LE PRINTEMPS DES MACHINES

Olivier Perriquet

Colloque « *Archéologie des Médias, Écologies de l'Attention* »
Centre Culturel International de Cerisy, 30 mai - 6 juin 2016
http://www.ccic-cerisy.asso.fr/media16.html

L'intelligence artificielle a subi une série de phases d'hibernation, dont l'une des plus importantes a fait suite aux prédictions pessimistes de Marvin Minsky quant à la capacité des réseaux de neurones artificiels à rivaliser avec des méthodes algorithmiques. Mais les succès récents du paradigme connexionniste, illustré par « l'apprentissage profond » (*deep learning, machine learning*), semblent augurer du retour d'un printemps ensoleillé pour une telle approche de l'intelligence artificielle, inspirée du vivant. Par ailleurs, l'objet même de l'intelligence artificielle s'est déplacé. Le modèle de la machine intelligente est passé de celui du joueur d'échec, trouvant son paroxysme avec la victoire médiatique de l'ordinateur Deep Blue ou du programme AlphaGo, à celui du petit enfant partant explorer le monde qui l'entoure. La difficulté des jeux combinatoires (les échecs, les grammaires des chatbots, etc.) est en effet bien inférieure à celle qui consiste à acquérir les rudiments d'un langage inconnu, en engageant son corps et non ses seules facultés intellectuelles. Il existe à cet endroit un déplacement épistémique, que je me propose d'explorer, dont le prix à payer est celui d'une opacité constitutive.

De Deep Blue au *deep learning*

Nous avons assisté récemment à la victoire très médiatisée d'un programme informatique, nommé AlphaGo (Silver 2016), contre Lee Sedol, joueur de go professionnel coréen, considéré comme l'un des meilleurs au monde. Cette victoire en rappelle une autre, elle aussi spectaculaire et également d'une grande portée symbolique dans le développement d'une intelligence non-humaine, celle de Deep Blue sur Garry Kasparov aux échecs en 1997 (Hsu 2002). Le jeu de go résistait depuis une vingtaine d'années aux efforts des chercheurs en intelligence artificielle pour plusieurs raisons, en particulier du fait d'une combinatoire beaucoup plus élevée qu'aux échecs car le goban est très grand, comparé à l'échiquier, et le nombre de coups possibles est par conséquent plus important, mais également à cause des principes de jeu et de la façon dont s'élaborent les stratégies au go, où les positions occupées comportent plus d'ambivalence qu'aux échecs, car elles participent à une conquête de territoires dont l'intérieur et l'extérieur n'apparaissent que progressivement au cours de la partie, rendant ainsi la définition même de ce qui constitue le contour d'un territoire délicate à établir (le go est un jeu constructif où l'on se partage un espace commun ; les échecs sont un jeu destructif où il s'agit d'anéantir l'adversaire...). Vingt ans après les échecs, le go est donc désormais tombé aux mains des machines.

Pour autant la machine ne *pense* pas, serions-nous tentés de nous rassurer. Elle n'est pas à proprement parler « attentive » à son environnement (dans la diversité toujours émergente de ce dernier) : elle traite des informations pré-canalisées pour elle, en se contentant d'y réagir mécaniquement. C'est simplement sa puissance de calcul qui l'emporte sur l'intelligence créative dont fait preuve le joueur humain. Tout se passe comme si l'espace symbolique qu'elle est en mesure d'explorer dépassait les capacités cognitives humaines, et c'est sans doute effectivement le cas. Après tout, une simple calculatrice ne nous bat-elle pas à la multiplication ? Bien évidemment, la supériorité d'une intelligence artificielle à un jeu où les humains se mettent mutuellement à l'épreuve, confrontent leur inventivité et démontrent leurs talents stratégiques, suscite nettement plus de trouble. Mais pourquoi cette nouvelle étape est-elle si importante ? N'avions-nous pas déjà franchi ce cap emblématique avec Deep Blue, deux décennies plus tôt ?

