Les algorithmes d'aujourd'hui sont devenus tellement bons que vous pouvez directement travailler avec des perceptions brutes.

**Un gadget?**

Non, je ne dirais pas cela. Il y a quand même un processus de traitement d'images, des capteurs, des caméras, différents logiciels... Mais les capacités cognitives d'une telle machine sont bien en deçà de ce que l'observateur lambda pourrait considérer. La tendance actuelle en programmation de robots, l'apprentissage par renforcement, permet d'aller plus loin. C'est un modèle qui ressemble plus au processus d'apprentissage par essai-erreur des êtres humains. La machine adopte un comportement et elle lui donne un score en fonction de la réaction de l'environnement. Par exemple,

le robot dit bonjour avec déférence à un humain. Son interlocuteur réagit en souriant. La machine comprend que son comportement a été adéquat. Elle enregistre cette expérience qui pourra être reproduite. Ainsi donc, elle s'améliore toute seule. Désormais, des centaines de chercheurs utilisent ce paradigme-là.

**Des exemples d'applications concrètes?**

Prenez un jeu Atari des années 1980: cette intelligence par renforcement est reliée à une caméra qui observe l'image du jeu et le score. Sans rien connaître des règles, juste par essais-erreurs, la machine fait très rapidement le tour de la question pour arriver à de bons scores impensables pour tout être humain. Vous êtes aussi en contact avec cette forme d'IA chaque fois que vous utilisez un moteur de recherches sur le web! Le placement des publicités, l'ordre des liens sur ces moteurs de recherche est réalisé au travers de ce mécanisme d'apprentissage par renforcement.

**Faut-il laisser à la science-fiction l'idée qu'un jour, une machine pourrait disposer de toutes les potentialités d'un être humain, en termes de créativité et d'émotivité?**

Non, certainement pas! Il y a un terme qui définit ce stade ultime de la machine qui devient consciente: la «singularité». Certes, quand on regarde ce qui se fait actuellement au niveau algorithmique, on semble encore loin de cela.

Mais les sciences évoluent dans tous les domaines, notamment dans celui de la compréhension du fonctionnement du cerveau humain. Dans un avenir proche, on arrivera certainement à reproduire ce qu'on appelle le connectome, c'est-à-dire toutes les liaisons qui s'opèrent entre les neurones du cerveau, en ce compris les diverses dynamiques de ces neurones. La cartographie du cerveau et les connaissances de plus en plus approfondies de la physique de connexion ouvriront de nouvelles perspectives en terme d'IA.

**Comme celle de programmer des machines équivalentes au cerveau humain?**

Oui, on pourra faire le «mapping» du cerveau humain vers une machine. Une fois que vous avez un modèle physique de quelque chose, vous pouvez le programmer: nous allons donc vers une programmation à l'identique de ce qui se passe dans le cerveau de l'homme. Le jour où cela fonctionnera, la machine sera la même chose que nous. Elle pensera, au sens littéral du terme.

**Vous dites bien qu'elle sera «consciente» de son existence?**

Oui, elle sera de ce point de vue exactement comme les êtres humains.

**Si la machine a conscience d'elle-même, ne voudra-t-elle pas jouer ses propres cartes, prendre le pouvoir sur l'homme, voire le faire disparaître?**

Imaginez une personne humaine qui aurait une mémoire absolue, ce qui serait le cas de cette machine vu le développement du stockage d'information à bon marché. Imaginez qu'en plus, cette personne dispose d'une rapidité de traitement de l'information tout aussi absolue... Elle prendrait sans doute le pouvoir, elle deviendrait immensément riche, elle ferait des tas de trucs que d'autres ne savent pas faire. En tous cas, ce serait dans la nature de l'être humain de le faire.

**Donc...**

Une machine programmée tel un cerveau humain pourrait avoir les mêmes aspirations.

**Le philosophe des sciences Jean Staune parle, lui aussi, de cette «singularité» dans un ouvrage consacré aux «Clés du futur». Il écrit: «L'ordinateur qui aura atteint le niveau de l'intelligence humaine pourra travailler vingt-quatre heures sur vingt-quatre et sept jours sur sept à produire une version améliorée de lui-même (...), jusqu'à produire une super-intelligence, quelque chose qui nous surpassera de plusieurs ordres de grandeur, exactement de la même façon que nous surpassons les souris.»**

Conceptuellement, il a raison. Il n'est d'ailleurs pas le seul à pointer ce danger. A partir du moment où la machine arrivera à un certain niveau d'intelligence, elle pourra s'améliorer par elle-même, voire pour elle-même. Cet écueil de la singularité, on ne peut pas l'exclure, mais il n'empêche pas le fait que des tas de scientifiques ont désormais l'ambition de programmer une machine à l'identique du cerveau humain. Si ces recherches aboutissent, et je pense qu'elles aboutiront, on entrera dans un moment périlleux de l'histoire de l'humanité.

**En janvier 2015, plus de 700 scientifiques et chefs d'entreprises ont lancé un appel demandant un débat sur l'évolution prévisible de l'IA. Un des signataires, le physicien Stephen Hawking, déclarait alors: «Les formes primitives d'IA que nous avons déjà se sont montrées très utiles. Mais je pense que le développement d'une IA complète pourrait mettre fin à la race humaine.» Jean Staune écrit lui aussi que la «super-intelligence» pourrait être «la dernière invention de l'espèce humaine.»**

Même si ce n'est pas si évident de construire ces machines intelligentes, je comprends l'inquiétude. On ne peut nier que les conditions existent désormais pour arriver à la singularité. L'accélération technologique est venue d'internet, des réseaux sociaux, de ces moteurs de recherche qui avaient besoin de l'IA pour se différencier de leurs concurrents et traiter des quantités de données gigantesques. A partir du moment où il y a eu ce business, la communauté du «machine learning», qui était autrefois pauvre comme Job, a été intégrée dans les grosses boîtes comme Google. Cela a donné une autre impulsion à ce petit monde où s'activent des milliers de chercheurs qui rêvent de devenir riches comme Mark Zuckerberg. Il y a de plus en plus d'argent pour financer les innovations.

