théorèmes mathématiques, analysent les marchés financiers, prennent des décisions concernant les crédits bancaires, inspectent des moteurs de pompes, contrôlent des émulsions dans les aciéries... Bientôt les programmes d'intelligence artificielle guideront des missions spatiales, exploreront d'autres planètes et conduiront des camions sur les autoroutes.

Toutes ces tâches sont-elles des preuves d'«intelligence»? La performance des systèmes d'intelligence artificielle, tout comme la vitesse des avions, est incontestable. Toutefois, l'histoire de l'intelligence artificielle a souvent montré que les capacités semblaient perdre leur caractère de prouesse mentale dès qu'on parvenait à les mécaniser. On a oublié qu'à l'époque de Turing un «calculateur» était un être humain qui faisait des calculs arithmétiques et qu'il était alors évident pour tous que cela nécessitait de l'intelligence. Le mot calculateur désigne aujourd'hui une machine. L'exécution rapide et précise de calculs n'est plus considérée comme une preuve de capacité mentale, tout comme le terme «voler» a évolué pour désigner le cas, alors inconcevable, où des centaines de personnes assises tranquillement dans un avion volent à 1 000 kilomètres par heure, loin au-dessus des nuages. Newcomb, qui était un des calculateurs prodiges de son temps, refusa jusque sur son lit de mort d'admettre que les premiers avions «volaient».

L'arbitraire de la désignation culturelle peut être illustrée par l'expérience de la pensée qui consiste à remplacer une machine par quelque chose de mystérieux, mais de naturel. Bien qu'un chien ne puisse réussir le test de Turing, on admet que les chiens ont une certaine intelligence. On dit souvent que *Deeper Blue*, l'ordinateur de la Société *IBM* qui a battu le champion d'échecs Garry Kasparov, n'est pas réellement intelligent, mais imaginez un chien jouant aux échecs : s'il battait Kasparov, on le considérerait comme un chien remarquablement intelligent.

L'idée selon laquelle l'intelligence naturelle est une forme complexe de calcul n'est encore qu'une hypothèse. Toutefois, aucune raison claire ne permet de l'infirmer. Certains ont argumenté qu'elle n'expliquerait pas la conscience, mais, de toutes les théories actuelles sur la nature de la conscience, la théorie «calculatoire» semble la plus prometteuse. D'autres «théories» considèrent que la conscience résulte de propriétés physiques mystérieuses, mais elles ne proposent pas d'hypothèses testables. En considérant la vie mentale comme l'expression de calculs, l'intelligence artificielle fournit les éléments pour expliquer comment les symboles internes peuvent posséder une signification pour la machine et comment cette signification peut exercer une influence et être influencée par les relations causales qui existent entre la machine et son environnement.

Le but scientifique de l'intelligence artificielle est l'explication, en termes de calculs, de l'intelligence ou, plus largement, des capacités mentales elles-mêmes, et pas seulement une explication de l'esprit humain. Si cette démarche réussit, elle réfutera la spécificité de la pensée humaine et, par là, nous permettra de l'étendre et de la développer. Turing ne voulait pas souligner les différences entre les êtres pensants et les machines non pensantes ; il voulait les effacer. L'humanité ne serait pas amoindrie pour autant, car, somme toute, la compréhension des mouvements de l'air autour des ailes renforce l'admiration que nous portons au merveilleux vol des oiseaux. Il apparaît peut-être moins magique, mais sa complexité et ses raffinements sont impressionnants. Nous soupçonnons qu'il en est de même pour l'intelligence humaine. Si nos cerveaux sont vraiment des calculateurs biologiques, alors ce sont de remarquables ordinateurs.

Kenneth FORD est directeur du Centre informatique de la NASA. Patrick HAYES travaille à l'Institut pour la cognition humaine et artificielle, à l'Université de Floride.

Terry WINOYRAD et Fernando FLORES, *L'intelligence artificielle en question*, PUF, 1989.

Hans MORAVEC, *Une vie après la vie : les robots, avenir de l'intelligence*, Odile Jacob, 1992.

Jacques PITRAT, *De la machine à l'intelligence*, Hermès, 1995.

Daniel CREVIER, *À la recherche de l'intelligence artificielle*, Flammarion, 1997.

L'intelligence des réseaux de neurones

JEAN-PIERRE NADAL

Représentations simplifiées de parties du système nerveux des animaux, les réseaux de neurones résolvent des problèmes informatiques et éclairent certaines caractéristiques du fonctionnement cérébral.

