

# Chapitre 1

## Introduction

*Où l'on tente d'expliquer les raisons qui nous amènent à considérer l'intelligence artificielle comme un sujet digne d'intérêt et où l'on essaie de cerner sa nature exacte, question qu'il importe de trancher avant d'aller plus avant.*

Nous disons de nous que nous sommes des *Homo sapiens*, autrement dit des sages, en raison de l'importance que nous attribuons à notre **intelligence**. Pendant des millénaires, nous avons essayé de comprendre *le processus de la pensée*, à savoir comment un simple amas de chair peut percevoir, comprendre, prévoir et manipuler un monde bien plus étendu et complexe que lui-même. Le domaine de l'**intelligence artificielle**, ou IA, va encore plus loin : il tente non seulement de comprendre des entités intelligentes, mais aussi d'en *construire*.

L'IA est un des champs les plus récents parmi les sciences et l'ingénierie. Les travaux ont sérieusement débuté juste après la Seconde Guerre mondiale et le terme a été forgé en 1956. Avec la biologie moléculaire, l'IA est régulièrement citée en tant que domaine qu'auraient volontiers choisi les spécialistes d'autres disciplines. Un étudiant en physique est susceptible de se dire que toutes les grandes idées ont déjà été formulées par Galilée, Newton, Einstein et d'autres éminents scientifiques. À l'inverse, l'IA offre des perspectives pour plusieurs Einstein ou Edison à plein temps.

À l'heure actuelle, l'IA est composée d'une grande diversité de sous-disciplines allant des plus générales (apprentissage, perception) aux plus spécifiques (jouer aux échecs, démontrer des théorèmes mathématiques, écrire des poèmes, conduire un véhicule au milieu de la circulation et diagnostiquer des maladies). L'IA relève de toutes les tâches intellectuelles : c'est vraiment un domaine universel.

### 1.1 Définition de l'IA

S'il a été question de l'attrait exercé par l'IA, nous n'avons encore rien dit de sa *nature*. Les huit définitions de l'intelligence artificielle données dans les manuels (voir figure 1.1) s'ordonnent selon deux dimensions. Grossièrement, celles de la première ligne concernent *le processus de la pensée* et du *raisonnement* tandis que celles de la seconde ont trait au

Des systèmes qui pensent comme les humains	Des systèmes qui pensent rationnellement
« La tentative nouvelle et passionnante d'amener les ordinateurs à penser... [d'en faire] des <i>machines dotées d'un esprit</i> au sens le plus littéral. » (Haugeland, 1985)	« L'étude des facultés mentales grâce à des modèles informatiques. » (Charniak et McDermott, 1985)
« [L'automatisation d']activités que nous associons à la pensée humaine, des activités telles que la prise de décision, la résolution de problèmes, l'apprentissage... » (Bellman, 1978)	« L'étude des moyens informatiques qui rendent possibles la perception, le raisonnement et l'action. » (Winston, 1992)
Des systèmes qui agissent comme les humains	Des systèmes qui agissent rationnellement
« L'art de créer des machines capables de prendre en charge des fonctions exigeant de l'intelligence quand elles sont réalisées par des gens. » (Kurzweil, 1990)	« L'intelligence artificielle ( <i>computational intelligence</i> ) est l'étude de la conception d'agents intelligents. » (Poole <i>et al.</i> , 1998)
« L'étude des moyens à mettre en œuvre pour faire en sorte que des ordinateurs accomplissent des choses pour lesquelles il est préférable de recourir à des personnes pour le moment. » (Rich et Knight, 1991)	« L'IA... étudie le comportement intelligent dans des artefacts. » (Nilsson, 1998)

Figure 1.1 : Quelques définitions de l'intelligence artificielle, regroupées en quatre catégories.

*comportement*. Les définitions de la colonne de gauche évaluent la réussite par rapport aux performances *humaines* tandis que celles de la colonne de droite la mesurent à l'aune d'un concept *idéal* de l'intelligence, que nous appellerons **rationalité**. Un système est rationnel s'il opère de manière appropriée compte tenu de ce qu'il sait.