En réalité, si les deux programmes – AlphaGo (2015) et Deep Blue (1997) – procèdent d'une façon similaire en explorant la combinatoire des coups possibles, leurs approches sont essentiellement différentes. Les deux jeux n'offrent d'ailleurs pas la même résistance. Dans le cas des échecs, les rapports de force s'expriment au travers de la succession des coups qui vont rendre possible ou non l'occupation de telle ou telle position et, en définitive, la prise de la pièce maîtresse de l'adversaire, tandis qu'au go, nous sommes face à un problème d'une difficulté équivalente à celui de la reconnaissance de formes. Au jeu d'échecs, la machine n'a qu'à classer des coups potentiellement prévisibles au sein d'une combinatoire préétablie tandis qu'au jeu de go elle doit faire émerger des figures (potentiellement inédites) à partir d'un fond. Alors que Deep Blue cherche à optimiser une fonction de coût qui reflète numériquement l'avantage du joueur, en utilisant un algorithme (Minimax) et une méthode heuristique (Alphabeta) qui sont devenus un classique de la théorie des jeux, le programme AlphaGo utilise des méthodes d'apprentissage par réseaux de neurones artificiels, et notamment des techniques d'apprentissage profond (*deep learning*) dont les résultats spectaculaires ces dernières années ont surpris jusqu'aux experts du domaine. Nous avons ainsi d'un côté un algorithme, de l'autre un réseau de neurones, deux méthodes aux présupposés idéologiques opposés, correspondant à deux paradigmes distincts en intelligence artificielle.

Un réseau de neurones n'est pas un algorithme

Le Minimax utilisé par Deep Blue est un *algorithme*. Il détermine un coup optimal en minimisant et maximisant alternativement une fonction d'évaluation selon une procédure mathématique explicite. Le terme « algorithme », qui vient du nom d'un mathématicien persan ayant vécu au VIII^{e} siècle – Muhammad Ibn Mūsā al-Khuwārizmī – traduit en latin par *algoritmi*, désigne une méthode mathématique qui peut s'énoncer de manière symbolique et univoque, et s'appliquer mécaniquement en vue d'obtenir un résultat précis ou de résoudre un problème donné. L'algorithme d'Euclide, qui décrit un procédé automatique de calcul du plus grand diviseur commun entre deux nombres entiers, en est un bon exemple. À l'origine liée aux nombres entiers, la notion d'algorithme s'est étendue par la suite à la manipulation d'objets de plus en plus complexes, tels que du texte, des images, des formules logiques, des structures mathématiques, des objets physiques, etc. Les amateurs de Rubik's cube savent par exemple résoudre le célèbre casse-tête par une série de mouvements qu'ils appliquent automatiquement, ressemblant à des énoncés de la forme :

$L^2(ED)(R'L)F^2(RL')(ED)R^2$; $y(R'UR'U')(R'U'R'U)(RUR^2)$; $y^2R^2(ED)(RL')B^2(R'L)(ED)L^2$ [...]

Une expression formelle de ce type évoque à l'évidence le code des langages de programmation. Le petit nombre d'opérations permises et le caractère systématique de leur enchaînement amène en effet à concevoir un jeu d'écriture qui en simplifie l'expression, et cette graphie compacte et synthétique nous rappelle précisément que, derrière la rhétorique ou les manipulations qui les expriment, les opérations en jeu sont bien de nature symbolique.

Un *réseau de neurones artificiel* est un modèle informatique inspiré du fonctionnement des neurones biologiques, mais dont le principe d'action ne s'exprime pas sous la forme d'un algorithme. Il s'agit d'un artefact construit pour être utilisé comme on se servirait d'un marteau, d'une fourchette ou d'un microscope... Ce sont sa fonction et ses possibilités d'action sur le réel qui le définissent en premier lieu, comme c'est le cas pour un outil (le terme anglais d'*affordance* traduit assez bien cette idée). Les neurones naturels, dont s'inspirent ces modèles, sont des cellules nerveuses qui transmettent un signal bioélectrique et dont on peut schématiser le fonctionnement ainsi : chaque neurone est connecté en entrée à un certain nombre d'autres neurones, dont il reçoit les influx nerveux, et en sortie à un petit groupe de neurones, à qui il transmet son influx, sous la forme de signaux intermittents. Si le neurone est excité au-delà d'un certain seuil par la somme des influx qu'il reçoit, il émet une décharge le long de son axone vers les neurones auxquels il est relié. Un neurone « formel », version abstraite et simplifiée du neurone biologique, est défini par la donnée d'un certain nombre d'entrées, d'une unique sortie, et d'une fonction de transfert simple qui détermine sous quelle condition et à partir de quelle valeur le neurone doit activer sa sortie. Un réseau de neurones artificiel est un assemblage de neurones formels réalisé par le câblage de certaines sorties sur certaines entrées.

