

Une IA ou des IA ? Représentations et relations avec les IA

Par **Arnaud de LA FORTELLE**

Centre de Robotique, MINES ParisTech – PSL

Cette simple question : « Une ou bien des intelligences artificielles ? » nous introduit dans la complexité de notre relation avec la notion même d'intelligence artificielle (IA). Comme pour l'homme, il n'y a bien entendu pas une seule intelligence et déjà la définition du concept se révèle difficile. Le fameux test de Turing – qui déciderait quand une intelligence artificielle serait du même niveau que l'homme – est de plus en plus contesté et ne donne aucune définition, au mieux il procède par la voie d'une comparaison et d'une analogie. Mais ce n'est pas tant le concept qui nous intéresse ici, que sa représentation qui influe sur notre usage de la technologie et notre relation avec celle-ci.

Dans un premier temps, nous aborderons les représentations, techniques ou artistiques, qui peuvent inconsciemment biaiser nos conceptions de ce qu'est, ou plutôt de ce que fait, l'intelligence artificielle. Dans un deuxième temps, nous confronterons ces représentations à la réalité des IA d'aujourd'hui. Il existe dans la manière dont nous construisons ces IA – comme nos autres outils, matériels et conceptuels – de nombreuses structures humaines qui façonnent les modalités dont les IA réagissent ou interagissent avec le monde. De plus, la réalité du déploiement des IA aujourd'hui est la pluralité, ce qui conduit à une vision substantiellement différente des représentations habituelles. Et ainsi se pose la question des interactions entre IA : pouvons-nous envisager ce que seraient des échanges entre plusieurs IA ? Et ce que cela signifierait pour les interactions entre humains et IA ? Finalement, ce raisonnement aboutit à se poser la question de toutes les interactions, non seulement entre ou avec les IA, mais aussi entre humains, avec des transformations probablement plus profondes encore qu'Internet et les réseaux sociaux, mais encore peu claires et où nous aurons certainement nos propres choix à faire.

Les représentations de l'intelligence artificielle

L'intelligence artificielle est assez systématiquement présentée au singulier. L'art propose de nombreuses images d'une telle IA, le plus souvent sans considération pour les humains (HAL dans le film *2001, Odyssée de l'espace*), voire leur étant ouvertement et agressivement hostile (Skynet dans le film *Terminator*). Parfois paradoxalement étouffante et protectrice (Colossus dans le film *Le Cerveau d'acier*). Le singulier engendre l'idée d'une concentration de puissance (omniscience et omnipotence) qui, au mieux, déséquilibre le monde. C'est une voie bien classique de construction d'un cadre narratif, mais le point est que ces représentations fictionnelles passent dans notre perception de la réalité, bien souvent inconsciemment. L'intelligence artificielle fascine par son omnipotence.

Si l'on regarde la manière dont le domaine a évolué depuis des décennies, il est en effet frappant de lire les attributs que nous avons prêtés à nos programmes : langages, calcul formel, systèmes experts, apprentissage profond. Les mots sont importants dans une relation, et ceux que nous choisissons pour décrire nos avancées en la matière sont révélateurs à la fois d'attentes et de craintes très fortes à leur sujet. Pourtant, la réalité est un peu différente et les scientifiques œuvrant dans ces domaines n'ont d'ailleurs jamais eu la naïveté de croire ce que leurs propres

mots sous-entendaient. L'apprentissage profond (*deep learning*) induit l'idée d'une profonde connaissance quand il ne fait référence qu'à la profondeur d'un réseau de neurones artificiels, le nombre de ses couches, ce qui est une esquisse plutôt simpliste de cerveau ; et l'apprentissage consiste à optimiser les paramètres de différentes fonctions (formant un neurone artificiel) pour aligner les prédictions du réseau de neurones avec une (très grande) base de données. Même si l'on est justement impressionné par les avancées que ces techniques ont permises, on est très loin d'une vraie intelligence. Typiquement, ces programmes savent bien détecter des humains dans des images ou des vidéos, mais ne comprennent pas qu'une affiche (par exemple une publicité) ne peut pas être un humain, ce qui peut poser un problème quand on fait appel aux IA pour conduire des voitures. Ce qu'il faut retenir ici, c'est que la technique elle-même soutient, par son langage, une représentation qui met en valeur la puissance de l'intelligence artificielle.

