

Pierre Jourlin

LA BOÎTE TRANSLUCIDE

UN ÉCLAIRAGE

SUR L'INTELLIGENCE ARTIFICIELLE



Editions
Universitaires
Avignon

Collection dirigée par Aurélia Barrière

Conception et mise en page : Catherine Julia

Correction du manuscrit : Sarah Jourden

Illustration de couverture : Pierre Jourlin, licence cc by-sa

ISBN : 978-2-35768-127-9

© 2021, Avignon Université



Éditions Universitaires d'Avignon

74, rue Louis Pasteur

84029 Avignon cedex 1 - France

Introduction : de la « matière grise »

Durant la majeure partie de mes études et dès les débuts de ma vie professionnelle, les expressions « intelligence artificielle », « apprentissage automatique », ainsi que leurs variantes anglophones étaient déjà discutées dans les milieux techniques ou technophiles, chez les *geeks* comme chez les ***nerds***¹ (Peyron, 2012). Mais, depuis quelques années, ces technologies sont effectivement utilisées – consciemment ou non – par un très large public et, dans le même temps, un nombre toujours croissant de disciplines académiques se saisissent du sujet : philosophie, économie, droit, sociologie, etc. Or, on présente souvent l'intelligence artificielle (IA) comme une boîte opaque, ce qui n'est pas vraiment démenti par les spécialistes, en particulier ceux qui développent des réseaux

1. Les mots en gras italique suivis d'un astérisque sont expliqués dans le lexique de fin de volume, pages 77-79.

de neurones artificiels (RNA). C'est naturellement une image qui peut générer des inquiétudes parfois justifiées et qui, sans aucun doute, fait couler beaucoup d'encre dans une grande variété de domaines et de disciplines. Cependant, si en tant qu'enseignant-chercheur en informatique, je reconnais bien volontiers que d'autres que moi auront plus de lumière à projeter sur les usages de cette boîte opaque, je vais tenter d'apporter ma contribution en l'éclairant un peu de l'intérieur. Ce livre vise ainsi un double objectif : constituer une introduction relativement brève, mais suffisamment large pour des personnes désirant se lancer dans des études en informatique ; faciliter le travail pluridisciplinaire pour des équipes de chercheuses et de chercheurs spécialisés dans d'autres domaines, mais concernés par les évolutions de la société dite numérique.

L'intelligence artificielle n'est pas un concept facile à cerner. En effet, lorsque nous évoquons les organes ou membres artificiels (cœur, jambe, bras, etc.), il n'est pas toujours nécessaire de définir précisément la liste des fonctions qui doivent être reproduites, réparées ou améliorées. Ces dernières sont en général bien identifiées. En revanche, c'est beaucoup plus difficile quand il s'agit

d'un organe tel que le cerveau, tant il est capable de réaliser une variété phénoménale de tâches. De plus, ces dernières ne sont pas toutes accessibles à la conscience ; réalisées inconsciemment, elles nous paraissent élémentaires, alors qu'un examen plus approfondi peut révéler de grands niveaux de complexité.

Prenons le langage : nous pouvons considérer que les mots que nous utilisons pour communiquer nos idées à autrui sont des symboles qui font référence à des objets concrets ou abstraits, à des représentations du réel et aux relations qu'ils entretiennent entre eux. Souvent, les mots sont ambigus : en fonction du contexte d'un énoncé ou de l'individu à l'origine de l'énoncé, un même mot ou une même séquence de mots peut faire référence à des choses très différentes. Même lorsque l'on a levé l'ambiguïté, il n'est pas toujours possible de lister tous les exemples correspondant à une définition et encore moins tous les exemples sortant du champ d'une définition. Ainsi, les contours du mot « intelligence » peuvent évoluer au gré des avancées des sciences et de la technologie, perdre ou gagner des représentants au fur et à mesure que la compréhension de notre propre fonctionnement cognitif s'élargit.

Par ailleurs, la démarche consistant pour une « intelligence » à s'autodéfinir ou à s'auto-comprendre peut donner lieu à quantité de productions philosophiques, mathématiques et même artistiques – et j'encourage vivement les lectrices et lecteurs à les découvrir au travers de l'œuvre de Douglas Hofstadter (Hofstadter, 1979).

De mon côté, je ne peux prétendre faire un long exposé sur le fonctionnement du cerveau humain, mais il me paraît indispensable de présenter les quelques éléments les plus généraux afin d'éclairer les problématiques que se proposent de traiter les technologies dites d'intelligence artificielle.

Olivier Houdé décrit l'architecture cognitive comme pouvant se décliner en « quatre principes fondamentaux et emboîtés : l'objet, le nombre, la catégorisation et le raisonnement » (Houdé, 2018, p. 12). On sait aussi qu'à l'un de ces plus bas niveaux, le cerveau est un immense réseau, composé de 85 milliards de neurones interconnectés par des centaines de milliards de ***synapses**** et de ***cellules gliales**** et qui interagissent entre eux par l'envoi (ou non) de signaux (Agid, 2018). Pour autant, il reste encore un grand nombre de zones d'ombre. Par exemple, si l'on connaît l'existence des cellules gliales depuis longtemps, leur rôle

indispensable dans les processus cognitifs commence à peine à être reconnu et étudié.

D'un autre point de vue, il est clair que certains niveaux, comme le raisonnement, sont accessibles à l'introspection (on dit aussi la **métacognition***). Nous sommes en effet capables d'analyser nos propres raisonnements, de constater les failles de nos théories et de les réviser. C'est même l'un des sujets de la philosophie, des mathématiques et de la logique depuis des millénaires. Cependant, de nombreux niveaux sont inaccessibles à la conscience et, même si nous avons réussi à en percer quelques secrets grâce, entre autres, à la psychologie expérimentale, puis à l'imagerie cérébrale, nous n'en avons encore qu'une représentation très floue et parcellaire (Dehaene, 2020).