En pratique, pour construire un réseau de neurones, on fixe d'abord sa topologie, c'est-à-dire l'agencement des connections. Le réseau comporte alors plusieurs entrées et au moins une sortie ainsi que de nombreuses connections internes, pouvant être organisées hiérarchiquement, en niveaux successifs – c'est le cas des réseaux convolutifs du *deep learning* (Schmidhuber 2015). Cette première étape est largement empirique. On doit ensuite réaliser une phase d'apprentissage, et il faut alors disposer d'un corpus d'exemples qu'on présentera à l'entrée du réseau. Cette seconde étape procède de façon incrémentale, en ajustant à chaque pas les pondérations des influx nerveux dans le réseau. Certaines méthodes d'apprentissage consistent à modifier par exemple l'efficacité des connections de façon à renforcer la connectivité des neurones dont on observe une activation synchrone – selon le principe *« neurons that fire together, wire together »* (Hebb 1949). Il s'agit, en définitive, de rendre le réseau apte à une tâche donnée : par exemple la reconnaissance d'un type de caractère (le corpus d'apprentissage est constitué dans ce cas d'un ensemble de caractères aux formes variées, dont on connaît déjà l'appartenance ou non à la classe à reconnaître), ou bien la classification d'un ensemble d'images en différentes catégories, l'inférence de concepts à partir d'un texte, la détection et le suivi d'objets dans un environnement, etc. Une fois entraîné, le réseau est opérationnel et rigide, et spécialisé dans une tâche donnée : il fonctionne comme un système opaque, comportant des entrées, où on lui présente des données, et des sorties, où il exprime un prédicat.

Les saisons de l'Intelligence Artificielle

L'intelligence artificielle s'est organisée au cours de son développement en deux paradigmes, dont on trouve les prémisses dans l'article séminal d'Alan Turing, *Computing Machinery and Intelligence* (Turing 1950), et dont chacune des deux approches évoquées plus haut est emblématique : le paradigme « logique » ou « fonctionnaliste », où un problème s'exprime sous forme symbolique et sa résolution sous forme algorithmique ; et le paradigme « biomimétique » ou « connexionniste », où la solution d'un problème passe par la construction et l'entraînement d'un modèle, inspiré du vivant. Nous pouvons dire, au risque d'être un peu schématique, qu'ils ont eu le vent en poupe à tour de rôle. La première approche est liée au développement des sciences cognitives, dont le postulat est qu'il est possible d'isoler un niveau autonome, celui des représentations, indépendamment des processus physiques en jeu et du support sur lequel elles s'inscrivent, tandis que la seconde est issue de la cybernétique, qui s'est toujours refusée à faire cette abstraction. Chacun de ces deux paradigmes a subi une série d'« hivers », dont l'un des plus importants a fait suite aux prédictions pessimistes de Marvin Minsky quant à la capacité des réseaux de neurones artificiels à rivaliser avec des méthodes algorithmiques (Minsky 1969), qui avait précipité à l'époque l'arrêt des recherches connexionnistes pendant dix ans. Or la vogue connexionniste actuelle et les progrès de la vie artificielle pourraient nous conduire aujourd'hui à revenir sur le jugement de l'intelligence artificielle symbolique et à rouvrir le débat philosophique sur la nature de l'information (Triclot 2008).

Bien qu'en pratique les réseaux de neurones artificiels soient généralement implémentés sur le même type de *hardware* que les algorithmes (préférentiellement les processeurs des cartes graphiques, dont l'architecture est optimisée pour la multiplication matricielle), ces modèles font circuler en premier lieu des signaux, qui nous ramènent très directement à la matérialité au travers de laquelle s'inscrit l'information, et non des signes, qui se seraient alors émancipés de leur nature physique. Lorsque des représentations internes semblent éventuellement émerger dans les différentes couches des réseaux de neurones convolutifs, et dès lors que celles-ci présentent des similarités avec ce qui est observé dans le cortex visuel humain (Kriegeskorte 2015), ce ne sont que des motifs visuels aux contours diffus, qui sont au mieux *pré-symboliques* : nous sommes sur le terrain de la reconnaissance de formes, non sur celui de la représentation symbolique.