Nous voudrions également souligner deux autres attributs importants de l'intelligence artificielle d'aujourd'hui : les données et la puissance de calcul (admirons encore ici l'utilisation du mot « puissance »). L'apprentissage artificiel (*machine learning*) tire son efficacité des gigantesques bases de données (des millions de livres, des milliards d'images) qui lui permettent d'optimiser ses réponses lors d'une phase d'optimisation (l'apprentissage) extrêmement gourmande en temps et en ressources. On peut donc avoir le sentiment que cet apprentissage conduirait à privilégier quelques rares acteurs possédant toutes les données et capables d'entraîner une IA qui serait omnisciente. De même, le besoin de puissances extrêmes de calcul abonde dans ce sens. Outre l'association que nous pouvons en faire avec le thème du *Big Brother*, il nous dirige encore une fois vers une conception au singulier de l'intelligence artificielle.

Réalité : structures et interactions

Tranchant résolument avec les paragraphes précédents, considérons maintenant les intelligences artificielles déployées aujourd'hui, ou plutôt les produits et services qui reposent sur ces techniques. Les premiers qui viennent à l'esprit sont les *chatbots* qui ont été largement disséminés par les pionniers mondiaux : Siri, Google Voice, Alexa et les autres. Ce sont effectivement des services centralisés, reposant sur un profilage des utilisateurs, qui ont à la fois la capacité de traiter des questions en langage naturel (un problème notoirement compliqué pour lequel les IA ont permis de gros progrès) et de gérer des associations d'idées afin de produire des réponses pertinentes à un utilisateur donné. Ces services sont généralement également associés à certains produits selon un modèle économique désormais bien maîtrisé. Mais il existe de nombreuses variations de ces services (banques, cinémas...) qui ne reposent pas toutes sur ces acteurs. Ensuite, il existe des services que l'on n'oserait pas nommer des intelligences, mais qui reposent sur les mêmes techniques, comme pour la prise de photos : on parlerait plutôt de filtre, mais c'est bien un réseau de neurones qui apprend à transformer des images brutes en des images qui nous plaisent. Ici, notons qu'un tel réseau se trouve dans chaque *smartphone* récent : on peut commencer à parler de pluralité des IA.

On voit ainsi apparaître un déploiement répandu des techniques d'IA pour proposer aux utilisateurs des services agrégés et adaptés à leurs besoins. La clef est donc bien entendu d'avoir accès aux désirs des utilisateurs, et aux données qui permettent de les prédire (d'où le succès des modèles d'affaires liés à la publicité prédictive). Suivant le chemin pris par ces données, on aura des modèles singuliers, si l'on centralise toutes ces données, ou des modèles pluriels si ces dernières restent locales, par exemple sur les *smartphones*. Il existe des projets visant à déployer de telles IA sur les *smartphones*, puisqu'ils en ont les capacités aujourd'hui : on aurait donc un traitement de nos données personnelles « chez nous » avec un accès aux données externes (achats, réservations, questions...) sur Internet *via* cette IA. Un modèle plutôt sympathique qui pêche essentiellement aujourd'hui par son manque de modèle d'affaires en comparaison des services

poussés par les géants du web. Mais on voit que la pluralité des IA est une réalité qui a des conséquences intéressantes.

Outre la pluralité de la localisation des données et des calculs, nous assistons à la spécialisation des IA en fonction de nos propres attentes. Qu'est-ce que j'attends d'une IA à ma disposition ? C'est une question qui obtient des réponses très variées, depuis le planificateur qui organise mes transports, mes finances ou mon travail de façon à profiter au mieux de mon temps, jusqu'au *coach* qui me stimule et me conforte. Au fur et à mesure que les IA déployées aujourd'hui se perfectionneront, nous devrions donc les voir apprendre et s'adapter à des situations très différentes. Or il est bien clair que cette faculté d'apprentissage est au cœur des capacités des IA : on aura alors autant d'IA que d'humains (peut-être même davantage), et aussi de multiples IA pour toutes sortes d'activités (cinéma, restaurant, voiture, bus, robots...). Chaque IA devra se développer de façon différenciée, et les données qui auront conduit à cette évolution ne seront pas partagées. Les IA seront nombreuses et finalement très hétérogènes.