Historiquement, ces quatre approches de l'IA ont toutes été suivies, chacune par différentes personnes avec différentes méthodes. On s'attend à ce qu'une approche centrée sur l'humain fasse partie des sciences molles, en mettant en jeu des observations et des hypothèses sur le comportement humain. Une approche rationaliste<sup>1</sup> fera appel à une combinaison de mathématiques et d'ingénierie. Les tenants de chaque approche ont autant décrié leurs concurrents qu'ils les ont aidés. Examinons de plus près chacune d'elles.

1. Précisons qu'en faisant une distinction entre les comportements *humains* et *rationnels*, nous ne suggérons pas que les humains sont nécessairement « irrationnels » au sens où ils seraient « émotionnellement instables » ou « insensés ». Nous nous contenterons de signaler que nous ne sommes pas parfaits : tous les joueurs d'échecs ne sont pas des grands maîtres ; et malheureusement, tout le monde n'obtient pas la meilleure note à l'examen. Pour une étude des erreurs de raisonnement systématiquement commises par les humains, voir Kahneman *et al* (1982).

### 1.1.1 Agir comme des humains : le test de Turing

Le **test de Turing**, proposé par Alan Turing (1950), vise à fournir une définition satisfaisante et opérationnelle de l'intelligence. Un ordinateur réussit le test si, après avoir posé un certain nombre de questions écrites, un questionneur humain est dans l'incapacité de dire si les réponses écrites proviennent d'une personne ou d'un ordinateur. Le chapitre 26 étudie le test en détail et cherche à savoir si l'on peut dire qu'un ordinateur serait vraiment intelligent en cas de succès. Pour l'instant, contentons-nous de noter que programmer un ordinateur pour passer un test rigoureux ouvre beaucoup de chantiers. L'ordinateur devrait posséder les fonctionnalités suivantes :

- le **traitement du langage naturel**, qui lui permettra de communiquer sans problème ;
- la **représentation des connaissances**, grâce à laquelle il stockera ce qu'il sait ou entend ;
- le **raisonnement automatisé**, qu'il emploiera pour répondre aux questions et tirer des conclusions en utilisant les informations mémorisées ;
- l'**apprentissage**, qui lui permettra de s'adapter à de nouvelles circonstances, de détecter des invariants et de les extrapoler.

Le test de Turing évite délibérément toute interaction physique directe entre le questionneur et l'ordinateur, car la simulation *physique* d'une personne n'est pas indispensable pour l'intelligence. Cependant, le **test de Turing complet** inclut un signal vidéo qui permet au questionneur de tester les capacités perceptives du sujet, ainsi que la possibilité pour le questionneur de passer des objets physiques « par un guichet ». Afin de réussir le test de Turing complet, l'ordinateur doit être doté :

- d'un dispositif de **vision artificielle** pour percevoir des objets ;
- d'une capacité **robotique** pour manipuler des objets et se déplacer.

Les six domaines précités constituent la majeure partie de l'IA et Turing a eu le mérite de concevoir un test qui demeure valide soixante ans plus tard. Néanmoins, les chercheurs en IA se sont peu préoccupés de construire des programmes capables de passer le test de Turing, car ils ont jugé plus important d'étudier les principes sous-jacents à l'intelligence que de tenter d'imiter celle-ci. La quête du « vol artificiel » a réussi lorsque les frères Wright et d'autres précurseurs ont cessé d'imiter les oiseaux pour utiliser des souffleries et s'intéresser à l'aérodynamique. L'ingénierie aéronautique ne se donne pas pour objectif de mettre au point « des machines qui volent exactement comme les pigeons au point que les pigeons eux-mêmes s'y méprennent ».