Un réseau de neurones, dans sa phase d'apprentissage, est un objet plastique, propriété qui le distingue à nouveau d'un algorithme. Les réseaux naturels se développent par des processus de morphogenèse (il existe au cours de leur maturation des moments privilégiés où ils présentent une plasticité plus élevée [Prochiantz 1989]) et il est vraisemblable que le biomimétisme soit poussé à l'avenir jusqu'à ce stade avec la fabrication de réseaux de neurones artificiels, *in vivo* tout autant qu'*in silico,* qui soient sujets à deux sortes de plasticité : une plasticité qu'on pourrait dire de type *software*

où, l'architecture du réseau étant préalablement fixée, l'apprentissage transforme la circulation du flux d'information au sein du réseau, et une autre de type *hardware*, qui descendrait un étage plus bas et modifierait la topologie du réseau. Certaines recherches, copieusement financées par les fonds destinés à la conception de technologies militaires, visent également au développement de supports matériels dédiés imitant les tissus biologiques vivants comme le projet SyNAPSE (DARPA 2016) ou directement implémentés sur des supports biologiques (Warwick 2010) dont la miniaturisation et la faible consommation énergétique sont hors de portée des *hardware* actuels.

Si ce sont bien des algorithmes qui régissent la propagation des signaux dans un réseau de neurones, ses performances ne peuvent être évaluées par les critères mathématiques avec lesquels on analyse ces algorithmes, qui ne font que l'agencer. Les réseaux de neurones sont jugés en définitive par leur efficacité pratique et leur conception relève aujourd'hui à certains égards d'une pratique d'ingénieur plus que de mathématicien. Leur opacité caractéristique, que les cybernéticiens qualifient de « boîte noire », conduit au retour d'une forme d'empirisme et le printemps de l'approche connexionniste pourrait bien précipiter un automne de la pensée symbolique, suspendant au dessus de nos têtes pensantes une épée de Damoclès, car de la boîte obscure à l'obscurantisme, il n'y a que le pas de l'idéologie à franchir.

Corps et opacité

Dans son article fondateur, Alan Turing pose la question suivante : « les machines peuvent-elles penser ? ». Mais il prévient le lecteur qu'il se gardera bien de définir chacun des deux termes : *machine* et *pensée*. Il propose de substituer à cette interrogation une mise en œuvre effective de l'intelligence, traditionnellement considérée comme un attribut, sous la forme d'un protocole (un test) restituant au contraire à l'intelligence son caractère performatif. Mais le *test de Turing* se déroule sous la forme d'un échange verbal via une interface de chat et continue à ne considérer l'intelligence que sous son versant symbolique. Il est même emblématique d'une telle conception. Les robots autonomes construits à la même époque par les premiers cybernéticiens, aux noms aussi poétiques qu'évocateurs – les « tortues cybernétiques » de William Grey Walter, le « renard électronique » d'Albert Ducrocq, le « robot écureuil » de Jack Koff – sont au contraire dotés d'une très faible capacité d'intelligence symbolique. En revanche, tout en étant livrés à eux-même, ils sont capables de se conduire en vue d'atteindre un but (Tchernia 1958). Montées sur roulettes, ces machines aux allures de jouets technologiques bricolés avec du *Meccano*, avaient en effet la propriété de se diriger spontanément vers une source lumineuse lorsqu'elle se présentait tout en restant soumises aux contingences imprévisibles de leur environnement physique (rappelons que le terme « cybernétique », provenant du grec *κυβερνητική* qui signifie « art de gouverner », a la même racine que le mot « gouvernail »). Ne conviendrait-il pas ici de parler d'*attention*, plutôt que d'*intelligence*, c'est-à-dire de cette faculté de tendre son « esprit » vers un objet défini, à l'exclusion de tout autre ; et en suivant les pas de Turing, c'est-à-dire sans circonscrire avec une précision définitive cette qualité d'attention que l'on prête aux machines, de s'attacher surtout à montrer ce qui la rend visible, reconnaissable ?

Cette forme d'intelligence, attentive, ne peut plus se contenter de faire circuler (calculer, computer, classer, bloquer, traduire) des informations, mais doit apprendre à établir des correspondances entre des données au sein desquelles elle doit elle-même faire émerger ce qui est suffisamment pertinent pour devenir une information. On passe ainsi d'une intelligence conçue comme une force de computation d'informations à une intelligence conçue comme une forme d'attention – une *attention artificielle* biomimétique capable de prendre en charge la fonction attentionnelle caractéristique des êtres vivants, en tant que leur survie dépend de leur capacité à s'ajuster à leur environnement vital. Passant du paradigme logique au paradigme biomimétique, l'intelligence artificielle ne devient plus « animale » qu'en se déployant comme attention, tout en s'opacifiant.