En dépit de cette diversité, on peut imaginer que les techniques sous-jacentes resteront relativement communes. Les réseaux de neurones sont organisés en une architecture (couches, liens entre couches...) qui est souvent très liée aux objectifs que l'on souhaite atteindre. Le cerveau est un organe bien plus complexe et dynamique que tout ce qu'on peut imaginer aujourd'hui avec des réseaux de neurones artificiels, mais passons ces considérations et admettons des capacités décuplées pour les IA. Nous l'avons vu, il existe, dans la construction des IA aussi bien que dans leur apprentissage (les données), des structures humaines, implicites ou explicites, par exemple le langage ou la mémoire de nos demandes et actions. Ces structures devraient transparaître dans les interactions (données demandées ou produites par les IA). Et puisque toute l'évolution initiale des IA sera d'optimiser leur réponse à nos questions, on peut imaginer qu'il ne devrait pas y avoir une trop grande distance entre les interactions que les IA pourront avoir avec nous ou entre elles.

Relations entre IA

On sait qu'il y a des liens très forts entre le langage, les structures de notre pensée et l'intelligence en général. Il est sans doute un peu présomptueux d'imaginer que ceci s'applique à des machines sans adaptation. En revanche, de ce que nous avons présenté, il semble réaliste que demain de nombreuses IA interagissent : on pourrait sous-traiter l'organisation d'un anniversaire au restaurant, avec choix de date et de menu, ce qui requiert de multiples aller-retours entre des IA qui organiseraient les activités et préférences de chaque entité, invités ou restaurateurs. La façon dont ceci se mettrait en place reste obscure, mais ce sera probablement un mélange de langage naturel (puisque'il faudra bien que les IA le maîtrisent) et d'échanges de données. La seule certitude, c'est que de nombreuses IA, diverses, devront interagir.

Ce que feront les IA de leurs interactions est difficile à évaluer. Le scénario optimiste est que cela conduise à une coopération ; le scénario pessimiste est que cette interaction produise, et ceci encore plus vite que chez les humains, des structures de domination entre IA. La réalité devrait se situer entre les deux, suivant la volonté et l'influence que nous aurons. Les limites sont extrêmement difficiles à placer, car la notion de nombre – qui est capitale dans les interactions – est certainement différente pour une IA et pour un humain. Avec combien d'autres une IA pourrait-elle se mettre en relation ? En quoi ceci changera la structure même des relations IA-IA ? La réponse est très ouverte. Mais cette structure interne au monde des IA aura évidemment un impact sur la relation des IA avec les hommes.

Conclusion

Cette notion d'évolution dans le monde des IA, qui semble inéluctable, a conduit à deux théories relativement différentes. L'une, portée par le transhumanisme, est l'idée de singularité (*The Technological Singularity*) : le processus s'emballera et à un moment – unique et sans retour – les deux mondes se déconnecteront (les adeptes de cette théorie en font la victoire de l'IA, mais on peut en faire d'autres lectures). La seconde est la coévolution : les humains seront entraînés dans la symbiose avec les IA, et les deux « vivront » ensemble. Dans les deux cas, il me semble que l'on retourne ici à nos représentations fictionnelles. S'il est certain que les outils et les concepts que l'homme a forgés ont contribué à le changer assez profondément, on peut en conclure que les IA, par une interaction bien plus forte avec nous qu'un marteau ou même qu'un réseau social, devraient également nous transformer sérieusement. Mais en faire une évolution ultime, comme une apocalypse ou un avènement rédempteur, paraît davantage tenir de la foi que de l'analyse : beaucoup ont prédit la venue du surhomme, du fait de l'évolution ou des révolutions, mais il me semble qu'il n'est pas encore parmi nous, n'en déplaît à certains.

La conclusion que nous souhaitons donner de cette analyse, c'est qu'il y a un vrai enjeu à considérer la pluralité des intelligences humaines et artificielles. C'est une voie que nous avons déjà commencée à emprunter, même si nous n'en sommes qu'aux balbutiements des IA. Sans prêter aux IA de demain plus que nous ne devons, reconnaissons que c'est un miroir pour nous-mêmes. Il s'agit bien d'une évolution, que ce soit à travers l'usage des IA, usage qui pourrait se transformer en relation forte (regardons ce qui se passe avec les réseaux sociaux), ou à travers la société qui pourrait réguler ou détourner ces usages. Et comme dans nos sociétés nous souhaitons éviter la concentration de tous les pouvoirs en une seule entité, réfléchissons à distribuer convenablement les rôles des IA et apprenons à développer des relations saines, sans être captifs de représentations biaisées.