### 1.1.2 Penser comme des humains : l'approche cognitive

Pour pouvoir dire qu'un programme donné pense comme un humain, il faut pouvoir déterminer comment pensent les êtres humains en pénétrant à *l'intérieur* des rouages de l'esprit. Il existe trois moyens d'y parvenir : l'introspection – la tentative de se saisir de ses propres pensées ; les expériences psychologiques – observer une personne dans ses comportements ; et l'imagerie cérébrale – observer le cerveau en fonctionnement. Dès lors qu'on dispose d'une théorie de l'esprit suffisamment précise, il devient envisageable d'exprimer cette théorie sous la forme d'un programme informatique. Si les comportements du programme en termes d'entrées-sorties correspondent à ceux des humains, c'est le signe que certains de ces mécanismes sont également susceptibles d'opérer chez les humains. Par exemple, Allen Newell et Herbert Simon, qui ont développé le GPS, le *General Problem Solver* (Newell et Simon, 1961), ne se sont pas contentés de seulement créer un programme qui résolvait correctement des problèmes : ils ont cherché en outre à comparer les étapes de son raisonnement à celles de

sujets humains confrontés aux mêmes problèmes. Le domaine interdisciplinaire des **sciences cognitives** combine les modèles informatiques de l'IA et les techniques expérimentales de la psychologie dans le but d'élaborer des théories précises et vérifiables du fonctionnement de l'esprit humain.

Les sciences cognitives constituent un domaine fascinant en soi, qui justifie qu'on lui ait consacré plusieurs ouvrages et au moins une encyclopédie (Wilson et Keil, 1999). Nous n'essaierons pas ici de décrire l'état du savoir quant aux processus cognitifs de l'homme. Nous signalerons de temps à autre les similitudes ou les différences entre les techniques de l'IA et la cognition humaine. Ces recherches ne peuvent être qualifiées de scientifiques que si elles font appel à des expérimentations sur des humains ou des animaux. Nous laisserons cela à d'autres ouvrages, car nous supposons que le lecteur ne dispose que d'un ordinateur pour mener des expériences à bien.

Aux débuts de l'IA, les deux approches étaient souvent confondues : un auteur avançait qu'un algorithme accomplissait bien une tâche et en *concluait* qu'il constituait un bon modèle du fonctionnement de l'esprit humain, ou *vice versa*. Les auteurs contemporains distinguent ces deux types d'assertions ; cette distinction a permis tant à l'IA qu'aux sciences cognitives de se développer plus rapidement. Ces deux disciplines continuent à se fertiliser mutuellement, notamment dans le domaine de la vision, qui incorpore des enseignements de la neuropsychologie aux modèles informatiques.

### 1.1.3 Penser rationnellement : les « lois de la pensée »

Le philosophe grec Aristote est l'un des premiers à avoir essayé de codifier « le bien-penser », autrement dit les procédés des raisonnements irréfutables. Ses **sylogismes** proposaient des modèles de structures argumentatives qui aboutissaient toujours à des conclusions vraies, dès lors qu'on leur fournissait des prémisses vraies ; par exemple : « Socrate est un homme, tous les hommes sont mortels, donc Socrate est mortel. » Ces lois de la pensée étaient supposées régir les opérations de l'esprit ; leur étude a ouvert le domaine de la **logique**.

Les logiciens du XIX<sup>e</sup> siècle ont mis au point une notation précise pour les assertions relatives à l'ensemble des objets constituant le monde et aux relations qui les lient. (On peut comparer celle-ci avec la notation arithmétique usuelle qui est seulement destinée aux assertions sur les *nombres*.) Dès 1965, il existait des programmes qui pouvaient, en principe, résoudre *tout* problème soluble formulé avec la notation logique. (Bien que s'il n'existe aucune solution, le programme puisse en chercher une indéfiniment.) En IA, cette tradition, dite **logiciste**, mise sur des programmes de ce genre pour créer des systèmes intelligents.