La physicalité du support de l'information, reprenant aujourd'hui une importance de tout premier plan, comme en écho à un mouvement similaire ayant eu lieu vers la fin des années 1980 et qu'on avait nommé la *thèse de l'embodiment* (Brooks 1990), remet au goût du jour les recherches en vie artificielle et en robotique, c'est-à-dire les formes les plus incarnées de la recherche en intelligence

artificielle. Le modèle de la machine intelligente qui était celui du joueur d'échec, trouvant son paroxysme avec la victoire médiatique de l'ordinateur Deep Blue, laisse sa place à celui du petit enfant partant explorer le monde qui l'entoure, vérifiant que le carré rentre bien dans le carré, le rond dans le rond, ou l'étoile dans l'étoile... L'ambition monte alors d'un cran car la difficulté des jeux combinatoires est fort inférieure à celle qui consiste à acquérir les rudiments d'un langage inconnu en engageant son corps et non ses seules facultés intellectuelles. Le jeu de l'enfant, si l'on veut, déclasse en complexité celui de l'adulte.

Ce retour au corps (perception, reconnaissance de motifs, préhension, voire émotions, apprentissage, importance des formes) préfiguré par les avancées spectaculaires des réseaux de neurones artificiels face à l'intelligence artificielle symbolique, comme refaisant surface, nous engage vraisemblablement vers une cohabitation prochaine avec d'autres entités, munies de facultés comparables aux nôtres et avec lesquelles il nous faudra interagir non plus de main à outil mais... de pair à pair, d'attention opaque à attention opaque, de corps à corps. Et ce constat s'accompagne d'un certain vertige, car si la loi empirique de Moore (celle qui affirme que les capacités de stockage et de calcul doublent tous les deux ans) continue à être vérifiée dans les années à venir, nous aurions théoriquement la capacité matérielle de réaliser un réseau de neurones de la puissance de celui d'un cerveau humain vers le milieu du XXI^{e} siècle. On peut prédire qu'à ce stade, une pensée consciente d'elle-même trouvera à s'exprimer dans cet esprit non-humain dont la corporéité nous est aujourd'hui encore inconnue, et nous retrouverions alors, à l'étage le plus haut, la pensée symbolique perdue quelques étages plus bas, mais ce serait au prix d'une opacité constitutive, qui semble être la condition de l'émergence de la conscience.

BIBLIOGRAPHIE

BROOKS Rodney Allen. 1990. « Elephants Don't Play Chess ». Robotics and Autonomous Systems, Volume 6, Issues 1–2, p. 3-15.

DARPA. 2016. Projet SyNAPSE – http://www.artificialbrains.com/darpa-synapse-program.

HEBB Donald Olding. 1949. The Organization of Behavior. New York, Wiley.

HSU Feng-hsiung. 2002. Behind Deep Blue. Building the Computer that Defeated the World Chess Champion. Princeton University Press.

KRIEGESKORTE Nikolaus. 2015. « Deep neural networks: a new framework for modelling biological vision ». The Annual Review of Vision Science, Volume 1, p. 417-46.

MINSKY Marvin & PAPERT Seymour. 1969. Perceptrons. An Introduction to Computational Geometry. Cambridge, MIT Press.

PROCHIANTZ Alain. 1989. La Construction du cerveau. Paris, Hachette.

SCHMIDHUBER Juergen. 2015. « Deep Learning in Neural Networks: An Overview ». Neural Networks, Volume 61, p. 85-117.

SILVER David et al. 2016. « Mastering the game of Go with deep neural networks and tree search ». Nature, Volume 529, Issue 7587, p. 484-489.

TCHERNIA Pierre. 1958. Émission « Répondez, Monsieur X » (à la rencontre d'Albert Ducrocq, 29 novembre 1958). Collection: RTF / ORTF – http://www.ina.fr/video/CPF86648196

TRICLOT Mathieu. 2008. Le moment cybernétique: La constitution de la notion d'information. Champ Vallon.

TURING Alan. 1950. « Computing machinery and intelligence ». Mind, New Series, Volume 59, No. 236, p. 433-460.

WARWICK Kevin et al. 2010. « Controlling a Mobile Robot with a Biological Brain ». Defence Science Journal, Volume 60, No. 1, pp. 5-14.