**Ray Kurzweil, le patron du moteur de recherche de Google, estime que «dans trente ans, l'IA sera un milliard de fois plus puissante que l'ensemble des cerveaux humains»...**

C'est bien plus préoccupant que cela! Il faut comprendre qu'à partir du moment où la machine deviendrait plus intelligente que l'homme – juste une seule fois! – elle pourrait se reproduire et accroître son intelligence sans plafond. A partir du moment où cette singularité existe, tout peut aller très vite.

**On ne parle pas de quelque chose qui pourrait se passer dans cent ans?**

Non... Dix ans, peut-être? Tout va dans le sens de créer cette singularité. Des capteurs de plus en plus performants, des études du cerveau de plus en plus poussées...

**Dix ans?**

Cela ne va pas nécessairement arriver, mais c'est du domaine du possible. **Ne faudrait-il pas que se tienne un débat planétaire sur ce qu'on veut faire ou pas avec l'IA?**

Le problème, quand on parle de «singularité», c'est que le débat est peu

**«NOUS ALLONS VERS UNE PROGRAMMATION À L'IDENTIQUE DE CE QUI SE PASSE DANS LE CERVEAU DE L'HOMME»**



*Damien Ernst, chercheur en intelligence artificielle.*

audible. Les gens pensent que c'est de la science-fiction.

**Avez-vous le sentiment que la communauté des chercheurs qui œuvrent dans le domaine de l'IA travaille encore dans le sens de l'intérêt général?**

J'ai fréquenté pendant très longtemps cette communauté scientifique du «machine learning» et je dois avouer que, parfois, je me pose des questions. Je me demande pourquoi on a fait tout cela. Quel est la place de l'être humain dans ce projet technologique? Il est un fait que certains chercheurs ne réfléchissent pas aux conséquences potentielles de ce qu'ils font. Ils travaillent dans des laboratoires extrêmement centralisés et sont fort mis sous pression. De plus, ils sont souvent très jeunes... Dans le monde de l'énergie auquel je me consacre plus actuellement, on rencontre des écologistes et autres bobos qui réfléchissent au monde de demain. Des idées percolent et c'est bénéfique pour les scientifiques et pour la société. Dans la recherche médicale, il y a toujours, in fine, un contact avec le patient, avec au surplus des règles éthiques qui ont été fixées pour éviter de faire n'importe quoi. Mais dans la recherche en IA, ces limites éthiques n'existent pas, tandis que la connexion des communautés de développeurs et de programmeurs avec «l'humanité» est très faible.

**C'est effrayant!**

Oui.

**Permettez-moi d'insister: cette idée de la machine qui prend son destin en main et se débarrasse de ses créateurs humains, ce n'est pas du roman de science-fiction?**

Comment réagirait-elle si elle arrivait à la conclusion qu'elle n'a pas besoin de l'humanité pour survivre? Il faudrait qu'elle se trouve un avantage à être amicale et «humaine»... A vrai dire, l'avantage en question, je ne le vois pas. Si la machine se rend compte qu'il y a concurrence et que l'homme pourrait l'arrêter, le combat durera cinq minutes. Dans ce monde où tout est connecté, cela ira très vite. Et c'est elle qui l'emportera!

**Cela nous renvoie à «2001, l'odyssée de l'espace»?**

Oui, la machine qui se sent trahie, qui se rebelle... Saut qu'en réalité, si on crée cette «superintelligence» dont nous parlons, les choses iront bien plus rapidement que dans le scénario du film. Ses auteurs n'avaient pas encore imaginé la multiplication des capteurs, l'internet... Cela dit, il est fondamental de discuter de la singularité, mais cela ne saurait exclure un débat indispensable sur toute la période chaotique préalable, celle des bouleversements sociaux énormes qui seront causés par l'automatisation grandissante de la plupart des métiers.

**Pourquoi ne parle-t-on pas plus de ces questions cruciales pour notre avenir?**

Parce que la compétition pour mettre au point les technologies de demain domine tout. C'est un cercle vicieux: on fait toujours plus de recherches dans le «high tech» pour faire face à la situation que le «high tech» a créée. Le monde s'est lancé dans une fuite en avant.

**Rencontrez-vous parfois des politiciens qui sont conscients des enjeux dont nous avons parlé lors de cet entretien?**

Non, il n'y a aucune réflexion à ce sujet dans le monde politique belge. Le «progress» technologique est si rapide que ses contemporains n'en perçoivent pas toutes les dimensions. Quoi qu'il en soit, c'est un débat qu'il faudrait porter dans une instance planétaire, à l'ONU, par exemple! Dans les années 1950, les hommes ont décidé qu'il ne fallait pas construire certaines armes, celles par exemple qui menaceraient la biosphère. C'était des armes, mais aujourd'hui, c'est plus difficile parce que ceux qui maîtrisent ces technologies, des machines à faire de l'argent, ont un pouvoir d'influence énorme sur le monde politique. ■

*Entretien Michel Bouffoux*