Ainsi, la recherche en intelligence artificielle doit accepter que l'intelligence soit un mécanisme qui émerge de la complexité de cet immense enchevêtrement de cellules, sans pouvoir aller significativement plus loin dans sa compréhension.

Par exemple, s'il est possible d'observer et de décrire avec une certaine précision le fonctionnement d'une cellule du cerveau, le fait que les quatre principes fondamentaux de la cognition

soient emboîtés n'est pas sans conséquence. Les illusions dites d'« optique », très connues, sont autant d'exemples qui montrent que notre perception et notre représentation du monde peuvent vite s'éloigner de la réalité lorsque certaines conditions sont réunies. Les illusions auditives sont généralement moins connues, mais tout aussi surprenantes.

L'effet Shepard, par exemple, est un peu l'équivalent acoustique de l'enseigne de barbier ou de la vis sans fin. En diminuant progressivement le volume d'une note au fur et à mesure qu'elle monte en fréquence et en augmentant simultanément le volume de la note située à l'octave en dessous, on peut donner l'illusion qu'une note monte à l'infini, éternellement. Il existe même des illusions créées par la perception simultanée de signaux auditifs et visuels naturels, mais contradictoires.

Par exemple, pour obtenir un effet McGurk, on synchronise la vidéo d'une personne qui prononce le son [ga] avec la bande-son de cette même personne prononçant [da]. Lorsque l'on regarde ensuite cette vidéo en boucle, c'est le son [da] qui est perçu si nos yeux fixent l'image, et le son original [ba] si nous regardons ailleurs.



WIKIMEDIA/DOMAINE PUBLIC

FIGURE 1 :

Enseigne
de barbier,
environ 1938.

Plus d'une centaine de biais dits « cognitifs » ont été identifiés. Ils montrent que la qualité de notre raisonnement peut être considérablement dégradée dans des circonstances particulières qui mettent en jeu nos émotions, ou parfois par de simples confusions liées aux différents niveaux de perception et d'analyse. Paradoxalement, cette architecture complexe permet de concevoir des représentations extrêmement rigoureuses, par exemple en logique ou en mathématiques. Nous verrons par la suite qu'elles nous ont permis de concevoir des machines à traiter de l'information tout aussi rigoureuses.

Néanmoins, il faut remarquer qu'une définition purement fonctionnelle de l'intelligence serait forcément normative et limitative. La diversité des formes d'intelligence est très vaste et, même à l'intérieur d'un milieu social, il est possible d'en trouver des formes très atypiques, comme le montre l'écrivain et philosophe Josef Schovanec, autiste Asperger, dans son livre intitulé *Nos intelligences multiples* (Schovanec, 2018).

Il semble donc assez évident qu'il sera extrêmement difficile de définir cette propriété du cerveau humain et *a fortiori* de la mesurer. Pourtant, le fameux quotient intellectuel (QI), qui a été

inventé par Alfred Binet et Théodore Simon en 1905 pour repérer les enfants nécessitant un enseignement adapté, a été largement dévoyé et présenté comme un outil de mesure de l'intelligence, instrumentalisé dans le passé à des fins racistes, sexistes et eugénistes. Encore aujourd'hui, le QI semble malheureusement perçu comme une mesure de l'intelligence par le grand public. Je me permets de renvoyer les lectrices et lecteurs qui souhaiteraient approfondir la question à l'un des meilleurs ouvrages de vulgarisation que j'aie pu lire sur le sujet, qui s'intitule *La mal-mesure de l'homme* de S.J. Gould (Gould, 1997).

L'expression « intelligence artificielle » est apparue quelques années après la publication de l'article d'Alan M. Turing intitulé « Computer machinery and intelligence » (Turing, 1950). Ce mathématicien britannique avait déjà une grande notoriété à l'époque, car dès 1936, il avait défini les bases théoriques de l'informatique avec un modèle abstrait, une machine « universelle », capable de répondre non pas à un problème particulier mais à une infinité de problèmes, et qui prendra son nom : « la machine de Turing ». Quelques années plus tard, ses travaux permirent le développement du dispositif Enigma qui

rendit possible le décryptage des messages échangés par les armées nazies et contribua grandement à la victoire des Alliés lors de la Seconde Guerre mondiale.

Or, dans son article de 1950, Alan M. Turing explique qu'il est impossible de déterminer si un être vivant peut « penser » et *a fortiori* si une machine peut être « intelligente ». Il propose donc de chercher la réponse à une autre question : la machine serait-elle capable, dans un jeu d'imitation, de nous faire croire qu'elle est un être humain ? Ce jeu d'imitation, qui prendra le nom de « test de Turing », se décline évidemment en une myriade de sous-problèmes liés à la grande diversité des informations que notre cerveau peut traiter. Par conséquent, depuis les années 1960, c'est un peu ce Graal de la recherche en informatique que l'on s'efforce d'atteindre. Un chemin important a été parcouru durant les soixante années qui ont suivi la publication de l'article de Turing ; nous verrons dans le prochain chapitre les approches dont le succès n'a, à ce jour, pas été démenti.

Pour en savoir plus

- DEHAENE S., *Le code de la conscience*, Paris, Odile Jacob, 2014.
- HOFSTADTER D.R., *Gödel, Escher, Bach : les brins d'une guirlande éternelle*, Paris, Dunod, 2008 (édition originale, New York, 1979).
- HOUDÉ O., *Les 100 mots de la psychologie*, Paris, PUF, 2018.