Cette approche se heurte à deux grands obstacles. Tout d'abord, il est difficile d'isoler une connaissance informelle et de l'exprimer dans les termes formels requis par la notation logique, notamment lorsque la connaissance n'est pas certaine à 100 %. En second lieu, il existe une grande différence entre le fait de résoudre un problème « en principe » et de le faire dans la pratique. Même des problèmes qui ne portent que sur quelques centaines de faits peuvent épuiser toutes les ressources de calcul d'un ordinateur en l'absence de directives indiquant les raisonnements à essayer en priorité. Ces deux obstacles, auxquels se heurtera *toute* tentative d'élaboration de systèmes de raisonnement artificiels, sont apparus pour la première fois dans la tradition logiciste.

### 1.1.4 Agir rationnellement : l'approche de l'agent rationnel

Un **agent** est simplement une entité qui agit (« *agent* » vient du latin *agere*, faire). Bien sûr, tous les programmes informatiques calculent quelque chose, mais les agents informatiques sont supposés faire plus : fonctionner de manière autonome, percevoir leur environnement, persister pendant une période prolongée, s'adapter au changement et créer et poursuivre des objectifs. Un **agent rationnel** est un agent qui agit de manière à atteindre la meilleure solution ou, dans un environnement incertain, la meilleure solution prévisible.

Dans le cadre d'une approche de l'IA subordonnée aux « lois de la pensée », l'accent est mis sur la validité des inférences. La capacité à élaborer des inférences correctes fait parfois *partie* de la nature d'un agent rationnel, car l'un des comportements rationnels possibles consiste à conclure logiquement qu'une action donnée permettra d'atteindre des objectifs déterminés, puis à agir conformément à cette conclusion. À l'inverse, la capacité à inférer correctement n'englobe pas *toute* la rationalité car, dans certaines situations, il n'y a aucune décision à prendre que l'on puisse déterminer avec certitude, pourtant il faut en prendre une. Il y a également des actes rationnels dont on ne peut pas dire s'ils reposent sur des inférences. Par exemple, s'écarter d'un poêle brûlant est une action réflexe qui se montre généralement plus efficace qu'une action décidée après mûre réflexion, donc plus lente.

Toutes les qualités requises par le test de Turing permettent également à un agent d'agir rationnellement. La représentation des connaissances et le raisonnement permettent à un agent de prendre de bonnes décisions. On doit pouvoir générer des phrases compréhensibles en langage naturel pour évoluer dans une société complexe. On a besoin d'apprendre non seulement à des fins d'érudition, mais également parce que ça améliore notre capacité à déterminer un comportement plus efficace.

L'approche de type agent rationnel a deux avantages sur les autres. Premièrement, elle est plus générale que l'approche par les « lois de la pensée » parce que la validité des inférences n'est que l'un des nombreux mécanismes qui permettent d'accéder à la rationalité. Deuxièmement, elle convient mieux au développement scientifique que les approches fondées sur le comportement ou la pensée humaine. La rationalité prise comme norme est mathématiquement définie et totalement générique, et peut être « déployée » pour générer des agents qui la réalisent, et ce de manière prouvable. À l'inverse, le comportement humain est bien adapté à un environnement spécifique, et est défini disons ... par l'ensemble des choses que les humains font. *En conséquence, cet ouvrage se concentre sur les principes généraux des agents rationnels et sur les composants qui permettent de les construire.* En dépit de l'apparente simplicité avec laquelle on peut énoncer le problème, nombre de difficultés surgissent dès lors qu'on essaie de le résoudre. Le chapitre 2 s'attarde sur plusieurs d'entre elles.

Un point important à garder en tête : on verra sous peu qu'il est inenvisageable d'atteindre une rationalité si parfaite qu'elle donne la possibilité de toujours se comporter au mieux dans des environnements complexes. Les besoins en calcul sont simplement trop élevés. Cependant, dans la plus grande partie de ce livre, nous prendrons pour hypothèse de travail que la rationalité parfaite est un bon point de départ pour l'analyse. Elle simplifie le problème et fournit le cadre approprié pour la majorité des éléments au fondement du domaine. Les chapitres 5 et 17 traitent explicitement du problème de la **rationalité limitée**, c'est-à-dire de l'action appropriée lorsqu'on ne dispose pas de suffisamment de temps pour effectuer tous les calculs souhaitables.