Depuis 15 ans, des chercheurs de plusieurs disciplines – mathématiques, physique, biologie, etc. – étudient le comportement «intelligent» à l'aide de systèmes qui ont une architecture et un fonctionnement analogues à ceux de groupes de cellules dans le cerveau. Ces «réseaux de neurones formels», que nous désignerons désormais sous le nom de réseaux de neurones, résolvent des problèmes informatiques (reconnaissance de formes, par exemple) ou modélisent la cognition animale et humaine. Pour reproduire les mécanismes qui sont à l'origine de certaines capacités telles qu'apprendre, s'orienter, se déplacer, etc., on étudie le fonctionnement de réseaux de neurones à qui l'on fait exécuter des algorithmes, ou «calculs», qui semblent être ceux des groupes de neurones réels du cerveau.

Pour désigner ces calculs qui n'ont rien d'explicitement numérique, on emploie le terme de computation, mot d'origine latine, fort justement repris par la langue anglaise. Les sciences expérimentales sur lesquelles s'appuient ces travaux théoriques sont les neurosciences, notamment les neurosciences computationnelles ; ces disciplines tentent de relier l'organisation et les propriétés physico-chimiques des systèmes nerveux à leur fonction cognitive. La psychologie expérimentale, notamment la psychophysique et la psychologie cognitive participent également à cette recherche.

Naturellement, les spécialistes des neurosciences computationnelles ne prétendent pas que l'intelligence s'explique exclusivement par les propriétés de réseaux de neurones. Toutefois, il est bien démontré expérimentalement que des réseaux, c'est-à-dire des assemblées de neurones plutôt que des neurones individuels, jouent un rôle fonctionnel dans le cerveau. Par exemple, les neurophysiologistes ont démontré que, chez le rat, un groupe de neurones particulier code l'orientation de la tête ; l'activité de chaque neurone de ce groupe dépend de l'angle de la tête, mais celui-ci n'est correctement décrit que par l'activité de tous les neurones du groupe.

La première grande étude théorique des réseaux de neurones fut effectuée par le neurologue américain W. McCulloch et par le mathématicien W. Pitts en 1943. McCulloch reconnut ultérieurement qu'ils avaient voulu considérer le cerveau comme une machine universelle, au sens défini par le mathématicien britannique Alan Turing en 1937, c'est-à-dire un système capable de faire, en principe, n'importe quel calcul. McCulloch et Pitts montraient qu'on peut réaliser une machine de Turing avec des réseaux de neurones. Les «neurones» qu'ils envisageaient n'étaient pas de véritables cellules nerveuses, mais des modèles simplifiés de ces cellules : des «neurones formels», qui sont des «automates à seuil» ; ces petites unités qui additionnent les stimulations qu'elles reçoivent, produisent la valeur 0 ou 1 selon que cette somme est inférieure ou supérieure à un seuil *(voir la figure 2)*.

Certains psychologues considérèrent alors que l'on pouvait comprendre l'intelligence d'un cerveau par analogie avec les capacités d'un ordinateur. Cette idée reste aujourd'hui à la base de certaines approches de l'intelligence, en psychologie, en philosophie et en intelligence artificielle, mais l'étude des réseaux de neurones s'est développée, d'abord dans les années 1960, puis dans les années 1980, avec un point de vue tout différent.

En effet, un ordinateur classique a pour principale propriété d'être généraliste : on peut le programmer pour lui faire effectuer n'importe quelle tâche. Au contraire, un cerveau ressemble plutôt à un assemblage de machines spécialisées (le cortex visuel assure la vision, le cortex olfactif l'olfaction, le cortex moteur la commande des muscles...). Ces structures ont la faculté d'évoluer, de s'adapter à l'environnement. Les spécialistes des neurosciences computationnelles cherchent quelles sont les propriétés cognitives à la portée d'un réseau de neurones où la structure et la fonction sont intimement mêlées, où la frontière entre programme et matériel s'estompe. La faculté d'apprentissage étant l'une des plus fascinantes propriétés des systèmes vivants, la perspective d'en obtenir une modélisation par les réseaux de neurones est un attrait majeur.

La fonction dans la structure

La mémoire d'un réseau de neurones réside dans les poids, ou efficacités synaptiques, qui caractérisent les connexions entre les neurones : si un signal est émis par un neurone *A*, il est transmis à un neurone *B* avec une amplitude *a*, où *a* est ce que l'on nomme le poids de la connexion du neurone *A* au neurone *B*.

Dans les applications informatiques (par exemple, lors d'une tâche de recon-